Published in final edited form as:
J Proteome Res. 2010 December 3; 9(12): 6647–6654. doi:10.1021/pr100798r.

Metabolome-Wide Association Study Identifies Multiple Biomarkers that Discriminate North and South Chinese Populations at Differing Risks of Cardiovascular Disease: INTERMAP Study

Ivan K. S. Yap[*,†,#], **Ian J. Brown**[†,#], **Queenie Chan**[†,#], **Anisha Wijeyesekera**[*,†], **Isabel Garcia-Perez**[*,†], **Magda Bictash**[*,†], **Ruey Leng Loo**[*], **Marc Chadeau-Hyam**[†], **Timothy Ebbels**[*], **Maria De Iorio**[†], **Elaine Maibaum**[*], **Liancheng Zhao**[‡], **Hugo Kesteloot**[§], **Martha L. Daviglus**[¶], **Jeremiah Stamler**[¶], **Jeremy K. Nicholson**[*,‖], **Paul Elliott**[†,‖], and **Elaine Holmes**[*,‖]

[*]Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, UK

[†]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, UK

[‡]Cardiovascular Institute and Fu Wai Hospital, Chinese Academy of Medical Sciences, Beijing, China

[§]Department of Public Health, Division of Epidemiology, Akademisch Ziekenhuis St. Rafael, Leuven, Belgium

[¶]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

[‖]MRC-HPA Centre for Environment and Health, Imperial College London, UK

Abstract

Rates of heart disease and stroke vary markedly between north and south China. A ^{1}H NMR spectroscopy-based Metabolome-Wide Association approach was used to identify urinary metabolites that discriminate between southern and northern Chinese population samples, to investigate population biomarkers that might relate to the difference in cardiovascular disease risk. NMR spectra were acquired from two 24-hour urine specimens per person for 523 northern and 244 southern Chinese participants in the INTERMAP Study of macro/micronutrients and BP. Discriminating metabolites were identified using Orthogonal Partial Least Squares Discriminant Analysis and assessed for statistical significance with conservative Family Wise Error Rate <0.01 to minimise false positive findings. Urinary metabolites significantly (P <1.2×10^{-16} to 2.9×10^{-69}) higher in northern than southern Chinese populations included dimethylglycine, alanine, lactate, branched-chain amino acids (isoleucine, leucine, valine), *N*-acetyls of glycoprotein fragments (including uromodulin), *N*-acetyl neuraminic acid, pentanoic/heptanoic acid, methylguanidine; metabolites significantly (P <1.1×10^{-12} to 2×10^{-127}) higher in the south were gut microbial co-

Corresponding authors: Prof. Elaine Holmes Section of Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ, UK Tel: +44 (0) 20 7594 3220 Fax: +44 (0) 20 7594 3226 elaine.holmes@imperial.ac.uk. Prof. Paul Elliott MRC-HPA Centre for Environment and Health Department of Epidemiology & Biostatistics School of Public Health St Mary's Campus Imperial College London Norfolk Place London W2 1PG, UK Tel: +44 207 594 3328 Fax: +44 207 602 2759 p.elliott@imperial.ac.uk.
[#]Equal contributions

Abbreviations: 1D, ^{1}H-NMR, ATE, BEST, BMI, BP, CVD, INTERMAP, kcal, MFA, MWA, NMR, OPLS-DA, PCA, PFA, SFA, STOCSY, TSP

metabolites (hippurate, 4-cresyl sulphate, phenylacetylglutamine; 2-hydroxyisobutyrate), succinate, creatine, *scyllo*-inositol, prolinebetaine, *trans*-aconitate. These findings indicate the importance of environmental influences (e.g., diet), endogenous metabolism and mammalian-gut microbial co-metabolism, that may help explain north-south China differences in cardiovascular disease risk.

Keywords

^{1}H-NMR; blood pressure; epidemiology; gut microbial co-metabolites; INTERMAP; metabolome wide association; metabonomics; nutrition

INTRODUCTION

Cardiovascular diseases (CVD), mainly coronary heart disease and stroke, are as a group the leading cause of death worldwide.[1] Patterns of CVD vary markedly across regions of the world; in western countries coronary heart disease is the leading CVD, whereas in East Asian countries, including China, cerebrovascular diseases predominate.[1] There are also marked geographical patterns within countries. Specifically, the rates of stroke and heart disease are higher in north than south China.[2, 3] Major modifiable risk factors for stroke and coronary heart disease include raised blood pressure (BP), raised serum cholesterol and cigarette smoking.[2, 4–6] There are higher smoking rates and markedly higher levels of systolic and diastolic blood pressure in north than south China, which are likely to contribute to differences in CVD rates, though serum cholesterol levels are similar.[2, 7] We have shown previously in the INTERMAP Study (International Study of Macro/micronutrients and Blood Pressure) that, compared to the south, northern Chinese had higher body mass index (BMI), less favorable diet including lower calcium, magnesium and phosphorus intakes, higher 24-hour urinary sodium and lower urinary potassium excretion; together these factors accounted for the north-south BP differences.[7] Furthermore, a metabolome wide association (MWA) study found different patterns of urinary metabolite excretion between north and south China based on proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy.[8] Here we identify urinary metabolites that discriminate north and south Chinese population samples, giving further insights into possible environmental, endogenous metabolic and gut microbial influences that may help explain north-south differences in CVD risk.

