Published in final edited form as:

Nature. 2008 May 15; 453(7193): 396–400. doi:10.1038/nature06882.

Human metabolic phenotype diversity and its association with diet and blood pressure

Elaine Holmes[1,*], **Ruey Leng Loo**[1,2,*], **Jeremiah Stamler**[3], **Magda Bictash**[1,2], **Ivan K. S. Yap**[1,2], **Queenie Chan**[2], **Tim Ebbels**[1], **Maria De Iorio**[2], **Ian J. Brown**[2], **Kirill A. Veselkov**[1], **Martha L. Daviglus**[3], **Hugo Kesteloot**[4], **Hirotsugu Ueshima**[5], **Liancheng Zhao**[6], **Jeremy K. Nicholson**[1], and **Paul Elliott**[2]

[1]Biomolecular Medicine, Division of Surgery, Oncology, Reproductive Biology and Anaesthetics (SORA), Faculty of Medicine, Imperial College London, South Kensington Campus,London SW7 2AZ, UK.

[2]Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, London W2 1PG, UK.

[3]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.

[4]Department of Public Health, Division of Epidemiology, Akademisch Ziekenhuis St Rafael, Leuven B-3000, Belgium.

[5]Department of Health Science, Shiga University of Medical Science, Otsu, Shiga 520-2192, Japan.

[6]Department of Epidemiology, Fu Wai Hospital and Cardiovascular Institute, Chinese Academy of Medical Sciences, Beijing 100037, China.

Metabolic phenotypes are the products of interactions among a variety of factors—dietary, other lifestyle/environmental, gut microbial and genetic[1–3]. We use a large-scale exploratory analytical approach to investigate metabolic phenotype variation across and within four human populations, based on ^{1}H NMR spectroscopy. Metabolites discriminating across populations are then linked to data for individuals on blood pressure, a major risk factor for coronary heart disease and stroke (leading causes of mortality worldwide[4]). We analyse spectra from two 24-hour urine specimens for each of 4,630 participants from the INTERMAP epidemiological study[5], involving 17 population samples aged 40–59 in China, Japan, UK and USA. We show that urinary metabolite excretion patterns for East Asian and western population samples, with contrasting diets, diet-related major risk factors, and

 Correspondence and requests for materials should be addressed to J.K.N. (j.nicholson@imperial.ac.uk) or P.E. (p.elliott@imperial.ac.uk).

Author Contributions The INTERMAP study was conceived by J.S., P.E. and Rose Stamler (deceased); INTERMAP urinary amino acids study was by J.S., P.E., M.L.D. and H.K.; INTERMAP metabonomics study was by J.K.N. and P.E., with E.H., and J.S., M.L.D. The manuscript was written by P.E., J.K.N., E.H. and J.S.; analyses were done by R.L.L., M.B., I.K.S.Y., Q.C. and I.J.B. T.E., M.D.I., and K.V. provided statistical and analytical support. H.U. and L.Z. were responsible for data collection. All authors reviewed and approved the manuscript.

[*]These authors contributed equally to this work.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

coronary heart disease/stroke rates, are significantly differentiated ($P < 10^{-16}$), as are Chinese/Japanese metabolic phenotypes, and subgroups with differences in dietary vegetable/animal protein and blood pressure[6]. Among discriminatory metabolites, we quantify four and show association ($P < 0.05$ to $P < 0.0001$) of mean 24-hour urinary formate excretion with blood pressure in multiple regression analyses for individuals. Mean 24-hour urinary excretion of alanine (direct) and hippurate (inverse), reflecting diet and gut microbial activities[2,7], are also associated with blood pressure of individuals. Metabolic phenotyping applied to high quality epidemiological data offers the potential to develop an area of aetiopathogenetic knowledge involving discovery of novel biomarkers related to cardiovascular disease risk.

Prehypertensive and hypertensive blood pressure (BP) is prevalent among a majority of middle-aged and older adults in most countries, and is a major risk factor perpetuating the cardiovascular disease epidemic[8]. The goal of the INTERMAP metabonomic study[1] is to develop metabolic phenotyping approaches to elucidate aetiopathogenetic mechanisms underlying the global BP problem and related disorders[8]. It aims to identify urinary metabolites that discriminate across population/subgroup strata defined by geographic or dietary criteria, and assess—for individuals—independent relationships of these metabolites to BP. The basic concepts are: (1) for population/subgroup strata with differing coronary heart disease (CHD)/stroke rates or BP levels, the differences are largely attributable to lifestyles, especially diet; (2) these differences across strata are reflected in urinary metabolite patterns and specific metabolic biomarkers; and (3) for individuals, these biomarkers may relate independently to their BP. We use a technology platform that is analytically unbiased and detects a wide variety of metabolites from dietary, gut microbial and host metabolism sources in one analytical sweep[1], thus maximizing opportunity for novel biomarker discovery. The numbers of individuals in our population samples are as follows: China, $n = 832$; Japan, $n = 1{,}138$; UK, $n = 496$; and USA, $n = 2{,}164$.

