SUPPLEMENTARY INFORMATION

Optimized Phenotypic Biomarker Discovery and Confounder Elimination via Covariate-Adjusted Projection to Latent Structures from Metabolic Spectroscopy Data

Authors
Joram M. Posma†‡*, Isabel Garcia-Perez‡§, Timothy M. D. Ebbels†, John C. Lindon†, Jeremiah Stamler†, Paul Elliott‡⊥, Elaine Holmes†⊥# and Jeremy K. Nicholson†⊥#

Affiliations
† Division of Integrative Systems Medicine and Digestive Diseases, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ, London, United Kingdom
‡ Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, W2 1PG, London, United Kingdom
§ Investigative Medicine, Department of Medicine, Faculty of Medicine, Imperial College London, W12 0NN, London, United Kingdom
∥ Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA
⊥ MRC-PHE Centre for Environment and Health, School of Public Health, Faculty of Medicine, Imperial College London, W2 1PG, London, United Kingdom
# MRC-NIHR National Phenome Centre, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ, London, United Kingdom

* Correspondence should be addressed to Joram Posma (jmp111@ic.ac.uk) and/or Jeremy Nicholson (j.nicholson@imperial.ac.uk), fax: +44 (0)20 7594 3226.
Supplementary Materials and Methods

Glossary of mathematical operations

PLS algorithm

Random Matrix Theory algorithm

Generation of simulated data sets

Supplementary Figures

Fig. S1. The effect size and overlap of simulated datasets.

Fig. S2. PCA pairs plot showing the structure of the simulated data.

Fig. S3. Score plots of external validation sets for the unadjusted model.

Fig. S4. Score plots of external validation sets for the adjusted model.

Fig. S5. Mean spectrum and consistently contributing metabolites labelled for unadjusted model.

Fig. S6. Statistical identification of metabolites.

Fig. S7. Correlation network.

Fig. S8. Dietary sodium/potassium ratio and physical activity.

Supplementary Tables

Table S1. Results of consistently contributing variables for dataset with 61% overlap.

Table S2. Results of consistently contributing variables for dataset with 41% overlap.

Table S3. Descriptive data of the Chinese INTERMAP population.

Table S4. Consistently contributing metabolites (and multiplicities) for OPLS (Yap et al.), OSC-PLS (new framework) and CA-(O)PLS.

Table S5. Abbreviations and full names of metabolites in the metabolic reaction network.
Glossary of mathematical operations

- $X_{ij}$: element of matrix $X$ on $i^{th}$ row and $j^{th}$ column
- $X^T$: transposition of matrix $X$
- $X^{-1}$: matrix inverse of $X$
- $\|X\|$: 2-norm of $X$
- $\text{rank}(X)$: rank of matrix $X$
- $0_{np}$: zero matrix of $n$ rows and $p$ columns
- $I$: identity matrix
- $\beta$: regression coefficients
- $r \leftarrow \text{corr}$: calculation of (Pearson) correlation matrix
- $\Sigma \leftarrow \text{cov}$: calculation of covariance matrix
- $\text{eig}$: eigenvalue decomposition
- $\lambda$: eigenvalues
- $\sim$: sampled from
- $\phi$: normal distribution
- $\epsilon$: small positive constant
- $\circ$: element-wise operation
- $\forall$: for all
- $\#$: number of
- $\delta_{ij}$: Kronecker delta of $i$ and $j$

**PLS algorithm** – To avoid the computationally expensive inflating step in non-iterative partial least squares (1) (NIPALS) we use a modified version (2) of the original SIMPLS algorithm (3) to deal with a data matrix ($X$) where the number of samples ($n$) is much smaller than the number of variables ($p$), $n<p$. The modification is regarding the first step, where the original SIMPLS (proposed for $n>p$), in the case of $n<p$, would result in a square matrix of order $p$. In the modified algorithm, a sample-sample association matrix ($A$) is defined as $A=XX^T$ (line 1, see pseudo-code below), a square matrix of order $n$, and both $X$ and $Y$ are assumed to be centered and, if required, scaled. The latent variables are calculated in an iterative manner until a stop criterion (specified number of components or the maximum number of $n$ components) is reached (lines 4–13). The algorithm gives the same results as the NIPALS algorithm for a univariate response, however with a significant time gain (3). Same as for
NIPALS, this algorithm can also be used to do OSC-PLS, for which the data matrix \( X \) is replaced by the orthogonal signal corrected (4) matrix \( X_{osc} \).

```plaintext
1: \( A \leftarrow XX^T \)
2: \( Y_0 \leftarrow Y \)
3: \( k \leftarrow 1 \)
4: while \( k \leq n \) do
5: \( R_k \leftarrow YY^TAY \)
6: \( t \leftarrow AR_k \)
7: \( T_k \leftarrow \frac{t}{\sqrt{t^Tt}} \sqrt{n-1} \)
8: \( P \leftarrow X^TT \)
9: \( C_k \leftarrow A T_k \)
10: \( Y \leftarrow Y - \frac{T_k}{T_k^TC_k} C_k^TY \)
11: \( \beta_k \leftarrow \left( R\left( \left( \frac{n-1}{C^TR} \right)^T I_k \right) \right)^T X \right)^T T^TY_0 \)
12: \( k \leftarrow k + 1 \)
end while
```

**Random Matrix Theory** – The number of un-correlated components (\( \nu \)) from the auto-correlation matrix of the regression coefficients is estimated using Random Matrix Theory (5) (RMT). RMT numerically computes a Jacobian matrix (J) that calculates the theoretical eigenvalue distribution of \( r_B \) (lines 1–27). The partial J can be used to estimate \( \nu \) (lines 28–29, see pseudo-code below).
1: \( r_\beta \leftarrow \text{corr}(\beta) \)
2: \( V, \lambda_\beta \leftarrow \text{eig}(r_\beta) \)
3: \( i \leftarrow 1 \)
4: for \( k = 1 \) to \( n \) do
5: for \( j = k \) to \( n \) do
6: \( r_V \leftarrow r_\beta \)
7: \( r_{V_{kj}} \leftarrow r_{V_{kj}} + \epsilon \)
8: \( r_{V_{jk}} \leftarrow r_{V_{jk}} + \epsilon \)
9: \( Q, \lambda_V \leftarrow \text{eig}(r_V) \)
10: \( d_\lambda \leftarrow \frac{\lambda_V - \lambda_\beta}{\epsilon} \)
11: \( D \leftarrow \frac{V^T(Q - V)}{\epsilon} \)
12: \( f \leftarrow 1 \)
13: for \( h = 1 \) to \( n \) do
14: for \( g = h \) to \( n \) do
15: \( P_f \leftarrow D_{hg} \)
16: \( f \leftarrow f + 1 \)
17: end for
18: end for
19: for \( h = 1 \) to \( n \) do
20: \( J_{hi} \leftarrow d_\lambda \)
21: end for
22: for \( h = (n + 1) \) to \( \frac{n(n+1)}{2} \) do
23: \( J_{hi} \leftarrow P_{h-n} \)
24: end for
25: \( i \leftarrow i + 1 \)
26: end for
27: end for
28: \( u \leftarrow \max(J_{kj}) \forall 1 \leq k \leq n, 1 \leq j \leq (i - 1) \)
29: \( \nu \leftarrow \#(\lambda_\beta > u) \)
Simulated data – In order to compare our method to standard PLS and OSC-PLS we simulated 100 data sets with 50 case ($Y_i=1$) and 50 control ($Y_i=0$) objects. In the data generation we added two confounding effects, one which was non-orthogonal to $Y$ ($r = \pm 0.36$) and another which was almost orthogonal to $Y$ ($r = 0.04$). The Pearson correlation between the two confounders was $r = \pm 0.2$. From the total of 2000 variables in each data set sampled from a normal distributions with unit variance ($\sigma^2 = \sigma = 1$), $X_{ij} \sim \varphi(\mu, \sigma^2=1)$, 10% of variables were sampled using a different mean for one class to induce a class-separation, these are considered to be true positives. The same goes for both confounders; with the addition that there is a 25% overlap in affected variables between each confounder and the 10% of variables affected by the class separation. The difference in mean between distributions (effect size) was 1 (61% overlap) and 1.645 (41% overlap) for the first 50 and the second 50 data sets, respectively (see Supplementary Figure 1). In total there are $2^3$ unique classes, and the class of each object $i$ is coded as a 3-tuple $S_{xyz}$, where $x$, $y$ and $z$ indicate whether it is case or control, affected or not by confounder 1 and affected or not by confounder 2, respectively. The same goes for the 2000 variables, and each variable $j$ has a 3-tuple code $V_{xyz}$, again where $x$, $y$ and $z$ indicate whether the variable is affected by the case/control status, confounder 1 and confounder 2, respectively. Additionally, the effect of each factor ($Y$, confounder 1, confounder 2) was assigned a sign for each variable at random, where for variable $j$, $a_j$, $b_j$ and $c_j$ are the signs for $Y$, confounder 1 and confounder 2, respectively. This results in a total of $3^3$ unique types of variables with different effects for the covariates.