METHODS

Population Samples and Field Methods (1996–1999)

INTERMAP is an international population-based cross-sectional study on relations of multiple dietary factors to BP among 4,680 men and women ages 40 to 59 years from 17 diverse population samples in China, Japan, United Kingdom and the United States. Three population samples were included from China, all rural: two in the north (Beijing, N=272 and Shanxi, N=289) and one in the south (Guangxi, N=278).[9] Data collected according to a common protocol include eight BP measurements, four 24-hour dietary recalls and two timed 24-hour urine collections per person. Each participant attended four times, with two visits on consecutive days, and a further two visits on consecutive days on average three weeks later. Blood pressure was measured twice per visit with a random-zero sphygmomanometer. Measurements of height and weight were obtained at two visits. Each participant provided two timed 24-hour urine collections, with both start and end done at the research center, between the first and second, and third and fourth clinic visits. Borate preservative was added to the urine specimen bottles prior to collection.[10] Urine volume was measured and aliquots obtained and stored at −20°C, then air-freighted on dry ice to the

Central Laboratory (Leuven, Belgium) for urinary biochemistry (sodium, potassium, calcium, magnesium, creatinine) and amino acid analysis by ion-exchange chromatography; frozen aliquots were also sent from Leuven to Imperial College London for analysis by ^{1}H NMR spectroscopy. Dietary data were collected at each visit by a trained interviewer with use of the multi-pass 24-hour recall method. All foods, drinks and supplements consumed in the previous 24 hours were recorded.[11] Institutional ethics committee approval was obtained for each site, and all participants gave written informed consent.

Preparation of Urine Specimens for ^{1}H NMR Spectroscopy

Urine specimens were thawed completely prior to mixing. 500 μL of urine were mixed with 250 μL of phosphate buffer for urinary pH stabilization (pH 7.4) and 75 μL of the sodium 3-trimethylsilyl-(2,2,3,3-$^2H_4$)-1-propionate (TSP) in D_2O. TSP served as a chemical shift reference, and D_2O served as a field-frequency lock for the NMR spectrometer. The resulting solution was then transferred into a 96-well plate and left to stand for 10 min. before centrifuging at 1,500 g for a further 10 min. to remove any precipitate prior to NMR analysis.

^{1}H NMR Spectroscopic Analysis of Urine Specimens

The urine specimens were analyzed by ^{1}H NMR spectroscopy at 600 MHz using a Bruker DRX600 spectrometer (Bruker Biospin, Rheinstetten, Germany) operating in flow injection mode. Urine specimens were automatically delivered to the spectrometer by a Gilson robot incorporated into the Bruker Efficient Sample Transfer (BEST) system. One-dimensional (1D) ^{1}H NMR spectra of urine were acquired using a standard 1D pulse sequence (recycle delay-90°-t_1-90°-t_m-90°-acquisition) with water presaturation during both the recycle delay (2s) and the mixing time, t_m, of 150ms. The 90° pulse length was set to ~10μs and an acquisition time of 2.73s was used. In total, 64 transients were collected into 32K data points using a spectral width of 20 ppm.

^{1}H NMR Data Pre-Processing

All free induction decays were multiplied by an exponential function equivalent to a 0.3 Hz line-broadening factor prior to Fourier transformation. The spectra were referenced and corrected for phase and baseline distortion. The spectral region δ 4.5 – 6.4 containing residual water and urea resonances was removed prior to normalisation by probabilistic quotient method.[12] The remaining spectrum (δ 0.5 – 9.5, excluding δ 4.5 – 6.4) was digitized to 7,100 variables (bin width 0.005 ppm). Principal component analysis (PCA) was performed on the pareto-scaled NMR dataset to identify metabolic outliers. PCA was done separately for first and second urine specimens, and participants whose scores mapped outside of the 95% Hotelling's T^2 ellipse[13] in either collection were excluded. Of the 839 Chinese INTERMAP participants, ^{1}H NMR spectra were not acquired for one of the two urine collections from seven participants and a further 65 participants were excluded following the PCA outlier analysis, leaving 767 individuals for the present report: 523 in the north (Beijing, 256; Shanxi, 267) and 244 in the south (Guangxi).

Statistical Methods

Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) (MATLAB version 7.3.1, MathWorks, Natick, MA)[14, 15] with unit variance scaling was used to identify metabolites discriminating northern and southern population samples, based on models constructed from one predictive and two orthogonal components. Analyses were done separately for first and second 24-hour urine collections. Discriminatory ability of the models was assessed by Q^2_Y statistic, i.e., the percent variance of the NMR data explained by geographic location (north-south), using 7-fold cross-validation. Mean north-south

differences in peak intensity for 7,100 spectral variables were assessed for statistical significance using a conservative Family Wise Error Rate of <0.01 to minimise false positive findings, corresponding to $P < 4 \times 10^{-6}$ for group mean north-south differences by Student's t, for each of the two urine collections considered separately.[16] Each discriminatory metabolite comprises multiple spectral variables at 0.005 ppm resolution; the minimum *P*-value for mean north-south differences in peak intensity among the spectral variables assigned to a particular metabolite was obtained separately for the first and second urine collections. This gave a ranking of the discriminatory strength of the metabolites.

Structural identification of discriminatory metabolites was achieved by 2-dimensional NMR experiments[17], statistical total correlation spectroscopy (STOCSY)[18], addition of known standards to the urine specimens, solid phase extraction chromatography and mass spectrometry.

Dietary data were converted to nutrient intakes (83 nutrients) with use of enhanced country-specific food tables, standardized across countries by the Nutrition Coordinating Center, University of Minnesota.[11] Measurements/person were averaged for BP and nutrient variables across the four visits; for 24-hour urinary sodium and potassium excretion, across the two collections. North-south differences in BP, BMI, nutrient intakes, urinary electrolyte excretion and questionnaire data were assessed for statistical significance by Student's t-test or X^2 test (SAS version 9.1, SAS Institute Inc, Cary, NC).