A schematic summarizing data-analysis strategy is shown in Supplementary Fig. 1a, b. From hierarchical clustering analysis (HCA) with group average linkage applied to the probabilistic quotient normalized[9] median ^{1}H NMR spectrum (Methods), we find that East Asian and western populations have well-differentiated metabolic phenotypes (Fig. 1). Results fo rfirst and second urine specimens show highly similar clustering order (Fig. 1 and Supplementary Fig. 2a), as do HCA dendrograms generated using the single-linkage method (Supplementary Fig. 2b, c). Geographic metabolic differences are greater than gender differences. Metabolic phenotypes of southern (Guangxi) and northern (Beijing and Shanxi) Chinese are also differentiated; those of UK and USA population samples overlap. These findings are consistent with principal components analysis (PCA) results (Fig. 2a–d and Supplementary Fig. 3a–d). The plots of PCA scores show similarity of median urine metabolite profiles by gender, with separate subclusters for China (north and south), Japan and the two western population samples. Analyses of spectroscopic data sets from first and second urine specimens are highly consistent (Supplementary Fig. 4a), as are analyses limited to normal weight, non-diabetic participants (Supplementary Fig. 4b), and analyses repeated after removal of metabolic outliers (individual data shown in Supplementary Fig. 5) using the 95% Hotelling's T^2 statistic[10] (Methods and Supplementary Fig. 6a, b).

We examine pairwise comparisons across countries (except for UK and USA which are poorly discriminated) and show significant differences (Hotelling's T^2, $P < 10^{-16}$) using orthogonally filtered partial least squares discriminant analysis (O-PLS-DA)[11], after exclusion of metabolic outliers. Japanese living in Japan and Japanese Americans are also differentiated ($P < 10^{-16}$). Discriminatory metabolites (Supplementary Table 1a, b) are identified from the O-PLS-DA coefficients on the basis of four criteria: P value, the rank of the P value, stability of rank and regression coefficient strength from a bootstrap resampling procedure (Methods). They include metabolites of predominantly dietary origin, for example, amino acids, creatine and trimethylamine-*N*-oxide; compounds related to energy metabolism (acetylcarnitine, tricarboxylic acid cycle intermediates); and dicarboxylic acids (for example, suberate). We also find that population gut microbial-mammalian co-metabolites[2] are discriminatory, for example, hippurate, phenylacetylglutamine and methylamines; we have previously shown structural differences in Chinese and American gut microbial speciation and direct linkage of microbial composition to metabolic phenotype[12].

Participants consuming dietary protein predominantly from vegetable or from animal sources (subgroups differing in BP levels[6]) are also differentiated (Hotelling's T^2 $P < 10^{-13}$) for East Asian and western samples considered separately (Fig. 3a, b and Supplementary Fig. 7a, b). Significant discriminatory metabolites (Fig. 3c, d and Supplementary Fig. 7c,d) closely correspond to those from the pairwise country comparisons.

We then quantify four discriminatory metabolites (alanine, formate, hippurate and *N*-methylnicotinate) from the ^{1}H NMR spectra (Methods and Supplementary Fig. 1b) and analyse these with respect to all other spectral variables usingO-PLS regression[11]. Largest r^2values (first urine specimens), other than for intra-molecular correlations, are (1) for alanine, with 2-oxoglutarate, reflecting close metabolic linkage via glutamate-pyruvate transaminase activity[13]; (2) for formate, with alanine, explained by pyruvate/Co-A metabolism; (3) for hippurate with *N*-methylnicotinate; and (4) for *N*-methylnicotinate with hippurate, reflecting common or related renal transporter/secretion mechanisms. We also find significant correlations ($|r| \geq 0.10$, $P < 10^{-9}$) with other variables (Supplementary Table 2): (1) alanine, positively with energy intake, dietary cholesterol, body mass index, 24-h urinary Na^+ and K^+ and Na^+/K^+ ratio; inversely with alcohol intake; (2) formate, positively with energy intake, 24-h urinary Na^+ and K^+; (3) hippurate, positively with dietary fibre, Mg^{2+}, phosphorus, 24-h urinary Na^+ and K^+; inversely with alcohol intake and urinary Na^+/K^+ ratio; and (4) *N*-methylnicotinate, positively with dietary Mg^{2+}, 24-h urinary Na^+ and K^+; inversely with urinary Na^+/K^+ ratio.Strongest correlationsare with 24-h urinary Na^+ excretion for alanine (Pearson $r = 0.39$) and formate ($r = 0.37$), and with 24-h urinary K^+ excretion for hippurate ($r = 0.40$) (when excreted as the sodium salt, hippurate can cause an aldosterone-mediated increase in K^+ excretion[14]).

In multiple linear regression models (four per metabolite for each of systolic and diastolic BP), accounting for the key non-dietary and dietary/urinary excretion variables associated with BP[15], we find significant inverse associations of formate with both systolic and diastolic BP (all eight models, Table 1); also of hippurate in six models, and a significant direct association of alanine with BP in five models (Table 1). Regression estimates are

similar or larger with analyses restricted to 'non-intervened' individuals[16], that is, people without special diet/nutritional supplements or diagnosis/treatmentfor cardiovascular disease or diabetes (Supplementary Table 3). Technical errors and reliability estimates of quantified metabolites are provided in Supplementary Information and Supplementary Table 4.