Generation of simulated data – The data for each variable is drawn from specific normal distributions (with $\sigma^2 = \sigma = 1$) depending on the sample, effect size (es) and effect signs ($a_j$, $b_j$ and $c_j$). Data for object $i$ and a variable $j$ which is unaffected by any of the covariates is simply drawn from a normal distribution as follows:

$$X_{ij} \sim N(0,\sigma)$$

Data for variable $j$ which only contains information of case/control status is calculated using:

$$X_{ij} = \delta_{y,1bc} \times N(a_j \times es,\sigma) + \delta_{y,0bc} \times N(0,\sigma)$$

In this equation only one Kronecker Delta has a value of 1 depending on object $i$ 3-tuple. The same goes for variables which only contain information from confounders 1 and 2:
\[ X_{ij} = \delta_{i,a1c} \times N(b_j \times es, \sigma) + \delta_{i,a0c} \times N(0, \sigma) \]

\[ X_{ij} = \delta_{i,a01} \times N(c_j \times es, \sigma) + \delta_{i,a00} \times N(0, \sigma) \]

The equations are expanded when multiple effects play a role for a variable. Data for object \( i \) for variables with information on \( Y \) and confounder \( 1 \) is generated as follows:

\[ X_{ij} = \delta_{i,11c} \times \left( N(a_j \times es, \sigma) + N(b_j \times es, \sigma) \right) + \delta_{i,10c} \times N(a_j \times es, \sigma) + \delta_{i,01c} \times N(b_j \times es, \sigma) + \delta_{i,00c} \times N(0, \sigma) \]

Note again that only one Kronecker Delta has a value of 1 in this equation. The \( c \) in the Kronecker Deltas indicates that the value for confounder 2 from the object 3-tuple has no effect here as the variable is unaffected by confounder 2. The same goes for variables affected by \( Y \) and confounder 2 (where confounder 1 plays no role, as indicated by the \( b \) in the Kronecker Deltas):

\[ X_{ij} = \delta_{i,a11} \times \left( N(b_j \times es, \sigma) + N(c_j \times es, \sigma) \right) + \delta_{i,a10} \times N(b_j \times es, \sigma) + \delta_{i,a01} \times N(c_j \times es, \sigma) + \delta_{i,a00} \times N(0, \sigma) \]

And also for variables where \( Y \) plays no role (indicated by a in the Kronecker Deltas), but both confounder 1 and 2 play a role:

\[ X_{ij} = \delta_{i,a11} \times \left( N(b_j \times es, \sigma) + N(c_j \times es, \sigma) \right) + \delta_{i,a10} \times N(b_j \times es, \sigma) + \delta_{i,a01} \times N(c_j \times es, \sigma) + \delta_{i,a00} \times N(0, \sigma) \]

Last, the data for variables affected by \( Y \), confounder 1 and confounder 2 is drawn for the following distributions:

\[ X_{ij} = \delta_{i,111} \times \left( N(a_j \times es, \sigma) + N(b_j \times es, \sigma) + N(c_j \times es, \sigma) \right) + \delta_{i,110} \times \left( N(a_j \times es, \sigma) + N(b_j \times es, \sigma) \right) + \delta_{i,101} \times \left( N(b_j \times es, \sigma) + N(c_j \times es, \sigma) \right) + \delta_{i,100} \times N(0, \sigma) \]

\[ X_{ij} \times N(a_j \times es, \sigma) + \delta_{i,011} \times \left( N(b_j \times es, \sigma) + N(c_j \times es, \sigma) \right) + \delta_{i,010} \times N(b_j \times es, \sigma) + \delta_{i,001} \times N(c_j \times es, \sigma) + \delta_{i,000} \times N(0, \sigma) \]

All of these individual equations can be combined into a single equation to generate the data. Only one Kronecker Delta for the variable \( j \) and one Kronecker Delta for the object \( i \) have a value of 1, hence ultimately only a few values are used to generate \( X_{ij} \).
We refer to **Supplementary Figure 2** for an example of the multivariate data structure of the simulated data sets shown by principal component analysis.

If a variable with 3-tuple of type $V_{1yz}$ is found to be significant it is considered a true positive finding, as these variables contain information of case/control status. All other variables ($V_{0yz}$) are considered to be false positives if they are found to be significantly contributing to the models. In order to compare CA-(O)PLS with PLS and OSC-PLS for the simulated data sets, the same seed for the random number generator was used, to ensure the random sampling was performed exactly the same for all methods. All simulated data sets are available upon request.
**SUPPLEMENTARY FIGURES**

