SUPPLEMENTARY INFORMATION

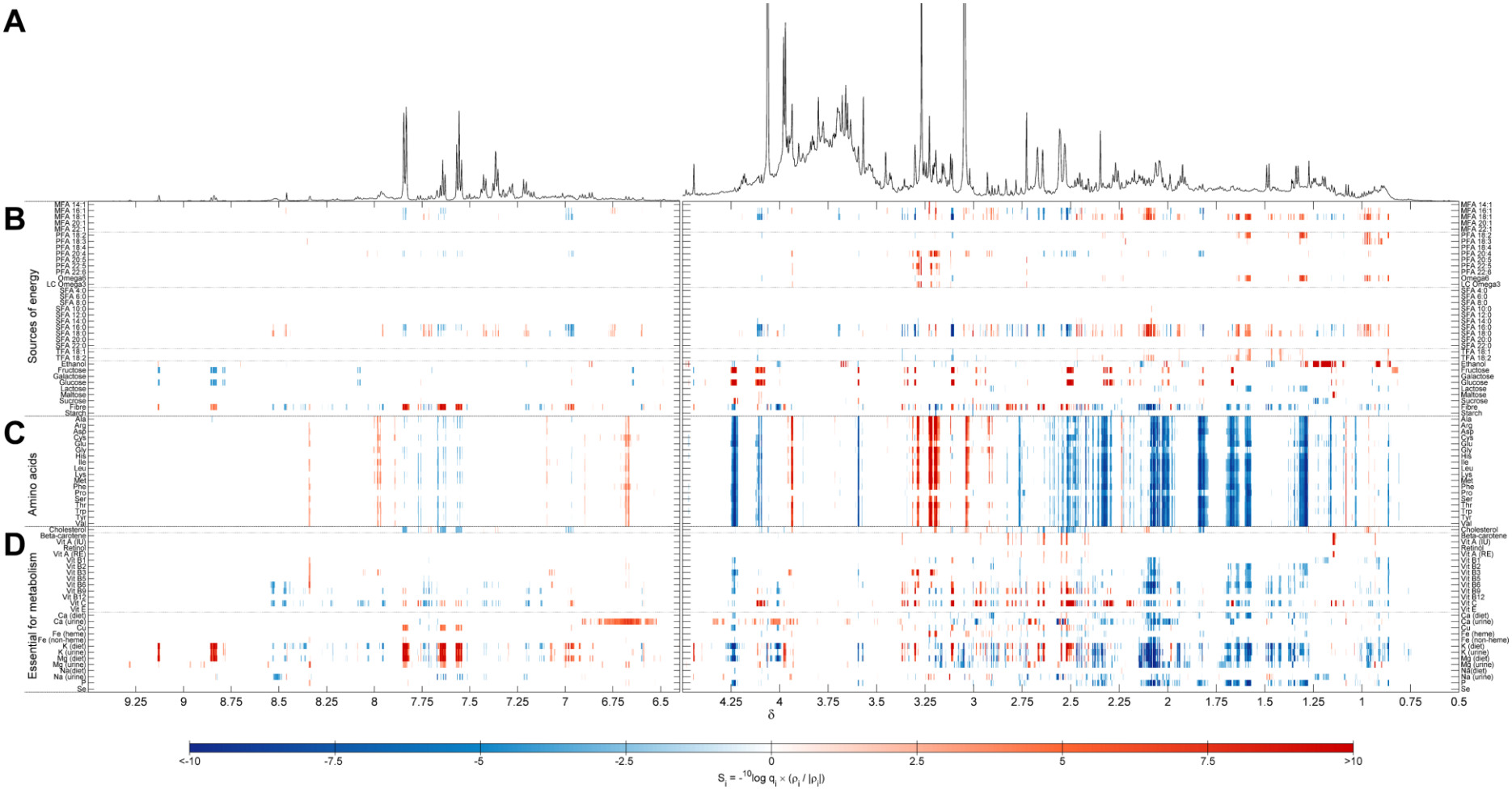
Title: Nutriome-metabolome relationships provide insights into dietary intake and metabolism

Authors: Joram M. Posma [1,2], Isabel Garcia-Perez [3], Gary Frost [3], Ghadeer S. Aljuraiban [4,5], Queenie Chan [5,6], Linda Van Horn [7], Martha Daviglus [8], Jeremiah Stamler [7], Elaine Holmes [3,9,10,11,*], Paul Elliott [2,5,6,9,12,13,*] , Jeremy K. Nicholson [10,11,*]

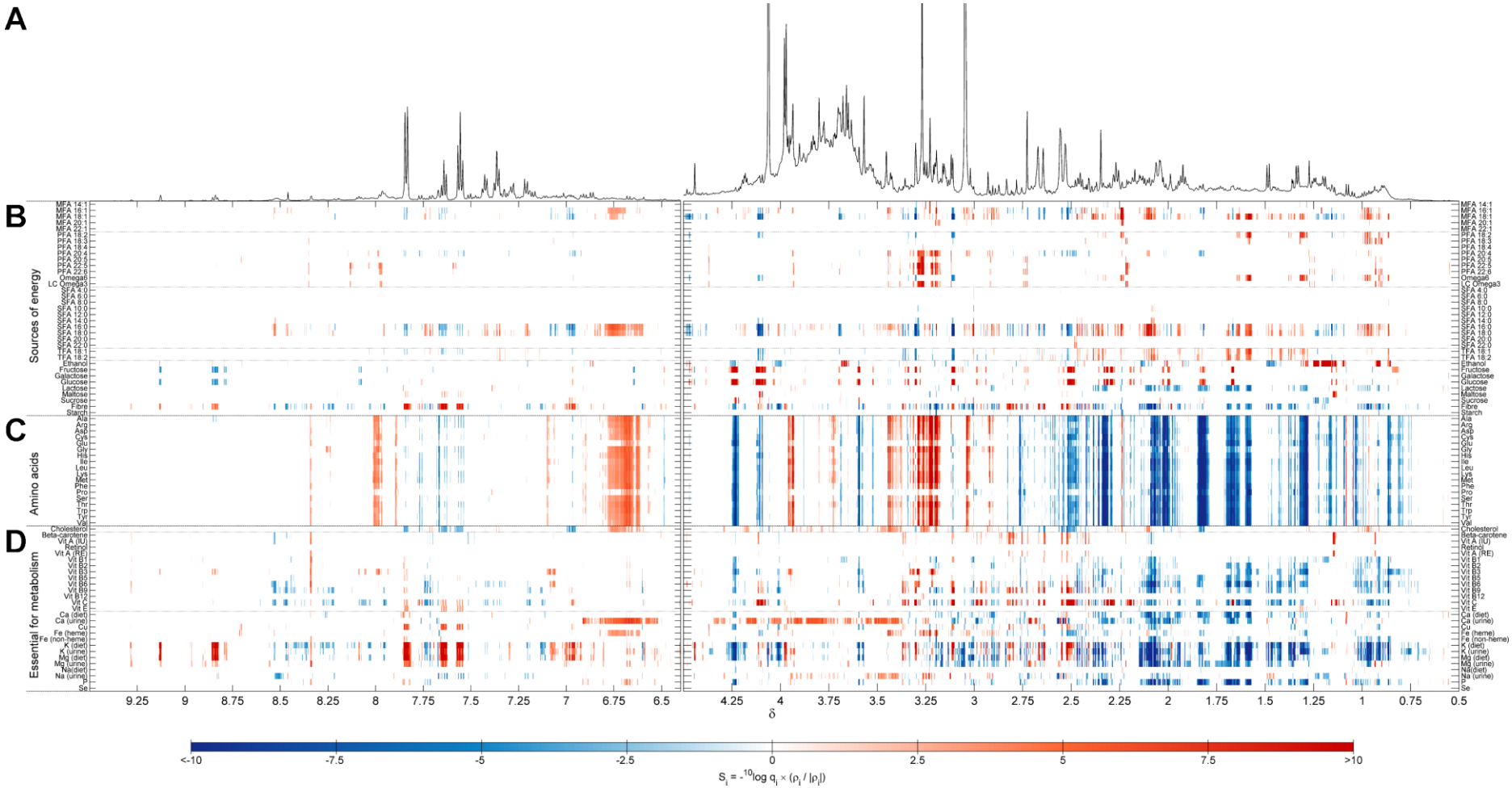
Affiliations: [1] Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, South Kensington Campus, Imperial College London, SW7 2AZ, U.K.; [2] Health Data Research UK-London, U.K.; [3] Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Hammersmith Campus, Imperial College London, W12 0NN, U.K.; [4] The Department of Community Health Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia; [5] Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, St. Mary's Campus, Imperial College London, W2 1PG, U.K.; [6] MRC Centre for Environment and Health, School of Public Health, Faculty of Medicine, St. Mary's Campus, Imperial College London, W2 1PG, U.K; [7] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, U.S.A.; [8] Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL 60612; [9] UK Dementia Research Institute, Faculty of Medicine, Hammersmith Campus, Imperial College London, W12 0NN, U.K.; [10] Division of Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; [11] The Australian National Phenome Center, Harry Perkins Institute, Murdoch University, WA 6150, Australia; [12] National Institute for Health Research Imperial Biomedical Research Centre, St. Mary's Campus, Imperial College London, W2 1PG, U.K; [13] British Heart Foundation Centre of Research Excellence at Imperial, Imperial College London, W2 1PG, U.K.

Correspondence: jeremy.nicholson@murdoch.edu.au, p.elliott@imperial.ac.uk, elaine.holmes@murdoch.edu.au.

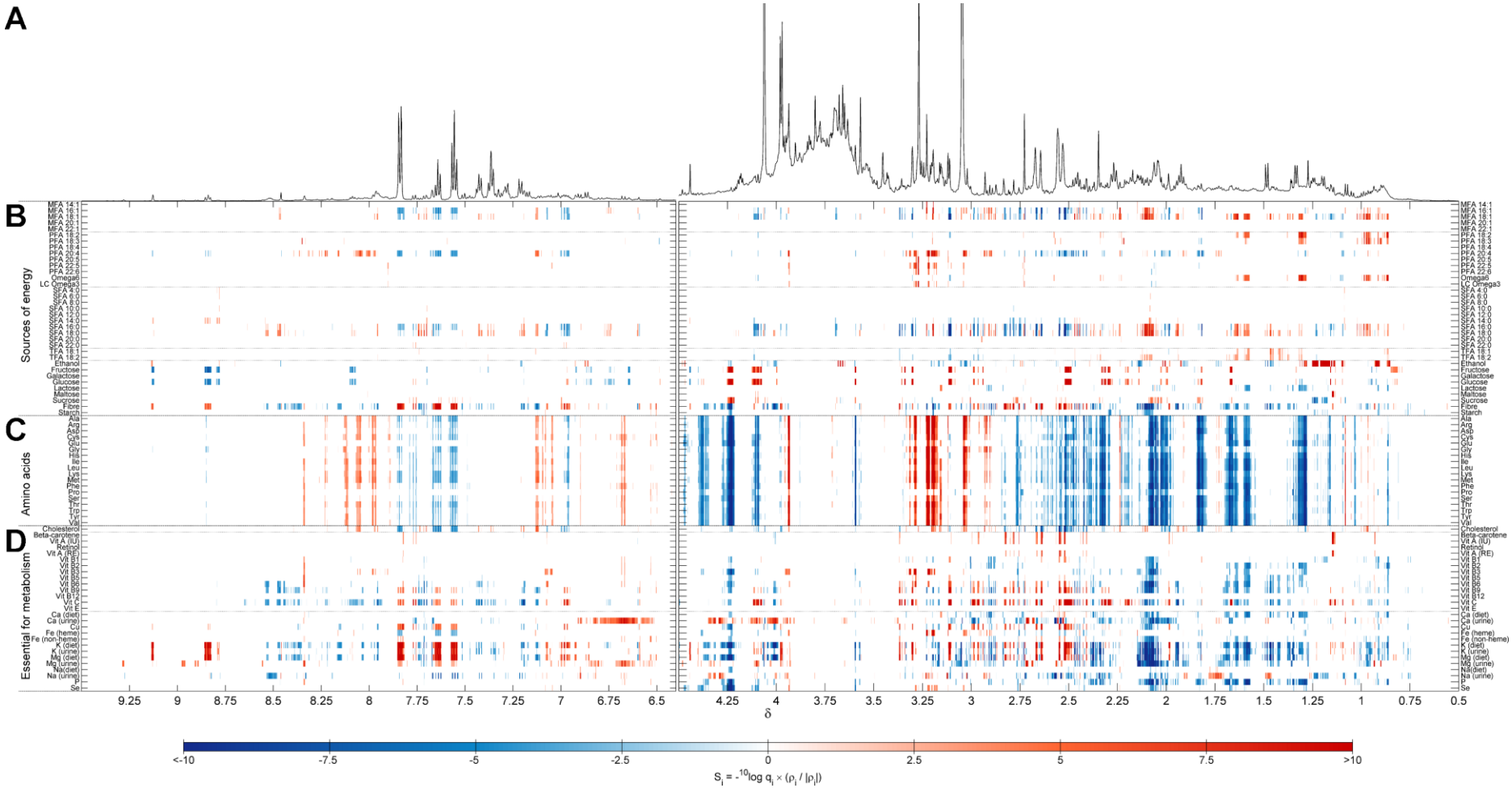
SUPPLEMENTARY FIGURES



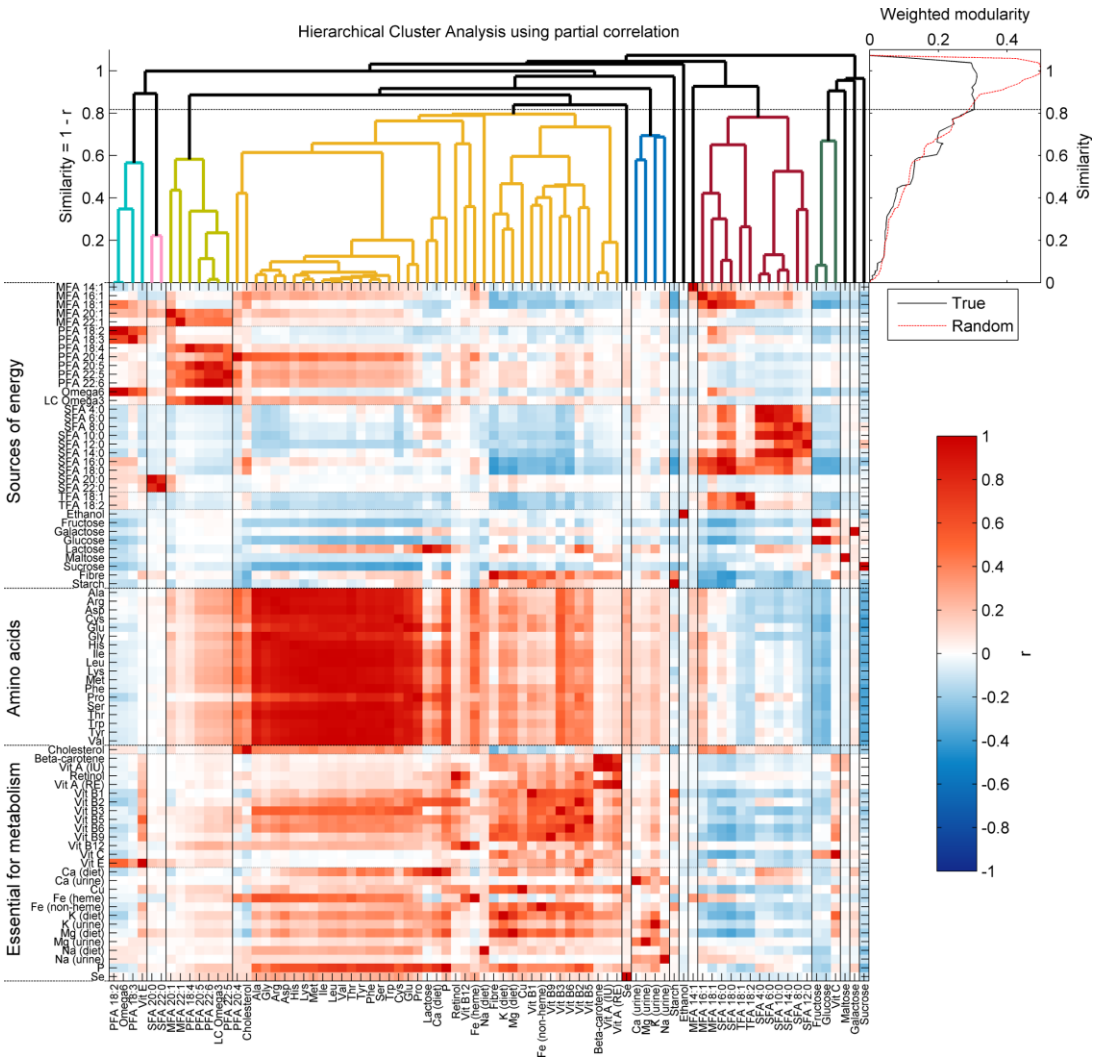
Supplementary Figure 1. The metabolic influence potential of 80 nutrients in the U.S. INTERMAP population. (**A**) Median ^{1}H-NMR spectrum of both visits. (**B**, **C**, **D**) The skyline projection of the significant associations ($-^{10}\log(q)\times\text{sign}(\rho)$, bounded to limits -10 and 10, where q is the false discovery rate q-value) for all nutrients for both visits combined (least significant visit shown for each variable); positive associations are shown in shades of red and inverse associations in shades of blue. (**B**) Nutrients that are energy providing (fatty acids, alcohol, sugars, fibre, starch), (**C**) dietary amino acids and (**D**) are essential for supporting metabolism (cholesterol, vitamins and derivatives, minerals).



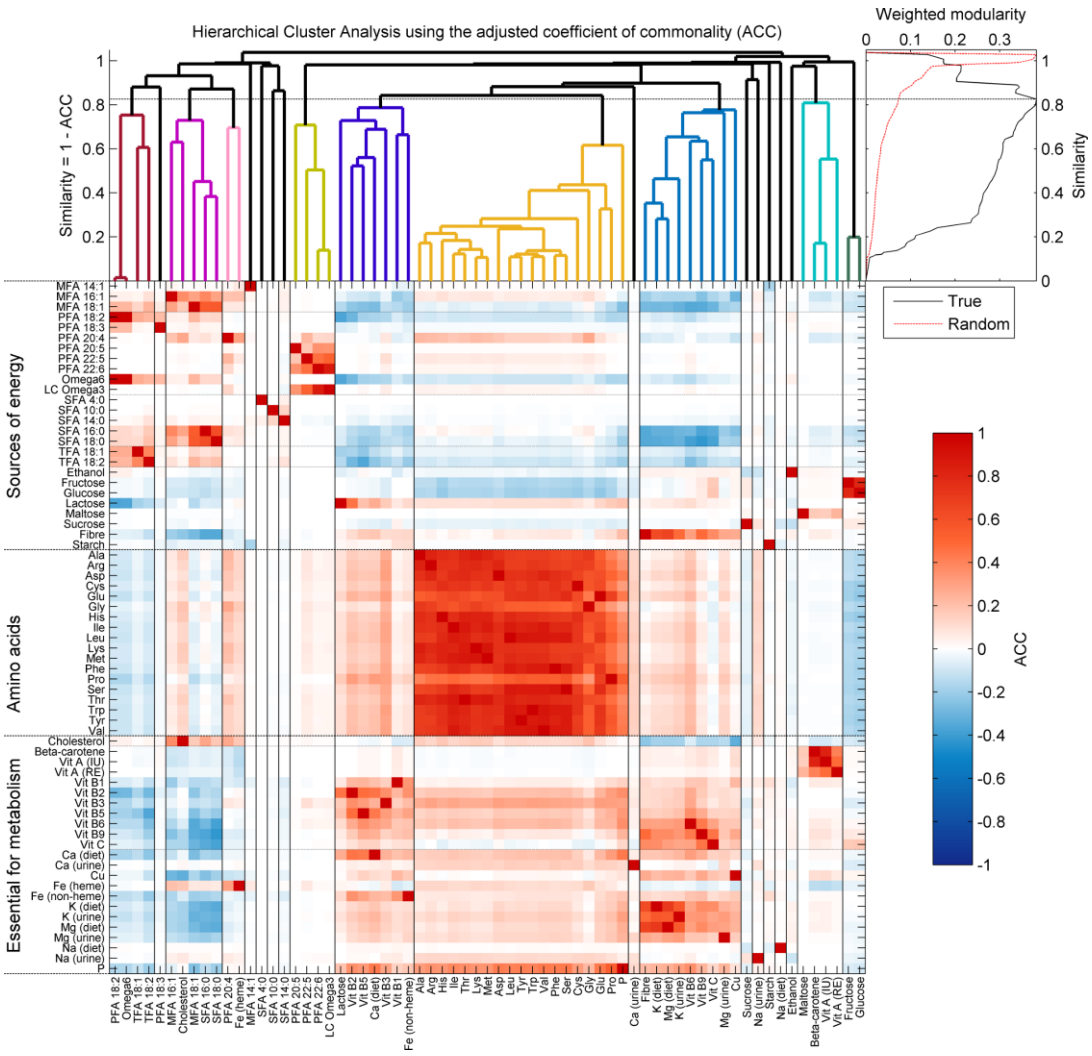
Supplementary Figure 2. The metabolic signatures of 80 nutrients in the U.S. INTERMAP population (first visit). (**A**) Median ^{1}H-NMR spectrum of the first visit. (**B**, **C**, **D**) The skyline projection of the significant associations ($-^{10}\log(q)\times\text{sign}(\rho)$, bounded to limits -10 and 10, where q is the false discovery rate q-value) for all nutrients for the first visit; positive associations are shown in shades of red and inverse associations in shades of blue. (**B**) Nutrients that are energy providing (fatty acids, alcohol, sugars, fibre, starch), (**C**) dietary amino acids and (**D**) are essential for supporting metabolism (cholesterol, vitamins and derivatives, minerals).