Using large-scale metabolic phenotyping, we have identified novel candidate urinary biomarkers related to BP. Endogenous formate is largely the product of one-carbon metabolism via the activities of mitochondrial and cytosolic serine hydroxymethyl transferases, and the tetrahydrofolate pathway[17]. Formate is also produced as one by product of fermentation of dietary fibre by the gut microbiome[18]. It is involved in active Cl^- reabsorption at the apical proximal tubule via the CFEX anion exchanger under inhibitory control of the serinethreonine kinase WNK4; gain-of-function mutations in *WNK4* cause pseudohypoaldosteronism type II, a mendelian disorder associated with hypertension[19]. We show that urinary formate and urinary Na^+ excretion are positively correlated. Given the central importance of NaCl in control of BP and the rise of BP with age[15,20], our findings suggest a previously unrecognized role for formate in BP regulation.

The inverse association of hippurate (benzoyl glycine) with BP may reflect physiological connections with diet[7] and gut microbial activity[2]. Availability of calories from the diet is also modulated by gut microbes in human obesity[21], which in turn relates to BP[15]. We previously reported that dietary alanine is higher in people consuming a predominantly animal compared with a predominantly vegetable diet[6], consistent with our findings here of a direct association of urinary alanine excretion with BP. Also in experimental animal models, alanine modulates cardiovascular responses to circulating catecholamines and increases BP[22].

Cross-population metabolic differences shown here add a new dimension to the decades-long knowledge of East–West contrasting patterns of diet, diet-related major risk factors and CHD/stroke mortality (Supplementary Information, Supplementary Figs 8a, b and 9a, b, and Supplementary Tables 5 and 6). We have shown that urinary metabolic phenotyping across populations/subgroups at differing risks of CHD/stroke and high BP identifies novel candidate biomarkers that relate to BP of individuals. This may provide the basis for a new 'metabolome-wide association' approach in molecular epidemiology to help understand the complex interactions of lifestyles, environment and genes that determine major diseases in the twenty first century.

METHODS SUMMARY

INTERMAP is an international standardized population-based epidemiological investigation of diet and BP[5]. We collected four in-depth 24-h multi-pass dietary recalls, eight BP measurements, anthropometric and questionnaire data and obtained two timed 24-h urine specimens, on average three weeks apart, from each individual according to standard protocol. We performed 600 MHz ^{1}H NMR spectroscopy on these samples; estimated within-specimen reproducibility was > 98% from blinded analysis of 8% specimens split in the field[1]. Scans (64) for each spectrum were acquired using standard parameters and pre-processing algorithms[23]; spectra were reduced to 7,100 variables by integrating spectral

intensity in segments (width in chemical shift δ 0.001) corresponding to the regions δ = 0.5–9.5 (excluding δ = 4.5–6.4 containing the residual water and urea resonances). We performed HCA and PCA using the median NMR specrum for each of 34 gender-specific population samples. The Hotelling T^2 statistic (95% criterion)[10], calculated from PCA analyses of all first and second 24-h urine spectra, was used to remove metabolic outliers (n = 575), enabling finer spectral detail. For the remaining 4,055 individuals, we used O-PLS-DA[11] to detect patterns of metabolites differentiating pairs of populations/subgroups. All models were computed separately for first and second 24-h urine specimens. For four quantified metabolites, we estimated technical error[5] from the 8% split samples and intra-individual reliability[24] from comparison of first and second 24-h urinary excretion values. Correlation-regression analyses were performed using standard methods[6].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

METHODS

Storage, preparation and ^{1}H NMR spectroscopic analysis of urine.

Urine aliquots (containing boric acid preservative) were stored at −30 ° C, thawed before use, and ^{1}H NMR spectra acquired using a standard one-dimensional pulse sequence with water suppression (Bruker Avance 600 spectrometer operating at 600.29 MHz in flow-injection mode[23]). Spectra were analysed in segments of width 0.6 Hz (less than the 1 Hz frequency resolution), thus retaining all structural information. We normalized the data to total spectral area to remove outliers. In all other models, the probabilistic quotient method[9] was used to expose finer detail in the metabolic profile, as it is relatively unaffected by residual outlying samples.

Data analyses.

Data were available for 4,630 of the 4,680 INTERMAP participants (Supplementary Fig. 1a). For population/subgroup analyses, we used data separately from the first and second urine specimens since averaging them loses spectral resolution due to slight shifts in peak registration. For individual-level analyses, for example, regression with BP, we used the mean of the two 24-h urinary values to increase precision[24].

HCA and PCA.

We used the median spectrum of each gender-specific population sample for each of the two 24-h urine specimens. For HCA we used Pirouette (version 3.1.1, Infometrix Inc.) with euclidean distance, and both group average and single linkages. We computed PCA models in Simca P+ (version 11, Umetrics) using sevenfold cross-validation[25] to select the number of components.

Detection and removal of metabolic outliers.

PCA modelling of all 4,630 participants revealed outlying groups due to high levels of urinary glucose, trimethylamine-*N*-oxide, ethanol, acetaminophen and their metabolites (Supplementary Fig. 5). To remove outliers, we normalized the spectra to total area and

applied Pareto scaling (dividing each variable by the square root of the standard deviation). This method weights the variables such that high concentration metabolites do not dominate the model, while avoiding noise amplification. Participants whose scores mapped outside of the 95% Hotelling T^2 ellipse[10] in either first or second specimens were excluded (n = 575), leaving n = 4,055 individuals.