RESULTS

Metabolite Excretion Patterns in Northern and Southern Chinese Participants

The median urinary 600 MHz ^{1}H NMR spectrum of the Chinese population samples from the first urine collection is shown in Figure 1 and spectral differences between north and south China are shown in Figure 2, separately for first and second collections. Discriminatory metabolites together with their *P*-values are listed in Table 1.

The OPLS-DA cross-validated scores plots (Figures 2A and 2C) of the ^{1}H NMR urine data showed clear discrimination between the northern and southern Chinese populations along the predictive component. The corresponding O-PLS-DA coefficients plots (Figures 2B and 2D) indicated that the northern Chinese had systematically different urinary metabolic profiles from the southern Chinese (Q^2_y statistic 82.5% for first urine collection, 82.8% for the second collection). Urinary metabolites significantly ($P < 1.2 \times 10^{-16}$ to 2.9×10^{-69}) higher in northern compared with southern Chinese populations included dimethylglycine, alanine, lactate, branched-chain amino acids (isoleucine, leucine, valine), *N*-acetyls of glycoprotein fragments (including uromodulin), *N*-acetyl neuraminic acid, pentanoic/heptanoic acid and methylguanidine; metabolites significantly ($P < 1.1 \times 10^{-12}$ to 2×10^{-127}) higher in the south were gut microbial cometabolites (hippurate, 4-cresyl sulphate, pheylacetylglutamine; 2-hydroxyisobutyrate), succinate, creatine, *scyllo*-inositol, prolinebetaine and *trans*-aconitate (Table 1).

Cardiovascular Disease Risk Factor and Dietary Differences between Northern and Southern Chinese Participants

Mean systolic/diastolic BP was higher in northern (123.8/75.5 mm Hg) compared with southern Chinese population samples (115.4/68.2 mm Hg), $P = 2.9 \times 10^{-10}$ systolic and 3.8×10^{-21} diastolic (Table 2). While no previous heart attacks were reported in either sample, prevalence of other doctor diagnosed heart diseases (5.5% vs. 1.2%, $P = 0.01$) and stroke (2.1% vs. 0.0%, $P = 0.05$) were higher in the north. Prevalence of diabetes was low and did not significantly differ (1.1% in the north vs. 0% in the south, $P = 0.22$). Body mass

index and smoking rates were higher in northern compared with southern participants; physical activity was lower in the north.

Concerning dietary intakes, mean energy, vegetable protein, starch, and omega-3 polyunsaturated fatty acid (predominantly plant-based α-linolenic acid) intakes were higher north than south (Table 2). Saturated, monounsaturated, and omega-6 polyunsaturated (predominantly linoleic acid) fatty acid intakes were lower in northern Chinese, as were total protein and animal protein (approximately one third of the intake in the south). Amino acid intakes were lower in the northern Chinese, with the exception of glutamic acid and proline intakes, both higher in the north. Mean vitamin and mineral intakes (including vitamin C, dietary potassium and urinary potassium excretion) were also lower in northern Chinese participants, with the exception of vitamin E and selenium (higher in the north) and iron (no difference). Mean dietary sodium intake and urinary sodium excretion were higher in the north.

Sensitivity Analyses

Findings from OPLS-DA analyses of the study sample including 65 metabolic outliers were comparable with the main findings (Q^2_y statistic 78.5% for first urine collection, 80.2% for the second collection); the same discriminatory metabolites were identified (Supporting Information Table S1).

Compared to participants included in the main analysis, excluded individuals were more likely to be diabetic (15.4% vs. 0.8%, $P = 8.5\times10^{-15}$) and had higher mean alcohol intake (16.6 g/24-h vs. 8.0 g/24-h, $P = 0.002$). There were no differences in mean blood pressure levels, BMI, or the prevalence of heart disease or stroke.

DISCUSSION

Main findings of this study are marked differences in the urinary metabolite profiles of north and south Chinese population samples, reflecting differences in diet, endogenous metabolism and mammalian gut microbial co-metabolites.[19–23] Hippurate (excretion higher in the south than north) is formed predominantly by hepatic glycine conjugation of gut microbial-derived benzoate, produced from plant phenolics.[21] Gut microbiota also extensively catabolize protein and aromatic amino acids, including phenylalanine and tyrosine, to form phenylacetylglutamine and 4-cresyl sulfate.[19, 24] The gut microbiota facilitate host energy recovery from dietary sources[25] by providing refined control mechanisms of energy recovery through catabolism of otherwise poorly digestible nutrients, e.g., resistant starch. This feature of the gut microbiota has recently been implicated in human obesity[26–28] a risk factor for raised BP[8, 29] and CVD[30]; we have also reported inverse associations of urinary hippurate excretion with BP of individuals.[8] These results offer potential mechanisms by which the gut microbiota could directly influence CVD risk.

Excretion of 2-hydroxyisobutyrate was higher in the southern population sample. 2-hydroxyisobutyrate is derived from microbial degradation of dietary proteins and is associated with the presence of some microbial species such as *Faecalibacterium prausnitzii* in the colon.[31] Correspondingly, southern Chinese INTERMAP participants had higher total and animal protein intake than in the north.[7] Greater excretion of urinary alanine and lactate observed in the northern Chinese may be linked to their starch-rich diet. Zuppi *et al.* found that individuals consuming a diet high in carbohydrate had greater urinary excretion of citrate, lactate, alanine and glycine.[32] We recently reported that in INTERMAP urinary alanine is directly associated with BP.[8] Increased urinary alanine excretion has also been reported in hyperglycaemic dogs and human diabetes models – possibly reflecting changes in liver gluconeogenesis and kidney metabolism[33–35] – providing a further possible link to