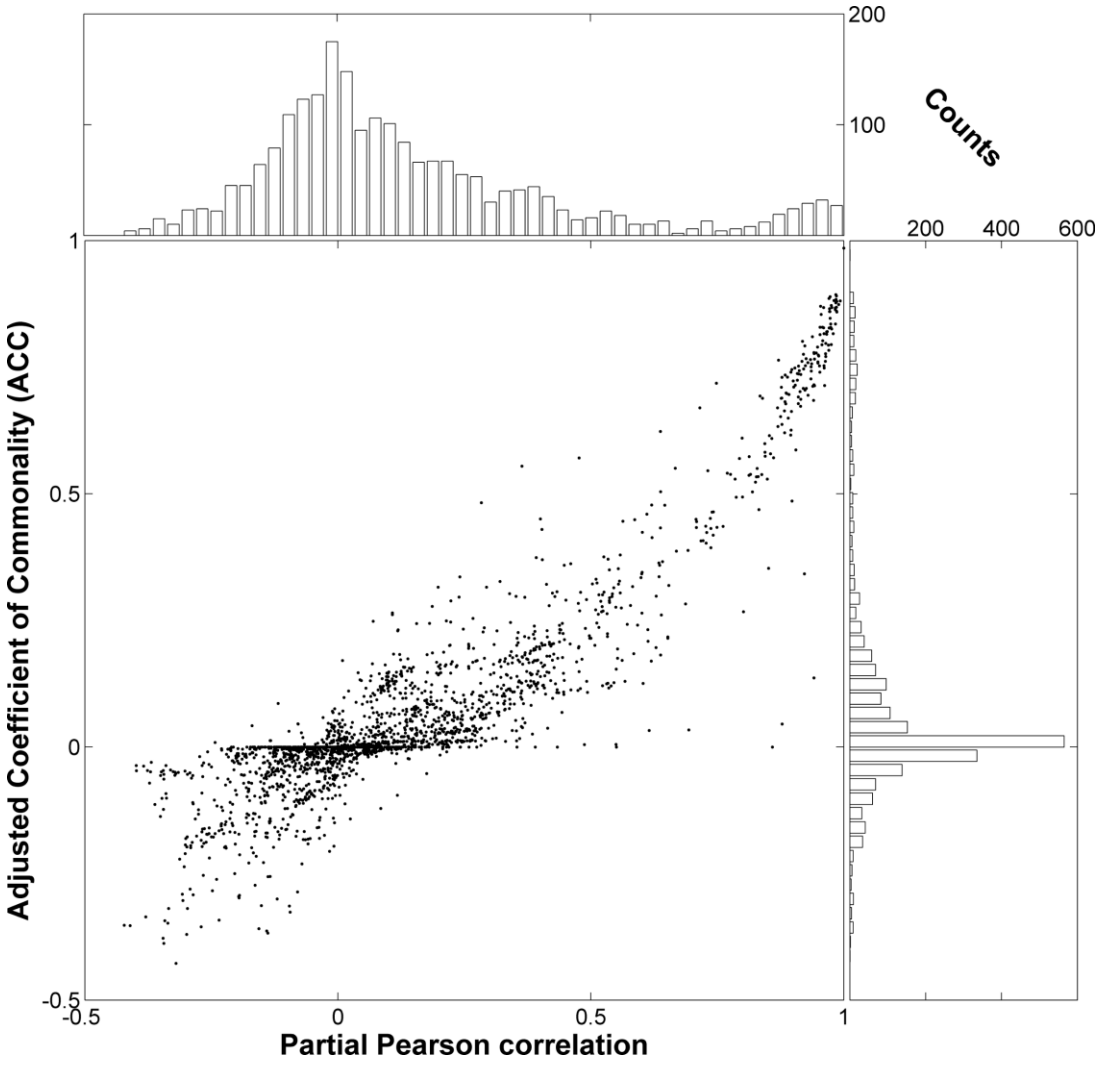
Supplementary Figure 3. The metabolic signatures of 80 nutrients in the U.S. INTERMAP population (second visit). (**A**) Median ^{1}H-NMR spectrum of the second visit. (**B**, **C**, **D**) The skyline projection of the significant associations ($-^{10}\log(q)\times\text{sign}(\rho)$, bounded to limits -10 and 10, where q is the false discovery rate q-value) for all nutrients for the second visit; positive associations are shown in shades of red and inverse associations in shades of blue. (**B**) Nutrients that are energy providing (fatty acids, alcohol, sugars, fibre, starch), (**C**) dietary amino acids and (**D**) are essential for supporting metabolism (cholesterol, vitamins and derivatives, minerals).



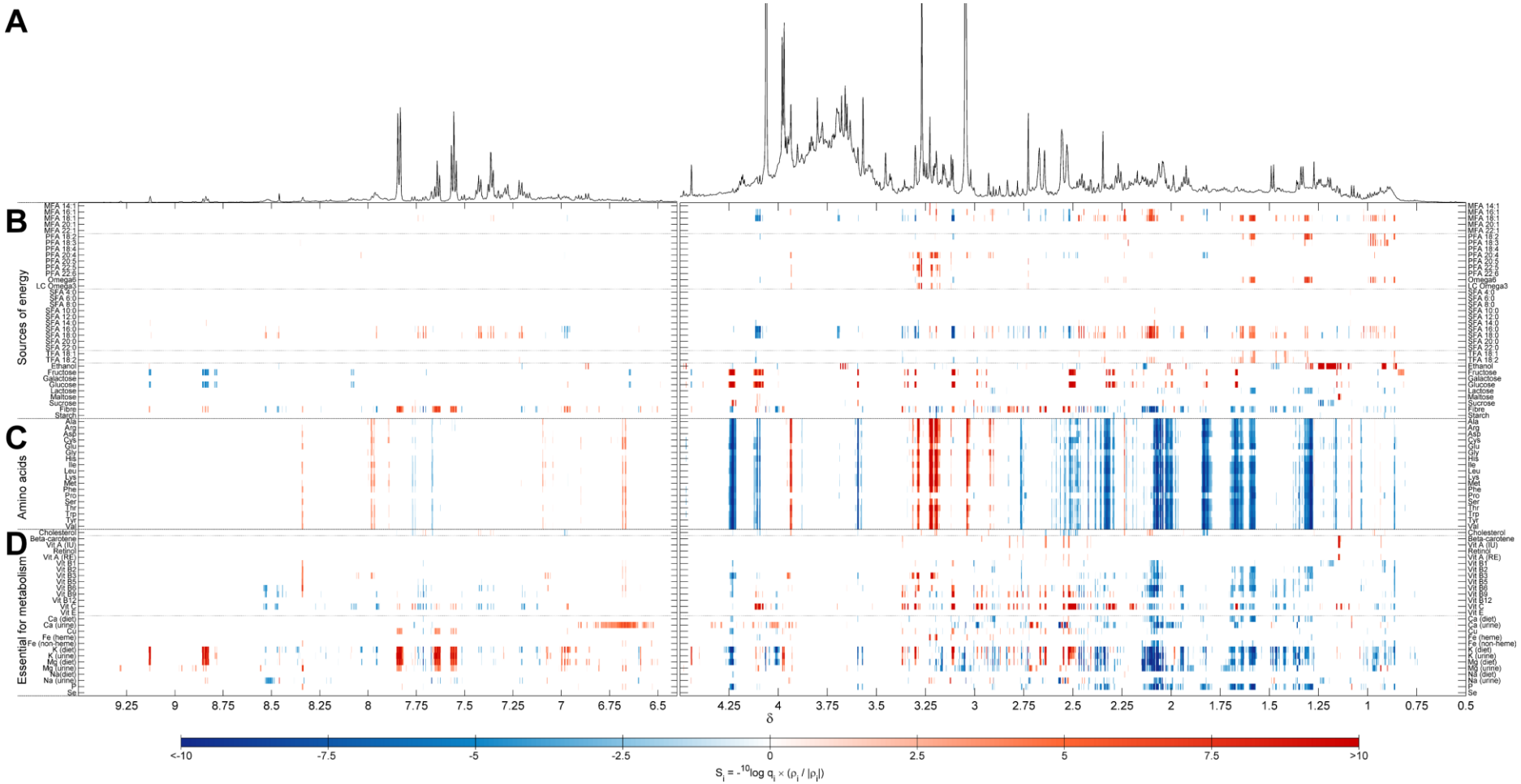
Supplementary Figure 4. Similarity between nutrients based on the dietary data of the U.S. INTERMAP population by means of partial correlation (adjusted for age, sex and population sample). The smallest (in absolute sense) partial correlation between two nutrients across both visits was used. All 80 nutrients were included and ordered as in **Supplementary Figure 1** on the y-axis, and ordered based on hierarchical clustering on the x-axis. The optimal number of clusters was found to be 14 and was calculated by comparing the modularity of the network with 1,000 random networks of the same degree structure. The random network with the highest modularity is shown for comparison.



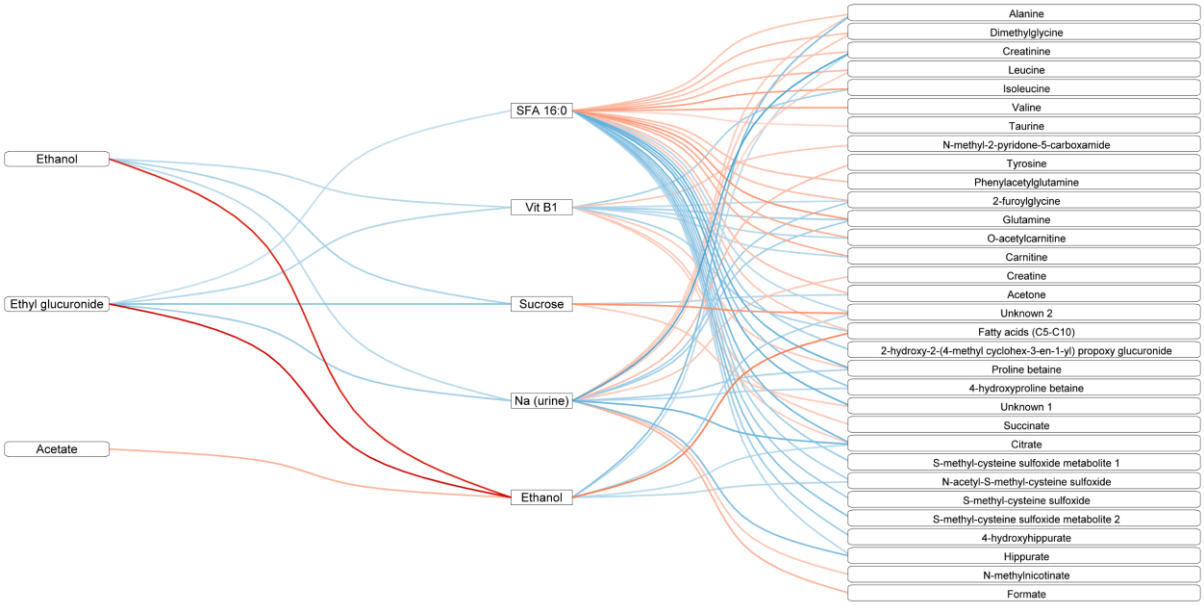
Supplementary Figure 5. Functional similarities between nutrients based on their urinary metabolic signatures. Similarity between nutrients is based on their correlation with urinary metabolic profiles in the U.S. INTERMAP population (n=1,848) by means of the Adjusted Coefficient of Commonality (ACC, see **Methods**). The 67 nutrients with significant associations for both visits were included and ordered as in **Supplementary Figure 1** on the y-axis, the clustering of dietary data based on the metabolic profiles is shown on the x-axis. The optimal number of clusters was found to be 20 and was calculated by comparing the modularity of the network with 1,000 random networks of the same degree structure. The random network with the highest modularity is shown for comparison. It shows the network structure is distinctly different from random networks of the same degree structure (red line for modularity).



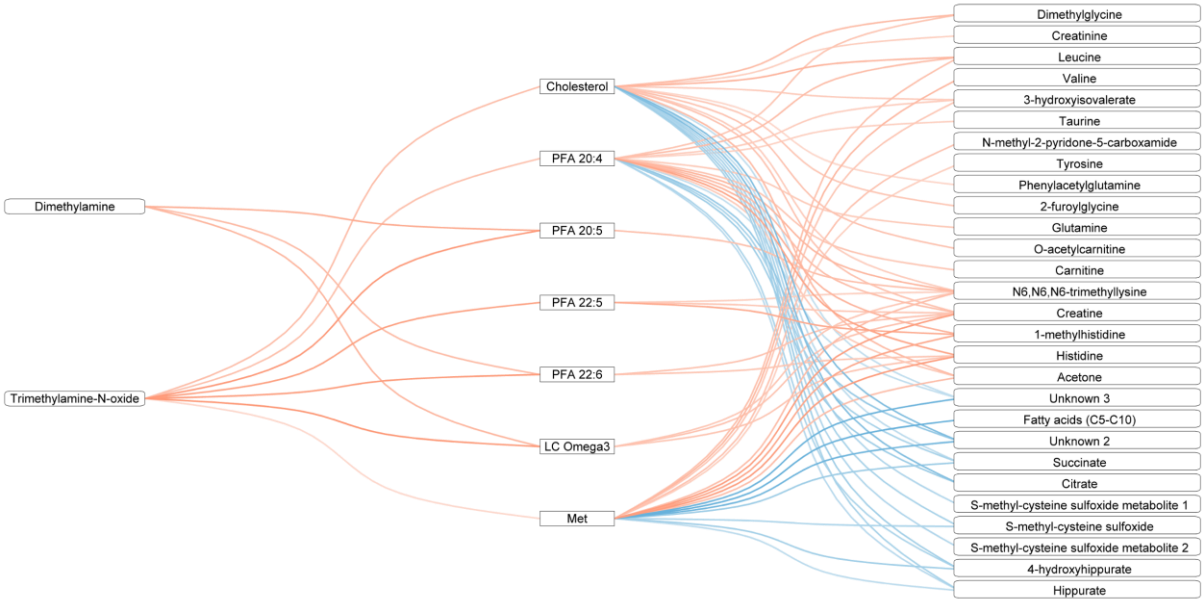
Supplementary Figure 6. Direct comparison of the partial correlations and adjusted coefficient of commonality (ACC). Comparison between partial correlation of all nutrients (from **Supplementary Figure 4**, shown on x-axis) and ACC of all nutrient pairs (from **Supplementary Figure 5**, shown on y-axis).



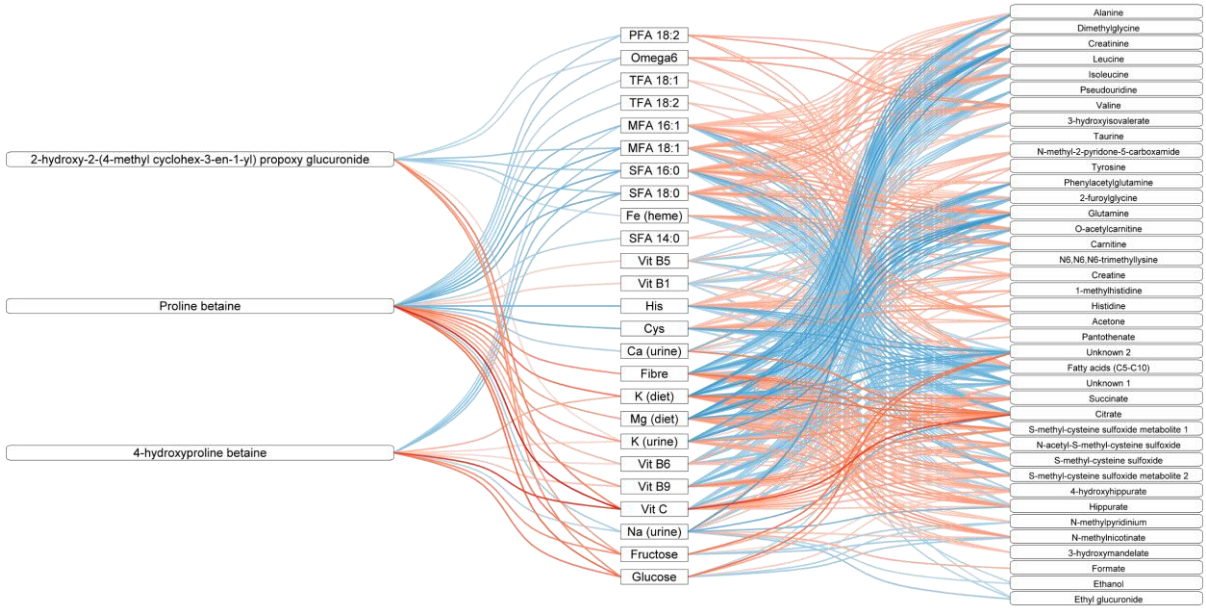
Supplementary Figure 7. The metabolic influence potential of 80 nutrients in the U.S. INTERMAP population (adjusted for BMI and physical activity in addition to age, sex and population sample as in Supplementary Figure 1). (**A**) Median ^{1}H-NMR spectrum of both visits. (**B**, **C**, **D**) The skyline projection of the significant associations ($-^{10}\log(q)\times\text{sign}(\rho)$, bounded to limits -10 and 10, where q is the false discovery rate q-value) for all nutrients for both visits combined (least significant visit shown for each variable); positive associations are shown in shades of red and inverse associations in shades of blue. (**B**) Nutrients that are energy providing (fatty acids, alcohol, sugars, fibre, starch), (**C**) dietary amino acids and (**D**) are essential for supporting metabolism (cholesterol, vitamins and derivatives, minerals).