O-PLS-DA[11] and selection of discriminatory metabolites.

We used an in-house O-PLS-DA algorithm in MATLAB 7.3.1 (MathWorks) to establish pairwise models between populations/subgroups, with variables scaled to unit variance. We used bootstrap resampling to identify discriminatory variables. Our method of metabolite identification used results from urinary collections: a metabolite was only considered discriminatory (Supplementary Table 1) if it was significantly associated and ranked among the top metabolites for both specimens. At each iteration, a bootstrap sample of the same size as the full sample was constructed by sampling at random with replacement. The sample was used to compute an O-PLS-DA model and the corresponding regression coefficients b_i were obtained, representing the contribution of the ith metabolic variable to between-group discrimination. This procedure was repeated 250 times; the resulting sampling distribution generated the bootstrap standard deviation of each regression coefficient, b_i, for calculation of an approximate Student's t statistic and corresponding P value, P_i, for the significance of the ith coefficient (following the procedure of ref. 26). The P values at each bootstrap iteration were ranked, and for each coefficient, the median and width of the 95% confidence interval of the ranks across the 250 bootstrap samples were calculated. For each pairwise model, the ith metabolic variable was then considered discriminatory if, separately for both first and second 24-h urinary specimens, the following criteria were met: (1) the coefficients b_i were significant, P_i, $7 < 10^{-6}$ (corresponding to $P < 0.05$ after Bonferroni correction for 7,100 spectral variables; this is conservative given the correlation structure between spectral variables); (2) the median ranks for the ith coefficient were in the top 5%; (3) the width of the confidence intervals of the ranks were in the bottom 5%; (4) the coefficients b_i were in the top 60%.

Quantitation of metabolites, reliability and regression against blood pressure.

Mean concentrations of four metabolites (alanine, formate, hippurate and *N*-methylnicotinate) were quantified from the NMR spectra of first and second urine specimens. We used an automated method modified from ref. 27, and excluded specimens where the estimation procedure failed or the values fell outside the method tolerance limits, for either urine specimen. Results were calibrated to the creatinine peak (CH_2 $\delta = 4.06$) and then to creatinine concentration measured externally using the Jaffé method (Supplementary Information). We calculated 24-h urinary excretion (mmol per 24 h) by multiplying urinary concentrations by urinary volume. Technical error[5] for each quantified metabolite was calculated from the 8% split specimens. To compare the NMR findings with independent analyses, we calculated the Spearman rank correlation coefficients relating our results (first urine specimens) for alanine (n = 4,232) with those from ion exchange chromatography[28], and, for hippurate (n = 124), with gas chromatography mass spectrometry[29].

Observed regression coefficients relating urinary variables to BP are attenuated because of within-person variability[24] in metabolite excretion. We estimated within and between-person variance for urinary metabolites from one-way ANOVA. Ratios of within to between-person variance, λ, were calculated for eight country/gender-specific subgroups and pooled, weighted by degrees of freedom. We averaged the two 24-h urinary values for each quantified metabolite, and estimated percentage of the theoretical regression coefficient (K_{xx}) for univariate regression by the formula $K_{xx} = 2/(2 +. \lambda) \times 100$ (ref. 24). We used multiple regression to relate mean metabolite concentrations to mean systolic and diastolic BP using SAS (version 9.1, SAS Institute) with adjustment for potential confounders, with and without body mass index[6]. We fitted regression models by country and pooled coefficients across countries, weighted by inverse of the variance, to estimate overall association, and tested for heterogeneity of country-specific coefficients[6]. We express regression coefficients as mm Hg per 2 s.d. higher urinary metabolite excretion, from pooled within-country standard deviations (from one-way ANOVA)[6].

Structural characterization of metabolites.

We used available spectral databases and chemical addition experiments to aid structural identification of discriminatory metabolites. For the remaining metabolites, we used statistical total correlation spectroscopy[30] and solid phase extraction chromatography coupled with NMR[29].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

INTERMAP is supported by the US National Heart, Lung, and Blood Institute (RO1 HL50490 and RO1 HL084228); the Chicago Health Research Foundation; and national agencies in Japan (the Ministry of Education, Science, Sports, and Culture), China and the UK. The funders had no role in the design and conduct of the study, or in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript. The INTERMAP study has been accomplished through the work of the staff at the local, national and international centres. A partial listing of colleagues is in ref. 5. We thank M. Rantalanein, O. Cloarec, E. Want and O. Beckonert (Imperial College London) for their assistance with the statistical and NMR analyses; and P. Oefner and H. Kaspar (University of Regensburg) for gas chromatography mass spectrometry analyses.