CVD. In addition, urinary excretion of branched-chain amino acids such as valine and isoleucine were higher in the northern population sample. Differences in branched-chain amino acid excretion could reflect north-south differences in gluconeogenesis secondary to differences in levels of physical activity; in northern Chinese where physical activity levels are lower, a greater proportion of these branched-chain amino acids may be excreted in the urine rather than utilized as fuel. Differences in urinary levels of tricarboxylic acid cycle intermediate succinate may indicate differences in renal energy metabolism between the southern and northern populations.[36, 37]

Higher levels of urinary creatine were observed in the southern population sample. Creatine is found predominantly in red muscle tissue in the form of creatine phosphate.[3839] About half of the daily required creatine is synthesized from glycine, arginine and methionine, but creatine can also be obtained directly from consumption of creatine-rich foods such as meat and fish.[38] We previously reported high urinary creatine excretion with high meat intake[39]; thus, higher urinary excretion of creatine in the southern Chinese may reflect the ~3 times higher animal protein intake in southern compared to northern Chinese. Higher creatine excretion in southern Chinese may also reflect higher levels of physical activity with increased turnover in the creatine/creatinine pathway.[7, 38]

Differences in other metabolites may be dietary in origin. Higher levels of *scyllo*-inositol were observed in the southern population sample. *Scyllo*-inositol is converted from dietary *myo*-inositol[40] and is also present as an osmolyte found in deep-sea animals[41, 42]. Prolinebetaine and *trans*-aconitate levels were also higher in southern compared to northern China. Prolinebetaine, an osmoprotectant, has recently been identified as a biomarker of citrus consumption and is associated with high vitamin C intake.[43]*Trans*-aconitate is highly correlated with potassium concentration in plant leaves[44] and can be metabolised by bacteria.[45] Our findings for prolinebetaine and *trans*-aconitate are consistent with dietary origins, as vitamin C, potassium intake and urinary potassium excretion were all higher in the south.

Although this study is limited to 3 rural samples (two northern, one southern), our findings are broadly consistent with the results of previous studies of north-south differences in BP, BMI, sodium, potassium and other dietary and lifestyle factors[46–50]. This lends credence to the north-south differences in metabolite profiles reported here. In addition there are genetic differences between Han Chinese north and south.[51, 52] In summary, we found multiple urinary metabolites that discriminate southern and northern Chinese population samples that are at markedly different risks of CVD. These metabolite differences were reflective of dietary and gut microbial differences, which may help explain geographical differences in CVD risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The INTERMAP Study has been supported by grants R01-HL050490 and R01-HL084228 from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA; by the Chicago Health Research Foundation; and by national agencies in China, Japan (the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Scientific Research [A], No. 090357003), and the UK (a project grant from the West Midlands National Health Service Research and Development, and grant R2019EPH from the Chest, Heart and Stroke Association, Northern Ireland). The INTERMAP study has been accomplished through the work of the staff at local, national and international centers. A partial listing of colleagues is in ref. 11. We thank Dr. Y. Wu (Chinese Academy of Medical Sciences, Beijing, China), Dr. L. Zhu (Guangxi Medical University, Nanning, China), Dr. D. Guo (Yu County Hospital, Shanxi, China), Chinese Local Center Co-ordinators/Principal Investigators. NMR

signal processing and multivariate in-house software was developed by Dr. O. Cloarec, Dr. T. Ebbels, Dr. Kirill A. Veselkov, Dr. H. Keun and Dr. M. Rantalainen (Biomolecular Medicine, Department of Surgery and Cancer, Imperial College London).

REFERENCES

1. Lopez, AD.; Mathers, CD.; Ezzati, M.; Jamison, DT.; Murray, CJL. Global Burden of Disease and Risk Factors. Oxford University Press; New York: 2006.
2. Zhou B, Zhang H, Wu Y, Li Y, Yang J, Zhao L, X. Z. Ecological analysis of the association between incidence and risk factors of coronary heart disease and stroke in Chinese populations. CVD Prevention. 1998; 1:207–16.
3. Liu M, Wu B, Wang WZ, Lee LM, Zhang SH, Kong LZ. Stroke in China: epidemiology, prevention, and management strategies. Lancet Neurol. 2007; 6(5):456–64. [PubMed: 17434100]
4. Lawes CM, Vander Hoorn S, Law MR, Elliott P, MacMahon S, Rodgers A. Blood pressure and the global burden of disease 2000. Part II: estimates of attributable burden. J Hypertens. 2006; 24(3): 423–30. [PubMed: 16467640]
5. Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, Halsey J, Qizilbash N, Peto R, Collins R. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. Lancet. 2007; 370(9602):1829–39. [PubMed: 18061058]
6. Teo KK, Ounpuu S, Hawken S, Pandey MR, Valentin V, Hunt D, Diaz R, Rashed W, Freeman R, Jiang L, Zhang X, Yusuf S. Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. Lancet. 2006; 368(9536):647–58. [PubMed: 16920470]
7. Zhao L, Stamler J, Yan LL, Zhou B, Wu Y, Liu K, Daviglus ML, Dennis BH, Elliott P, Ueshima H, Yang J, Zhu L, Guo D. Blood pressure differences between northern and southern Chinese: role of dietary factors: the International Study on Macronutrients and Blood Pressure. Hypertension. 2004; 43(6):1332–7. [PubMed: 15117915]
8. Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, Ebbels T, De Iorio M, Brown IJ, Veselkov KA, Daviglus ML, Kesteloot H, Ueshima H, Zhao L, Nicholson JK, Elliott P. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008; 453(7193):396–400. [PubMed: 18425110]
9. Stamler J, Elliott P, Dennis B, Dyer AR, Kesteloot H, Liu K, Ueshima H, Zhou BF. INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). J Hum Hypertens. 2003; 17(9):591–608. [PubMed: 13679950]
10. Smith LM, Maher AD, Want EJ, Elliott P, Stamler J, Hawkes GE, Holmes E, Lindon JC, Nicholson JK. Large-scale human metabolic phenotyping and molecular epidemiological studies via ^{1}H NMR spectroscopy of urine: investigation of borate preservation. Anal Chem. 2009; 81(12):4847–56. [PubMed: 19453167]
11. Dennis B, Stamler J, Buzzard M, Conway R, Elliott P, Moag-Stahlberg A, Okayama A, Okuda N, Robertson C, Robinson F, Schakel S, Stevens M, Van Heel N, Zhao L, Zhou BF. INTERMAP: the dietary data--process and quality control. J Hum Hypertens. 2003; 17(9):609–22. [PubMed: 13679951]
12. 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. 2006; 78(13):4281–90. [PubMed: 16808434]
13. Hotelling H. The generalization of Student's ratio. Ann Math Stat. 1931; 2:360–78.
14. Trygg J, Wold S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. J Chemometr. 2003; 17(1):53–64.
15. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, Nicholson JK, Holmes E. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^{1}H NMR spectroscopic metabonomic studies. Anal Chem. 2005; 77(2):517–26. [PubMed: 15649048]
16. Chadeau-Hyam M, Ebbels TMD, Brown IJ, Chan Q, Stamler J, Huang CC, Daviglus ML, Ueshima H, Zhao L, Holmes E, Nicholson JK, Elliott P, De Iorio M. Metabolic profiling and the