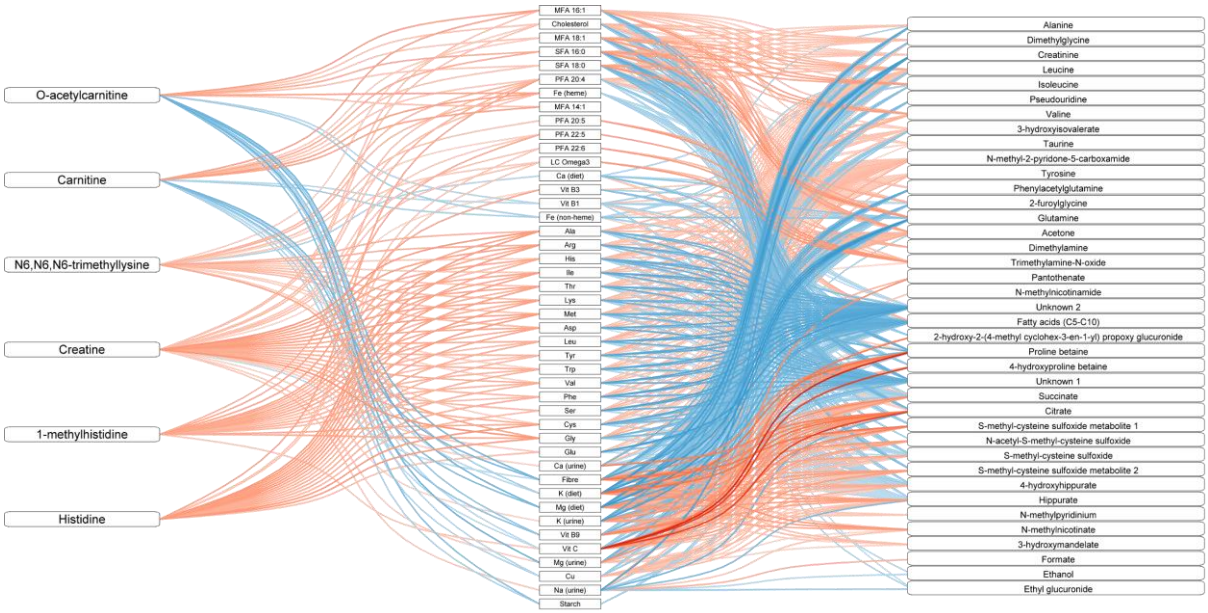
Supplementary Figure 8. Targeted hierarchical tripartite graph of connectivities of associations from cluster M9. Cluster M9 consists of urinary ethanol, ethyl glucuronide and acetate and it shows associated nutrients, and metabolites associated with those nutrients. These plots can be explored interactively using the NutriomeXplorer available free of charge.



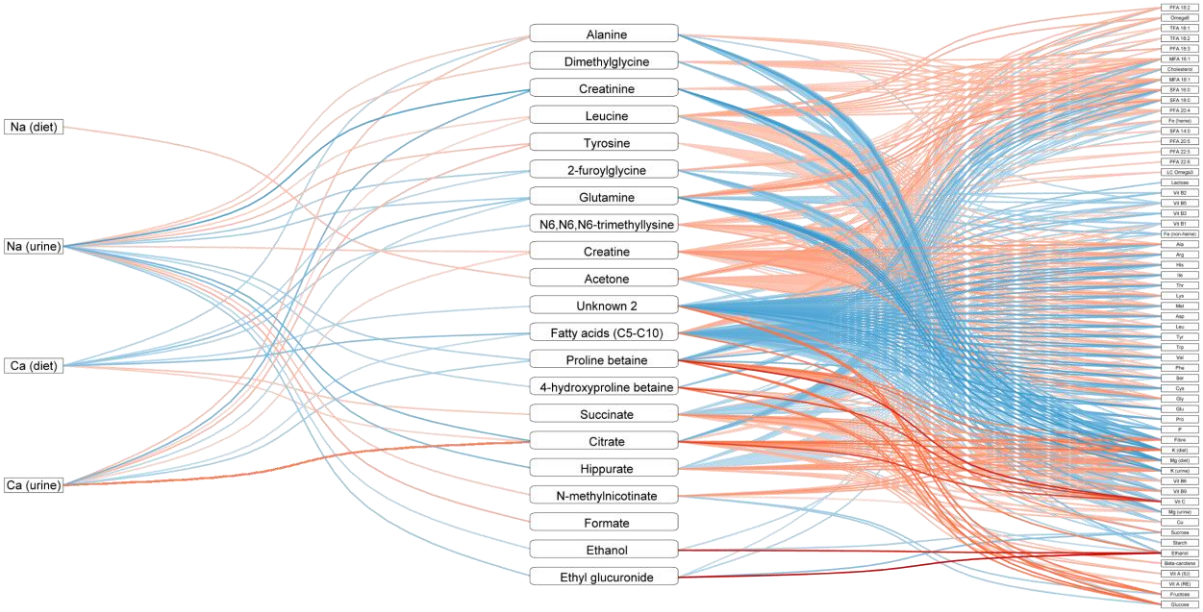
Supplementary Figure 9. Targeted hierarchical tripartite graph of connectivities of associations from cluster M4. Cluster M4 consists of urinary trimethylamine-*N*-oxide and dimethylamine and it shows associated nutrients, and metabolites associated with those nutrients. These plots can be explored interactively using the NutriomeXplorer available free of charge.



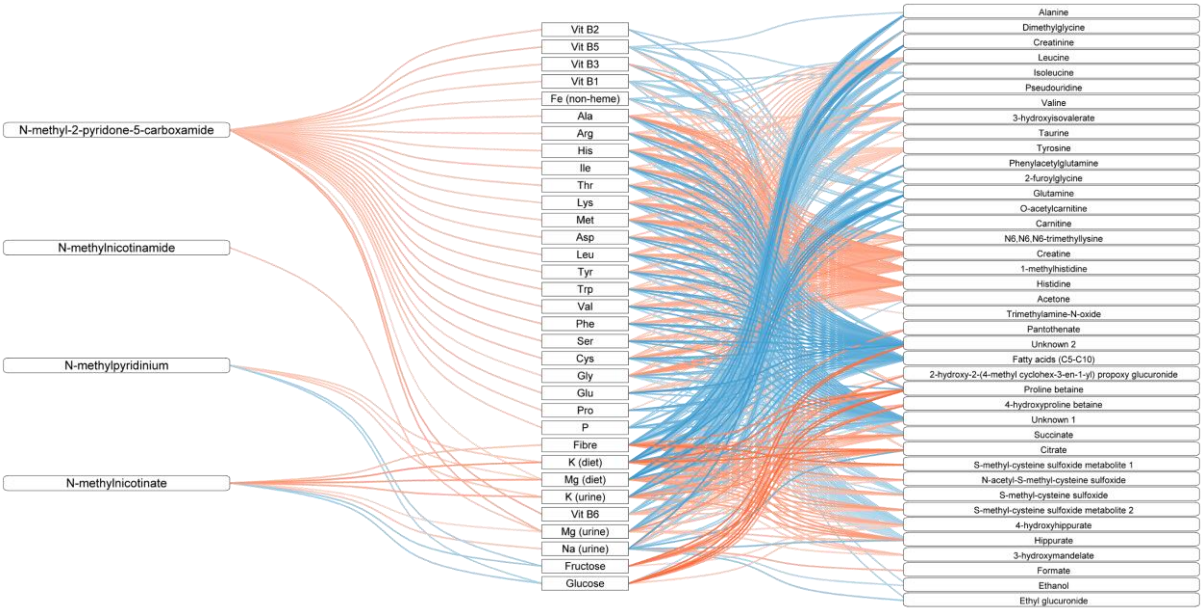
Supplementary Figure 10. Targeted hierarchical tripartite graph of connectivities of associations of citrus fruit metabolites. Citrus fruit metabolites include proline betaine, 4-hydroxyproline betaine and 2-hydroxy-2-(4-methyl cyclohex-3-en-1-yl)propoxyglucuronide and it shows associated nutrients, and metabolites associated with those nutrients. These plots can be explored interactively using the NutriomeXplorer available free of charge.



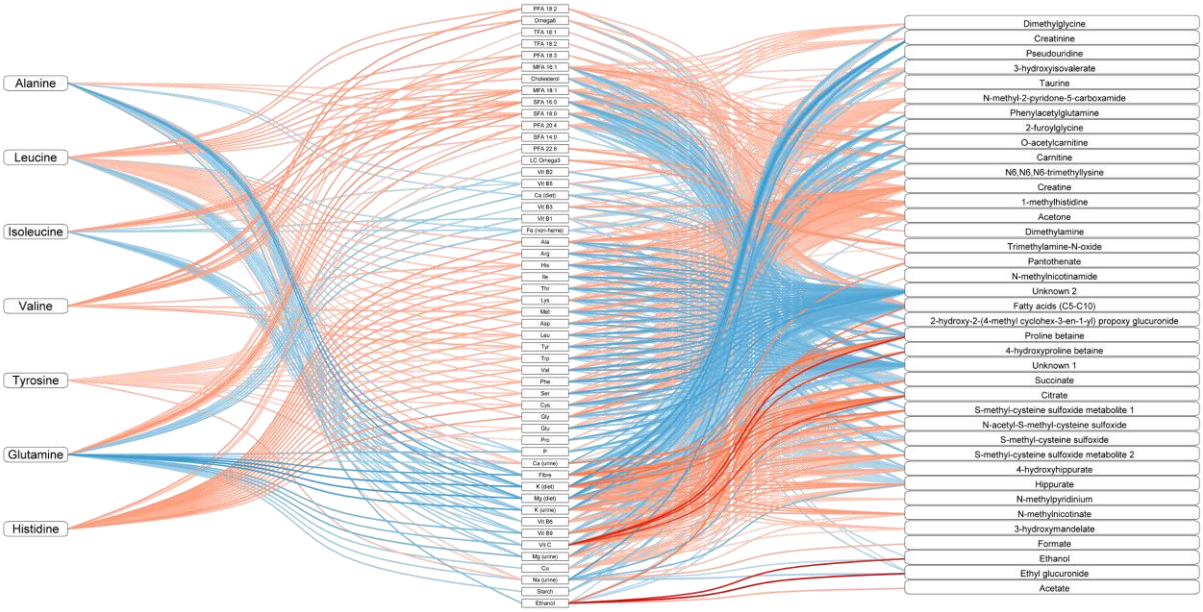
Supplementary Figure 11. Targeted hierarchical tripartite graph of connectivities of associations from cluster M3. Metabolites include those that related to meat intake (*O*-acetylcarnitine, carnitine, *N6,N6,N6*-trimethyllysine, creatine, 1-methylhistidine and histidine) and it shows associations with nutrients, and metabolites associated with those nutrients. These data can be explored interactively using the NutriomeXplorer available free of charge.



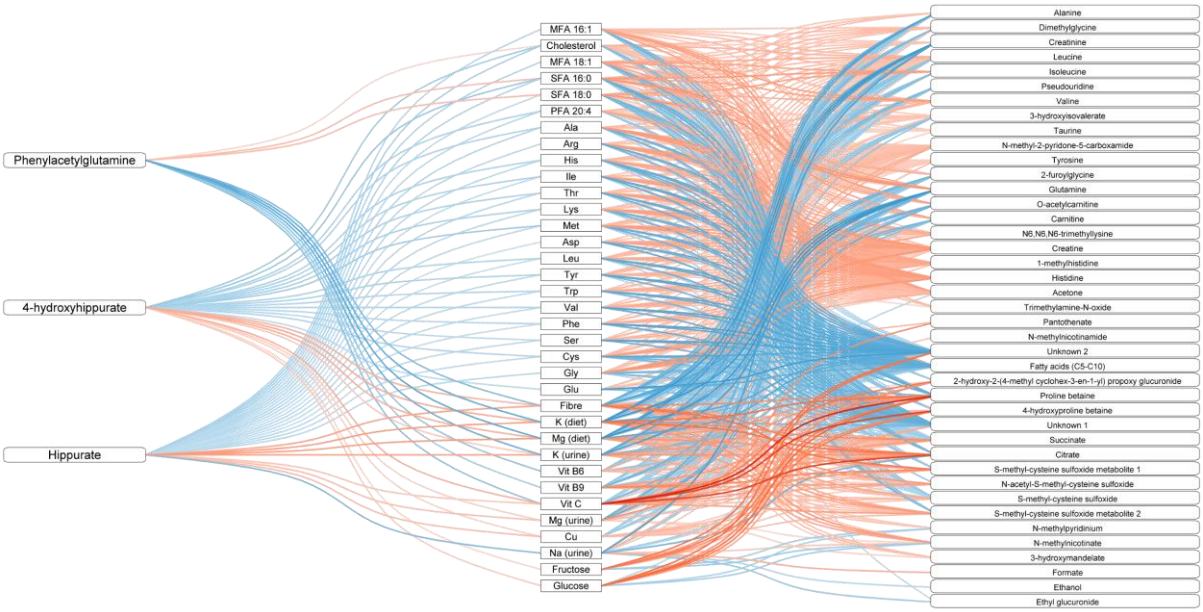
Supplementary Figure 12. Targeted hierarchical tripartite graph of connectivities of associations for dietary and urinary sodium and calcium. It shows associated urinary metabolites, and nutrients associated with those metabolites for dietary and urinary sodium and calcium. These plots can be explored interactively using the NutriomeXplorer available free of charge.



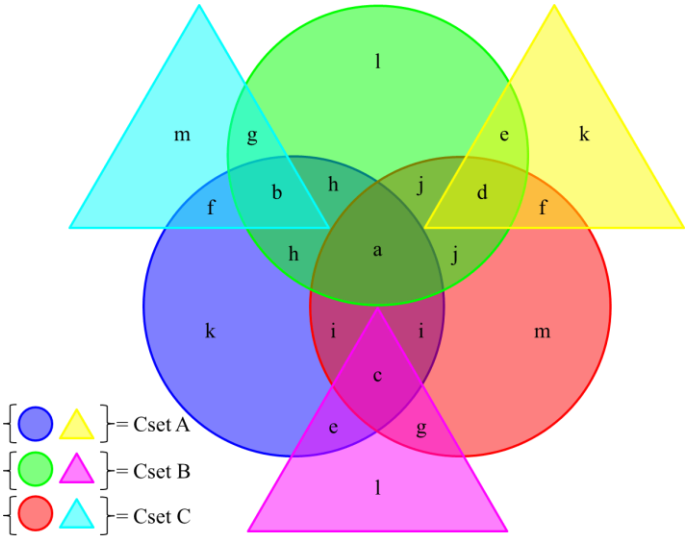
Supplementary Figure 13. Targeted hierarchical tripartite graph of connectivities of associations of niacin-related metabolites. Niacin-related metabolites (*N*-methyl-2-pyridone-5-carboxamide, *N*-methylniconamide, *N*-methylpyridinium and *N*-methylniconate) and their associated nutrients are shown, and metabolites associated with those nutrients. These data can be explored interactively using the NutriomeXplorer available free of charge.



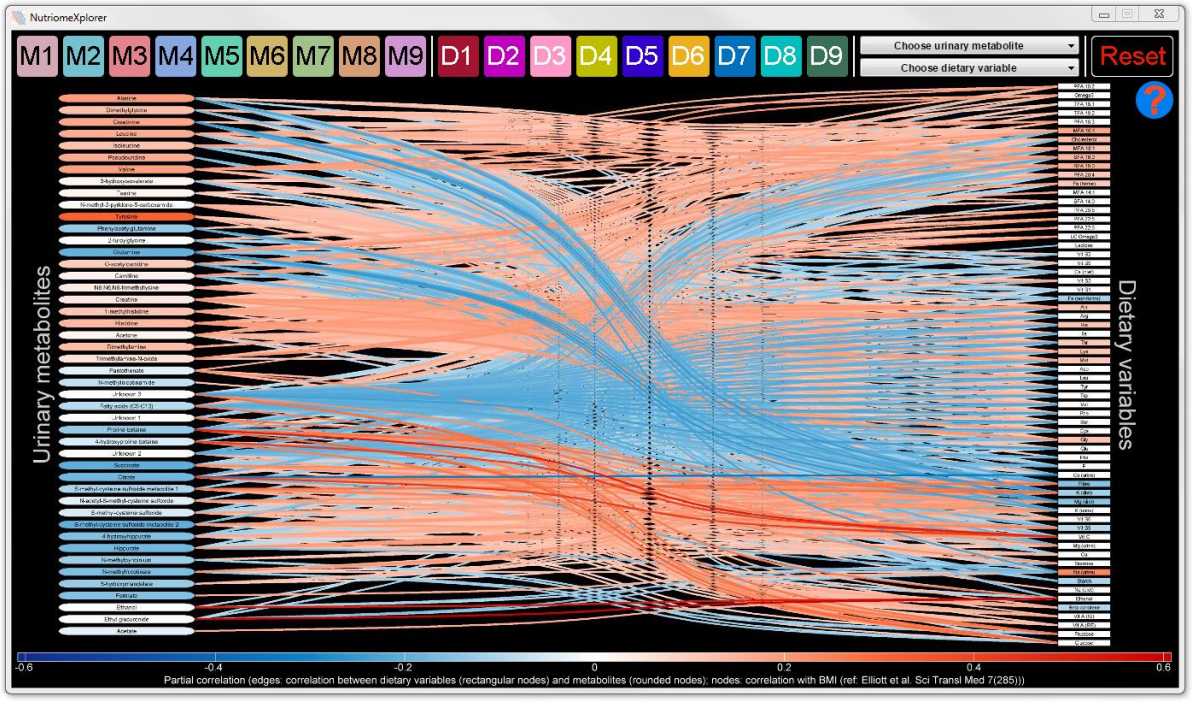
Supplementary Figure 14. Targeted hierarchical tripartite graph of connectivities of associations of urinary amino acids. Associations of urinary amino acids alanine, leucine, isoleucine, valine, tyrosine, glutamine and histidine with nutrients are shown, and metabolites associated with those nutrients. These plots can be explored interactively using the NutriomeXplorer available free of charge.



Supplementary Figure 15. Targeted hierarchical tripartite graph of connectivities of associations of gut microbial (co-)metabolites. It shows associations of phenylacetylglutamine, 4-hydroxyhippurate and hippurate with nutrients, and metabolites associated with those nutrients. These data can be explored interactively using the NutriomeXplorer available free of charge.



Supplementary Figure 16. Schematic representation of the overlap between 3 charged binary sets (csets A, B, C).



Supplementary Figure 17. Starting layout of the NutriomeXplorer software. Toolbar includes buttons to display different clusters of metabolites or nutrients as defined in the main text (**Figure 2**) and **Supplementary Information** (**Supplementary Figure 5**).

SUPPLEMENTARY TABLES

Supplementary Table 1. All nutrients with significantly and reproducibly associated urinary metabolites listed and the direction of association for each metabolite. The significance of each association (as shown in **Figure 2**) is indicated in the table by colours, with black indicating associations with $q<10^{-2}$, green indicating $q<10^{-5}$, orange indicating $q<10^{-7.5}$ and red indicating any association more significant than $q=10^{-10}$. 'Unidentified ^{1}H-NMR signals' indicates there were significant associations with multiple variables, however none could be (tentatively) identified. 'n.s.' indicates none of the 7,100 ^{1}H-NMR spectral variables were found to be associated with the nutrient.