References

1. Dumas ME et al. Assessment of analytical reproducibility of ^{1}H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. Anal. Chem 78, 2199–2208 (2006). [PubMed: 16579598]
2. Nicholson JK, Holmes E & Wilson ID Gut microorganisms, mammalian metabolism and personalized health care. Nature Rev. Microbiol 3, 431–438 (2005). [PubMed: 15821725]
3. Sabeti PC et al. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918 (2007). [PubMed: 17943131]
4. Murray CJ & Lopez AD Mortality by cause for eight regions of the world: global burden of disease study. Lancet 349, 1269–1276 (1997). [PubMed: 9142060]
5. Stamler J et al. INTERMAP: Background, aims, design, methods, and descriptive statistics (non-dietary). J. Hum. Hypertens 17, 591–608 (2003). [PubMed: 13679950]
6. Elliott P et al. Association between protein intake and blood pressure: the INTERMAP study. Arch. Intern. Med 166, 79–87 (2006). [PubMed: 16401814]

7. Mulder TP, Rietveld AG & van Amelsvoort JM Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. Am. J. Clin. Nutr 81 (Suppl.), 256S–260S (2005). [PubMed: 15640488]
8. Elliott P & Stamler J in Coronary Heart Disease Epidemiology: From Aetiology to Public Health 2nd edn (eds Marmot M & Elliott P) 751–768 (Oxford Univ. Press, Oxford, UK, 2005).
9. Dieterle F, Ross A, Schlotterbeck G & Senn H Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. Anal. Chem 78, 4281–4290 (2006). [PubMed: 16808434]
10. Hotelling H The generalization of Student's ratio. Ann. Math. Stat 2, 360–378 (1931).
11. Trygg J & Wold S Orthogonal projections to latent structures (O-PLS).J. Chemometr 16, 119–128 (2002).
12. Li M et al. Symbiotic gut microbes modulate human metabolic phenotypes. Proc. Natl Acad. Sci. USA 105, 2117–2122 (2008). [PubMed: 18252821]
13. Chen S-H & Giblett ER Polymorphism of soluble glutamic-pyruvic transaminase: a new genetic marker in man. Science 173, 148–149 (1971). [PubMed: 5581908]
14. Lin S-H, Lin Y-F & Halperin ML Hypokalaemia and paralysis. Q. J. Med 94, 133–139 (2001).
15. Intersalt Co-operative Research Group. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Br. Med. J 297, 319–328 (1988). [PubMed: 3416162]
16. Elliott P et al. Dietary phosphorus and blood pressure. International study of macro and mcro-nutrients and blood pressure. Hypertension 51, 669–675 (2008). [PubMed: 18250363]
17. Gregory JF et al. Primed, constant infusion with 2H3 serine allows in vivo kinetic measurement of serine turnover, homocysteine remethylation and transsulfuration processes in human one-carbon metabolism. Am. J. Clin. Nutr 72, 1535–1541 (2000). [PubMed: 11101483]
18. Samuel BS & Gordon JI A humanized gnotobiotic mouse model of hostarchaeal-bacterial mutualism. Proc. Natl Acad. Sci. USA 103, 10011–10016 (2006). [PubMed: 16782812]
19. Kahle KT et al. WNK4 regulates apical and basolateral Cl^{-} flux in extrarenal epithelia. Proc. Natl Acad. Sci. USA 101, 2064–2069 (2004). [PubMed: 14769928]
20. Elliott P et al. Change in salt intake affects blood pressure of chimpanzees: Implications for human populations. Circulation 116, 1563–1568 (2007). [PubMed: 17785625]
21. Ley RE, Turnbaugh PJ, Klein S & Gordon JI Human gut microbes associated with obesity. Nature 444, 1022–1023 (2006). [PubMed: 17183309]
22. Conlay LA, Maher TJ & Wurtman RJ Alanine increases blood pressure during hypotension. Pharmacol. Toxicol 66, 415–416 (1990). [PubMed: 2371250]
23. Holmes E et al. Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. Anal. Chem 79, 2629–2640 (2007). [PubMed: 17323917]
24. Grandits GA et al. Method issues in dietary data analysed in the Multiple Risk Factor Intervention Trial. Am. J. Clin. Nutr 65 (Suppl.), 211S–227S (1997). [PubMed: 8988939]
25. Wold S Cross-validatory estimation of number of components in factor and principal components models. Technometrics 20, 397–405 (1978).
26. Martens H & Martens M Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Qual. Prefer 11, 5–16 (2000).
27. Crockford DJ et al. Curve fitting method for direct quantitation of compounds in complex biological mixtures using ^{1}H NMR: Application in metabonomic toxicology studies. Anal. Chem 77, 4556–4562 (2005). [PubMed: 16013873]
28. Fekkes D, Voskuilen-Kooyman A, Jankie R & Huijmans J Precise analysis of primary amino acids in urine by an automated high-performance liquid chromatography method: Comparison with ion-exchange chromatography.J. Chromatogr. B 744, 183–188 (2000).
29. Lenz EM & Wilson ID Analytical strategies in metabonomics. J. Proteome Res. 443, 443–458 (2007).

30. Cloarec O et al. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. Anal. Chem 77, 1282–1289 (2005). [PubMed: 15732908]