metabolome-wide association study: significance level for biomarker identification. J Proteome Res. 2010 Just Accepted Manuscript, doi: 10.1021/pr1003449.

17. Holmes E, Loo RL, Cloarec O, Coen M, Tang H, Maibaum E, Bruce S, Chan Q, Elliott P, Stamler J, Wilson ID, Lindon JC, Nicholson JK. Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. Anal Chem. 2007; 79(7):2629–40. [PubMed: 17323917]
18. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. Anal Chem. 2005; 77(5):1282–9. [PubMed: 15732908]
19. Diaz E, Ferrandez A, Prieto MA, Garcia JL. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol Mol Biol Rev. 2001; 65(4):523–69. table of contents. [PubMed: 11729263]
20. Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Microbiol. 2005; 3(5):431–8. [PubMed: 15821725]
21. Schwab AJ, Tao L, Yoshimura T, Simard A, Barker F, Pang KS. Hepatic uptake and metabolism of benzoate: a multiple indicator dilution, perfused rat liver study. Am J Physiol Gastrointest Liver Physiol. 2001; 280(6):G1124–36. [PubMed: 11352805]
22. Yap IK, Li JV, Saric J, Martin FP, Davies H, Wang Y, Wilson ID, Nicholson JK, Utzinger J, Marchesi JR, Holmes E. Metabonomic and microbiological analysis of the dynamic effect of vancomycin-induced gut microbiota modification in the mouse. J Proteome Res. 2008; 7(9):3718–28. [PubMed: 18698804]
23. Yap IK, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. Urinary Metabolic Phenotyping Differentiates Children with Autism from Their Unaffected Siblings and Age-Matched Controls. J Proteome Res. 2010; 9(6):2996–3004. [PubMed: 20337404]
24. Selmer T, Andrei PI. p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of p-cresol. Eur J Biochem. 2001; 268(5):1363–72. [PubMed: 11231288]
25. Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A. 2004; 101(44):15718–23. [PubMed: 15505215]
26. Backhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc Natl Acad Sci U S A. 2007; 104(3):979–84. [PubMed: 17210919]
27. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006; 444(7122):1022–3. [PubMed: 17183309]
28. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006; 444(7122):1027–31. [PubMed: 17183312]
29. Elliott P, Stamler J, Nichols R, Dyer AR, Stamler R, Kesteloot H, Marmot M. Intersalt revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations. Intersalt Cooperative Research Group. BMJ. 1996; 312(7041):1249–53. [PubMed: 8634612]
30. Pischon T, Boeing H, Hoffmann K, Bergmann M, Schulze MB, Overvad K, van der Schouw YT, Spencer E, Moons KG, Tjonneland A, Halkjaer J, Jensen MK, Stegger J, Clavel-Chapelon F, Boutron-Ruault MC, Chajes V, Linseisen J, Kaaks R, Trichopoulou A, Trichopoulos D, Bamia C, Sieri S, Palli D, Tumino R, Vineis P, Panico S, Peeters PH, May AM, Bueno-de-Mesquita HB, van Duijnhoven FJ, Hallmans G, Weinehall L, Manjer J, Hedblad B, Lund E, Agudo A, Arriola L, Barricarte A, Navarro C, Martinez C, Quiros JR, Key T, Bingham S, Khaw KT, Boffetta P, Jenab M, Ferrari P, Riboli E. General and abdominal adiposity and risk of death in Europe. N Engl J Med. 2008; 359(20):2105–20. [PubMed: 19005195]
31. Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Wei H, Chen Y, Lu H, Zuo J, Su M, Qiu Y, Jia W, Xiao C, Smith LM, Yang S, Holmes E, Tang H, Zhao G, Nicholson JK, Li L, Zhao L. Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci U S A. 2008; 105(6):2117–22. [PubMed: 18252821]