Nutrient	Direction	Associated urinary metabolites
myristoleic acid (MFA 14:1) (%kcal)	↑	carnitine, O-acetylcarnitine
	↓	
palmitoleic acid (MFA 16:1) (%kcal)	↑	acetone, glutamine, O-acetylcarnitine, carnitine, Isoleucine, leucine, valine, creatine, dimethylglycine, creatinine, taurine, 2-furoylglycine
	↓	unknown (1.82m, 3.52m), citrate, *S*-methyl-cysteine sulfoxide metabolite (2.80), *S*-methyl-cysteine sulfoxide, proline betaine, *S*-methyl-cysteine sulfoxide metabolite (2.76), 4-hydroxyproline betaine, hippurate
oleic acid (MFA 18:1) (%kcal)	↑	isoleucine, valine, acetone, fatty acids (C5-C10), leucine, glutamine, dimethylglycine, *O*-acetylcarnitine, alanine, phenylacetylglutamine, creatinine, carnitine, 2-furoylglycine
	↓	proline betaine, citrate, 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, unknown (1.82m, 3.52m), 4-hydroxyproline betaine, *S*-methyl-cysteine sulfoxide metabolite (2.80), unknown (3.59s, 3.89, 4.25), hippurate
gadoleic acid (MFA 20:1) (%kcal) [1]	↑	acetone [1], histidine [1], leucine [1]
	↓	
erucic acid (MFA 22:1) (%kcal)	↑	n.s.
	↓	n.s.
linoleic acid (PFA 18:2) (%kcal)	↑	valine, fatty acids (C5-C10), leucine, acetone
	↓	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, proline betaine
linolenic acid (PFA 18:3) (%kcal)	↑	valine, fatty acids (C5-C10), leucine
	↓	
stearidonic acid (PFA 18:4) (%kcal) [1]	↑	trimethylamine-*N*-oxide [1]
	↓	
arachidonic acid (PFA 20:4) (%kcal)	↑	histidine, leucine, glutamine, dimethylglycine, creatine, *N6*,*N6*,*N6*-trimethyllysine, trimethylamine-*N*-oxide, 3-hydroxyisovalerate, acetone, taurine
	↓	unknown (1.82m, 3.52m), citrate, succinate, unknown (3.59s, 3.89, 4.25), hippurate, 4-hydroxyhippurate

eicosapentaenoic acid (PFA 20:5) (%kcal)	↑	trimethylamine-*N*-oxide, dimethylamine, *N6*,*N6*,*N6*-trimethyllysine
	↓	
docosapentaenoic acid (PFA 22:5) (%kcal)	↑	trimethylamine-*N*-oxide, histidine, creatine, *N6*,*N6*,*N6*-trimethyllysine
	↓	
docosahexaenoic acid (PFA 22:6) (%kcal)	↑	trimethylamine-N-oxide, dimethylamine, creatine, histidine
	↓	
omega-6 PFA (%kcal)	↑	valine, fatty acids (C5-C10), leucine, acetone
	↓	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, proline betaine
long chain omega-3 PFA (%kcal)	↑	trimethylamine-N-oxide, dimethylamine, creatine, *N6*,*N6*,*N6*-trimethyllysine, histidine
	↓	
butyric acid (SFA 4:0) (%kcal)	↑	unidentified ^{1}H-NMR signals
	↓	unidentified ^{1}H-NMR signals
caproic acid (SFA 6:0) (%kcal) [1,2]	↑	unidentified ^{1}H-NMR signals [1,2]
	↓	
caprylic acid (SFA 8:0) (%kcal) [1]	↑	fatty acids (C5-C10) [1]
	↓	
capric acid (SFA 10:0) (%kcal)	↑	unidentified ^{1}H-NMR signals
	↓	unidentified ^{1}H-NMR signals
lauric acid (SFA 12:0) (%kcal)	↑	n.s.
	↓	n.s.
myristic acid (SFA 14:0) (%kcal)	↑	isoleucine, glutamine
	↓	proline betaine
palmitic acid (SFA 16:0) (%kcal)	↑	isoleucine, valine, glutamine, *O*-acetylcarnitine, carnitine, fatty acids (C5-C10), leucine, alanine, phenylacetylglutamine, acetone, dimethylglycine, creatinine, 2-furoylglycine, taurine
	↓	unknown (1.82m, 3.52m), citrate, proline betaine, *S*-methyl-cysteine sulfoxide metabolite (2.80), 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, 4-hydroxyproline betaine, 4-hydroxyhippurate, unknown (3.59s, 3.89, 4.25), hippurate
stearic acid (SFA 18:0) (%kcal)	↑	isoleucine, glutamine, fatty acids (C5-C10), valine, 2-furoylglycine, leucine, alanine, phenylacetylglutamine, dimethylglycine, *O*-acetylcarnitine, carnitine, acetone, creatinine
	↓	proline betaine, unknown (1.82m, 3.52m), citrate, *S*-methyl-cysteine sulfoxide metabolite (2.80), 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, 4-hydroxyproline betaine, hippurate

arachidic acid (SFA 20:0) (%kcal) [1,2]	↑	unidentified ^{1}H-NMR signals [1,2]
	↓	
behenic acid (SFA 22:0) (%kcal) [1,2]	↑	unidentified ^{1}H-NMR signals [1], 4-hydroxyhippurate [2]
	↓	
trans-octadecenoic acid (TFA 18:1) (%kcal)	↑	fatty acids (C5-C10), glutamine
	↓	proline betaine
trans-octadecadienoic acid (TFA 18:2) (%kcal)	↑	fatty acids (C5-C10), glutamine
	↓	proline betaine
alcohol (%kcal)	↑	fatty acids (C5-C10), ethanol, ethyl glucuronide, acetate
	↓	Alanine, creatinine, unknown (3.59s, 3.89, 4.25)
fructose (%kcal)	↑	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, proline betaine, 4-hydroxyproline betaine, unknown (3.59s, 3.89, 4.25), unknown (1.82m, 3.52m), 4-hydroxyhippurate
	↓	*N*-methylpyridinium, *N*-methylnicotinate
galactose (%kcal) [1]	↑	unidentified ^{1}H-NMR signals [1]
	↓	
glucose (%kcal)	↑	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, unknown (1.82m, 3.52m), proline betaine, 4-hydroxyproline betaine, unknown (3.59s, 3.89, 4.25), 4-hydroxyhippurate
	↓	acetone, *N*-methylpyridinium, *N*-methylnicotinate
lactose (%kcal)	↑	
	↓	fatty acids (C5-C10)
maltose (%kcal)	↑	unidentified ^{1}H-NMR signals
	↓	unidentified ^{1}H-NMR signals
sucrose (%kcal)	↑	unknown (3.59s, 3.89, 4.25), unknown (1.82m, 3.52m)
	↓	ethyl glucuronide, ethanol, acetone
fibre (g/1000kcal)	↑	succinate, citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide metabolite (2.80), *S*-methyl-cysteine sulfoxide, hippurate, proline betaine, 4-hydroxyhippurate, *N*-methylnicotinate
	↓	alanine, glutamine, creatinine, *O*-acetylcarnitine, carnitine, 2-furoylglycine, phenylacetylglutamine, dimethylglycine, fatty acids (C5-C10), isoleucine, leucine, 3-hydroxyisovalerate, unknown (3.59s, 3.89, 4.25), pseudouridine, creatine
starch (%kcal)	↑	
	↓	*O*-acetylcarnitine, acetone, glutamine, carnitine
alanine (%kcal)	↑	creatine, histidine, acetone, *N6*,*N6*,*N6*-trimethyllysine, leucine, 3-hydroxyisovalerate, taurine, 1-methylhistidine, *N*-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate

arginine (%kcal)	↑	creatine, *N6*,*N6*,*N6*-trimethyllysine, histidine, acetone, leucine, 3-hydroxyisovalerate, taurine, 1-methylhistidine, *N*-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate
aspartic acid (%kcal)	↑	creatine, histidine, acetone, *N6*,*N6*,*N6*-trimethyllysine, leucine, 1-methylhistidine, *N*-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), hippurate
cysteine (%kcal)	↑	creatine, histidine, leucine, *N*-methyl-2-pyridone-5-carboxamide, 3-hydroxyisovalerate, acetone, *N6*,*N6*,*N6*-trimethyllysine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, proline betaine, hippurate, 4-hydroxyhippurate
glutamic acid (%kcal)	↑	histidine, *N*-methyl-2-pyridone-5-carboxamide, acetone, creatine, tyrosine
	↓	fatty acids (C5-C10), unknown (3.59s, 3.89, 4.25), unknown (1.82m, 3.52m), 4-hydroxyhippurate
glycine (%kcal)	↑	creatine, *N6*,*N6*,*N6*-trimethyllysine, histidine, acetone, leucine, 3-hydroxyisovalerate, taurine, *N*-methyl-2-pyridone-5-carboxamide
	↓	unknown (3.59s, 3.89, 4.25), fatty acids (C5-C10), unknown (1.82m, 3.52m), succinate, hippurate
histidine (%kcal)	↑	creatine, histidine, leucine, acetone, 3-hydroxyisovalerate, *N6*,*N6*,*N6*-trimethyllysine, *N*-methyl-2-pyridone-5-carboxamide, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, proline betaine, hippurate, 4-hydroxyhippurate
isoleucine (%kcal)	↑	creatine, histidine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, *N6*,*N6*,*N6*-trimethyllysine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate
leucine (%kcal)	↑	creatine, histidine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate
lysine (%kcal)	↑	creatine, histidine, leucine, acetone, *N6*,*N6*,*N6*-trimethyllysine, 3-hydroxyisovalerate, 1-methylhistidine, *N*-methyl-2-pyridone-5-carboxamide, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate

methionine (%kcal)	↑	creatine, histidine, leucine, acetone, *N6*,*N6*,*N6*-trimethyllysine, *N*-methyl-2-pyridone-5-carboxamide, 3-hydroxyisovalerate, trimethylamine-N-oxide, 1-methylhistidine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, 4-hydroxyhippurate, hippurate
phenylalanine (%kcal)	↑	histidine, creatine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate
proline (%kcal)	↑	*N*-methyl-2-pyridone-5-carboxamide, tyrosine
	↓	fatty acids (C5-C10), unknown (3.59s, 3.89, 4.25) , unknown (1.82m, 3.52m)
serine (%kcal)	↑	creatine, histidine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), hippurate, 4-hydroxyhippurate
threonine (%kcal)	↑	creatine, histidine, leucine, acetone, *N6*,*N6*,*N6*-trimethyllysine, *N*-methyl-2-pyridone-5-carboxamide, 3-hydroxyisovalerate, 1-methylhistidine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate
tryptophan (%kcal)	↑	histidine, creatine, *N*-methyl-2-pyridone-5-carboxamide, leucine, acetone, *N6*,*N6*,*N6*-trimethyllysine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), succinate, hippurate, 4-hydroxyhippurate
tyrosine (%kcal)	↑	histidine, creatine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), hippurate, 4-hydroxyhippurate
valine (%kcal)	↑	histidine, creatine, acetone, *N*-methyl-2-pyridone-5-carboxamide, leucine, *N6*,*N6*,*N6*-trimethyllysine, tyrosine
	↓	fatty acids (C5-C10), unknown (1.82m, 3.52m), unknown (3.59s, 3.89, 4.25), hippurate, 4-hydroxyhippurate
cholesterol (mg/1000kcal)	↑	leucine, acetone, glutamine, dimethylglycine, trimethylamine-*N*-oxide, 3-hydroxyisovalerate, phenylacetylglutamine, creatine, creatinine, *N6*,*N6*,*N6*-trimethyllysine, *O*-acetylcarnitine, 2-furoylglycine
	↓	unknown (1.82m, 3.52m), citrate, succinate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *S*-methyl-cysteine sulfoxide metabolite (2.80), *S*-methyl-cysteine sulfoxide, proline betaine, unknown (3.59s, 3.89, 4.25), hippurate, 4-hydroxyhippurate
beta-carotene (µg/1000kcal)	↑	citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, succinate, *S*-methyl-cysteine sulfoxide metabolite (2.80)

	↓	
vitamin A (IU/1000kcal)	↑	citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76) , *S*-methyl-cysteine sulfoxide, succinate, *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide metabolite (2.80)
	↓	
retinol (µg/1000kcal)	↑	n.s.
	↓	n.s.
vitamin A (RE/1000kcal)	↑	*S*-methyl-cysteine sulfoxide metabolite (2.76), succinate, citrate, *S*-methyl-cysteine sulfoxide metabolite (2.80), *S*-methyl-cysteine sulfoxide
	↓	
thiamin (B1) (mg/1000kcal)	↑	succinate, citrate, proline betaine, *N*-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), isoleucine, ethanol, ethyl glucuronide, glutamine, *O*-acetylcarnitine, carnitine, 2-furoylglycine
riboflavin (B2) (mg/1000kcal)	↑	*N*-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), glutamine, unknown (3.59s, 3.89, 4.25)
niacin (B3) (mg/1000kcal)	↑	*N*-methyl-2-pyridone-5-carboxamide
	↓	unknown (3.59s, 3.89, 4.25), fatty acids (C5-C10), glutamine
pantothenic acid (B5) (mg/1000kcal)	↑	*N*-methyl-2-pyridone-5-carboxamide, citrate, proline betaine
	↓	fatty acids (C5-C10), isoleucine, alanine, glutamine, unknown (3.59s, 3.89, 4.25), 2-furoylglycine
pyridoxal 5-phosphate (B6) (mg/1000kcal)	↑	citrate, *N*-methyl-2-pyridone-5-carboxamide, succinate, proline betaine, *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide metabolite (2.80), 4-hydroxyproline betaine
	↓	glutamine, fatty acids (C5-C10), isoleucine, alanine, phenylacetylglutamine, 2-furoylglycine, leucine, unknown (3.59s, 3.89, 4.25), pseudouridine
folate (B9) (µg/1000kcal)	↑	citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, proline betaine, succinate, *S*-methyl-cysteine sulfoxide metabolite (2.80), *S*-methyl-cysteine sulfoxide, 4-hydroxyproline betaine
	↓	glutamine, carnitine, 2-furoylglycine, alanine, creatinine, *O*-acetylcarnitine, fatty acids (C5-C10), isoleucine, leucine, dimethylglycine, pseudouridine
cobalamin (B12) (mg/1000kcal) [1]	↑	trimethylamine-*N*-oxide [1]
	↓	
vitamin C (mg/1000kcal)	↑	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, succinate, citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *S*-methyl-cysteine sulfoxide metabolite (2.80), proline betaine, 4-hydroxyproline betaine, *N*-acetyl-*S*-methyl-cysteine sulfoxide, 4-hydroxyhippurate, *S*-methyl-cysteine sulfoxide, hippurate

	↓	glutamine, phenylacetylglutamine, 2-furoylglycine, fatty acids (C5-C10), leucine, alanine, pseudouridine, dimethylglycine, creatinine, *O*-acetylcarnitine, carnitine
vitamin E (mg/1000kcal) [1,2]	↑	fatty acids (C5-C10) [1], hippurate [1], unidentified [1]H-NMR signals [2]
	↓	*O*-acetylcarnitine [1], carnitine [1]
calcium (mg/1000kcal)	↑	succinate, citrate
	↓	fatty acids (C5-C10), alanine, glutamine, unknown (3.59s, 3.89, 4.25), 2-furoylglycine
calcium (mmol/24-hr)	↑	citrate, alanine, tyrosine, pantothenate
	↓	creatinine, 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide
copper (mg/1000kcal)	↑	citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *S*-methyl-cysteine sulfoxide metabolite (2.80), proline betaine, succinate, *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, hippurate
	↓	glutamine, creatinine, *O*-acetylcarnitine, carnitine, 2-furoylglycine, pseudouridine
iron (heme) (mg/1000kcal)	↑	*O*-acetylcarnitine, carnitine, *N6*,*N6*,*N6*-trimethyllysine, creatine
	↓	fatty acids (C5-C10), 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, unknown (1.82m, 3.52m), succinate, citrate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *S*-methyl-cysteine sulfoxide
iron (non-heme) (mg/1000kcal)	↑	citrate, *N*-methyl-2-pyridone-5-carboxamide
	↓	glutamine, fatty acids (C5-C10), isoleucine, leucine, *O*-acetylcarnitine, carnitine
magnesium (mg/1000kcal)	↑	succinate, citrate, *S*-methyl-cysteine sulfoxide metabolite (2.80), hippurate, *N*-methylnicotinate, 4-hydroxyhippurate, *S*-methyl-cysteine sulfoxide metabolite (2.76), *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, proline betaine, 3-hydroxymandelate, *N*-methylpyridinium, N-methyl-2-pyridone-5-carboxamide
	↓	fatty acids (C5-C10), alanine, phenylacetylglutamine, glutamine, dimethylglycine, creatinine, *O*-acetylcarnitine, carnitine, unknown (3.59s, 3.89, 4.25), 2-furoylglycine, pseudouridine, isoleucine, leucine, 3-hydroxyisovalerate
magnesium (mmol/24-hr)	↑	pantothenate, *N*-methyl-2-pyridone-5-carboxamide, proline betaine, succinate, hippurate, *N*-methylnicotinamide, 4-hydroxyhippurate, *N*-methylnicotinate, 3-hydroxymandelate
	↓	glutamine, creatinine, leucine, 3-hydroxyisovalerate, alanine, pseudouridine, fatty acids (C5-C10), *O*-acetylcarnitine, dimethylglycine, unknown (3.59s, 3.89, 4.25), 2-furoylglycine
phosphorus (mg/1000kcal)	↑	
	↓	fatty acids (C5-C10), unknown (3.59s, 3.89, 4.25), alanine, isoleucine, glutamine, 2-furoylglycine

potassium (mg/1000kcal)	↑	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, proline betaine, succinate, citrate, S-methyl-cysteine sulfoxide metabolite (2.76), *S*-methyl-cysteine sulfoxide metabolite (2.80), hippurate, *N*-methylnicotinate, *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide, 4-hydroxyproline betaine, 4-hydroxyhippurate, 3-hydroxymandelate, *N*-methylpyridinium
	↓	fatty acids (C5-C10), alanine, phenylacetylglutamine, glutamine, dimethylglycine, creatinine, unknown (3.59s, 3.89, 4.25), 2-furoylglycine, pseudouridine, leucine, isoleucine, 3-hydroxyisovalerate, *O*-acetylcarnitine, carnitine
potassium (mmol/24-hr)	↑	proline betaine, citrate, hippurate, *N*-methylnicotinate, succinate, 4-hydroxyhippurate, 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide, *S*-methyl-cysteine sulfoxide metabolite (2.80), 4-hydroxyproline betaine, *N*-methylpyridinium, 3-hydroxymandelate
	↓	fatty acids (C5-C10), glutamine, creatinine, isoleucine, 3-hydroxyisovalerate, unknown (3.59s, 3.89, 4.25), pseudouridine, leucine, alanine, 2-furoylglycine, phenylacetylglutamine, dimethylglycine
selenium (µg/1000kcal) [2]	↑	creatine [2], histidine [2], acetone [2]
	↓	unknown (3.59s, 3.89, 4.25) [2]
sodium (mg/1000kcal)	↑	acetone
	↓	
sodium (mmol/24-hr)	↑	formate, tyrosine, alanine, dimethylglycine, *N*-methylnicotinate
	↓	creatinine, hippurate, ethyl glucuronide, ethanol, phenylacetylglutamine, proline betaine, 2-furoylglycine, 4-hydroxyproline betaine

[1] Indicates it was only found significant in the first visit data.

[2] Indicates it was only found significant in the second visit data.

Supplementary Table 2. A list of identified metabolites, their significant chemical shifts, multiplicities and associated nutrients. Unknown metabolites are only included if the SubseT Optimization by Reference Matching[1] (STORM) analysis showed clear structural correlations. The significance of each association (as shown in **Figure 2**) is indicated in the table by colours, with black indicating associations with $q<10^{-2}$, green indicating $q<10^{-5}$, orange indicating $q<10^{-7.5}$ and red indicating $q<10^{-10}$.