Supplementary Figure 1. The effect size (es) determines the difference between the two sampling distributions (red and blue), the larger the effect size, the less overlap there is, thus the clearer the separation. The grey area highlights the overlap between the distributions. The left figure shows the resulting overlap (61%) for es=1, the right figure for es=1.645 (41%).
Supplementary Figure 2. Principal Component Analysis pairs plot of the first 5 components of an auto-scaled simulated dataset with 61% overlap between variables. The $R^2_X$ shows the variance of $X$ explained by each component. Labels: ● $S_{000}$, ● $S_{010}$, ● $S_{001}$, ● $S_{100}$, ● $S_{110}$, ● $S_{101}$ and ● $S_{111}$. 
Supplementary Figure 3. a) Score (prediction) plot of urine collection 2 predicted by the unadjusted model of urine collection 1. b) Score (prediction) plot of urine collection 1 predicted by the unadjusted model of urine collection 2. North Chinese samples (Beijing and Shanxi) are shown as red circles and South Chinese (Guangxi) as cyan crosses. The prediction accuracy and robustness of both models was good ($R^2=0.72/0.71$ and $Q^2=0.68/0.67$) and the external validation closely matched the internal (unbiased) validation, $Q^2_{ext}$ of 0.61 and 0.64 for models 1 and 2, respectively.
Supplementary Figure 4. Score plots of models adjusted for age, gender, BMI, (on medication for) HBP, smoking status, physical activity, Na:K ratio and total intake of fats. a) Score (prediction) plot of urine collection 2 predicted by the covariate adjusted model of urine collection 1. b) Score (prediction) plot of urine collection 1 predicted by the covariate adjusted model of urine collection 2. North Chinese samples (Beijing and Shanxi) are shown as red circles and South Chinese (Guangxi) as cyan crosses. The prediction accuracy and robustness of both models decreased compared to the unadjusted models, however the cross validation robustness is still good ($R^2=0.54/0.64$ and $Q^2=0.50/0.60$, for model 1/2) and the external validation closely matched the internal (unbiased) validation, $Q_{ext}^2$ of 0.46 and 0.57 for models 1 and 2, respectively.
Supplementary Figure 5. Top shows the average $^1$H NMR spectrum from the first urine collection samples. The bottom shows the contribution of each variable. In red, metabolites are shown that are significantly higher in the north (compared to the south), and in cyan metabolites significantly lower in the north are shown. For each variable the highest q-value across both models is shown. Labels: 1 – fatty acids (C5 – C10), 2 – 2-oxoisocaprate, 3 – leucine, 4 – valine, 5 – isoleucine, 6 – unknown (1.15 (s), 3.49 (d), 3.61 (d), 3.67 (m), 3.83 (m)), 7 – ethanol, 8 – ethyl glucuronide, 9 – lactate, 10 – 2-hydroxyisobutyrate, 11 – unknown (1.42 (d), 1.46 (d), 1.51 (d)), 12 – alanine, 13 – lysine, 14 – unknown (1.82 (m), 3.52 (s)), 15 – phenylacetylglutamine, 16 – N-acetyl-S-(1Z)-propenyl-cysteine-sulfoxide, 17 – N-acetyl neuraminic acid, 18 – glutamine, 19 – acetone, 20 – prolinebetaene, 21 – unknown (2.32 (d), 2.34 (d), 2.38 (d), 2.40 (d), 3.52 (m)), 22 – 4-cresyl sulfate, 23 – succinate, 24 – citrate, 25 – beta-aminoisobutyrate, 26 – dimethylamine, 27 – sarcosine, 28 – unknown methin metabolite (2.76 (s)), 29 – N-acetyl-S-methyl-L-cysteine sulfoxide, 30 – unknown methin metabolite (2.81 (s)), 31 – S-methyl-L-cysteine sulfoxide (methin), 32 – trimethylamine, 33 – dimethylglycine, 34 – unknown (1.84 (m), 2.78 (m), 2.95 (s), 3.36 (m), 3.59 (m), 3.62 (m)), 35 – creatine, 36 – creatinine, 37 – N6,N6,N6-trimethyllysine, 38 – dimethylsulfone, 39 – O-acetylcarnitine, 40 – choline, 41 – carnitine, 42 – histidine, 43 – taurine, 44 – scyllo-inositol, 45 – trans-aconitate, 46 – 4-hydroxyphenylacetate, 47 – glucose, 48 – glycine, 49 – 1-methylhistidine, 50 – 3-methylhistidine, 51 – guanidinoacetate, 52 – 4-hydroxyhippurate, 53 – unknown (3.96 (d), 7.30 (t), 7.42 (t)), 54 – hippurate, 55 – N-methyl nicotinic acid, 56 – N-methyl nicotinamide, 57 – tyrosine, 58 – 3-hydroxymandelate, 59 – tryptophan/tryptamine, 60 – pseudouridine, 61 – N-methyl-2-pyridone-5-carboxamide, 62 – formate.
Supplementary Figure 6. Statistical pseudo-spectra obtained with STORM for significant metabolites with multiple visible peaks in the NMR spectrum. From top to bottom, left to right: 2-oxoisocaproate, leucine, valine, unknown 1.15 (s), 3.49 (d), 3.61 (d), 3.67 (m), 3.83 (m), ethyl glucuronide, unknown 1.42 (d), 1.46 (d), unknown 1.82 (m), 3.52 (s), N-acetyl-S-(1Z)-propenyl-cysteine-sulfoxide 1.97 (dd) 2.03 (s), N-acetyl-S-(1Z)-propenyl-cysteine-sulfoxide 6.49 (dq) 6.65 (dq), glutamine, proline betaine, unknown 2.32 (d), 2.34 (d), 2.38 (d), 2.40 (d), 3.52 (m), sarcosine, unknown 1.84 (ddd?), 2.78 (?), 2.95 (s, 2xCH3), 3.36 (dd, 2xCH), 3.59 (m), 3.62 (m), creatine, histidine, dimethyl sulfone, O-acetylcarnitine, carnitine, taurine, 1-methylhistidine, 4-hydroxyhippurate, N-methyl nicotinic acid and tyrosine.
**Supplementary Figure 7.** Correlation network of urinary metabolites associated with the difference between north and south Chinese individuals and the confounding factors. Metabolites listed here are no longer significant after covariate adjustment, and the correlations here may indicate which confounders are associated with the metabolites. The confounding factors are shown in grey. The average Spearman rank correlations of both urine collections is only shown if it had the same sign for both urine collections and both urine collections passed the multiple testing Bonferroni threshold of $p<1.9\times10^{-4}$. 
Supplementary Figure 8. Dietary sodium-to-potassium ratio for both visits plotted against the physical activity level, defined as the number of self-reported hours a day of moderate-to-heavy physical activity. Both figures show that physical activity and Na:K-ratio are anti-correlated and different between northern and southern Chinese individuals.
**SUPPLEMENTARY TABLES**

**Supplementary Table 1.** Full results table showing the mean percentage (95% confidence intervals are given in parenthesis) of consistently-and-similarly contributing associations for the simulated data using an effect size of 1 (61% overlap). The type of variable is shown as $J_{x,y,z}$, where $x$, $y$ and $z$ are the influence of the response variable, confounder 1 and confounder 2, respectively. The value for $x$, $y$ and $z$ indicate the type of variable and how each factor influences the generated variable, e.g. $\pm J_{1,0,1}$ indicates the variable is influenced by the response variable and confounder 2 only, and that they have the same sign, and $\pm J_{1,-1,0}$ means that the variable is influenced by both the response variable and the first confounder, however the sign of both is inverse. The bottom section shows the total number of true positives (TP), false negatives (FN, type-II errors), false positives (FP, type-I errors) and true negatives (TN).

| Type of variable | PLSDA       | OSC-PLSDA  | C₁A-OPLS | C₂A-OPLS |
|------------------|-------------|------------|----------|----------|
| $\pm J_{1,1,1}$  | 96.36       | 91.60      | 91.63    | 88.99    |
|                  | (93.16 – 99.55) | (87.52 – 95.68) | (87.14 – 96.12) | (84.33 – 93.64) |
| $\pm J_{1,1,0}$  | 97.73       | 91.38      | 98.21    | 91.50    |
|                  | (96.68 – 98.79) | (88.32 – 94.44) | (97.47 – 98.95) | (88.45 – 94.54) |
| $\pm J_{1,1,-1}$ | 97.19       | 86.83      | 97.25    | 90.01    |
|                  | (95.06 – 99.31) | (81.56 – 92.11) | (95.42 – 99.08) | (85.14 – 94.88) |
| $\pm J_{1,0,1}$  | 98.96       | 99.64      | 94.10    | 99.37    |
|                  | (98.50 – 99.43) | (99.36 – 99.91) | (92.69 – 95.52) | (99.03 – 99.72) |
| $\pm J_{1,0,0}$  | 99.43       | 99.42      | 98.06    | 99.42    |
|                  | (99.24 – 99.62) | (99.24 – 99.60) | (97.57 – 98.54) | (99.24 – 99.60) |
| $\pm J_{1,0,-1}$ | 99.20       | 99.22      | 97.67    | 99.39    |
|                  | (98.77 – 99.63) | (98.78 – 99.66) | (96.86 – 98.49) | (98.98 – 99.79) |
| $\pm J_{1,-1,1}$ | 95.89       | 92.51      | 94.64    | 90.61    |
|                  | (93.42 – 98.37) | (88.32 – 96.70) | (91.67 – 97.61) | (86.08 – 95.15) |
| $\pm J_{1,-1,0}$ | 97.95       | 91.44      | 98.29    | 91.74    |
|                  | (97.11 – 98.79) | (88.67 – 94.21) | (97.60 – 98.99) | (89.03 – 94.46) |
| $\pm J_{1,-1,-1}$ | 96.33       | 85.95      | 98.12    | 90.70    |
|                  | (93.93 – 98.72) | (80.22 – 91.68) | (96.52 – 99.72) | (85.98 – 95.42) |
| $\pm J_{0,1,1}$  | 2.89        | 26.03      | 0        | 26.36    |
|                  | (1.82 – 3.96) | (22.46 – 29.60) | (0 – 0) | (22.89 – 29.84) |
| $\pm J_{0,1,0}$  | 4.08        | 30.86      | 0.39     | 28.51    |
|                  | (3.69 – 4.46) | (28.33 – 33.39) | (0.27 – 0.51) | (26.22 – 30.80) |
| $\pm J_{0,1,-1}$ | 4.14        | 28.29      | 0        | 29.23    |
|                  | (2.75 – 5.53) | (24.51 – 32.07) | (0 – 0) | (25.46 – 32.99) |
| $\pm J_{0,0,1}$  | 0.93        | 0.63       | 0.07     | 0.58     |
|                  | (0.76 – 1.10) | (0.47 – 0.78) | (0.02 – 0.12) | (0.44 – 0.72) |
| $\delta J_{0,0,0}$ | 1.12        | 0.61       | 0.33     | 0.59     |
|                  | (1.03 – 1.21) | (0.54 – 0.68) | (0.28 – 0.38) | (0.51 – 0.66) |
| $\pm \delta J_{1,3,2}$ (TP) | 98.85       | 97.27      | 97.54    | 97.34    |
|                  | (98.58 – 99.13) | (96.76 – 97.78) | (96.98 – 98.10) | (96.83 – 97.86) |
| FN = 1 – TP      | 1.15        | 2.73       | 2.46     | 2.66     |
|                  | (0.87 – 1.42) | (2.22 – 3.24) | (1.90 – 3.02) | (2.14 – 3.17) |
| $\pm \delta J_{0,3,2}$ (FP) | 1.35        | 3.09       | 0.31     | 2.91     |
|                  | (1.26 – 1.44) | (2.92 – 3.26) | (0.26 – 0.36) | (2.75 – 3.06) |
| TN = 1 – FP      | 98.65       | 96.91      | 99.69    | 97.09    |
|                  | (98.56 – 98.74) | (96.74 – 97.08) | (99.64 – 99.74) | (96.94 – 97.25) |
Supplementary Table 2. Full results table showing the mean percentage (95% confidence intervals are given in parenthesis) of consistently-and-similarly contributing associations for the simulated data using an effect size of 1.645 (41% overlap).