32. Zuppi C, Messana I, Forni F, Ferrari F, Rossi C, Giardina B. Influence of feeding on metabolite excretion evidenced by urine ^{1}H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard). Clin Chim Acta. 1998; 278(1):75–9. [PubMed: 9877127]
33. Balasse EO, De Graef J, Neef MA. Alanine turnover in normal and diabetic dogs. Horm Metab Res. 1985; 17(11):554–8. [PubMed: 4077021]
34. Nicholson JK, O'Flynn MP, Sadler PJ, Macleod AF, Juul SM, Sonksen PH. Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. Biochem J. 1984; 217(2):365–75. [PubMed: 6696735]
35. Terentyeva EA, Hayakawa K, Tanae A, Katsumata N, Tanaka T, Hibi I. Urinary biotinidase and alanine excretion in patients with insulin-dependent diabetes mellitus. Eur J Clin Chem Clin Biochem. 1997; 35(1):21–4. [PubMed: 9156561]
36. Nicholson JK, Timbrell JA, Sadler PJ. Proton NMR spectra of urine as indicators of renal damage. Mercury-induced nephrotoxicity in rats. Mol Pharmacol. 1985; 27(6):644–51. [PubMed: 2860559]
37. Simpson DP. Citrate excretion: a window on renal metabolism. Am J Physiol. 1983; 244(3):F223–34. [PubMed: 6338740]
38. Wyss M, Kaddurah-Daouk R. Creatine and creatinine metabolism. Physiol Rev. 2000; 80(3):1107–213. [PubMed: 10893433]
39. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, van der Ouderaa F, Bingham S, Cross AJ, Nicholson JK. Susceptibility of human metabolic phenotypes to dietary modulation. J Proteome Res. 2006; 5(10):2780–8. [PubMed: 17022649]
40. Candy DJ. Occurrence and metabolism of scylloinositol in the locust. Biochem J. 1967; 103(3): 666–71. [PubMed: 4383051]
41. Needham J. Scyllitol in selachian ontogeny. Biochem J. 1929; 23(3):319–23. [PubMed: 16744216]
42. Yancey PH, Rhea MD, Kemp KM, Bailey DM. Trimethylamine oxide, betaine and other osmolytes in deep-sea animals: depth trends and effects on enzymes under hydrostatic pressure. Cell Mol Biol (Noisy-le-grand). 2004; 50(4):371–6. [PubMed: 15529747]
43. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas ME, Kochhar S, Stamler J, Holmes E, Elliott P, Nicholson JK. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. Am J Clin Nutr. 2010; 92
44. Brauer D, Teel MR. Metabolism of *trans*-aconitic acid in maize : II. Regulatory properties of two compartmented forms of citrate dehydrase. Plant Physiol. 1982; 70(3):723–7. [PubMed: 16662564]
45. Altekar WW, Rao MR. Microbiological dissimilation of tricarballylate and *trans*-aconitate. J Bacteriol. 1963; 85:604–13. [PubMed: 14042938]
46. Huang Z, Wu X, Stamler J, Rao X, Tao S, Friedewald WT, Liao Y, Tsai R, Stamler R, He H, et al. A north-south comparison of blood pressure and factors related to blood pressure in the People's Republic of China: a report from the PRC-USA Collaborative Study of Cardiovascular Epidemiology. J Hypertens. 1994; 12(9):1103–12. [PubMed: 7852756]
47. Reynolds K, Gu D, Muntner P, Wu X, Chen J, Huang G, Duan X, Whelton PK, He J. Geographic variations in the prevalence, awareness, treatment and control of hypertension in China. J Hypertens. 2003; 21(7):1273–81. [PubMed: 12817173]
48. Stamler J. Dietary salt and blood pressure. Ann N Y Acad Sci. 1993; 676:122–56. [PubMed: 8489127]
49. Zhou B, Zhang X, Zhu A, Zhao L, Zhu S, Ruan L, Zhu L, Liang S. The relationship of dietary animal protein and electrolytes to blood pressure: a study on three Chinese populations. Int J Epidemiol. 1994; 23(4):716–22. [PubMed: 8002184]
50. Zhou BF, Wu XG, Tao SQ, Yang J, Cao TX, Zheng RP, Tian XZ, Lu CQ, Miao HY, Ye FM, et al. Dietary patterns in 10 groups and the relationship with blood pressure. Collaborative Study Group for Cardiovascular Diseases and Their Risk Factors. Chin Med J (Engl). 1989; 102(4):257–61. [PubMed: 2507238]
51. Chen J, Zheng H, Bei JX, Sun L, Jia WH, Li T, Zhang F, Seielstad M, Zeng YX, Zhang X, Liu J. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. Am J Hum Genet. 2009; 85(6):775–85. [PubMed: 19944401]

52. Xu S, Yin X, Li S, Jin W, Lou H, Yang L, Gong X, Wang H, Shen Y, Pan X, He Y, Yang Y, Wang Y, Fu W, An Y, Wang J, Tan J, Qian J, Chen X, Zhang X, Sun Y, Wu B, Jin L. Genomic dissection of population substructure of Han Chinese and its implication in association studies. Am J Hum Genet. 2009; 85(6):762–74. [PubMed: 19944404]

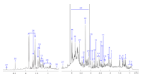

Figure 1.
Median urinary ^{1}H NMR spectrum of INTERMAP Chinese population samples, based on the first urine collection (N=747). **Key**: 1, Pentanoic/heptanoic acid; 2, Branched-chain amino acids (leucine, isoleucine, valine); 3, D-3-hydroxybutyrate; 4, Lactate; 5, 2-hydroxyisobutyrate; 6, Alanine; 7, Acetate; 8, *N*-acetyls of glycoprotein fragments (including uromodulin); 9, *N*-acetyl neuraminic acid; 10, Phenylacetylglutamine; 11, 4-cresyl sulfate; 12, Succinate; 13, Glutamine; 14, Citrate; 15, Dimethylamine; 16, Methylguanidine; 17, Trimethylamine; 18, Dimethylglycine; 19, Creatine; 20, Creatinine; 21, Prolinebetaine; 22, Trimethylamine *N*-oxide; 23, *Scyllo*-inositol; 24, Glycine; 25, Guanidinoacetate; 26, Hippurate; 27, *N*-methyl nicotinic acid; 28, *Trans*-aconitate; 29, Tyrosine; 30, Formate.