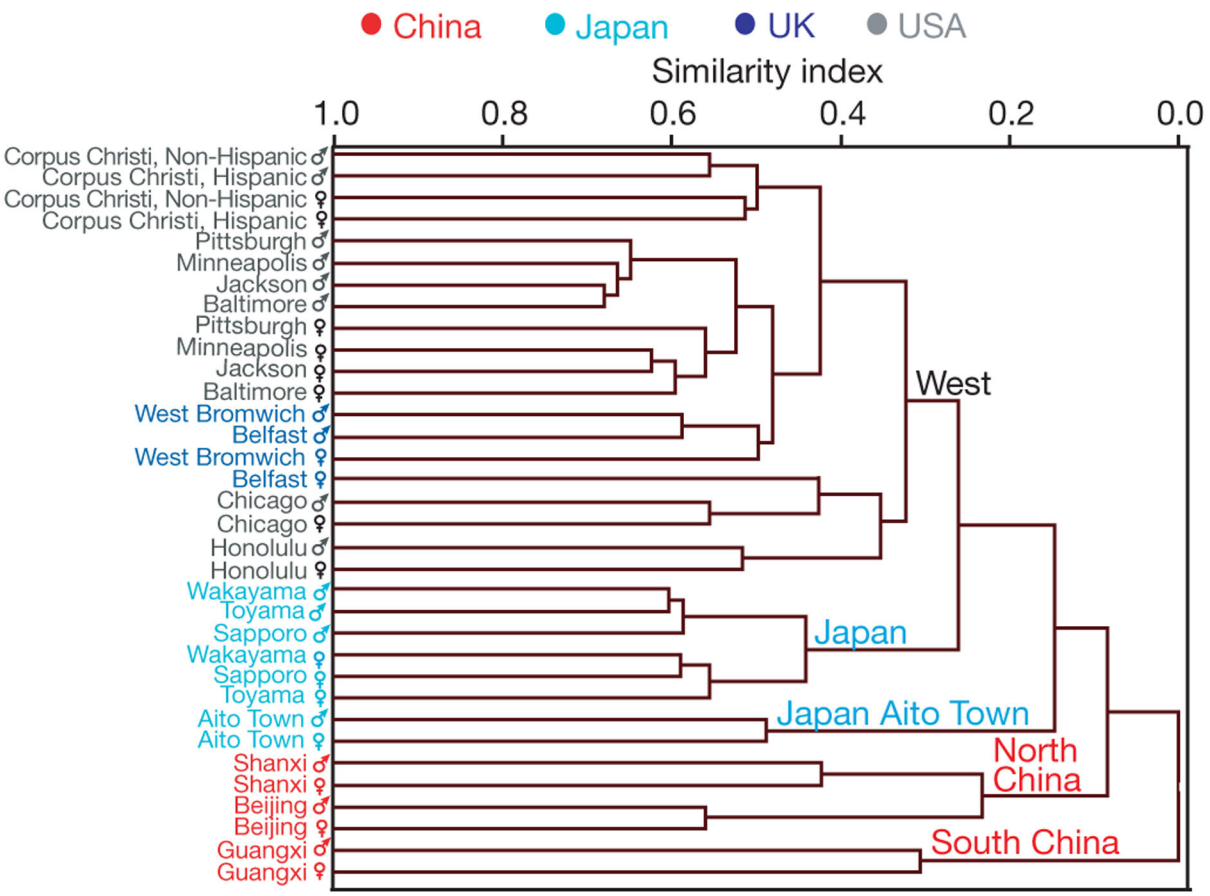

Figure 1 |. Hierarchical cluster analysis using group average linkage based on median ^{1}H NMR urine spectra, by population sample and gender (*n* = 4,630).

Data for first 24-h urinary specimens. The hierarchical cluster analysis (HCA) algorithm produces a dendrogram showing the overall similarity/dissimilarity between population samples. Similarity index is normalized to intercluster distance. The similarity index measures the multivariate distance between clusters. A similarity of one indicates zero distance between clusters; a value of zero indicates the maximum intercluster separation seen in the data. Each branch of the dendrogram defines a subcluster; population samples within subclusters are more similar to each other than to those in other subclusters.