| Type of variable | PLSDA       | OSC-PLSDA   | \(C_1\)-A-OPLS | \(C_2\)-A-OPLS |
|------------------|-------------|-------------|-----------------|-----------------|
| \(\pm \delta J_{1,1,1}\) | 100 | 98.33 | (96.42 – 100) | (91.12 – 100) |
| \(\pm \delta J_{1,1,0}\) | 100 | 98.83 | 100 | 98.95 |
| \(\pm \delta J_{1,1,1}\) | 100 | 93.61 | 100 | 97.15 |
| \(\pm \delta J_{1,0,1}\) | 100 | 100 | (99.68 – 100) | (100 – 100) |
| \(\pm \delta J_{1,0,0}\) | 100 | 100 | 100 | 100 |
| \(\pm \delta J_{1,0,1}\) | 100 | 100 | 100 | 100 |
| \(\pm \delta J_{1,1,1}\) | 100 | 98.13 | 100 | 98.32 |
| \(\pm \delta J_{1,1,0}\) | 100 | 96.28 | 100 | 96.81 |
| \(\pm \delta J_{0,1,1}\) | 3.00 | 36.59 | 0 | 37.42 |
| \(\pm \delta J_{0,1,0}\) | 4.10 | 42.99 | 0.11 | 39.34 |
| \(\pm \delta J_{0,1,1}\) | 3.26 | 38.32 | 0 | 39.83 |
| \(\pm \delta J_{0,0,1}\) | 0.81 | 0.44 | 0 | 0.39 |
| \(\pm \delta J_{0,0,0}\) | 0.81 | 0.33 | 0.14 | 0.30 |
| \(\pm \delta J_{1,1,1}\) (TP) | 100 | 99.66 | 99.99 | 99.64 |
| FN = 1 – TP | 0 | 0.44 | 0.01 | 0.36 |
| \(\pm \delta J_{1,0,1}\) (FP) | 1.07 | 3.83 | 0.13 | 3.55 |
| TN = 1 – FP | 98.93 | 96.17 | 99.87 | 96.45 |
| \(\pm \delta J_{0,1,1}\) | (98.84 – 99.01) | (96.05 – 96.28) | (99.85 – 99.90) | (96.33 – 96.57) |
**Supplementary Table 3.** Descriptive statistics of the Chinese INTERMAP population showing the mean (s.e.) or percentage and significance* for different factors for both visits.

|                          | North n = 523 | South n = 244 | p-value† |
|--------------------------|--------------|---------------|---------|
| **Participants:**        |              |               |         |
| Gender (% men)           | 49.14%       | 47.13%        | 6.04e-01|
| Age (years)              | 48.81 (5.86) | 48.91 (5.64)  | 8.23e-01|
| Education (years)        | 5.40 (2.96)  | 5.47 (2.70)   | 7.40e-01|
| **SBP (mmHg) – mean of visit 1** | 124.83 (19.63) | 116.98 (14.95) | 1.79e-09† |
| **SBP (mmHg) – mean of visit 2** | 122.71 (18.78) | 113.73 (12.75) | 3.35e-14† |
| **DBP (mmHg) – mean of visit 1** | 75.60 (11.34)  | 69.09 (8.99)  | 9.04e-17† |
| **DBP (mmHg) – mean of visit 2** | 75.42 (10.77)  | 67.27 (7.48)  | 1.11e-30† |
| Hypertension, HBP        | 8.99%        | 0.41%         | 4.93e-06† |
| HBP or HBP medication    | 22.75%       | 5.33%         | 2.61e-09† |
| **Body mass index (kg m$^{-2}$)** | 23.79 (3.49)  | 21.77 (2.57)  | 2.23e-18† |
| Diabteses Mellitus, DM   | 1.15%        | 0%            | 9.30e-02 |
| Diagnosed heart disease  | 7.65%        | 1.23%         | 3.19e-04† |
| CVD/HBP/DM/lipid meds.   | 12.62%       | 2.87%         | 1.82e-05† |
| **Smoking**              | 41.87%       | 22.54%        | 1.95e-07† |
| Physical activity (hours/day) | 4.55 (3.57)  | 8.85 (2.03)   | 7.07e-78† |
| Alcohol drinking         | 56.41%       | 52.87%        | 3.59e-01|

**Visit 1 dietary intakes**

|                          | North n = 523 | South n = 244 | p-value† |
|--------------------------|--------------|---------------|---------|
| Energy intake (kcal/day)  | 2131.33 (675.86) | 1980.82 (620.17) | 2.48e-03 |
| Na (mg/1000kcal)         | 2323.45 (823.26) | 1361.66 (609.41) | 3.03e-59† |
| K (mg/1000kcal)          | 900.77 (222.97)  | 1034.33 (287.57) | 4.23e-10† |
| Na:K-ratio               | 2.69 (1.06)    | 1.37 (0.65)    | 6.17e-78† |
| Cholesterol intake (mg/1000kcal) | 84.04 (112.11) | 96.51 (77.70)  | 7.46e-02 |
| Total fat intake (%kcal) | 18.83 (6.81)   | 22.87 (6.91)   | 1.96e-13† |
| Alcohol intake (%kcal)   | 2.22 (5.35)    | 2.87 (6.88)    | 1.92e-01|
| Fibre intake (g/1000kcal) | 14.23 (4.20)  | 14.17 (5.14)   | 8.54e-01|

**Visit 2 dietary intakes**

|                          | North n = 523 | South n = 244 | p-value† |
|--------------------------|--------------|---------------|---------|
| Energy intake (kcal/day)  | 2028.05 (607.81) | 1943.36 (617.23) | 7.60e-02 |
| Na (mg/1000kcal)         | 2313.50 (782.81) | 1217.47 (550.87) | 3.27e-82† |
| K (mg/1000kcal)          | 873.03 (172.57)  | 951.69 (203.07)  | 2.66e-07† |
| Na:K-ratio               | 2.72 (0.98)    | 1.32 (0.60)    | 4.18e-96† |
| Cholesterol intake (mg/1000kcal) | 84.46 (114.85) | 89.00 (70.88)  | 5.02e-01 |
| Total fat intake (%kcal) | 18.64 (6.88)   | 22.32 (7.01)   | 2.92e-11† |
| Alcohol intake (%kcal)   | 2.13 (5.46)    | 2.90 (7.20)    | 1.38e-01|
| Fibre intake (g/1000kcal) | 14.50 (4.20)  | 13.55 (5.22)   | 1.33e-02|

*Calculated using two-sample t-test or $\chi^2$-test as appropriate. †Significant according to a Bonferroni threshold of $p<7.7\times10^{-4}$, for both visits.
Supplementary Table 4. Metabolites that differ significantly between northern and southern Chinese individuals for the study of Yap et al., our unadjusted OSC-PLSDA model and the CA-(O)PLSDA model. Differences between the unadjusted and adjusted model are due to the covariate adjustment. A letter indicates in what the group higher relative concentrations were found, ‘N’ for northern Chinese and ‘S’ for southern Chinese, a ‘–’ indicates the metabolite was not found to be significant.