Metabolite	Chemical shift (multiplicity)		Associated nutrients
fatty acids (C5-C10)	0.86 (m), 1.27-1.33 (ms), 1.58 (m), 2.19 (m)	↑	ethanol, MFA 18:1, PFA 18:2, omega-6, SFA 16:0, SFA 18:0, PFA 18:3, TFA 18:1, TFA 18:2
		↓	Ala, Arg, Asp, Cys, Glu, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val, K (diet), K (urine), Mg (diet) , P, Gly, Ca (diet), lactose, fibre, vit B2, vit B6, vit C, vit B1, vit B3, vit B5, vit B9, Fe (heme), Fe (non-heme), Mg (urine)
pantothenate	0.90 (s), 0.93 (s)	↑	Mg (urine), Ca (urine)
isoleucine	0.94 (t), 1.01 (d)	↑	MFA 18:1, SFA 16:0, SFA 18:0, MFA 16:1, SFA 14:0
		↓	K (urine), Mg (diet), fibre, vit B6, K (diet), vit B1, vit B5, vit B9, Fe (non-heme), P
leucine	0.96 (2d)	↑	MFA 18:1, PFA 18:2, PFA 18:3, omega-6, MFA 16:1, PFA 20:4, SFA 16:0, SFA 18:0, Ala, Cys, Gly, His, Lys, Met, Thr, cholesterol, Arg, Asp, Ile, Leu, Phe, Ser, Trp, Tyr, Val
		↓	K (diet), Mg (diet), fibre, vit C, K (urine), Mg (urine) , vit B6, vit B9, Fe (non-heme)
valine	0.99 (d), 1.04 (d)	↑	MFA 18:1, PFA 18:2, PFA 18:3, omega-6, SFA 16:0, SFA 18:0, MFA 16:1
2-hydroxy-2-(4-methyl cyclohex-3-en-1-yl) propoxy glucuronide	1.16 (s), 1.67 (s), 3.52 (2s)	↑	fructose, glucose, vit C, K (diet), K (urine)
		↓	MFA 18:1, SFA 16:0, SFA 18:0, PFA 18:2, omega-6, Fe (heme), Ca (urine)
ethanol	1.18 (t), 3.65 (q)	↑	ethanol
		↓	sucrose, vit B1, Na (urine)
ethyl glucuronide	1.23 (t), 4.48 (d)	↑	ethanol
		↓	sucrose, Na (urine), vit B1
3-hydroxyisovalerate	1.27 (s)	↑	Ala, Gly, PFA 20:4, Arg, Cys, His, Lys, Met, Thr, cholesterol
		↓	K (urine), fibre, K (diet), Mg (diet), Mg (urine)
alanine	1.48 (d)	↑	SFA 16:0, SFA 18:0, MFA 18:1, Ca (urine)
		↓	fibre, K (diet), Mg (diet), ethanol, vit B6, vit B9, vit C, Ca (diet), K (urine), Mg (urine), P, vit B5, Na (urine)
unknown (1)	1.82 (m), 3.52 (s)	↑	glucose, fructose, sucrose
		↓	MFA 16:1, SFA 16:0, Ala, Arg, Asp, Cys, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, Val, SFA 18:0, Glu, Gly, Cholesterol, MFA 18:1, PFA 20:4, Pro, Fe (heme)
acetate	1.92 (s)	↑	ethanol
phenylacetylglutamine	1.93 (m), 2.13 (m), 2.27 (t), 3.68 (q), 4.19 (dd), 7.36 (t), 7.43 (t)	↑	SFA 16:0, SFA 18:0, MFA 18:1, cholesterol
		↓	K (diet), Mg (diet), fibre, vit C, vit B6, K (urine), Na (urine)

Metabolite	Chemical shift (multiplicity)		Associated nutrients
glutamine	2.14 (m), 2.46 (m)	↑	MFA 16:1, SFA 16:0, SFA 18:0, MFA 18:1, PFA 20:4, cholesterol, SFA 14:0, TFA 18:1, TFA 18:2
		↓	fibre, vit B9, vit C, K (diet), K (urine), Mg (diet), Mg (urine), vit B6, Cu, Fe (non-heme), starch, vit B1, vit B2, vit B3, vit B5, Ca (diet), P
O-acetylcarnitine	2.15 (s), 3.19 (s)	↑	MFA 16:1, SFA 16:0, Fe (heme), MFA 18:1, SFA 18:0, MFA 14:1, cholesterol
		↓	fibre, Mg (diet), starch, vit B9, vit B1, vit C, Cu, Fe (non-heme), K (diet), Mg (urine)
acetone	2.24 (s)	↑	MFA 16:1, MFA 18:1, Ala, Gly, SFA 16:0, Arg, Asp, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Tyr, Val, cholesterol, PFA 18:2, PFA 20:4, omega-6, SFA 18:0, Cys, Glu, Trp, Na (diet)
		↓	glucose, sucrose, starch
proline betaine	2.30 (m), 2.50 (m), 3.11 (s), 3.30 (s), 4.10 (m)	↑	fructose, glucose, vit B9, vit C, K (diet), K (urine), fibre, vit B6, Cu, Mg (diet), Mg (urine), vit B1, vit B5
		↓	MFA 18:1, SFA 16:0, SFA 18:0, MFA 16:1, PFA 18:2, omega-6, SFA 14:0, TFA 18:1, TFA 18:2, Cys, His, cholesterol, Na (urine)
succinate	2.41 (s)	↑	fibre, vit C, K (diet), Mg (diet), vit B9, K (urine), vit B6, Mg (urine), beta-carotene, vit A (IU), vit A (RE), vit B1, Ca (diet), Cu
		↓	Ala, Gly, Lys, Met, PFA 20:4, Arg, Cys, His, Ile, Leu, Phe, Thr, Trp, cholesterol, Fe (heme)
citrate	2.54 (d), 2.65 (d)	↑	fibre, vit B9, vit C, Ca (urine), K (diet), K (urine), Mg (diet), beta-carotene, vit A (IU), vit B6, Cu, vit A (RE), vit B1, vit B5, Ca (diet), Fe (non-heme)
		↓	SFA 16:0, MFA 16:1, MFA 18:1, SFA 18:0, cholesterol, PFA 20:4, Fe (heme)
dimethylamine	2.72 (s)	↑	PFA 20:5, PFA 22:6, long chain omega-3
S-methyl-cysteine sulfoxide metabolite	2.76 (s)	↑	fibre, vit B9, vit C, K (diet), beta-carotene, vit A (IU) , Mg (diet), vit A (RE), Cu
		↓	SFA 16:0, SFA 18:0, MFA 16:1, cholesterol, Fe (heme)
N-acetyl-S-methyl-cysteine sulfoxide	2.78 (s)	↑	fibre, vit B9, vit C, K (diet), beta-carotene, Mg (diet), vit A (IU), vit B6, Cu
		↓	SFA 16:0, SFA 18:0
S-methyl-cysteine sulfoxide metabolite	2.80 (s)	↑	fibre, vit C, K (diet) , Mg (diet), vit B9, Cu, beta-carotene, vit A (IU), vit A (RE), vit B6, K (urine)
		↓	SFA 16:0, SFA 18:0, MFA 16:1, MFA 18:1, cholesterol
S-methyl-cysteine sulfoxide	2.83 (s)	↑	fibre, K (diet), beta-carotene, vit A (IU), vit B9, vit C, Mg (diet), vit A (RE), Cu
		↓	MFA 16:1, SFA 16:0, SFA 18:0, cholesterol, Fe (heme)

Metabolite	Chemical shift (multiplicity)		Associated nutrients
dimethylglycine	2.93 (s)	↑	MFA 18:1, PFA 20:4, SFA 16:0, SFA 18:0, cholesterol, MFA 16:1, Na (urine)
		↓	K (diet), Mg (diet), fibre, vit B9, vit C, K (urine), Mg (urine)
creatine	3.05 (s), 3.93 (s)	↑	Ala, Arg, Asp, Cys, Gly, His, Ile, Leu, Lys, Met, Thr, Phe, Ser, Trp, Tyr, Val, MFA 16:1, PFA 20:4, PFA 22:5, PFA 22:6, long chain omega-3, Glu, cholesterol, Fe (heme)
		↓	fibre
creatinine	3.06 (s), 4.06 (s)	↑	SFA 16:0, MFA 16:1, MFA 18:1, SFA 18:0, cholesterol
		↓	fibre, Ca (urine), K (diet), K (urine), Mg (diet), Mg (urine), Na (urine) , vit B9, Cu, ethanol, vit C
N6,*N6*,*N6*-trimethyllysine	3.12 (s)	↑	Arg, Gly, Ala, PFA 20:4, Asp, Lys, Met, Thr, Fe (heme), PFA 20:5, PFA 22:5, long chain omega-3, Cys, His, Ile, Trp, Val, cholesterol
		↓	Ca (diet)
histidine	3.14 (dd), 3.25 (dd), 4.00 (dd), 7.09 (s), 7.84 (s)	↑	PFA 20:4, Ala, Arg, Asp, Cys, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr, Val, Ser, PFA 22:5, Glu, PFA 22:6, long chain omega-3
1-methylhistidine	3.21 (2d), 3.29 (2d), 3.69 (s), 3.92 (t), 7.02 (s), 7.85 (s)	↑	Ala, Arg, Asp, Lys, Met, Thr
carnitine	2.44 (dd), 3.23 (s), 3.43 (m)	↑	MFA 14:1, MFA 16:1, SFA 16:0, Fe (heme), SFA 18:0, MFA 18:1
		↓	fibre, Mg (diet), vit B9, starch, vit B1, vit C, Cu, Fe (non-heme), K (diet)
taurine	3.26 (t), 3.44 (t)	↑	MFA 16:1, PFA 20:4, SFA 16:0, Ala, Arg, Gly
trimethylamine-*N*-oxide	3.27 (s)	↑	PFA 20:5, PFA 22:5, PFA 22:6, long chain omega-3, PFA 20:4, cholesterol, Met
4-hydroxyproline betaine	3.34 (s)	↑	glucose, fructose, vit C, K (diet), vit B9, vit B6, K (urine)
		↓	MFA 18:1, SFA 16:0, SFA 18:0, MFA 16:1, Na (urine)
unknown (2)	3.59 (s), 3.89, 4.25	↑	fructose, glucose, sucrose
		↓	Ala, Arg, Asp, Cys, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val, K (diet), Mg (diet), P, vit B3, K (urine), fibre, MFA 18:1, PFA 20:4, SFA 16:0, ethanol, cholesterol, vit B2, vit B5, vit B6, Ca (diet), Mg (urine)
4-hydroxyhippurate	3.94 (s), 6.97 (d), 7.77 (d)	↑	fibre, vit C, K (diet), Mg (diet), K (urine), fructose, glucose, Mg (urine)
		↓	SFA 16:0, Met, PFA 20:4, Ala, Cys, Glu, His, Ile, Leu, Lys, Phe, Ser, Thr, Trp, Tyr, Val, cholesterol
hippurate	3.98 (d), 7.55 (t), 7.65 (t), 7.84 (d)	↑	fibre, K (diet), K (urine), Mg (diet), vit C, Cu, Mg (urine)
		↓	Na (urine), MFA 16:1, MFA 18:1, PFA 20:4, SFA 16:0, SFA 18:0, Ala, Arg, Asp, Cys, Gly, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, Val, cholesterol

Metabolite	Chemical shift (multiplicity)		Associated nutrients
N-methylpyridinium	4.40 (s), 8.79 (d)	↑	K (diet), K (urine), Mg (diet)
		↓	fructose, glucose
N-methylnicotinate	4.45 (s), 8.10 (t), 8.88 (t), 9.11 (s)	↑	K (diet), K (urine), Mg (diet), fibre, Na (urine), Mg (urine)
		↓	fructose, glucose
N-methyl nicotinamide	4.48 (s), 8.19 (t), 8.90 (d), 8.97 (d), 9.28 (s)	↑	Mg (urine)
N-methyl-2-pyridone-5-carboxamide	3.65 (d), 6.67 (d), 7.83 (dd), 8.34 (d)	↑	vit B6, Mg (urine), Cys, Glu, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val, vit B2, vit B5, Ala, Arg, Asp, Gly, His, Lys, vit B1, vit B3, Fe (non-heme), Mg (diet), P
tyrosine	6.90 (d), 7.19 (d)	↑	Ca (urine), Na (urine), Cys, Glu, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val
3-hydroxymandelate	6.85 (d), 6.92 (t), 6.99 (d), 7.31 (t)	↑	K (diet), Mg (diet), K (urine), Mg (urine)
2-furoylglycine	6.65 (dd), 7.19 (d), 7.71 (d)	↑	SFA 18:0, SFA 16:0, MFA 16:1, MFA 18:1, cholesterol
		↓	fibre, K (diet), Mg (diet), vit B9, vit C, vit B6, K (urine), vit B1, vit B5, Ca (diet), Cu, Mg (urine), P
pseudouridine	7.67 (s)	↓	K (diet), Mg (diet), K (urine), fibre, vit C, Mg (urine), vit B6, vit B9, Cu
formate	8.46 (s)	↑	Na (urine)

Supplementary Table 3. The partial intraclass correlation (ICC) for each ^{1}H NMR urinary metabolite and nutrient. Data shown related to the first and second urine collection data and two dietary records (obtained on average 3 weeks apart), adjusted for age, sex and population sample. The median 3-week partial ICC across all 7,100 ^{1}H NMR spectral variables for the U.S. population (n=1,848) is 0.325 (IQR: 0.120-0.450). The thick line indicates the variables that have ICCs that fall in the top quartile of ^{1}H NMR ICCs. 25/46 urinary metabolites (54.4%) (coloured in orange) are in the top quartile of ICCs compared to 17/80 (21.3%) nutrients (coloured in blue).

Data	Variable	3-week pICC	*P*-value	2.5% CI	97.5% CI
NMR	Pantothenate	0.793	$< 4.94\times10^{-324}$	0.78	0.81
NMR	*N*-methylnicotinate	0.707	1.80×10^{-280}	0.68	0.73
NMR	Tyrosine	0.702	8.45×10^{-275}	0.68	0.72
Diet	Ca (urine)	0.701	9.39×10^{-274}	0.68	0.72
NMR	*N*-methyl-2-pyridone-5-carboxamide	0.661	2.79×10^{-233}	0.63	0.69
NMR	Ethyl glucuronide	0.634	2.04×10^{-208}	0.61	0.66
Diet	Ethanol	0.631	1.43×10^{-206}	0.60	0.66
NMR	Phenylacetylglutamine	0.623	9.93×10^{-200}	0.59	0.65
NMR	3-hydroxyisovalerate	0.611	5.13×10^{-190}	0.58	0.64
Diet	Mg (urine)	0.598	6.18×10^{-180}	0.57	0.63
NMR	3-hydroxymandelate	0.596	7.77×10^{-179}	0.57	0.62
NMR	Dimethylglycine	0.590	7.10×10^{-174}	0.56	0.62
Diet	K (urine)	0.586	1.99×10^{-171}	0.56	0.62
Diet	Fructose	0.585	1.10×10^{-170}	0.55	0.61
Diet	Lactose	0.576	2.37×10^{-164}	0.54	0.61
NMR	Ethanol	0.569	1.33×10^{-159}	0.54	0.60
NMR	Alanine	0.567	4.09×10^{-158}	0.54	0.60
Diet	Glucose	0.562	4.21×10^{-155}	0.53	0.59
NMR	Leucine	0.562	9.76×10^{-155}	0.53	0.59
NMR	Acetate	0.559	6.04×10^{-153}	0.53	0.59
NMR	Glutamine	0.555	2.30×10^{-150}	0.52	0.59
Diet	Fibre	0.548	9.22×10^{-146}	0.52	0.58
NMR	Proline betaine	0.544	2.71×10^{-143}	0.51	0.58
NMR	Hippurate	0.543	1.44×10^{-142}	0.51	0.57
NMR	Formate	0.538	1.63×10^{-139}	0.50	0.57
Diet	Mg (diet)	0.529	2.08×10^{-134}	0.50	0.56
NMR	Creatinine	0.526	2.93×10^{-132}	0.49	0.56
Diet	Ca (diet)	0.513	3.57×10^{-125}	0.48	0.55
Diet	P	0.510	2.07×10^{-123}	0.48	0.54
Diet	K (diet)	0.505	1.28×10^{-120}	0.47	0.54
NMR	*N*-methylpyridinium	0.493	2.54×10^{-114}	0.46	0.53
Diet	Sucrose	0.490	9.06×10^{-113}	0.45	0.52
NMR	Taurine	0.482	9.30×10^{-109}	0.45	0.52
NMR	Citrate	0.481	5.00×10^{-108}	0.45	0.52
NMR	4-hydroxyhippurate	0.480	1.15×10^{-107}	0.44	0.51
Diet	MFA 16:1	0.476	1.84×10^{-105}	0.44	0.51
NMR	Acetone	0.472	1.14×10^{-103}	0.44	0.51
Diet	SFA 16:0	0.471	4.46×10^{-103}	0.43	0.51
NMR	Creatine	0.462	5.98×10^{-99}	0.43	0.50
Diet	Vit B6	0.462	9.77×10^{-99}	0.43	0.50
NMR	*S*-methyl-cysteine sulfoxide metabolite 1	0.462	1.15×10^{-98}	0.43	0.50

Data	Variable	3-week pICC	*P*-value	2.5% CI	97.5% CI
Diet	Fe (non-heme)	0.458	8.57×10^{-97}	0.42	0.49
NMR	2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide	0.448	3.07×10^{-92}	0.41	0.48
NMR	Valine	0.446	2.02×10^{-91}	0.41	0.48
Diet	Vit B2	0.445	4.86×10^{-91}	0.41	0.48
Diet	Galactose	0.445	5.59×10^{-91}	0.41	0.48
Diet	PFA 18:3	0.437	2.67×10^{-87}	0.40	0.47
Diet	SFA 18:0	0.436	8.55×10^{-87}	0.40	0.47
NMR	Isoleucine	0.435	1.41×10^{-86}	0.40	0.47
Diet	Starch	0.428	1.34×10^{-83}	0.39	0.46
Diet	MFA 18:1	0.426	7.43×10^{-83}	0.39	0.46
NMR	2-furoylglycine	0.425	2.67×10^{-82}	0.39	0.46
NMR	Histidine	0.424	7.12×10^{-82}	0.39	0.46
Diet	Pro	0.420	2.74×10^{-80}	0.38	0.46
NMR	Fatty acids (C5-C10)	0.419	1.04×10^{-79}	0.38	0.46
NMR	4-hydroxyproline betaine	0.415	4.66×10^{-78}	0.38	0.45
Diet	VitC	0.415	5.08×10^{-78}	0.38	0.45
Diet	Na (urine)	0.412	3.95×10^{-77}	0.37	0.45
Diet	Ser	0.410	3.01×10^{-76}	0.37	0.45
Diet	Vit B5	0.406	1.42×10^{-74}	0.37	0.44
NMR	Succinate	0.405	2.29×10^{-74}	0.37	0.44
Diet	Glu	0.404	6.76×10^{-74}	0.37	0.44
Diet	Val	0.404	7.60×10^{-74}	0.37	0.44
Diet	Vit B1	0.401	1.44×10^{-72}	0.36	0.44
Diet	Phe	0.399	7.45×10^{-72}	0.36	0.44
Diet	Tyr	0.399	7.87×10^{-72}	0.36	0.44
NMR	Unknown 2	0.397	3.38×10^{-71}	0.36	0.43
Diet	Trp	0.395	1.77×10^{-70}	0.36	0.43
Diet	Leu	0.395	1.95×10^{-70}	0.36	0.43
Diet	Ile	0.387	1.77×10^{-67}	0.35	0.43
Diet	Thr	0.378	2.48×10^{-64}	0.34	0.42
Diet	Vit B9	0.377	9.34×10^{-64}	0.34	0.42
NMR	*O*-acetylcarnitine	0.376	2.58×10^{-63}	0.34	0.41
Diet	SFA 6:0	0.374	6.41×10^{-63}	0.33	0.41
Diet	Met	0.373	1.56×10^{-62}	0.33	0.41
Diet	Cys	0.373	2.08×10^{-62}	0.33	0.41
Diet	Vit B3	0.370	2.16×10^{-61}	0.33	0.41
NMR	Unknown 1	0.366	5.20×10^{-60}	0.33	0.40
Diet	His	0.364	1.91×10^{-59}	0.32	0.40
Diet	Asp	0.364	2.33×10^{-59}	0.32	0.40
Diet	Lys	0.364	2.75×10^{-59}	0.32	0.40
Diet	Ala	0.361	2.00×10^{-58}	0.32	0.40
Diet	Arg	0.351	4.15×10^{-55}	0.31	0.39
Diet	Gly	0.347	6.63×10^{-54}	0.31	0.39
NMR	*N*-acetyl-*S*-methyl-cysteine sulfoxide metabolite 2	0.341	6.09×10^{-52}	0.30	0.38
Diet	SFA 14:0	0.341	6.97×10^{-52}	0.30	0.38
NMR	*N6,N6,N6*-trimethyllysine	0.341	8.15×10^{-52}	0.30	0.38
Diet	SFA 4:0	0.341	9.58×10^{-52}	0.30	0.38
Diet	SFA 22:0	0.338	4.72×10^{-51}	0.30	0.38

Data	Variable	3-week pICC	*P*-value	2.5% CI	97.5% CI
Diet	PFA 18:2	0.337	1.26×10^{-50}	0.30	0.38
Diet	Omega6	0.337	1.40×10^{-50}	0.30	0.38
Diet	SFA 20:0	0.332	4.19×10^{-49}	0.29	0.37
Diet	Cholesterol	0.328	8.34×10^{-48}	0.29	0.37
Diet	Vit E	0.326	2.87×10^{-47}	0.28	0.37
Diet	SFA 10:0	0.323	1.37×10^{-46}	0.28	0.36
NMR	*S*-methyl-cysteine sulfoxide	0.307	4.59×10^{-42}	0.27	0.35
Diet	TFA 18:1	0.279	7.76×10^{-35}	0.24	0.32
Diet	Na (diet)	0.272	5.21×10^{-33}	0.23	0.31
Diet	TFA 18:2	0.266	1.46×10^{-31}	0.22	0.31
NMR	*S*-methyl-cysteine sulfoxide metabolite 2	0.263	6.33×10^{-31}	0.22	0.30
NMR	1-methylhistidine	0.262	1.11×10^{-30}	0.22	0.30
Diet	Cu	0.255	3.03×10^{-29}	0.21	0.30
NMR	Pseudouridine	0.253	1.28×10^{-28}	0.21	0.29
Diet	Beta-carotene	0.246	2.71×10^{-27}	0.20	0.29
Diet	Vit A (IU)	0.235	6.09×10^{-25}	0.19	0.28
NMR	*N*-methylnicotinamide	0.233	1.94×10^{-24}	0.19	0.28
Diet	PFA 20:4	0.232	2.85×10^{-24}	0.19	0.27
NMR	Carnitine	0.226	4.23×10^{-23}	0.18	0.27
Diet	SFA 12:0	0.217	2.13×10^{-21}	0.17	0.26
Diet	MFA 20:1	0.216	3.18×10^{-21}	0.17	0.26
Diet	Fe (heme)	0.197	6.15×10^{-18}	0.15	0.24
Diet	Vit A (RE)	0.191	6.46×10^{-17}	0.15	0.23
Diet	PFA 22:6	0.180	3.63×10^{-15}	0.14	0.22
Diet	SFA 8:0	0.173	2.93×10^{-14}	0.13	0.22
Diet	LC Omega3	0.172	5.29×10^{-14}	0.13	0.22
Diet	Vit B12	0.156	7.79×10^{-12}	0.11	0.20
Diet	PFA 22:5	0.150	4.90×10^{-11}	0.10	0.19
NMR	Dimethylamine	0.149	6.68×10^{-11}	0.10	0.19
Diet	Maltose	0.146	1.40×10^{-10}	0.10	0.19
Diet	PFA 20:5	0.139	1.07×10^{-9}	0.09	0.18
Diet	MFA 14:1	0.132	5.72×10^{-9}	0.09	0.18
Diet	Retinol	0.111	9.23×10^{-7}	0.07	0.16
NMR	Trimethylamine-*N*-oxide	0.093	3.29×10^{-5}	0.05	0.14
Diet	MFA 22:1	0.074	7.47×10^{-4}	0.03	0.12
Diet	Se	0.051	1.36×10^{-2}	0.01	0.10
Diet	PFA 18:4	0.035	6.88×10^{-2}	-0.01	0.08

Supplementary Table 4. Full/common names accompanying all metabolites abbreviated shown in the network (Extended Data Figure 2).