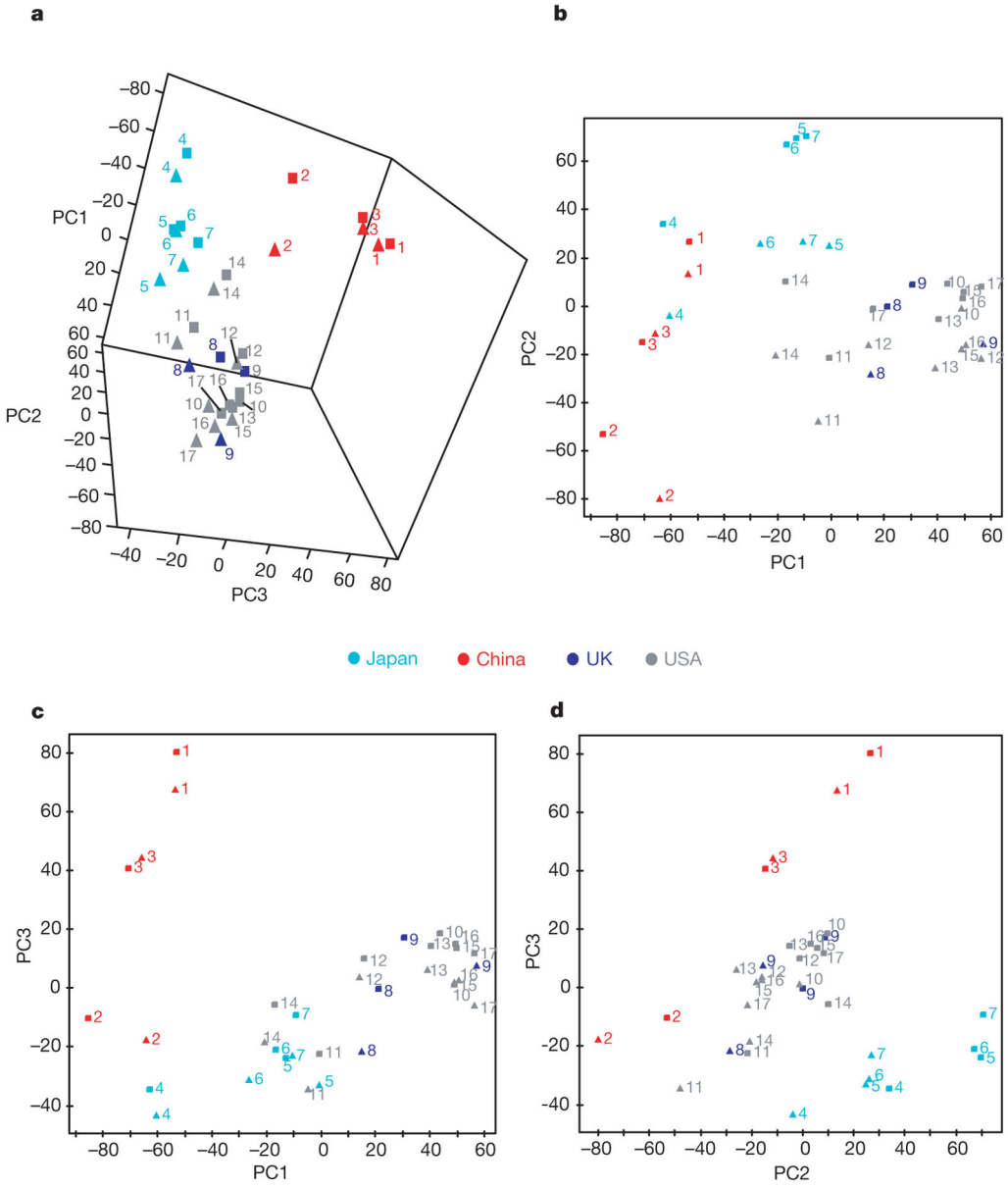

Figure 2 |. Plots of cross-validated principal components analysis scores (n = 4,630). a, Pseudo three-dimensional plot for principal components (PC) 1–3; **b**, PC2 versus PC1; **c**, PC3 versus PC1; **d**, PC3 versus PC2. Median ^{1}H NMR spectra of the first 24-h urine specimens stratified by country and by gender, female (triangles) and male (squares). R^2_X = 74.2% (percentage variation in the NMR data explained by the model); Q^2_X 5 49.6% (percentage variation in the NMR data predictable by the model from cross validation). The cross-validated scores values for the first three components are available in Supplementary Information. Symbols in **b–c** as in **a**.

Key: 1, Beijing; 2, Guangxi; 3, Shanxi; 4, Aito Town; 5, Sapporo; 6, Toyama; 7, Wakayama; 8, Belfast; 9, West Bromwich; 10, Baltimore; 11, Chicago; 12, Corpus Christi Hispanic; 13, Corpus Christi non-Hispanic; 14, Honolulu; 15, Jackson; 16, Minneapolis; 17, Pittsburgh.