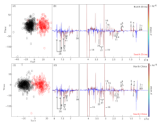

Figure 2.

(A) Cross-validated OPLS-DA scores plot derived from urinary NMR spectra of northern and southern Chinese population samples, based on the first urine collection; (B) Covariance plot showing color-coded significance of urinary metabolite differences between northern and southern Chinese populations, based on the first urine collection. Mean north-south differences in peak intensity for 7,100 spectral variables were assessed for statistical significance using Family Wise Error Rate <0.01, corresponding to $P < 4 \times 10^{-6}$ for group mean north-south differences by Student's t, for the two urine collections considered separately; (C) Cross-validated OPLS-DA scores plot derived from urinary NMR spectra of northern and southern Chinese population samples, based on the second urine collection; (D) Covariance plot showing color-coded significance of urinary metabolite differences between northern and southern Chinese populations, based on the second urine collection. **Key**: 1, Pentanoic/heptanoic acid; 2, Branched-chain amino acids; 4, Lactate; 5, 2-hydroxyisobutyrate; 6, Alanine; 8, *N*-acetyls of glycoprotein fragments (including uromodulin); 9, *N*-acetyl neuraminic acid; 11, 4-cresyl sulfate; 12, Succinate; 16, Methylguanidine; 18, Dimethylglycine; 19, Creatine; 21, Prolinebetaine; 23, *Scyllo*-inositol; 10, Phenylacetylglutamine; 26, Hippurate; 28, *Trans*-aconitate.

Table 1

^{1}H NMR-derived metabolites that differ significantly* between northern and southern Chinese participants

		minimum *P*-value†	
metabolite	**chemical shifts, ppm (multiplicity)**	**1st collection**	**2nd collection**
Higher in the north			
N-acetyls of glycoprotein fragments‡	1.95 – 2.04 (s)	2.9×10^{-69}	1.6×10^{-61}
Alanine	1.48 (d); 3.79 (q)	1.4×10^{-58}	5.2×10^{-64}
Branched-chain amino acids§	0.96 – 1.00 (overlapped resonances)	3.3×10^{-56}	2.0×10^{-41}
Unknown 1	1.82 (m)	1.2×10^{-51}	1.2×10^{-16}
N-acetyl neuraminic acid	2.06 (s)	1.5×10^{-47}	1.4×10^{-36}
Lactate	1.32 (d); 4.11 (q)	8.8×10^{-47}	1.4×10^{-43}
Pentanoic/heptanoic acid	0.86 – 0.89 (m) (overlapped broad resonances)	1.5×10^{-31}	5.5×10^{-22}
Dimethylglycine	2.93 (s); 3.72 (s)	1.1×10^{-26}	1.7×10^{-20}
Methylguanidine	2.84 (s)	7.0×10^{-21}	6.1×10^{-17}
Higher in the south			
Creatine	3.03 (s); 3.93 (s)	7.4×10^{-120}	2.0×10^{-127}
Prolinebetaine	3.11 (s); 3.31 (s)	3.5×10^{-86}	7.4×10^{-73}
2-hydroxyisobutyrate	1.36 (s)	6.2×10^{-54}	6.0×10^{-98}
Phenylacetylglutamine	0.89 (m); 1.33 (m); 1.55 (m); 1.92 (m); 2.11 (m); 2.26 (t); 3.66 (q); 4.18 (m); 7.36 (m)‖, 7.42 (m)‖	5.5×10^{-44}	5.9×10^{-54}
Scyllo-inositol	3.35 (s)	6.8×10^{-40}	9.5×10^{-60}
Succinate	2.41 (s)	9.3×10^{-38}	4.7×10^{-54}
4-cresyl sulfate	2.35 (s); 7.20 (m)#; 7.28 (m)#	5.6×10^{-30}	1.3×10^{-27}
Hippurate	3.98 (d); 7.55 (m)‖; 7.64 (m)‖; 7.84 (m)‖	1.2×10^{-26}	3.0×10^{-20}
Trans-aconitate	3.45 (s); 6.59 (s)	1.3×10^{-13}	1.1×10^{-12}

Abbreviations: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet.

*Mean north–south differences in peak intensity for 7100 spectral variables were assessed for statistical significance using Family Wise Error Rate < 0.01, corresponding to $P < 4 \times 10^{-6}$ for group mean north-south differences by Student's t, for the two urine collections considered separately.

†Minimum *P*-values for mean north–south differences in peak intensity among the spectral variables assigned to a particular metabolite, obtained separately for first and second urine collections, give a ranking of the discriminatory strength of the metabolites.

‡including uromodulin.

§isoleucine, leucine and valine.

‖AA'BB'C spin system.

#AA'BB' spin system.