Abbreviation in network	Full/common name
1,2-Diacylglycerol	1,2-Diacylglycerol
1MeHis	1-Methylhistidine
2,5diOH-PhAc	2,5-Dihydroxyphenylacetate
2,5diOH-pyridine	2,5-Dihydroxypyridine
2Am-adipate	2-Aminoadipate
2Am-adipate 6semial.	2-Aminoadipate 6-semialdehyde
2Am-benzoate	2-Aminobenzoate
2deH-pantoate	2-Dehydropantoate
2-FuroylGly	2-Furoylglycine
2-Ketovaline	2-Oxoisovalerate
2Me1OHBu-TPP	2-Methyl-1-hydroxybutyl-thiamin diphosphate
2Me1OHPr-TPP	2-Methyl-1-hydroxypropyl-thiamin diphosphate
2MeBt-CoA	2-Methylbutanoyl-coenzyme A
2O-butyrate	2-Oxobutyrate
2O-glutarate	2-Oxoglutarate
2OHEt-TPP	2-(*alpha*-Hydroxyethyl)thiamine diphosphate
2-Oxoisocaproate	4-Methyl-2-oxopentanoate
2PY	*N*-Methyl-2-pyridone-5-carboxamide
3Cx1OHPrTPP	Succinate semialdehyde-thiamin diphosphate
3cyano-Ala	3-Cyanoalanine
3HIV	3-Hydroxyisovalerate
3I-Tyr	3-Iodotyrosine
3Me1OHBu-TPP	3-Methyl-1-hydroxybutyl-thiamin diphosphate
3Me2O-pentanoate	3-Methyl-2-oxopentanoate
3MeBt-CoA	Isovaleryl-coenzyme A
3OH-Asp	3-Hydroxyaspartate
3OH-mandelate	3-Hydroxymandelate
3O-propanoate	3-Oxopropanoate
4-Cresol	4-Cresol
4OH-benzoate	4-Hydroxybenzoate
4OH-benzoyl-CoA	4-Hydroxybenzoyl-coenzyme A
4OH-hippurate	4-Hydroxyhippurate
4OH-PhAc	4-Hydroxyphenylacetate
4OH-Ph-acetal.	4-Hydroxyphenylacetaldehyde
4OH-ProBet	4-Hydroxyproline betaine
5Ad-2Am-adipate	5-Adenylyl-2-aminoadipate
5MeTHF	5-Methyltetrahydrofolate
6CxLys	2,6-Diaminoheptanedioate
Ac P	Acetyl phosphate
Ac-Ad	Acetyl adenylate
Ac-CoA	Acetyl coenzyme A
Acetal.	Acetaldehyde
Acetate	Acetate

Abbreviation in network	Full/common name
AcetoAc	Acetoacetate
AcetoAc-CoA	Acetoacetyl coenzyme A
Acetone	Acetone
ACV	*delta*-(2-Aminoadipyl)cysteinylvaline
ADP-ribose	Adenosine diphosphate ribose
Ad-Succ	Adenylosuccinate
Ala	Alanine
αisoPr-malate	*alpha*-Isopropylmalate
AMP	Adenosine monophosphate
Anserine	Anserine
Arg	Arginine
Arg-succinate	Argininosuccinate
Asp	Aspartate
Benzoate	Benzoate
Benzoyl-CoA	*S*-Benzoyl-coenzyme A
β-Ala	*beta*-Alanine
β-Ribosylnicotinate	Nicotinate ribonucleoside
Betaine	Betaine
Bt-CoA	Butyryl-coenzyme A
Butyrate	Butyrate
C10:0	Caprate
C5:0	Valerate
C6:0	Caproate
C7:0	Enanthate
C8:0	Caprylate
C9:0	Pelargonate
Carbamoyl P	Carbamoyl phosphate
Carnitine	Carnitine
Carnosine	Carnosine
Catechol	1,2-Dihydroxybenzene
CDP-diacylglycerol	Cytidine-diphosphate diacylglycerol
Choline	Choline
Chorismate	Chorismate
Citrate	Citrate
Citrulline	Citrulline
Citryl-CoA	Citryl-coenzyme A
CMP	Cytidine-5-monophosphate
CoA	Coenzyme A
Creatine	Creatine
Creatinine	Creatinine
Cys	Cysteine
CysGly	Cysteinylglycine
Cystathionine	Cystathionine
Deamido-NAD	Nicotinic acid adenine dinucleotide
deH-Ala	Dehydroalanine

Abbreviation in network	Full/common name
δ1-Piperideine-6Cxate	*delta*1-Piperideine-6carboxylate
DeP-CoA	Dephosphocoenzyme A
diH-LipE	Enzyme *N6*-(dihydrolipoyl)lysine
DiiodoTyr	3,5-Diiodotyrosine
DMA	Dimethylamine
DMA-CoA	3-Methylcrotonyl-coenzyme A
DMG	Dimethylglycine
EA	Ethanolamine
EA P	Phosphoethanolamine
Ethanol	Ethanol
Ethyl glucuronide	Ethyl glucuronide
FA	Fatty acid
Formate	Formate
Fuma-acetoAc	4-Fumarylacetoacetate
Fumarate	Fumarate
GABA	*gamma*-Aminobutyric acid
γGluCys	*gamma*-Glutamylcysteine
γGlu-S-(hercyn-2-yl)CSO	*gamma*-Glutamyl-*S*-(hercyn-2-yl)cysteine sulfoxide
Gln	Glutamine
Glu	Glutamate
Glucuronate	Glucuronate
Gly	Glycine
Glyoxylate	Glyoxylate
GSH	Glutathione
GuanidinoAc	Guanidinoacetate
Hercynine	*Nalpha,Nalpha,Nalpha*-Trimethylhistidine
Hippurate	Hippurate
His	Histidine
HMG-CoA	Hydroxymethylglutaroyl coenzyme A
Ile	Isoleucine
IMP	Inosine monophosphate
IsoBt-CoA	2-Methylpropionyl-coenzyme A
Isocitrate	Isocitrate
Lactate	Lactate
Lecithin	Phosphatidylcholine
Leu	Leucine
LipE	Enzyme *N6*-(lipoyl)lysine
Lys	Lysine
Malate	Malate
Malonyl-CoA	Malonyl coenzyme A
Malyl-CoA	Malyl-coenzyme A
Me-corrinoid	Methylcorrinoid
MeOH	Methanol
Methanal	Methanal
Methyl-CoM	Methylcoenzyme M

Abbreviation in network	Full/common name
MMA	Methylamine
NAAG	*N*-Acetylaspartylglutamate
NAAGG	*N*-Acetylaspartylglutamylglutamate
NAcAsp	*N*-Acetylaspartate
NAc-citrulline	*N*-Acetylcitrulline
NAcOrn	*N*-Acetylornithine
NAcSMCSO	*N*-Acetyl-*S*-methylcysteine sulfoxide
NAD	Nicotinamide adenine dinucleotide
Niacin	Nicotinate
Nicotinamide	Nicotinamide
Nicotinate ribonucleotide	Nicotinate ribonucleotide
NMe-GABA	4-Methylaminobutyrate
NMeGlu	*N*-Methylglutamate
NMe-nicotinamide	*N*-Methylnicotinamide
NMe-pyridinium	*N*-Methylpyridinium
NMNA	*N*-Methylnicotinate, trigonelline
OAc-carnitine	*O*-Acetylcarnitine
OAcSer	*O*-Acetylserine
Orn	Ornithine
OSucc-hSer	*O*-Succinylhomoserine
OxaloAc	Oxaloacetate
Oxalyl-CoA	Oxalyl-coenzyme A
PAG	Phenylacetylglutamine
Palmitate	Palmitate
Palmitoyl-CoA	Palmitoyl-coenzyme A
Pantetheine 4P	Phosphopantetheine
Pantothenate	Pantothenate
Pantothenate 4P	4-Phosphopantothenate
Pantothenoyl-Cys 4P	4-Phosphopantothenoylcysteine
PhAc	Phenylacetate
Phenol	Hydroxybenzene
PLP	Pyridoxal 5-phosphate
Pp-Ad	Propionyladenylate
Pp-CoA	Propionyl coenzyme A
ProBet	Proline betaine
Propanoate	Propanoate
Psi	Pseudouridine
Psi 5P	Pseudouridine 5-phosphate
PtdEA	Phosphatidylethanolamine
Ptd-inositol	Phosphatidylinositol
PtdSer	Phosphatidylserine
Pyruvate	Pyruvate
Quinolinate	Pyridine-2,3-dicarboxylate
Ribose	Ribose
Ribose 5P	Ribose 5-phosphate

Abbreviation in network	Full/common name
Ribosylamine 5P	5-Phosphoribosylamine
Ribulose 5P	Ribulose 5-phosphate
S(2MeBt)diH-LipE	*S*-(2-Methylbutanoyl)-dihydrolipoamide-E
S(2MePp)diH-LipE	*S*-(2-Methylpropionyl)-dihydrolipoamide-E
S(3MeBt)diH-LipE	*S*-(3-Methylbutanoyl)-dihydrolipoamide-E
S-(Hercyn-2-yl)-CSO	*S*-(Hercyn-2-yl)cysteine sulfoxide
Saccharopine	Saccharopine
Sarcosine	Sarcosine
Ser	Serine
SMCSO	*S*-methylcysteine sulfoxide
SMCSO-M1	*S*-methylcysteine sulfoxide metabolite 1
SMCSO-M2	*S*-methylcysteine sulfoxide metabolite 2
S-Succ-diH-LipE	*S*-Succinyldihydrolipoamide-E
Succ semial.	Succinate semialdehyde
Succ-CoA	Succinyl coenzyme A
Succinate	Succinate
Sulfoacetal.	Sulfoacetaldehyde
Taurine	Taurine
Thr	Threonine
TMA	Trimethylamine
TMAO	Trimethylamine-*N*-oxide
TPP	Thiamin diphosphate
TriMe-Lys	*N6,N6,N6*-Trimethylysine
Tyr	Tyrosine
Tyramine	Tyramine
UDP	Uridine 5-diphosphate
UDPglucose	Uridine diphosphate glucose
UDPglucuronate	Uridine diphosphate glucuronate
UDP-xylose	Uridine diphosphate xylose
UMP	Uridine monophosphate
Uracil	Uracil
Urea	Urea
Ureidoglycolate	Ureidoglycolate
Uridine	Uridine
Val	Valine

Supplementary Table 5. Model statistics for prediction of healthy and unhealthy dietary patterns in the U.S. population over a 3-week period. The top and bottom quartiles of the NRF9.3 index, DASH-nutrient, OMNIHEART-carbohydrate, OMNIHEART-MFA and OMNIHEART-protein scores for the U.S. population were used to define healthy and unhealthy dietary patterns in this population. The 46 metabolites identified here were used to classify the dietary patterns in the U.S. data, from the first urine collection, using a Monte-Carlo Cross-Validated (MCCV) Projections to Latent Structures model. This model was used to classify the same U.S. population at the second urine collection (test data) based on their urinary data alone.

Dietary score [a]	Data	R^2_Y [b]	Q^2_Y [b]	AUROC [b]	TPR [b]	TNR [b]	Accuracy [b]
Nutrient Rich Food index	Training model (U.S. 1st urine)	0.43		0.89	0.78	0.84	81.2%
	Test data (U.S. 2nd urine)		0.40	0.88	0.76	0.84	79.9%
DASH-nutrient score	Training model (U.S. 1st urine)	0.30		0.83	0.79	0.72	74.9%
	Test data (U.S. 2nd urine)		0.27	0.81	0.75	0.70	71.8%
OMNIHEART-carbohydrate score	Training model (U.S. 1st urine)	0.33		0.85	0.76	0.78	77.1%
	Test data (U.S. 2nd urine)		0.30	0.83	0.75	0.77	75.6%
OMNIHEART-MFA score	Training model (U.S. 1st urine)	0.33		0.85	0.80	0.73	76.1%
	Test data (U.S. 2nd urine)		0.31	0.83	0.77	0.72	74.3%
OMNIHEART-protein score	Training model (U.S. 1st urine)	0.31		0.84	0.76	0.75	75.6%
	Test data (U.S. 2nd urine)		0.29	0.83	0.75	0.74	74.2%

a The cut-offs for the bottom and top quartiles for the U.S. population were: 28.35 and 46.76 (for NRF), 1.0 and 3.0 (DASH-nutrient), 0.5 and 3.0 (OMNIHEART-carbohydrate), 1.0 and 3.0 (OMNIHEART-MFA), 0.5 and 2.5 (OMNIHEART-protein), respectively.

b R^2_Y: goodness of fit (training data); Q^2_Y: goodness of prediction (test set data); AUROC: area-under-receiver-operator-curve; TPR: true positive rate, defined as the number of participants correctly predicted as having a healthy dietary pattern based on their urinary metabolites divided by the total number of participants with a dietary score in the top quarter; TNR: true negative rate, defined as the number of participants correctly predicted as having an unhealthy dietary pattern based on their urinary metabolites divided by the total number of participants with a dietary score in the bottom quarter; accuracy is calculated as number of participants correctly predicted as having healthy or unhealthy dietary patterns divided by the total number of participants.

Supplementary Table 6. Descriptive statistics of the U.S. INTERMAP population, the participants included in the data set and the dietary outliers, estimated using two techniques, are shown. The mean (median) and standard deviation, or number and percentage, are shown for different descriptors. Last column shows the *P*-value of a two-sample t-test or χ^2-test, as appropriate, of the comparison between the participants included in the data set and the dietary outliers.

	Included in the data set (n = 1,848)	**Dietary outliers (n = 184)**	
Descriptor	Mean (median) ± s.d. or n (%)		*P*-value
Age (years)	49.1 (49) ± 5.4	49.6 (50) ± 5.4	2.31×10^{-1}
Sex (% men)	954 (51.6%)	86 (46.7%)	2.06×10^{-1}
Center – Baltimore	225 (89.6%)	26 (10.4%)	4.42×10^{-1}
Center – Chicago	278 (91.1%)	27 (8.9%)	8.94×10^{-1}
Center – Corpus Christi, Hispanic	208 (89.3%)	25 (10.7%)	3.44×10^{-1}
Center – Corpus Christi, Non-Hispanic	225 (91.8%)	20 (8.2%)	6.04×10^{-1}
Center – Honolulu	245 (96.8%)	8 (3.2%)	4.81×10^{-4}
Center – Jackson	217 (86.1%)	35 (13.9%)	4.28×10^{-3}
Center – Minneapolis	226 (91.5%)	21 (8.5%)	7.47×10^{-1}
Center – Pittsburgh	224 (91.1%)	22 (8.9%)	9.48×10^{-1}
Systolic blood pressure (BP) (mm Hg)	117.8 (116.8) ± 13.47	121.6 (119.4) ± 15.31	1.47×10^{-3}
Diastolic BP (mm Hg)	73.3 (72.8) ± 9.65	73.9 (72.9) ± 9.82	3.90×10^{-1}
Body Mass Index ($kg \times m^{-2}$)	28.6 (27.8) ± 5.60	31.7 (30.4) ± 7.15	4.99×10^{-8}
Diagnosed Diabetes Mellitus (%)	133 (7.2%)	19 (10.3%)	1.24×10^{-1}
Diagnosed heart condition/disease (%)	158 (8.5%)	19 (10.3%)	4.15×10^{-1}
Smoker (%)	280 (15.2%)	36 (19.6%)	1.15×10^{-1}
Physical activity (hrs/day)	3.2 (2.0) ± 3.1	3.7 (2.8) ± 3.4	5.69×10^{-2}
On drug for high BP, CVD or lipids (%)	483 (26.1%)	70 (38.0%)	5.38×10^{-4}
Reported energy intake (kcal/day), visit 1	2311 (2209) ± 740	1626 (1405) ± 770	2.42×10^{-24}
Reported energy intake (kcal/day), visit 2	2307 (2197) ± 748	1564 (1368) ± 677	8.15×10^{-33}
Estimated energy requirement (kcal/day)	2198 (2182) ± 357	2289 (2251) ± 409	4.18×10^{-3}
Reported protein intake (kcal/day), visit 1	350 (326) ± 125	256 (236) ± 116	4.45×10^{-21}
Estimated protein intake from urea excretion (kcal/day), visit 1	275 (264) ± 80	269 (264) ± 81	4.08×10^{-1}
Reported protein intake (kcal/day), visit 2	352 (331) ± 127	251 (223) ± 128	2.02×10^{-20}
Estimated protein intake from urea excretion (kcal/day), visit 2	273 (264) ± 81	268 (262) ± 91	4.67×10^{-1}
Nutrient-Rich Food (NRF) index	39.0 (36.3) ± 15.3	37.6 (33.7) ± 15.4	2.34×10^{-1}

Supplementary Table 7. Nutrients for 1,848 U.S. INTERMAP participants included in this study. Nutrients are expressed as % of kcal, unless otherwise specified.