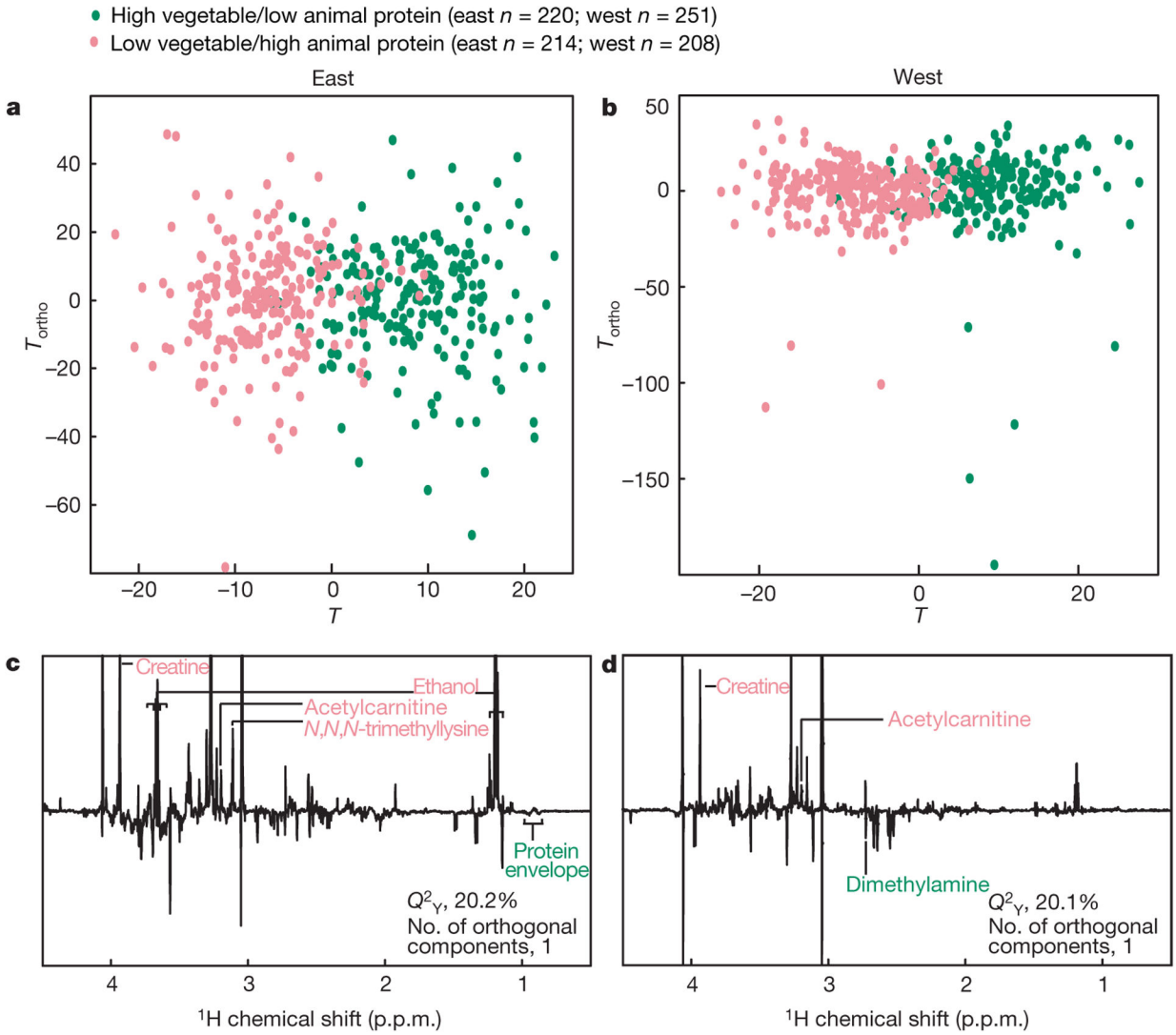

Figure 3 |. O-PLS-DA scores and loadings plots (bootstrap analyses) for participants reporting high vegetable/low animal protein and low vegetable/high animal protein intakes, first 24-h urinary specimens.

Plots (one orthogonal component) compare top and bottom quartiles, adjusted for sample, age and sex, from **a**, East Asian, and **b**, western population samples. Loadings plots from the O-PLS-DA bootstrap analyses are shown with discriminatory metabolites labelled (see Methods for metabolite selection criteria) for **c**, East Asian and **d**, western participants. Analyses are after removal of metabolic outliers using the 95% Hotelling's T^2 statistic in the initial PCA. The plots show the number of participants, the number of components used in each model and the Q^2_Y values (percentage variation in the protein subgroup assignment predictable by the model from cross validation).

Table 1 |

Estimated mean differences in systolic and diastolic BP

Urinary metabolite	A*				B†			
	Not adjusted for BMI‡		Adjusted for BMI‡		Not adjusted for BMI‡		Adjusted for BMI‡	
	Systolic blood pressure (mm Hg)							
Alanine	2.69	(6.06)	0.40	(0.92)	2.66	(5.54)	1.13	(2.43)
Formate	−1.19	(−2.62)	−1.42	(−3.29)	−1.94	(−3.92)	−1.04	(−2.20)
Hippurate	−2.10	(−4.85)	−1.63	(−3.95)	−1.72	(−3.70)	−0.82	(−1.83)
N-methylnicotinate	−0.09	(−0.21)	0.20	(0.49)	0.00	(0.00)	0.65	(1.53)
	Diastolic blood pressure (mm Hg)							
Alanine	1.57	(5.17)	0.17	(0.55)	1.58	(4.77)	0.61	(1.90)
Formate	−0.90	(−2.96)	−1.02	(−3.49)	−1.41	(−4.22)	−0.86	(−2.65)
Hippurate	−0.98	(−3.33)	−0.71	(−2.50)	−0.77	(−2.42)	−0.23	(−0.73)
N-methylnicotinate	−0.07	(−0.25)	0.09	(0.32)	−0.01	(−0.03)	0.37	(1.27)

Systolic and diastolic blood pressure differences per 12 s.d. difference in easch of four quantified urinary metabolites (mean of two 24-h urine values). Numbers in parentheses are Z scores, that is, regression coefficient divided by standard error (Z-score ≥ 1.96, $P < 0.05$; ≥ 2.58, $P < 0.01$; ≥ 3.29, $P < 0.001$;≥ 3.89, $P < 0.0001$). 2 s.d. difference for alanine = 0.34 mmol per 24 h (n = 4,232); formate = 0.29 mmol per 24 h (n 5 4,147); hippurate = 3.55 mmol per 24 h (n = 4,184); *N*-methylnicotinate = 0.41 mmol per 24 h (n = 4,081) (chemical shifts used for quantification: alanine, δ 1.48; formate, δ 8.45; hippurate, δ 7.85 and *N*-methylnicotinate, δ 4.44). Regression coefficients for individuals are pooled across countries (Methods); there is no evidence for cross-country heterogeneity in size of coefficients.

*A: Adjusted for age, sex, sample, special diet, supplement use, cardiovascular disease or diabetes mellitus diagnosis, physical activity (h per 24 h moderate or heavy activity), family history of high blood pressure.

†B: A + 7-day alcohol (g per 24 h) 1 urinary Na^+ (mmol per 24 h) 1 urinary K^+ excretion (mmol per 24 h).

‡Body mass index (kg m^{-2}).