| Metabolite                          | Yap et al. | Unadjusted | Adjusted |
|-------------------------------------|------------|------------|----------|
| Fatty acids (C5 – C10) 0.88 (m), 1.31 (m), 1.56 (m), 2.19 (m) | N          | N          | –        |
| 2-Oxoisocaproate 0.94 (d)           | –          | N          | N        |
| Leucine 0.97 (2d)                   | N          | N          | N        |
| Valine 0.99 (d), 1.05 (d)           | N          | N          | N        |
| Isoleucine 1.01 (d)                 | N          | N          | –        |
| UNK 1.15 (s), 3.49 (d), 3.61 (d), 3.67 (m), 3.83 (m) | –          | N          | N        |
| Ethanol 1.19 (t), 3.67 (q)          | –          | S          | –        |
| Ethyl glucuronide 1.23 (t)          | –          | S          | S        |
| Lactate 1.33 (d)                    | N          | N          | –        |
| 2-Hydroxyisobutyrate 1.36 (s)       | S          | S          | S        |
| UNK 1.42 (d), 1.46 (d)              | –          | N          | N        |
| Alanine 1.48 (d)                    | N          | N          | –        |
| Lysine 1.48 (m), 1.73 (m), 1.91 (m), 3.03 (t) | –         | S          | –        |
| UNK 1.82 (m), 3.52 (s)              | N          | N          | N        |
| Phenylacetylglutamine 1.92 (m), 2.11 (m), 2.27 (m), 3.67 (m), 4.19 (m), 7.36 (t), 7.43 (t) | S          | S          | –        |
| N-Acetyl glycoproteins 1.95–2.04    | N          | –          | –        |
| N-acetyl-S-(1Z)-propenyl-cysteine sulfoxide 1.96 (dd), 2.03 (s), 6.49 (dq), 6.65 (dq) | –         | N          | N        |
| N-Acetyl neuraminic acid 2.06 (s)   | N          | N          | –        |
| Glutamine 2.14 (m), 2.46 (m)        | –          | N          | N        |
| Acetone 2.24 (s)                    | –          | S          | S        |
| Proline betaine 2.30 (m), 2.51 (m), 3.11 (s), 3.30 (s), 3.55 (m), 4.08 (m) | S          | S          | S        |
| UNK 2.32 (d), 2.34 (d), 2.38 (d), 2.40 (d), 3.52 (m) | –         | N          | N        |
| 4-Cresyl sulfate 2.35 (s), 7.21 (d), 7.29 (d) | S          | S          | –        |
| Succinate 2.41 (s)                  | S          | S          | –        |
| Citrate 2.54 (d), 2.68 (d)          | S          | S          | –        |
| Beta-Aminoisobutyrate (3-aminoisobutyrate) 1.20 (d), 2.61 (m), 3.04 (2d), 3.11 (2d) | S          | –          | –        |
| Dimethylamine 2.72 (s)              | –          | N          | –        |
| Sarcosine 2.74 (s), 3.61 (s)        | –          | N          | N        |
| UNK methion metabolite 2.76 (s)     | –          | N          | –        |
| N-Acetyl-S-methyl-cysteine sulfoxide 2.78 (s) | –         | N          | –        |
| UNK methion metabolite 2.81 (s)     | –          | N          | –        |
| S-Methyl-cysteine sulfoxide (methiin) 2.84 (s) | N         | N          | –        |
| Trimethylamine 2.88 (s)             | –          | N          | –        |
| Dimethylglycine 2.93 (s)            | N          | N          | N        |
| UNK 1.84 (ddd?), 2.78 (t?), 2.95 (s, 2xCH3), 3.36 (dd, 2xCH), 3.59 (m), 3.62 (m) | –         | N          | N        |
| Creatine 3.04 (s), 3.93 (s)         | S          | S          | S        |
| Creatinine 3.06 (s), 4.06 (s)       | –          | S          | –        |
| N6,N6,N6-Trimethyllysine 3.12 (s)   | S          | S          | S        |
| Histidine 3.14 (dd), 3.25 (dd), 4.00 (dd), 7.09 (s), 7.84 (s) | –         | N          | N        |
| Dimethylsulfone 3.15 (s)            | –          | N          | N        |
| O-Acetylcarnitine 2.15 (s), 3.19 (s) | –          | S          | S        |
| Choline 3.20 (s), 3.52 (m), 4.07 (m) | –          | N          | –        |
| Carnitine 2.44 (dd), 3.23 (s), 3.43 (m) | –         | S          | S        |
| Taurine 3.26 (t), 3.42 (t)          | –          | S          | S        |
| Scylo-Inositol 3.34 (s)             | S          | S          | –        |
| Trans-Aconitate 3.45 (s), 6.59 (s)  | S          | S          | –        |
| 4-Hydroxyphenylacetate 3.45 (s), 6.87 (d), 7.17 (d) | –         | S          | –        |
| Glucose 3.25 (dd), 3.42 (m), 3.49 (m), 3.54 (dd), 3.74 (m), 3.84 (m), | –         | N          | –        |
### Metabolite a, b

| Metabolite                  | Yap et al. | Unadjusted | Adjusted |
|-----------------------------|------------|------------|----------|
| 3.91 (dd)                   | –          | N          | –        |
| Glycine 3.57 (s)            | –          | S          | S        |
| 1-Methylhistidine 3.21 (2d), 3.29 (2d), 3.69 (s), 3.92 (t), 7.02 (s), 7.85 (s) | –          | S          | –        |
| 3-Methylhistidine 3.17 (2d), 3.25 (2d), 3.69 (s), 7.00 (s), 7.62 (s) | –          | S          | –        |
| Guanidinoacetate 3.80 (s)   | –          | S          | –        |
| 4-Hydroxyhippurate 3.95 (s), 6.97 (d), 7.76 (d) | –          | S          | S        |
| UNK 3.96 (d), 7.30 (t), 7.42 (t) | –          | S          | –        |
| Hippurate 3.98 (d), 7.56 (t), 7.64 (t), 7.84 (d) | S          | S          | –        |
| N-Methyl nicotinamide 4.44 (s), 8.09 (t), 8.84 (t), 9.12 (s) | –          | N          | N        |
| N-Methyl nicotinamide 4.48 (s), 8.19 (t), 8.90 (d), 8.97 (d), 9.28 (s) | –          | S          | –        |
| 3-Hydroxymandelate 6.85 (d), 6.92 (t), 6.99 (d), 7.31 (t) | –          | S          | –        |
| Tyrosine 6.90 (d), 7.19 (d)  | –          | N          | N        |
| Trytophan/tryptamine 7.21 (t), 7.28 (t), 7.36 (s), 7.51 (d), 7.71 (d) | –          | S          | –        |
| Pseudouridine 7.67 (s)      | –          | S          | S        |
| N-Methyl-2-pyridone-5-carboxamide 3.65 (d), 6.67 (d), 7.83 (dd), 8.34 (d) | –          | S          | –        |
| Formate 8.46 (s)            | –          | N          | N        |

---

**a** The chemical shifts and multiplicities are listed for peaks from significantly associated metabolites. Peaks are only listed if they are in the range of the processed data (9.5–6.4 and 4.5–0.5 ppm) and showed clear correlation patterns. Multiplicity key is as follows: s – singlet, d – doublet, t – triplet, q – quartet, dd – doublet of doublets, dq – doublet of quartets, 2d – two doublets, m – (other) multiplet.

**b** Unknown metabolites, listed as UNK, are only included if statistical analyses using STORM showed sufficient evidence of peaks belonging to the same molecule with the most likely multiplicity labelled. STORMs recovered latent compounds are shown in the supplementary information.

**c** Overlapped signals of valeric (C5), caproic (C6), enanthic (C7), caprylic (C8), pelargonic (C9) and capric acid (C10).