Nutrients	Visit 1			Visit 2		
	Mean	SD	Median	Mean	SD	Median
MFA 14:1-myristoleic acid	0.03	0.04	0.01	0.03	0.04	0.01
MFA 16:1-palmitoleic acid	0.60	0.31	0.57	0.60	0.28	0.58
MFA 18:1-oleic acid	11.57	3.21	11.46	11.66	3.10	11.70
MFA 20:1-gadoleic acid	0.08	0.06	0.06	0.07	0.06	0.06
MFA 22:1-erucic acid	0.02	0.05	0.01	0.02	0.06	0.01
PFA 18:2-linoleic acid	6.28	2.45	5.94	6.26	2.37	5.99
PFA 18:3-linolenic acid	0.68	0.33	0.62	0.68	0.31	0.62
PFA 18:4-stearidonic acid	0.00	0.01	0.00	0.00	0.01	0.00
PFA 20:4-arachidonic acid	0.06	0.05	0.05	0.06	0.04	0.05
PFA 20:5-eicosapentaenoic acid	0.02	0.05	0.00	0.02	0.05	0.00
PFA 22:5-docosapentaenoic acid	0.01	0.02	0.00	0.01	0.02	0.00
PFA 22:6-docosahexaenoic acid	0.04	0.09	0.02	0.04	0.09	0.02
omega-6 PFA	6.34	2.45	6.00	6.32	2.38	6.04
long chain omega-3 PFA	0.07	0.15	0.02	0.07	0.15	0.03
SFA 4:0-butyric acid	0.18	0.14	0.15	0.18	0.15	0.15
SFA 6:0-caproic acid	0.09	0.07	0.07	0.09	0.08	0.07
SFA 8:0-caprylic acid	0.09	0.07	0.07	0.08	0.09	0.07
SFA 10:0-capric acid	0.17	0.11	0.15	0.16	0.11	0.14
SFA 12:0-lauric acid	0.38	0.44	0.23	0.36	0.39	0.23
SFA 14:0-myristic acid	0.88	0.47	0.82	0.88	0.47	0.80
SFA 16:0-palmitic acid	5.89	1.62	5.90	5.94	1.63	5.94
SFA 18:0-stearic acid	2.91	0.95	2.88	2.97	0.92	2.94
SFA 20:0-arachidic acid	0.01	0.03	0.00	0.01	0.02	0.00
SFA 22:0-behenic acid	0.02	0.04	0.00	0.01	0.04	0.00
TFA 18:1-trans-octadecenoic acid	1.62	0.86	1.48	1.70	0.88	1.54
TFA 18:2-trans-octadecadienoic acid	0.25	0.13	0.23	0.25	0.12	0.24
alcohol	1.83	4.10	0.04	1.67	3.90	0.05
fibre (g/1000kcal)	9.01	3.62	8.36	8.91	3.90	8.17
fructose	5.09	3.19	4.48	5.01	3.12	4.41
galactose	0.04	0.09	0.01	0.04	0.10	0.01
glucose	5.40	2.93	4.94	5.36	2.86	4.89
lactose	2.28	2.02	1.65	2.31	1.99	1.75
maltose	0.62	0.53	0.46	0.63	0.60	0.46
sucrose	10.67	5.55	9.81	10.87	5.49	10.06
starch	23.03	6.39	22.65	22.86	6.38	22.53
alanine	0.73	0.21	0.70	0.74	0.21	0.71

Nutrients	Visit 1			Visit 2		
	Mean	SD	Median	Mean	SD	Median
arginine	0.85	0.23	0.82	0.86	0.23	0.83
aspartic acid	1.31	0.34	1.27	1.32	0.35	1.28
cysteine	0.21	0.05	0.20	0.21	0.05	0.20
glutamic acid	3.01	0.63	2.96	3.02	0.64	2.96
glycine	0.65	0.19	0.63	0.66	0.20	0.63
histidine	0.43	0.11	0.42	0.43	0.12	0.42
isoleucine	0.70	0.18	0.68	0.71	0.18	0.68
leucine	1.19	0.29	1.16	1.20	0.29	1.16
lysine	1.03	0.31	0.98	1.04	0.31	1.00
methionine	0.35	0.10	0.34	0.35	0.10	0.34
phenylalanine	0.67	0.15	0.66	0.67	0.15	0.66
proline	0.99	0.22	0.97	0.99	0.22	0.97
serine	0.69	0.15	0.67	0.69	0.16	0.68
threonine	0.59	0.15	0.57	0.60	0.16	0.58
tryptophan	0.18	0.04	0.18	0.18	0.04	0.18
tyrosine	0.54	0.13	0.52	0.54	0.13	0.52
valine	0.78	0.19	0.76	0.79	0.19	0.76
cholesterol (mg/1000kcal)	128.94	67.67	113.07	131.16	69.38	115.97
beta-carotene (µg/1000kcal)	1867.14	2121.48	1131.10	2043.54	2781.23	1100.88
vitamin A (IU/1000kcal)	3846.23	3717.09	2645.95	4130.13	4801.33	2554.17
retinol (µg/1000kcal)	218.68	335.75	141.03	215.46	291.52	145.92
vitamin A (RE/1000kcal)	529.87	488.07	393.46	556.05	562.11	403.20
thiamin (mg/1000kcal)	0.87	0.24	0.83	0.86	0.23	0.83
riboflavin (mg/1000kcal)	0.92	0.29	0.86	0.91	0.29	0.86
niacin (mg/1000kcal)	11.52	3.28	11.06	11.51	3.35	10.99
pantothenic acid (mg/1000kcal)	2.27	0.85	2.14	2.26	0.85	2.12
vitamin B6 (mg/1000kcal)	0.90	0.33	0.85	0.91	0.33	0.86
folate (µg/1000kcal)	136.08	63.79	123.14	135.78	67.76	120.50
vitamin B12 (mg/1000kcal)	2.23	2.95	1.67	2.30	2.58	1.72
vitamin C (mg/1000kcal)	53.52	41.84	44.11	52.71	41.60	42.65
vitamin E (mg/1000kcal)	4.50	2.03	4.18	4.46	2.05	4.10
calcium (mg/1000kcal)	366.23	159.22	333.00	364.94	155.82	333.87
calcium (mmol/24-hr)	4.25	2.30	3.84	4.25	2.28	3.91
copper (mg/1000kcal)	0.67	0.24	0.63	0.66	0.25	0.62
iron (heme) (mg/1000kcal)	0.48	0.37	0.41	0.49	0.35	0.43
iron (non-heme) (mg/1000kcal)	7.32	3.03	6.61	7.34	3.19	6.49
magnesium (mg/1000kcal)	148.04	43.44	140.56	146.89	44.99	139.68
magnesium (mmol/24-hr)	4.29	1.68	4.07	4.31	1.75	4.08

Nutrients	Visit 1			Visit 2		
	Mean	SD	Median	Mean	SD	Median
phosphorus (mg/1000kcal)	589.87	139.30	574.87	590.75	137.20	573.17
potassium (mg/1000kcal)	1354.60	395.36	1298.59	1347.83	410.77	1297.71
potassium (mmol/24-hr)	59.49	23.02	56.04	57.83	22.90	54.60
selenium (µg/1000kcal)	60.25	47.32	56.62	59.65	18.13	57.02
sodium (mg/1000kcal)	1661.83	470.36	1606.60	1667.46	461.20	1609.17
sodium (mmol/24-hr)	164.46	66.45	155.37	165.46	70.68	153.72

SUPPLEMENTARY NOTES

Supplementary Note 1. Reproducibility of some nutrients may be lower than urinary metabolites.

We found that urinary potassium, a well-known dietary biomarker of fruit and vegetable intakes and diet quality[2], was relatively stable over the 3-week period (pICC=0.59, P=1.99×10^{-171}), with dietary potassium showing a lower value (pICC=0.51, P=1.28×10^{-120}) than its urinary counterpart. Yet two urinary metabolites associated with potassium, NMNA (pICC=0.71, P=1.80×10^{-280}) and 3-hydroxymandelate (pICC=0.60, P=7.77×10^{-179}), had higher pICCs than urinary potassium over the same period, while other associated metabolites such as proline betaine (pICC=0.54, P=2.71×10^{-143}) and hippurate (pICC=0.54, P=1.44×10^{-142}) had intermediate values, but still higher than for dietary potassium (pICC 95% confidence interval 0.47–0.54).

Supplementary Note 2. Discordance between biochemical pathway information and statistical relationships in the data.

Levels of metabolites participating in multiple biochemical reactions are influenced by a variety of enzymatic conversions creating extra degrees of freedom and therefore more variance in their measurements. This variance can obscure correlations between structurally similar metabolites and those in closely related pathways, whereas metabolites in simple linear chain conversions are more likely to be directly or inversely correlated. For example, the intermediate metabolites niacin and nicotinamide (**Figure 4A**) are involved in many different biochemical reactions and pathways (**Extended Data Figure 2**), which may explain the absence of correlations between NMNA and *N*-methylpyridinium with *N*-methylnicotinamide and 2PY. Moreover, the reaction involving niacin and nicotinamide is known to occur due to host genome enzymes (shown in purple boxes in **Figure 4**) as well as due to microbial genomes (shown in green) which adds further complexity.

N-methylnicotinate (NMNA, also known as trigonelline) and *N*-methylpyridinium have previously been linked to coffee consumption[3], whereas NMNA has also been reported to be a biomarker of peas consumption[4] and other foods that are rich in vitamin B3 (niacin). Another recent report linked 2-furoylglycine to coffee consumption[5]. We found that the urinary excretions of NMNA and *N*-methylpyridinium were correlated (cluster M8, **Figure 2**), whereas 2-furoylglycine separately clusters with phenylacetylglutamine (PAG) and glutamine (cluster M2). This is biologically plausible as *N*-methylpyridinium is a product of thermal degradation (Maillard reaction) of NMNA[3], whereas 2-furoylglycine is excreted in urine via a different mechanism[5] and these metabolites are not in close biological proximity in the human

metabolic reaction network (see **Extended Data Figure 2** for the full connected network and **Supplementary Table 4** for abbreviations and full names).

SUPPLEMENTARY DISCUSSION

Functional relationships between nutrients and health outcomes.

Certain urinary metabolites that we found associated with nutrients are have also been found in relation to health outcomes. For example, sodium is well-known to relate to blood pressure differences[6] and higher urinary calcium excretion associated with raised blood pressure[7]. Here we observed associations between urinary sodium and calcium with citrate and formate. Other studies have reported differences in these urinary metabolites in relation to renal function[8,9] (citrate and formate) and blood pressure[10] (formate). The association between urinary sodium and formate may be explained by sodium-cation transporters that work in parallel with solute carrier family 26 member 6 (slc26a6) renal tubular transporter proteins, which mediate the uptake of monovalent anions including formate through exchange processes with oxalate and chloride[11]. This points to a possible mechanism in renal tubules involving sodium and formate in blood pressure regulation. Citrate inhibits urinary crystallization of calcium salts by forming soluble (divalent) calcium-complexes and thereby reduces the potential for kidney stone formation including calcium-oxalate[12]. Citrate is filtered and reabsorbed by proximal tubular transporters that are dependent on sodium[13], although we did not find total citrate associated with (monovalent) sodium here. Increased circulating sodium results in higher concentrations of sodium and calcium in the urine. Urinary citrate-calcium complexation may help to reduce renal stone formation as it prevents precipitation of calcium salts in the kidney[14].

Another example relates to proline betaine which has been reported to have an inverse relationship with blood pressure and obesity[8,15]. Proline betaine is present in high concentrations in citrus fruits (particularly oranges), correlates with dietary vitamin C[15] and has been used as a biomarker for the assessment of dietary citrus fruit intake[16]. 2-hydroxy-2-(4-methylcyclohex-3-en-1-yl)propoxy glucuronide is another marker of citrus fruit intake that has previously only been tentatively identified[17]; here we confirm the structural elucidation from NMR and mass spectrometric data.

Host-gut microbial co-metabolites hippurate, 4-hydroxyhippurate and 3-hydroxymandelate correlated with each other possibly due to polyphenolic metabolism in the proximal colon (followed by subsequent glycine conjugation in mitochondria to form hippurate/4-hydroxyhippurate), whereas the gut microbial co-metabolite PAG has its origins in the distal colon (protein putrefaction); PAG correlated with glutamine (direct link in the network because glutamination is dependent on mitochondrial CoA activation) and 2-

furoylglycine but not with the above three gut microbial co-metabolites. The correlation of PAG with 2-furoylglycine cannot be explained from the metabolic reaction network map (**Extended Data Figure 2**). 2-furoate may compete with other compounds to be glycine conjugated in the mitochondria explaining this apparent 'wormhole connection'[18] between correlated metabolites that may reflect the common site (mitochondria) of co-enzyme A (CoA) mediated metabolic activation prior to amino acid conjugation[19].

Urinary metabolites can differentiate between healthy and unhealthy dietary patterns.

We also show that a model built using samples from the first urine collection accurately predict dietary patterns based on the urinary metabolites in the second urine collection, for the same individuals three weeks later. This further attests to the utility of urinary metabolic profiles for the reliable capture of dietary information. Nineteen of the 28 metabolites associated with different dietary patterns in our previous study[20] were also seen in our set of 46 nutrient-associated urinary metabolites. Of these, five (carnitine, *O*-acetylcarnitine, PAG, alanine and fatty acids – suggestive of a diet high in meat/animal protein intake) were associated with the diet least concordant with WHO healthy eating guidelines, while 14 (hippurate, 4-hydroxyhippurate, dimethylamine, trimethylamine-*N*-oxide, *N*-acetyl-*S*-methyl-cysteine sulfoxide, *S*-methyl-cysteine sulfoxide and its two (unidentified) metabolites, acetate, creatine, 2PY, NMNA and *N*-methylnicotinamide) were associated with the diet most concordant with WHO guidelines, reflecting higher intakes of fruits, vegetables and fish[20].

More details on strengths and limitations of the current study.

The INTERMAP study in U.S. and U.K. is cross-sectional with a short-term (3-week) follow-up for replication. The intake of nutrients was calculated from foods based on four multi-pass 24-hr dietary recall interviews using national food composition tables to convert foods to nutrients. While repeated dietary recall interviews provide more detailed and accurate information when compared to, for example, food-frequency questionnaires, they still rely on participant recall and are thus prone to reporting and other biases. We used two independent methods to eliminate obvious outliers, one based on spectroscopic data and the other based on concurrence between reported and estimated dietary intakes. Compared with those included, the excluded participants had significantly higher BMI but lower calculated energy intake from their 24-hr recalls, reflecting bias in their dietary data (see **Supplementary Discussion** below; **Supplementary Table 6**). BMI is an often used, but crude, measure of adiposity that highly correlates with the body fat percentage[21]. Other measures of adiposity, such as waist circumference or fat free mass, were unavailable in INTERMAP. In the 3-week follow-up no significant weight gain/loss was recorded in the participants. The associations

reported here were not materially altered when BMI and physical activity were included as additional covariates in the models (**Supplementary Figure 7**).

While 24-hr urine samples have the advantage over other methods (e.g. spot or overnight samples) in that they capture metabolic process information over an entire day, their collection is less practicable than other methods. Nonetheless, there are promising data to suggest that less burdersome methods such as cumulative samples overnight[20], collection of repeated spot urine samples[22] or timed spot urine samples[23] may provide a valid alternative means to approximate the 24-hr urinary metabolome and hence be reflective of nutrient intakes. Therefore, the results presented here may extend to different types of samples such as single 24-hr or multiple spot urine samples, although this will need to be tested in other studies.

^{1}H-NMR spectroscopy has been shown to be effective and reliable in the exploration of nutritional interventions and discovery of novel metabolic biomarkers associated with diet[20,24,25]. Although ^{1}H-NMR spectroscopy is less sensitive than mass spectrometry, it is exceptionally reproducible for measuring complex mixtures of metabolites in biofluids[26-28] and detects abundant metabolites with high dynamic ranges that relate to a variety of metabolic pathways[29]. Our approach is readily scalable to procedurally intense studies such as large-scale epidemiogic research, as it takes just 5 minutes to obtain a ^{1}H-NMR spectrum from a urine sample.

We provide three means by which the nutrient-metabolite associations can be explored. First, the heat map in **Figure 2** summarizes associations between each nutrient-metabolite pair visually. Second, **Supplementary Table 1** shows all significant associations with metabolites for each nutrient and **Supplementary Table 2** does the same by listing for each metabolite all significant associations with nutrients. Last, a standalone data visualization software program (NutriomeXplorer) allows deep exploration of the associations between nutrients and urinary metabolites and is supplied as a readily accessible resource for further direct interrogation of our data beyond the associations reported here (**Supplementary Figures 8–15**).

While we have reproduced our findings and assessed the stability of the metabolites, these data can not be used to define a dietary score (akin to DASH/NRF/OMNIHEART scores) based on urine measurements alone due to the cross-sectional (with short-term follow-up) design of the study. Any such endeavour should be validated in a controlled clinical trial to avoid introduction of bias/misreporting into a dietary score that can be applied solely based on measurements in urine. The nutrients and other chemical compounds found in foods have different kinetics of uptake and excretion of metabolic products in the urine. Some compounds such as proline betaine (relating to citrus fruits) and tartrate (relating to grapes) have been shown using kinetic studies to be cleared from the body within 2-8 hours after intake[15,30], while

for others, such as sodium, excretion may occur over two to three or more days[31], and for yet other variables the information from reaction kinetics is lacking. We included 24hr recall data for the day of, and the day before, the urine collection and therefore may have underestimated excretion of nutrient-related metabolites occurring over a longer timescale. Therefore, our approach based on urinary spectroscopic data reported here should ideally be used alongside conventional dietary measurements to reflect both short- and long-term exposures.

Dietary outliers underreport energy and protein intakes, and have higher BMI.

The U.S. INTERMAP population consists of 2,195 participants from 8 population samples. A total of 2,164 participants had complete dietary data and as well as ^{1}H-NMR data of two 24-hr urine samples. A total of 132 participants were excluded due to the ^{1}H-NMR data mapping outside the 95% Hotelling's T^2 ellipse[10]. In order to account for possible under- or over-reporting, and be left with a homogenous dataset for analysis, for the 2,032 participants that remain the ratios between the reported and estimated protein intake and the reported and estimated energy intakes (see **Methods**) were mapped as a multivariate distribution and the participants (n=184) that mapped outside the 95% confidence interval were excluded from data analysis (**Extended Data Figure 3**) (hereafter referred to as the 'dietary outliers'). This leaves a total of 1,848 U.S. individuals for the data analysis (hereafter referred to as the 'dataset').

The estimated energy requirement is slightly higher for the dietary outliers than it is for the 1,848 participants included in the analysis (P=4.18×10^{-3}), however the reported energy intakes for the dietary outliers are significantly lower (P=2.42×10^{-24} for visit 1 and P=8.15×10^{-33} for visit 2) compared to the reported energy intakes from the participants in the dataset (**Supplementary Table 5**). The estimated protein intake from the dietary outliers and the dataset is not significantly different, however the reported protein intakes (see **Methods**) are significantly lower (P=4.45×10^{-21} for visit 1 and P=2.02×10^{-20} for visit 2) in the dietary outliers. The significant underreporting taken together with the fact that the BMI for the dietary outliers is significantly higher (P=4.99×10^{-8}) compared to the data set is in accordance with estimated higher degrees of under-reporting in obese populations[32-34]. Other differences between dietary outliers and the 1,848 participants included in the analysis include higher systolic blood pressure (P=1.47×10^{-3}) and medication use (P=5.38×10^{-4}) in the dietary outlier group. The proportion of individuals from the Jackson center is significantly higher in the dietary outlier group, this population has previously been found to have higher blood pressure[35], whereas the Honolulu center has significantly less individuals in the dietary outlier group (P=4.81×10^{-4}). Nutrient intakes for both visits for the 1,848 participants are given in **Supplementary Table 6**.

SUPPLEMENTARY METHODS

Adjusted Coefficient of Commonality. Proofs of the Adjusted Coefficient of Commonality (ACC) properties. For a metric distance (d) to be proper it has to specify the following properties:

1) Non-negativity: $d \geq 0$
2a) Identity: $d(A, A) = 0$
2b) Definitiveness: $d(A, B) = 0 \ \ iff \ \ A \equiv B$
3) Symmetry: $d(A, B) \equiv d(B, A)$
4) Triangle inequality: $d(A, B) \leq d(A, C) + d(B, C) \quad \forall\, A, B, C$

Prior to proving properties 1-4 above, I write here the notations and operations that will be used in the proofs.