Table 2

Cardiovascular disease risk factor and dietary differences, mean or prevalence, between northern and southern Chinese participants

Variable	North China (N=523)		South China (N=244)		P-value*
	Mean or %	(SD)	Mean or %	(SD)	
Systolic blood pressure, mm Hg	123.8	18.6	115.4	13.0	2.9×10^{-10}
Diastolic blood pressure, mm Hg	75.5	10.6	68.2	7.6	3.8×10^{-21}
Male, %	49.1		47.1		0.60
Age, years	48.8	5.9	48.9	5.6	0.83
Education, years	5.4	3.0	5.5	2.7	0.75
Body mass index, kg/m^2	23.8	3.5	21.8	2.6	2.2×10^{-15}
Current smoker, %	41.9		22.5		1.9×10^{-7}
Physical activity, hours/day moderate or heavy activity	4.6	3.6	8.8	2.0	5.0×10^{-58}
Special diet, %	6.3		0.8		0.001 †
Dr diagnosed heart attack, %	0.0		0.0		-
Dr diagnosed other heart disease, %	5.5		1.2		0.01 †
Dr diagnosed stroke, %	2.1		0.0		0.05 †
Dr diagnosed diabetes, reported insulin use, oral antidiabetic, or diabetic diet, %	1.1		0.0		0.22 †
Current drinker, %	43.4		47.1		0.36
14-day alcohol, all, g/24-h	7.6	19.2	9.0	22.0	0.38
14-day alcohol, drinkers only, g/24-h	17.4	26.1	19.0	28.9	0.60
Energy, kcal/24-h	2080	586	1962	554	0.009
Total protein, %kcal	11.7	1.4	13.7	2.0	8.4×10^{-48}
Animal protein, %kcal	1.5	1.6	4.5	2.5	2.9×10^{-72}
Vegetable protein, %kcal	10.3	1.1	9.3	1.3	1.5×10^{-24}
Total SFA, %kcal	4.5	1.9	6.1	2.0	7.8×10^{-26}
Total MFA, %kcal	7.5	2.8	9.2	2.6	9.0×10^{-15}
Total PFA, %kcal	5.8	2.2	5.9	2.2	0.36
Omega-3 PFA, %kcal	0.70	0.35	0.23	0.12	4.3×10^{-74}
Omega-6 PFA, %kcal	5.1	2.2	5.7	2.1	0.0003

Variable	North China (N=523)		South China (N=244)		P-value*
	Mean or %	(SD)	Mean or %	(SD)	
Cholesterol, mg/1,000 kcal	84.2	93.0	92.8	61.0	*0.19*
Keys dietary lipid score‡	15.7	10.6	22.1	8.1	3.6×10^{-16}
Starch, %kcal	58.8	8.9	51.9	11.1	2.1×10^{-19}
Total dietary fiber, g/1,000 kcal	14.4	3.5	13.9	4.2	*0.08*
Estimated total sugar, %kcal	8.5	4.3	8.9	6.9	*0.44*
Vitamin A, IU/1,000 kcal	1528	1067	3539	2050	8.6×10^{-60}
Retinol, μg/1,000 kcal	34.2	49.3	70.8	114.3	9.1×10^{-10}
Beta-carotene, μg/1,000 kcal	849	630	1982	1249	1.4×10^{-53}
Vitamin C, mg/1,000 kcal	37.0	17.7	45.3	22.1	3.1×10^{-8}
Total vitamin E, mg ATE/1,000 kcal	5.5	1.6	4.9	1.7	1.1×10^{-5}
Calcium, mg/1,000 kcal	136.5	48.4	175.0	62.5	1.0×10^{-19}
Magnesium, mg/1,000 kcal	133.2	38.7	198.2	27.2	1.8×10^{-93}
Iron, mg/1,000 kcal	7.8	1.4	7.8	2.3	*0.57*
Phosphorus, mg/1,000 kcal	377.4	75.7	563.3	66.2	7.0×10^{-149}
Selenium, μg/1,000 kcal	17.4	3.9	14.6	4.4	1.7×10^{-17}
Dietary sodium, mg/1,000 kcal	2318	645	1290	493	5.2×10^{-84}
Dietary potassium, mg/1,000 kcal	887	159	993	202	1.2×10^{-14}
Urinary sodium, mmol/24-h	271.4	88.3	139.2	55.5	1.8×10^{-80}
Urinary potassium, mmol/24-h	37.0	11.5	40.6	14.1	*0.0002*
Ratio, urinary sodium/potassium	7.8	2.4	3.7	1.5	2.0×10^{-96}
Glutamic acid, %kcal	3.1	0.4	2.5	0.4	6.5×10^{-66}
Cystine, %kcal	0.25	0.03	0.30	0.04	1.9×10^{-67}
Proline, %kcal	0.93	0.14	0.42	0.15	5.5×10^{-227}
Phenylalanine, %kcal	0.56	0.07	0.69	0.10	5.6×10^{-77}
Serine, %kcal	0.55	0.07	0.67	0.09	6.3×10^{-67}
Glycine, %kcal	0.48	0.09	0.66	0.12	2.3×10^{-98}
Alanine, %kcal	0.54	0.09	0.79	0.13	8.0×10^{-138}
Histidine, %kcal	0.26	0.04	0.36	0.07	5.5×10^{-94}

J Proteome Res. Author manuscript; available in PMC 2011 December 3.

Variable	North China (N=523)		South China (N=244)		P-value*
	Mean or %	(SD)	Mean or %	(SD)	
Threonine, %kcal	0.39	0.07	0.57	0.10	2.4×10^{-126}
Methionine, %kcal	0.19	0.04	0.27	0.05	1.2×10^{-108}
Lysine, %kcal	0.45	0.13	0.78	0.19	7.6×10^{-122}

Special diet: weight loss, weight gain, vegetarian, salt-reduced, diabetic, fat-modified, or other; CVD-DM diagnosis: history of heart attack, other heart disease, stroke, or diabetes

Abbreviations: ATE, α-tocopherol equivalents; kcal, kilocalories; MFA, monounsaturated fatty acids; PFA, polyunsaturated fatty acids; SFA, saturated fatty acids

*From Student's t-test or χ^2 test

†From Yates' continuity adjusted χ^2 test

‡Calculated as 1.35 (2 SFA − PFA) + 1.5 cholesterol$^{0.5}$, where SFA, PFA and cholesterol are expressed as above