**d** There are some differences in assignments between the Yap et al. study and ours, Yap et al. list pentanoic/heptanoic acid opposed to fatty acids C5–C10, branch-chain amino acids together opposed to leucine, valine and isoleucine separately and methylguanidine instead of methiin.
Supplementary Table 5. List of abbreviations and full names for metabolites in the network shown in Figure 4.

| (Abbreviated) name | Full metabolite name |
|--------------------|----------------------|
| (3S)-Cit-CoA       | (3S)-Cityl coenzyme A|
| 1,2-Propanediol    | 1,2-Propanediol      |
| 1Me-His            | 1-Methylhistidine    |
| 2(OH-Et)TPP        | 2-Hydroxyethyl thiamin pyrophosphate |
| 2,3,4,5-TetraH-dipicolinate | 2,3,4,5-Tetrahydrodipicolinate |
| 2,3-DIH-2,3-diOH-benzoyl-CoA | 2,3-Dihydro-2,3-dihydroxybenzoyl coenzyme A |
| 2,3-DiOH-3Me-valerate | 2,3-Dihydroxy-3-methylvalerate |
| 2,3-DiOH-isovalerate | 2,3-Dihydroxyisovalerate |
| 2,4-DiOH-hept-2-enedioate | 2,4-Dihydroxy-hept-2-enedioate |
| 2,5-Dioxopentanoate | 2,5-Dioxopentanoate |
| 2,6-DiAm-pimelate  | 2,6-Diaminopimelate |
| 2-Aceto-2OH-butanoate | 2-Aceto-2-hydroxybutanoate |
| 2-Acetolactate     | 2-Acetolactate       |
| 2Am-6-oxohexanoate | 2-Amino-6-oxohexanoate |
| 2Am-6-oxopimelate  | 2-Amino-6-oxopimelate |
| 2Am-adipate        | 2-Amino adipate      |
| 2Am-malonate semi aldehyde | 2-Aminomalonate semialdehyde |
| 2-Butenoyl-CoA     | 2-Butenoyl coenzyme A|
| 2DeH-3deO-galactonate 6P | 2-Dehydro-3-deoxy-galactonate 6-phosphate |
| 2DeH-pantoate      | 2-Dehydro-pantoate  |
| 2DeO-ribose 5P     | 2-Deoxyribose 5-phosphate |
| 2HIB               | 2-Hydroxyisobutyrate |
| 2Me-1(OH-Bu)TPP    | 2-Methyl-1-hydroxybutyl thiamin pyrophosphate |
| 2Me-1(OH-Pr)TPP    | 2-Methyl-1-hydroxypropyl thiamin pyrophosphate |
| 2Me-Bt-CoA         | 2-Methylbutanoyl coenzyme A |
| 2Me-citrate        | 2-Methylcitrate      |
| 2OH-glutarate      | 2-Hydroxyglutarate   |
| 2-oxo-3deO-gluconate | 2-Oxo-3-deoxy-gluconate |
| 2-oxo-3deO-gluconate 6P | 2-Oxo-3-deoxy-gluconate 6-phosphate |
| 2-Oxoadipate       | 2-Oxoadipate         |
| 2-Oxobutyrate      | 2-Oxobutyrate        |
| 2-Oxoglutaramate   | 2-Oxoglutaramate     |
| 2-Oxoglutarate     | 2-Oxoglutarate       |
| 2-Oxoisocaproat e  | 2-Oxoisocaproate     |
| 2-oxoisovalerate   | 2-Oxoisovalerate     |
| 2Ph-acetamide      | 2-Phenylacetamide    |
| 3(Me-Bt)CoA        | 3-Methylbutanoyl coenzyme A |
| 3,4-diHPA          | 3,4-Dihydroxyphenylacetate |
| 3,4-diOH-benzoate  | 3,4-Dihydroxybenzoate |
| 3,4-diOH-Ph-acetaldehyde | 3,4-Dihydroxyphenylacetalddehyde |
| 3-Amino-isobutyrate | 3-Amino-isobutyrate  |
| 3Cx-1(OH-Pr)TPP    | 3-Carboxy-1-hydroxypropyl thiamin pyrophosphate |
| 3DeH-sphinganine   | 3-Dehydroxysphinganine |
| (Abbreviated) name                      | Full metabolite name                                      |
|----------------------------------------|----------------------------------------------------------|
| 3HM                                    | 3-Hydroxymandelate                                        |
| 3IsoPr-malate                          | 3-Isopropylmalate                                         |
| 3Me-(OH-Bu)TPP                        | 3-Methyl-1-hydroxybutyl thiamin pyrophosphate            |
| 3Me-2-oxopentanoate                   | 3-Methyl-2-oxopentanoate                                 |
| 3Me-His                               | 3-Methylhistidine                                         |
| 3OH-3Me-2-oxopentanoate               | 3-Hydroxy-3-methyl-2-oxopentanoate                       |
| 3OH-But-CoA                            | 3-Hydroxybutanoyl coenzyme A                             |
| 3-Oxopropanoate                       | 3-Oxopropanoate                                           |
| 3-Ureidopropionate                    | 3-Ureidopropionate                                        |
| 4Am-butryate                           | 4-Aminobutyrate                                           |
| 4-Cresol                              | 4-Cresol                                                  |
| 4CS                                    | 4-Cresyl sulfate                                          |
| 4-Guanidinobutanoate                  | 4-Guanidinobutanoate                                      |
| 4HPA                                   | 4-Hydroxyphenylacetate                                    |
| 4-Imidazolone-5-propanoate            | 4-Imidazolone-5-propanoate                               |
| 4OH-2-oxoglutarate                    | 4-Hydroxy-2-oxoglutarate                                 |
| 4OH-2-oxohexanoate                    | 4-Hydroxy-2-oxohexanoate                                 |
| 4OH-2-oxopimelate                     | 4-Hydroxy-2-oxopimelate                                  |
| 4OH-benzaldehyde                      | 4-Hydroxybenzaldehyde                                    |
| 4OH-benzoate                          | 4-Hydroxybenzoate                                         |
| 4OH-benzoyl-CoA                       | 4-Hydroxybenzoyl coenzyme A                              |
| 4OH-cinnamate                         | 4-Hydroxycinnamate                                        |
| 4OH-cinnamoyl-CoA                     | 4-Hydroxycinnamoyl coenzyme A                            |
| 4OH-hippurate                         | 4-Hydroxyhippurate                                        |
| 4OH-Ph-acetaldehyde                   | 4-Hydroxyphenylacetaldehyde                              |
| 4OH-phenacyl alcohol                  | 4-Hydroxyphenacyl alcohol                                 |
| 4-Oxobutanoate                        | 4-Oxobutanoate                                            |
| 5,10-Methylene-tetraH-methanopterin   | 5,10-Methylene-tetrahydrothromethanopterin               |
| 5,10-Methylene-THF                     | 5,10-Methylene-tetrahydrofolate                           |
| 5-Adenylyl-2Am-adipate                | 5-Adenylyl-2-amino adipate                               |
| 5DeH-4deO-glucarate                   | 5-Dehydro-4-deoxyglucarate                               |
| 5Me-barbiturate                       | 5-Methylbarbiturate                                       |
| 5Me-H4MPT                              | 5-Methyl-5,6,7,8-tetrahydrothromethanopterin             |
| 5Me-THF                               | 5-Methyltetrahydrofolate                                 |
| 6DeO-5-oxofructose 1P                 | 6-Deoxy-5-oxofructose 1-phosphate                         |
| Ac-adenylate                           | Acetyl adenylate                                          |
| Ac-choline                             | Acetylcholine                                             |
| Ac-CoA                                 | Acetyl coenzyme A                                         |
| Acetate                                | Acetate                                                   |
| Acetoacetate                           | Acetate                                                   |
| Acetoacetyl-CoA                        | Acetoacetyl coenzyme A                                    |
| Acetone                                | Acetone                                                   |
| AcP                                    | Acetylphosphate                                           |
| Ac-phenol                              | Acetylphenol                                              |
| Adenylosuccinate                      | Adenylosuccinate                                          |
| (Abbreviated) name       | Full metabolite name                                      |
|--------------------------|----------------------------------------------------------|
| ADMA                     | Asymmetric dimethylarginine                              |
| ADP                      | Adenosine 5-diphosphate                                  |
| ADP-glucose              | Adenosine 5-diphosphate glucose                          |
| Adrenaline               | Epinephrine                                              |
| Ala                      | Alanine                                                  |
| Allothreonine            | Allothreonine                                            |
| α-Glucose 6P             | alpha-Glucose 6-phosphate                                |
| αIsop-malate             | alpha-Isopropylmalate                                    |
| α-Linolenoyl-CoA         | alpha-Linolenoyl coenzyme A                              |
| Am-acetaldehyde          | Aminoacetaldehyde                                        |
| Am-adip.-S               | alpha-Aminoacidoyl-S-acyl enzyme                          |
| AMP                      | Adenosine 5-phosphate                                    |
| Anserine                 | Anserine                                                 |
| APS                      | Adenosine 5-phosphosulfate                               |
| Arabino-hex-3-ulose 6P   | Arabino-hex-3-ulose 6-phosphate                          |
| Arachidonate             | Arachidonate                                             |
| Arachidonyl-CoA          | Arachidonyl coenzyme A                                   |
| Arg                      | Arginine                                                 |
| Arg-succinate            | Argininosuccinate                                        |
| Arterenol                | Norepinephrine                                           |
| Asp                      | Aspartate                                                |
| Benzoate                 | Benzoate                                                 |
| Benzoyl-CoA              | Benzoyl coenzyme A                                       |
| Betaine                  | Betaine                                                  |
| βAla                     | beta-Alanine                                             |
| βAla-Arg                 | beta-Alanylarginine                                      |
| βAla-Lys                 | beta-Alanyllysine                                        |
| β-Fructose 6P            | Beta-fructose 6-phosphate                                |
| βMe-Mal-CoA              | Betamethylmalyl coenzyme A                               |
| Biocytin                 | N6- Biotinyl- lysine                                     |