Notations

A = charged binary set (hereafter referred to as 'cset') with values (-1, 0, 1) for each element
An element of a cset with value 0 means it is not in the set, and 1/-1 indicate a charged (signed) contribution

a_i = element i in cset A

∅ = empty set, a cset of which all elements are 0

A^- = additive inverse of cset A

$A + A^- = ∅$

Set measures

|A| = cardinality of cset A, number of elements in A that are non-zero

$|A| \equiv |A^-|$

$|∅| = 0$

¦A¦ = total number of elements of cset A

¦A¦ ≥ |A|

¦A¦ ≡ ¦B¦ ∀ A, B

Operations

|A∩B| = cardinality of the intersection of csets A and B, defined as the number of elements in A and B that are identical and also non-zero (i.e. $a_i = 1$ and $b_i = 1$ or $a_i = -1$ and $b_i = -1$)

|A∩B| ≡ |B∩A| intersection operator is commutative

$|A \cap B^-| \equiv |A^- \cap B|$

|A∩∅| = 0

|∅∩∅| = 0

|A∪B| = cardinality of the union of csets A and B, defined as the number of elements in A and B that are non-zero in A, in B or both A and B

$|A \cup B| \equiv |A| + |B| - |A \cap B| - |A \cap B^-|$

|A∪B| ≡ |B∪A| union operator is commutative

$|A \cup B| \equiv |A \cup B^-| \equiv |A^- \cup B|$

|A∪∅| = |A|

|∅∪∅| = 0

$|A \cap B| + |A \cap B^-| \leq |A \cup B|$

Similarity measure and distance

The adjusted coefficient of commonality (ACC):

$$ACC = \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|}$$

ACC distance (d):

$$d(A,B) = 1 - ACC = 1 - \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|}$$

Proof of property 1

$$d(A,B) = 1 - \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} \geq 0 \rightarrow 1 \geq \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|}$$

By definition all terms in the equation are non-negative, therefore, in order to prove the maximum value the ACC can be is indeed 1 we remove $A \cap B^-$ from the equation as it contributes negatively.

$$1 \geq \frac{|A \cap B|}{|A \cup B|}$$

The maximal value the nominator can obtain is if both sets are equal, thus when A∩B = A∪B, proving property 1.

Proof of property 2a

$$d(A,A) = 1 - \frac{|A \cap A| - |A \cap A^-|}{|A \cup A|} = 0 \rightarrow 1 - \frac{|A| - 0}{|A|} = 1 - \frac{|A|}{|A|} = 0$$

Proof of property 2b

$$d(A,B) = 1 - \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} = 0 \rightarrow \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} = 1 \rightarrow |A \cap B| - |A \cap B^-| = |A \cup B|$$

$$|A \cap B| - |A \cap B^-| = |A| + |B| - |A \cap B| - |A \cap B^-| \rightarrow |A \cap B| + |A \cap B| = |A| + |B|$$

$$2|A \cap B| = |A| + |B|$$

Given the fact:

$$|A \cap B| \leq min(|A|, |B|)$$

$$2 \times min(|A|, |B|) = |A| + |B|$$

Suppose the minimum cardinality is |A|:

$$2|A| = |A| + |B| \rightarrow |A| = |B|$$

Therefore the distance between A and B is only 0 if both sets are identical.

Proof of property 3

$$d(A,B) = d(B,A) \rightarrow 1 - \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} = 1 - \frac{|B \cap A| - |B \cap A^-|}{|B \cup A|}$$

By definition |A∩B| ≡ |B∩A| and |A∪B| ≡ |B∪A|, therefore the above is equivalent.

Proof of property 4

Case 1, A = ∅

By definition, the distance of a cset and an empty cset is 1. Assuming A is empty we get:

$$d(A,B) \leq d(A,C) + d(B,C)$$

$$1 \leq 1 + d(B,C) \rightarrow 0 \leq d(B,C)$$

This is property 1 and has already been proven. Naturally the same is true if B were empty.

Case 2, C = ∅

For assuming C is empty we get:

$$d(A,B) \leq 1 + 1 \rightarrow 1 - \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} \leq 2 \rightarrow \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} \geq -1$$

This reduced our problem to proving the lower bound of the ACC is indeed -1. So in order to minimize the left hand side, we assume |A∩B|=0, as it contributes positively to the ACC:

$$-\frac{|A \cap B^-|}{|A \cup B|} \geq -1 \rightarrow |A \cup B| \geq |A \cap B^-|$$

Which by definition is true as no intersection of two csets can be larger than the union of the two (|A∩B| + |A∩B$^-$| ≤ |A∪B|).

Case 3, A ≡ B

Same goes for when csets are equal, first we assume A≡B, making use of property 2:

$$d(A,B) \leq d(A,C) + d(B,C)$$

$$d(A,A) \leq d(A,C) + d(A,C)$$

$$0 \leq d(A,C) + d(A,C) \rightarrow 0 \leq 2 \times d(A,C) \rightarrow \frac{0}{2} \leq \frac{2 \times d(A,C)}{2} \rightarrow 0 \leq d(A,C)$$

This is property 1 and has already been proven.

Case 4, A ≡ C

Next is when A≡C:

$$d(A,B) \leq d(A,C) + d(B,C)$$

$$d(A,B) \leq d(A,A) + d(B,A)$$

$$d(A,B) \leq 0 + d(B,A) \rightarrow \ d(A,B) = d(B,A)$$

This is property 3 and has already been proven.

Case 5, A ≠ B ≠ C, and A ≠ ∅, B ≠ ∅ and C ≠ ∅

Next, we assume that all csets are non-empty and that they are all different.

$$d(A,B) \leq d(A,C) + d(B,C) \quad \forall A,B,C$$

$$1 - \left(\frac{|A \cap B| - |A \cap B^-|}{|A \cup B|}\right) \leq \left(1 - \left(\frac{|A \cap C| - |A \cap C^-|}{|A \cup C|}\right) + 1 - \left(\frac{|B \cap C| - |B \cap C^-|}{|B \cup C|}\right)\right)$$

$$\frac{|A \cap C| - |A \cap C^-|}{|A \cup C|} + \frac{|B \cap C| - |B \cap C^-|}{|B \cup C|} \leq \frac{|A \cap B| - |A \cap B^-|}{|A \cup B|} + 1$$

In **Supplementary Figure 16** the overlap between csets A, B and C is broken up in 13 parts (a–m); the lower case letters also denote the number of elements in each part as follows:

$$a = |A \cap B \cap C| \equiv |A^- \cap B^- \cap C^-|$$

$$b = |A \cap B \cap C^-| \equiv |A^- \cap B^- \cap C|$$

$$c = |A \cap B^- \cap C| \equiv |A^- \cap B \cap C^-|$$

$$d = |A^- \cap B \cap C| \equiv |A \cap B^- \cap C^-|$$

$$e = |A \cap B^-| - c - d$$

$$f = |A \cap C^-| - b - d$$

$$g = |B \cap C^-| - b - c$$

$$h = |A \cap B| - a - b$$

$$i = |A \cap C| - a - c$$

$$j = |B \cap C| - a - d$$

$$k = |A| - a - b - c - d - e - f - h - i$$

$$l = |B| - a - b - c - d - e - g - h - j$$

$$m = |C| - a - b - c - d - f - g - i - j$$

Furthermore, we define n as the union of the 3 csets:

$$n = a + b + c + d + e + f + g + h + i + j + k + l + m$$

We can now rewrite the inequality using equations for a-n as follows:

$$\frac{(a+c+i)-(b+d+f)}{n-l} + \frac{(a+d+j)-(b+c+g)}{n-k} \leq \frac{(a+b+h)-(c+d+e)}{n-m} + 1$$

$$\frac{a-b+c-d-f+i}{n-l} + \frac{a-b-c+d-g+j}{n-k} \leq \frac{a+b-c-d-e+h}{n-m} + 1$$

We now reduce the fractions to common denominators:

$$\frac{(n-k)(n-l)(n-m)(a-b+c-d-f+i)}{n-l} + \frac{(n-k)(n-l)(n-m)(a-b-c+d-g+j)}{n-k}$$

$$\leq \frac{(n-k)(n-l)(n-m)(a+b-c-d-e+h)}{n-m} + (n-k)(n-l)(n-m)$$

$$(n-k)(n-m)(a-b+c-d-f+i) + (n-l)(n-m)(a-b-c+d-g+j)$$

$$\leq (n-k)(n-l)(a+b-c-d-e+h) + (n-k)(n-l)(n-m)$$

$$(n^2 - kn - mn + km)(a-b+c-d-f+i) + (n^2 - ln - mn + lm)(a-b-c+d-g+j)$$

$$\leq (n^2 - kn - ln + kl)(a+b-c-d-e+h)$$

$$+ (n^3 - kn^2 - ln^2 - mn^2 + kln + kmn + lmn - klm)$$

Next, we move everything to the right side, order everything by powers of n and simplify:

$$0 \leq n^3 + n^2(-a + 3b - c - d - e + f + g + h - i - j - k - l - m)$$

$$+ n(-2bk + 2ck + ek - fk - hk + ik + kl + km)$$

$$+ n(-2bl + 2dl + el - gl - hl + jl + lm)$$

$$+ n(2am - 2bm - fm - gm + im + jm) + kl(a + b - c - d - e + h - m)$$

$$+ km(-a + b - c + d + f - i) + lm(-a + b + c - d + g - j)$$

We can rewrite the n^2 term as follows:

$$n^2(-a + 3b - c - d - e + f + g + h - i - j - k - l - m) = n^2(4b + 2f + 2g + 2h - n)$$

$$= n^2(4b + 2f + 2g + 2h) - n^3$$

Inserting it back into the inequality and simplifying gives:

$$0 \leq n^2(4b + 2f + 2g + 2h) + n(-2bk + 2ck + ek - fk - hk + ik + kl + km)$$

$$+ n(-2bl + 2dl + el - gl - hl + jl + lm)$$

$$+ n(2am - 2bm - fm - gm + im + jm) + kl(a + b - c - d - e + h - m)$$

$$+ km(-a + b - c + d + f - i) + lm(-a + b + c - d + g - j)$$

Given that a-n are all ≥0, we need to find terms that cancel each of the negative terms, first for negative 'n' terms:

$$\{-2bkn, -fkn, -hkn, -2bln, -gln, -hln, -2bmn, -fmn, -gmn\}$$

From we can rewrite the n^2 term as follows:

$$n^2(4b + 2f + 2g + 2h) = 4bn^2 + 2fn^2 + 2gn^2 + 2hn^2$$
$$= \cdots + 4bkn + 4bln + 4bmn + \cdots + 2fkn + 2fmn + \cdots + 2gln + 2gmn + \cdots$$
$$+ 2hkn + 2hln + \cdots$$

Subtracting the negative 'n' terms gives:

$$\ldots + 4bkn + 4bln + 4bmn + \cdots + 2fkn + 2fmn + \cdots + 2gln + 2gmn + \cdots + 2hkn + 2hln + \cdots$$
$$- 2bkn - fkn - hkn - 2bln - gln - hln - 2bmn - fmn - gmn$$
$$= \cdots + 2bkn + 2bln + 2bmn + \cdots + fkn + fmn + \cdots + gln + gmn + \cdots + hkn$$
$$+ hln + \cdots = x$$

This has cancelled out all negative 'n' terms:

$$0 \leq n^2(x) + n(2ck + ek + ik + kl + km) + n(2dl + el + jl + lm) + n(2am + im + jm)$$
$$+ kl(a + b - c - d - e + h - m) + km(-a + b - c + d + f - i)$$
$$+ lm(-a + b + c - d + g - j)$$

Next, we find 'n' terms that cancel the negative terms in the 'kl', 'km' and 'lm' terms:

$$\{-ckl, -dkl, -ekl, -klm, -akm, -ckm, -ikm, -alm, -dlm, -jlm\}$$

$$n(2ck + ek + ik + kl + km) + n(2dl + el + jl + lm) + n(2am + im + jm)$$
$$= \cdots + 2ckl + 2ckm + \cdots + ekl + \cdots + ikm + \cdots + klm + \cdots + 2dkl + \cdots + 2dlm$$
$$+ \cdots + jlm + \cdots + 2akm + 2alm + \cdots$$

Subtracting the negative terms from the 'n' terms gives:

$$\ldots + 2ckl + 2ckm + \cdots + ekl + \cdots + ikm + \cdots + klm + \cdots + 2dkl + \cdots + 2dlm + \cdots + jlm + \cdots$$
$$+ 2akm + 2alm + \cdots - ckl - dkl - ekl - klm - akm - ckm - ikm - alm$$
$$- dlm - jlm = \cdots + ckl + ckm + \cdots + dkl + \cdots + dlm + \cdots + akm + alm + \cdots$$
$$= y$$

This has cancelled out all remaining negative terms and inserting it back into the equality leaves us with only non-negative terms:

$$0 \leq n^2(x) + n(y) + kl(a + b + h) + km(b + d + f) + lm(b + c + g)$$

Since all a-n are ≥0, the triangle inequality has thus been proven and the ACC distance is valid.

SUPPLEMENTARY REFERENCES

1 Posma, J. M. *et al.* Subset Optimization by Reference Matching (STORM): An Optimized Statistical Approach for Recovery of Metabolic Biomarker Structural Information from (1)H NMR Spectra of Biofluids. *Analytical chemistry* **84**, 10694-10701, doi:10.1021/ac302360v (2012).

2 Mente, A., Irvine, E. J., Honey, R. J. D. & Logan, A. G. Urinary Potassium Is a Clinically Useful Test to Detect a Poor Quality Diet. *Journal of Nutrition* **139**, 743-749, doi:10.3945/jn.108.098319 (2009).

3 Lang, R. *et al.* Development of a Hydrophilic Liquid Interaction Chromatography-High-Performance Liquid Chromatography-Tandem Mass Spectrometry Based Stable Isotope Dilution Analysis and Pharmacokinetic Stud Bioactive Pyridines in Human Plasma and Urine after Coffee Consumption. *Analytical chemistry* **82**, 1486-1497, doi:Doi 10.1021/Ac902616k (2010).

4 Posma, J. M. *et al.* Integrated Analytical and Statistical Two-Dimensional Spectroscopy Strategy for Metabolite Identification: Application to Dietary Biomarkers. *Analytical chemistry* **89**, 3300-3309, doi:10.1021/acs.analchem.6b03324 (2017).

5 Heinzmann, S. S., Holmes, E., Kochhar, S., Nicholson, J. K. & Schmitt-Kopplin, P. 2-Furoylglycine as a Candidate Biomarker of Coffee Consumption. *J Agr Food Chem* **63**, 8615-8621, doi:10.1021/acs.jafc.5b03040 (2015).

6 Aburto, N. J. *et al.* Effect of lower sodium intake on health: systematic review and meta-analyses. *BMJ* **346**, f1326, doi:10.1136/bmj.f1326 (2013).

7 Kesteloot, H. *et al.* Relation of urinary calcium and magnesium excretion to blood pressure: The International Study Of Macro- And Micro-nutrients And Blood Pressure and The International Cooperative Study On Salt, Other Factors, And Blood Pressure. *American journal of epidemiology* **174**, 44-51, doi:10.1093/aje/kwr049 (2011).

8 Elliott, P. *et al.* Urinary metabolic signatures of human adiposity. *Sci Transl Med* **7**, 285ra262, doi:10.1126/scitranslmed.aaa5680 (2015).

9 Garcia-Perez, I. *et al.* Urinary Metabolic Phenotyping the slc26a6 (Chloride-Oxalate Exchanger) Null Mouse Model. *J Proteome Res* **11**, 4425-4435, doi:10.1021/pr2012544 (2012).

10 Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396-400, doi:10.1038/Nature06882 (2008).

11 Alper, S. L. & Sharma, A. K. The SLC26 gene family of anion transporters and channels. *Molecular aspects of medicine* **34**, 494-515, doi:10.1016/j.mam.2012.07.009 (2013).

12 Nicar, M. J., Skurla, C., Sakhaee, K. & Pak, C. Y. Low urinary citrate excretion in nephrolithiasis. *Urology* **21**, 8-14 (1983).

13 Hamm, L. L. Renal handling of citrate. *Kidney international* **38**, 728-735 (1990).

14 Simpson, D. P. Citrate excretion: a window on renal metabolism. *Am J Physiol* **244**, F223-234, doi:10.1152/ajprenal.1983.244.3.F223 (1983).

15 Heinzmann, S. S. *et al.* Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr* **92**, 436-443, doi:10.3945/ajcn.2010.29672 (2010).

16 Gibbons, H. *et al.* Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. *Molecular nutrition & food research* **61**, doi:10.1002/mnfr.201700037 (2017).

17 Pujos-Guillot, E. *et al.* Mass Spectrometry-based Metabolomics for the Discovery of Biomarkers of Fruit and Vegetable Intake: Citrus Fruit as a Case Study. *J Proteome Res* **12**, 1645-1659, doi:10.1021/pr300997c (2013).

18 Bang, J. W. *et al.* Integrative top-down system metabolic modeling in experimental disease states via data-driven Bayesian methods. *J Proteome Res* **7**, 497-503, doi:10.1021/pr070350l (2008).

19 Knights, K. M., Sykes, M. J. & Miners, J. O. Amino acid conjugation: contribution to the metabolism and toxicity of xenobiotic carboxylic acids. *Expert opinion on drug metabolism & toxicology* **3**, 159-168, doi:10.1517/17425255.3.2.159 (2007).
20 Garcia-Perez, I. *et al.* Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *The lancet. Diabetes & endocrinology* **5**, 184-195, doi:10.1016/S2213-8587(16)30419-3 (2017).
21 Micozzi, M. S., Albanes, D., Jones, D. Y. & Chumlea, W. C. Correlations of body mass indices with weight, stature, and body composition in men and women in NHANES I and II. *Am J Clin Nutr* **44**, 725-731 (1986).
22 Iwahori, T. *et al.* Six random specimens of daytime casual urine on different days are sufficient to estimate daily sodium/potassium ratio in comparison to 7-day 24-h urine collections. *Hypertension research : official journal of the Japanese Society of Hypertension* **37**, 765-771, doi:10.1038/hr.2014.76 (2014).
23 Wilson, T. *et al.* Spot and Cumulative Urine Samples Are Suitable Replacements for 24-Hour Urine Collections for Objective Measures of Dietary Exposure in Adults Using Metabolite Biomarkers. *The Journal of nutrition* **149**, 1692-1700, doi:10.1093/jn/nxz138 (2019).
24 Heinzmann, S. S. *et al.* Stability and Robustness of Human Metabolic Phenotypes in Response to Sequential Food Challenges. *J Proteome Res* **11**, 643-655, doi:10.1021/Pr2005764 (2012).
25 Stella, C. *et al.* Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**, 2780-2788, doi:10.1021/Pr060265y (2006).
26 Dumas, M. E. *et al.* Assessment of analytical reproducibility of H-1 NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. *Analytical chemistry* **78**, 2199-2208, doi:10.1021/Ac0517085 (2006).
27 Smith, L. M. *et al.* Large-Scale Human Metabolic Phenotyping and Molecular Epidemiological Studies-via H-1 NMR Spectroscopy of Urine: Investigation of Borate Preservation. *Analytical chemistry* **81**, 4847-4856, doi:10.1021/ac9004875 (2009).
28 Keun, H. C. *et al.* Analytical reproducibility in H-1 NMR-based metabonomic urinalysis. *Chem Res Toxicol* **15**, 1380-1386, doi:Doi 10.1021/Tx0255774 (2002).
29 Nicholson, J. K. *et al.* Metabolic phenotyping in clinical and surgical environments. *Nature* **491**, 384-392, doi:10.1038/nature11708 (2012).
30 Garcia-Perez, I. *et al.* An Analytical Pipeline for Quantitative Characterization of Dietary Intake: Application To Assess Grape Intake. *J Agric Food Chem* **64**, 2423-2431, doi:10.1021/acs.jafc.5b05878 (2016).
31 Luft, F. C., Fineberg, N. S. & Sloan, R. S. Overnight Urine Collections to Estimate Sodium-Intake. *Hypertension* **4**, 494-498, doi:Doi 10.1161/01.Hyp.4.4.494 (1982).
32 Heitmann, B. L. & Lissner, L. Dietary underreporting by obese individuals--is it specific or non-specific? *BMJ* **311**, 986-989 (1995).
33 Rennie, K. L., Coward, A. & Jebb, S. A. Estimating under-reporting of energy intake in dietary surveys using an individualised method. *The British journal of nutrition* **97**, 1169-1176, doi:10.1017/S0007114507433086 (2007).
34 Freisling, H. *et al.* Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *The British journal of nutrition* **107**, 910-920, doi:10.1017/S0007114511003564 (2012).
35 Stamler, J. *et al.* Dietary and urinary metabonomic factors possibly accounting for higher blood pressure of black compared with white americans: results of international collaborative study on macro-/micronutrients and blood pressure. *Hypertension* **62**, 1074-1080, doi:10.1161/HYPERTENSIONAHA.113.01810 (2013).