| Biotin                   | Biotin                                                   |
| Biotinyl-5-AMP           | Biotinyl 5-adenosine 5-phosphate                         |
| Carbamoyl P              | Carbamoyl phosphate                                      |
| Carnitine                | Carnitine                                                |
| Carnosine                | Carnosine                                                |
| Ceramide                 | N-Acylphosphosine                                        |
| Chenodiol                | Chenodeoxycholate                                        |
| Cholate                  | Cholate                                                  |
| Choline                  | Choline                                                  |
| Choline P                | Choline phosphate                                        |
| Choloyl-CoA              | Choloyl coenzyme A                                       |
| Chorismate               | Chorismate                                               |
| Citrate                  | Citrate                                                  |
| Citrulline               | Citrulline                                               |
| CMP                      | Cytidine 5-phosphate                                     |
| (Abbreviated) name | Full metabolite name |
|-------------------|----------------------|
| CMP-NANA          | Cytidine 5-phosphate N-acetyl neuraminic acid |
| CoA               | Coenzyme A           |
| Creatine          | Creatine             |
| Creatinine        | Creatinine           |
| Cys               | Cysteine             |
| CysGly            | Cysteinylglycine     |
| Cystathionine     | Cystathionine        |
| Cysteate          | Cysteate             |
| Cytidine          | Cytidine             |
| Deoxycholate      | Deoxycholate         |
| DeP-CoA           | Dephospho coenzyme A |
| DHF               | Dihydrofolate        |
| DiH-ceramide      | Dihydroceramide      |
| DiH-LipE          | Dihydrolipoamide E   |
| DiMe sulfone      | Dimethylsulfone      |
| DKHP              | 2-Deoxy-5-ketogluconic acid 6-phosphate |
| DMA               | Dimethylamine        |
| DMF               | Dimethylformamamide  |
| DMG               | Dimethylglycine      |
| Dopamine          | 3,4-Dihydroxyphenethylamine |
| Ethanal           | Ethanal              |
| Ethanol           | Ethanol              |
| Ethyl glucuronide | Ethyl glucuronide    |
| FA                | Fatty acid           |
| FA C5             | Valeric acid         |
| FA C6             | Caproic acid         |
| FA C7             | Enanthic acid        |
| FA C8             | Caprylic acid        |
| FA C9             | Pelargonic acid      |
| FA C10            | Capric acid          |
| Farnesyl PP       | Farnesyl pyrophosphate |
| For-kynurenine    | Formylkynurenine     |
| Formamide         | Formamide            |
| Formate           | Formate              |
| Fructose 1P       | Fructose 1-phosphate  |
| Fructose 6P       | Fructose 6-phosphate  |
| Fuculose 1P       | Fuculose 1-phosphate  |
| Fum-acetoacetate  | Fumarylacetoacetate  |
| Fumarate          | Fumarate             |
| γGluCys           | gamma-Glutamylcysteine |
| GAR               | Glycinamide ribonucleotide |
| Glca              | Glucosamine          |
| Glca P            | Glucosamine phosphate |
| GlcNAc            | N-Acetylglucosamine  |
| Gln               | Glutamine            |
| (Abbreviated) name          | Full metabolite name                                                                 |
|-----------------------------|--------------------------------------------------------------------------------------|
| Glu                         | Glutamine                                                                           |
| Gluconate                   | Gluconate                                                                           |
| Gluconate 6P                | Gluconate 6-phosphate                                                               |
| Gluco-1,5-lactone 6P        | Gluco-1,5-lactone 6-phosphate                                                       |
| Gluconolactone              | Gluconolactone                                                                      |
| Glucose                     | Glucose                                                                             |
| Glucose 1P                  | Glucose 1-phosphate                                                                 |
| Glucose 6P                  | Glucose 6-phosphate                                                                 |
| Glucosylceramide            | Glucosyl-N-acethylphosphosine                                                       |
| Glutaryl-CoA                | Glutaryl coenzyme A                                                                 |
| Gly                         | Glycine                                                                             |
| Glyceraldehyde 3P           | Glyceraldehyde 3-phosphate                                                          |
| Glycerone P                 | Glycerone phase                                                                     |
| Glycochenodeoxycholate      | Glycochenodeoxycholate                                                              |
| Glycocholate                | Glycocholate                                                                        |
| Glycodeoxycholate           | Glycodeoxycholate                                                                   |
| Glyoxylylate                | Glyoxylylitate                                                                       |
| GM3                         | (N-Acetylmuramoyl)galactosyl-N-acethylphosphosine                                    |
| GSH                         | Glutathione                                                                         |
| Guanidinoacetate            | Guanidinoacetate                                                                    |
| Hippurate                   | Hippurate                                                                           |
| His                         | Histidine                                                                           |
| HMG-CoA                     | Hydroxymethylglutaryl coenzyme A                                                    |
| Homocarnosine               | Homocarnosine                                                                       |
| Homocitrate                 | Homocitrate                                                                          |
| HTPA                        | (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate                                     |
| Hydouracil                  | Hydouracil                                                                          |
| Ile                         | Isoleucine                                                                          |
| Indole-3-acetamide          | Indole-3-acetamide                                                                  |
| Indoleacetaldehyde          | Indoleacetaldehyde                                                                  |
| Indoleacetate               | Indoleacetate                                                                       |
| Indoleglycerol P            | Indoleglycerol phosphate                                                             |
| IsoBt-CoA                   | Isobutanoyl coenzyme A                                                              |
| Isocitrinate                | Isocitrinate                                                                        |
| ISOe PP                     | Isopentenyl pyrophosphate                                                           |
| LacCer                      | Lactosylceramide                                                                    |
| Lactaldehyde                | Lactaldehyde                                                                        |
| Lactate                     | Lactate                                                                             |
| Lactose                     | Lactose                                                                             |
| Lecithin                    | Phosphatidylethanolcholine                                                           |
| Leu                         | Leucine                                                                             |
| Levulose                    | Levulose                                                                            |
| Linoleate                   | Linoleate                                                                           |
| Linolenate                  | Linolenate                                                                          |
| Linoleoyl-CoA               | Linoleoyl coenzyme A                                                                |
| (Abbreviated) name | Full metabolite name |
|-------------------|----------------------|
| LipE              | Lipoamide E          |
| Lys               | Lysine               |
| Malate            | Malate               |
| Mal-CoA           | Malyl coenzyme A     |
| Malonyl-CoA       | Malonyl coenzyme A   |
| Me-CoM            | Methylcoenzyme M     |
| Me-corrinoid      | Methylcorrinoid      |
| Me-glyoxal        | Methylglyoxal        |
| Me-isocitrate     | Methylisocitrate     |
| Me-malonate       | Methylmalonate       |
| Me-malonate semialdehyde | Methylmalonate semialdehyde |
| Me-malonyl-CoA    | Methylmalonyl coenzyme A |
| meso-2,6-DiAm-pimelate | meso-2,6-Diaminopimelate |
| Methanal          | Methanal             |
| Methanol          | Methanol             |
| Methiin           | S-Methyl- L-cysteine sulfoxide |
| MMA               | Methylamine          |
| N2Cit-N6Ac-N6OH-Lys | N2-Cityl-N6-acetyl-N6-hydroxylysine |
| N6,N6,N6-TriMe-Lys | N6,N6,N6-Trimethyllysine |
| N6-[(Indol-3-yl)Ac]-Lys | N6-[(Indol-3-yl)acetyl]-lysine |
| N6Ac-2,6-diAm-pimelate | N6-Acetyl-2,6-diaminopimelate |
| N6Ac-N6OH-Lys     | N6-Acetyl-N6-hydroxylysine |
| N6OH-Lys          | N6-Hydroxylysine     |
| NA                | Nicotinamide         |
| NA ribonucleotide | Niconamid ribonucleotide |
| NAc-Asp           | N-Acetylaspartate    |
| NAc-citrulline    | N-Acetylcitrulline   |
| NAc-glca 6P       | N-Acetylglucosamine 6-phosphate |
| NAc-mana          | N-Acetylmannosamine  |
| NAc-mana 6P       | N-Acetylmannosamine 6-phosphate |
| NAc-Methiin       | N-Acetyl-S-methyl-cysteine sulfoxide |
| NAcSPCSO          | N-Acetyl-S-(1Z)-propenyl-cysteine sulfoxide |
| NAD               | Nicotinamide adenine dinucleotide |
| NANA              | N-Acetyl neuraminic acid |
| NANA 9P           | N-Acetyl neuraminic acid 9-phosphate |
| N-Carbamoyl-Asp   | N-Carbamoylaspartate |
| N-For-Asp         | N-Formylaspartate    |
| N-For-Glu         | N-Formylglutamate    |
| N-Formimino-Glu   | N-Formiminoglutamate |
| NM2P5C            | N-Methyl-2-pyridone-5-carboxamide |
| NMNA              | N-Methyl nicotinic acid |
| NMND              | N-Methyl nicotinamide |
| Nonaprenyl-4OH-benzoate | Nonaprenyl-4-hydroxybenzoate |
| Nopaline          | N2-(1,3-Dicarboxypropyl) arginine |
| N-Ribosyl-NA      | N-Ribosylnicotinamide |
| (Abbreviated) name                  | Full metabolite name                                      |
|-------------------------------------|----------------------------------------------------------|
| N-Suc-2,6-diAm-pimelate             | N-Succinyl-2,6-diaminopimelate                           |
| N-Suc-Glu                           | N-Succinylglutamate                                      |
| Oxaloacetate                        | Oxaloacetate                                             |
| O-Ac-carnitine                      | O-Acetylcarnitine                                        |
| O-Suc-hSer                          | O-Succinylhomoserine                                     |
| Palmitoylcarnitine                  | Palmitoylcarnitine                                       |
| Palmitoyl-CoA                        | Palmitoyl coenzyme A                                     |
| Pantothenate                         | Pantothenate                                             |
| PAP                                 | Phosphoadenosine phosphate                               |
| PhAc                                | Phenylacetate                                            |
| PhAc-CoA                             | Phenylacetyl coenzyme A                                  |
| Phe                                 | Phenylalanine                                            |
| Phenol                              | Phenol                                                   |
| Pp-adenylate                        | Propanoyladenylate                                       |
| Proline betaine                     | Proline betaine                                          |
| Propanal                            | Propanal                                                 |
| Propanoyl-CoA                       | Propanoyl coenzyme A                                     |
| PRPP                                | 5-Phosphoribosyl 1-pyrophosphate                         |
| PtdSer                              | Phosphatidylserine                                       |
| Pyridoxal 5P                        | Pyridoxal 5-phosphate                                     |
| Pyruvate                            | Pyruvate                                                 |
| Rhamnulose 1P                       | Rhamnulose 1-phosphate                                   |
| Riba 5P                             | Ribosylamine 5-phosphate                                 |
| Ribose                              | Ribose                                                   |
| Ribose 5P                           | Ribose 5-phosphate                                       |
| Ribulose 5P                         | Ribulose 5-phosphate                                     |
| S-(OHMe)GSH                         | S-(Hydroxymethyl)glutathione                             |
| S-2(Me-Bt)-DiH-LipE                 | S-(2-Methylbutanoyl)-dihydrolipoamide E                  |
| S-2(Me-Pp)-DiH-LipE                 | S-(2-Methylpropanoyl)-dihydrolipoamide E                 |
| S-3(Me-Bt)-DiH-LipE                 | S-(3-Methylbutanoyl)-dihydrolipoamide E                  |
| Saccharopine                         | Saccharopine                                             |
| Sarcosine                           | Sarcosine                                                |
| scyllo-Inositol                     | scyllo-Inositol                                          |
| Seleno-Cys                           | Selenocysteine                                           |
| Selenocystathionine                 | Selenocystathionine                                      |
| Ser                                 | Serine                                                   |
| S-For-GSH                           | S-Formylglutathione                                      |
| Solanesyl PP                        | Solanesyl pyrophosphate                                  |
| Sphinganine                          | Sphinganine                                              |
| Sphingomyelin                        | Sphingomyelin                                            |
| S-Suc-DiH-LipE                      | S-Succinyl-dihydrolipoamide E                            |
| (Abbreviated) name       | Full metabolite name                                                                 |
|--------------------------|--------------------------------------------------------------------------------------|
| S-Suc-GSH                | $S$-Succinylglutathione                                                            |
| Succinate                | Succinate                                                                            |
| Suc-CoA                  | Succinyl coenzyme A                                                                 |
| Sucrose                  | Sucrose                                                                              |
| Sulfite                  | Sulfite                                                                              |
| SulfoAc-CoA              | Sulfoacetyl coenzyme A                                                             |
| Sulfoacetaldehyde        | Sulfoacetaldehyde                                                                    |
| Taurine                  | Taurine                                                                              |
| Taurochenodeoxycholate   | Taurochenodeoxycholate                                                               |
| Taurocholate             | Taurocholate                                                                         |
| Taurodeoxycholate        | Taurodeoxycholate                                                                    |
| TCE                      | Trichloroethylene                                                                     |
| TCE epoxide              | Trichloroethylene epoxide                                                            |
| Thr                      | Threonine                                                                            |
| TMA                      | Trimethylamine                                                                       |
| TPP                      | Thiamin pyrophosphate                                                                |
| trans-Aconitate          | $trans$-Aconitate                                                                    |
| Trehalose 6P             | Trehalose 6-phosphate                                                                |
| Trp                      | Tryptophan                                                                           |
| Tryptamine               | Tryptamine                                                                           |
| Tyr                      | Tyrosine                                                                             |
| Tyramine                 | Tyramine                                                                             |
| UDP                      | Uridine 5-diphosphate                                                                |
| UDP-Ara4FN               | Uridine 5-diphosphate 4-deoxy-4-formamido-$\beta$-arabinopyranose                   |
| UDP-galactose            | Uridine 5-diphosphate galactose                                                      |
| UDP-glucose              | Uridine 5-diphosphate glucose                                                       |
| UDP-NAc-glca             | Uridine 5-diphosphate $N$-acetyl-glucosamine                                        |
| UMP                      | Uridine 5-monophosphate                                                              |
| Undecaprenyl P           | Undecaprenyl phosphate                                                               |
| Undecaprenyl P $\alpha$-Ara4FN | Undecaprenyl phosphate 4-deoxy-4-formamido-$\alpha$-arabinopyranose   |
| Undecaprenyl PP          | Undecaprenyl pyrophosphate                                                           |
| Uracil                   | Uracil                                                                               |
| Urea                     | Urea                                                                                 |
| Urocanate                | Urocanate                                                                            |
| Val                      | Val                                                                                  |
| Vanillate                | 4-Hydroxy-3-methoxybenzoate                                                          |
REFERENCES

1. Wold, S.; Martens, H.; Wold, H., The Multivariate Calibration-Problem in Chemistry Solved by the Pls Method. Lecture Notes in Mathematics 1983, 973, 286-293.

2. Daszykowski, M.; Serneels, S.; Kaczmarek, K.; Van Espen, P.; Croux, C.; Walczak, B., TOMCAT: A MATLAB toolbox for multivariate calibration techniques. Chemometrics and Intelligent Laboratory Systems 2007, 85, (2), 269-277.

3. De Jong, S., Simpls - an Alternative Approach to Partial Least-Squares Regression. Chemometrics and Intelligent Laboratory Systems 1993, 18, (3), 251-263.

4. Eriksson, L.; Trygg, J.; Johansson, E.; Bro, R.; Wold, S., Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data. Analytica Chimica Acta 2000, 420, (2), 181-195.

5. Edelman, A.; Rao, N. R., Random matrix theory. Acta Numerica 2005, 14, 233-297.