Urinary 1H-NMR metabolomic fingerprinting reveals biomarkers of pulse consumption related to energy-metabolism modulation in a subcohort from the PREDIMED study

Francisco Madrid-Gambin[†], Rafael Llorach*[†], Rosa Vázquez-Fresno[†], Mireia Urpi-Sarda[†], Enrique Almanza-Aguilera[†], Mar Garcia-Aloy[†], Ramon Estruch[§#], Dolores Corella[⊥#] and Cristina Andres-Lacueva*[†]

[†]Biomarkers & Nutrimetabolomics Laboratory. Nutrition, Food Science and Gastronomy Department, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona, Spain.

[§]Department of Internal Medicine, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona, Spain.

[⊥]Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain.

[#]CIBER OBN, The Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid, Spain.

*E-mail: rafallorach@ub.edu. Tel: +34.934033798. *E-mail: candres@ub.edu. Tel: +34.934034840. Fax: +34.93403593.

20 Abstract

21 Little is known about the metabolome fingerprint of pulse consumption. The
22 study of robust and accurate biomarkers for pulse dietary assessment has great
23 value for nutritional epidemiology regarding health benefits and their
24 mechanisms. To characterize the fingerprinting of dietary pulses (chickpeas,
25 lentils and beans), spot urine samples from a subcohort from the PREDIMED
26 study were stratified, using a validated food frequency questionnaire. Non-pulse
27 consumers (≤ 4 g/day of pulse intake) and habitual pulse consumers (≥ 25
28 g/day of pulse intake) were analysed using a ^{1}H-NMR metabolomics approach
29 combined with multi- and univariate data analysis. Pulse consumption showed
30 differences through 16 metabolites coming from (i) choline metabolism, (ii)
31 protein-related compounds, and (iii) energy metabolism (including lower urinary
32 glucose). Stepwise logistic regression analysis was applied to design a
33 combined model of pulse exposure, which resulted in glutamine, dimethylamine
34 and 3-methylhistidine. This model was evaluated by receiver operating
35 characteristic curve (AUC > 90% in both training and validation sets). The
36 application of NMR-based metabolomics to pulse exposure highlighted new
37 candidates for biomarkers of pulse consumption, the role of choline metabolism
38 and the impact on energy metabolism, generating new hypotheses on energy
39 modulation. Further intervention studies will confirm these findings.

40

41 Keywords

42 pulses, legumes, metabolomics, NMR, choline metabolism, energy, biomarkers,
43 ROC curve

1. Introduction

The Mediterranean diet (MD) is a dietary pattern characterized by a high intake of vegetables, cereals, pulses, nuts, fish and olive oil, low intake of red meat and processed meat products, and low to moderate consumption of poultry, wine and dairy products.[1] Moreover, the MD has been demonstrated to be useful in the prevention of type 2 diabetes, obesity, inflammatory diseases, cardiovascular diseases (CVD) and even cancer.[2–5]

One of the components of the MD is pulses, which constitute an excellent food, providing protein, dietary fibre, many vitamins and minerals, as well as a great variety of phytochemicals.[6–8] Thus, they could contribute to the beneficial effects reported for this dietary pattern.[9] In addition, pulses are increasingly being recognized for their role in promoting good health.[6,10–12] Indeed, habitual pulse consumption is included in the main dietary guidelines worldwide, including the MD,[13] the Dietary Guidelines for Americans[14,15] and the Nordic Diet,[16] among others, and they are also advocated in view of their low environmental impact compared with other protein sources.[17]

Metabolomics is a powerful tool for identifying food exposure biomarkers in humans[18] and provides new information on dietary components and dietary patterns.[19] In this regard, the evaluation of dietary exposure through a combination of biomarkers enables a better understanding of compliance to a dietary exposure.[20] Moreover, little is known about the metabolome fingerprint from legume consumption either individually or as a complex food group, with only a few tentative biomarkers being described.[21,22]

67 Determining the changes in the urinary metabolome, new biomarkers of intake
68 and/or their effect may reveal potential modifications in diet-related physiology
69 both in healthy and diseased individuals.[23] Furthermore, metabolomic
70 approaches have been proposed for evaluating the relationship between
71 nutrition and health status.[24] In light of this connection, recent scientific
72 publications have pointed out the potential health benefits of legumes in chronic
73 diet-related diseases, such as CVD and type 2 diabetes mellitus.[6,8,25,26] Thus
74 the application of nutrimetabolomics to a high-cardiovascular-risk population
75 could provide new insights into this potential relationship.

76 In the present work, we compared the metabolome profiles of reported pulse
77 consumption in a free-living population to find putative biomarkers reflecting
78 intake and/or effect of intake. Analysis of individuals under free-living conditions
79 enables more representative data to be obtained on the metabolome
80 fingerprints of pulse consumers. In light of this, a better understanding of the
81 specific role of pulse consumption in terms of health benefits, beyond their
82 excellent nutritional profile, is expected. Therefore, the aim of the present study
83 was to investigate dietary pulse fingerprinting in spot urine using an untargeted
84 ^{1}H-NMR metabolomic approach on a free-living subcohort from the PREDIMED
85 study. For this purpose, we mainly focused on urinary biomarkers of a complex
86 pulse exposure comprising chickpeas, lentils and beans in a combined urinary
87 biomarker model.

88

89 ## 2. Material and methods

90 ### 2.1. PREDIMED subcohort study

For the present study, a subsample of 50 participants from the PREDIMED study (ISRCTN 35739639; http://www.predimed.org) was taken. The PREDIMED study is a large, parallel-group, multicentre, randomized and controlled clinical trial assessing the effects of an MD on the primary prevention of CVD. The trial protocol was conducted according to the Declaration of Helsinki and was approved by the Institutional Review Boards of all the centres involved. Briefly, free-living participants (55–80 years old) without CVD that fulfilled at least one of the two following criteria – type 2 diabetes mellitus or three or more major cardiovascular risk factors – were included for an MD supplemented either with extra virgin olive oil or mixed nuts.[27] The exclusion criteria were CVD, any severe chronic illness, drug or alcohol addiction, a history of allergy, or intolerance to olive oil or nuts. The subcohort consisted of a random sample of participants at high cardiovascular risk, recruited from the Barcelona and Valencia PREDIMED centres. The PREDIMED study design and 137-item validated food frequency questionnaires (FFQs) used have been reported elsewhere.[28,29] Data reported from the FFQs included information on total legume consumption, and disaggregated type of legume consumed.

2.2. Stratification of the study population

2.2.1. Defining potential consumers

Both the use of FFQs and the population stratification of a cohort of individuals by consumption have demonstrated an effective approach for the study of biomarkers of food consumption.[30–32] Participants were classified into two levels (consumers and non-consumers) of habitual intake of dietary pulse foods

(chickpeas, lentils or beans) based on the analysis of the validated FFQs (Supporting Information, Table S1). Intake of pulses was calculated as the sum of consumed chickpeas, lentils and beans. Non-pulse (NP) consumers were defined as subjects with sporadic or non-consumption (≤ 4.00 g/day) of pulses. Habitual pulse (HP) consumers were set at a consumption of ≥ 25.71 g/day, regularly. In order to explore global pulse consumption, individuals that did not consume the three kinds of pulses simultaneously were also excluded. Additionally, the condition of sporadic or non-intake of peas (≤ 4 g/day) was taken into consideration, since the features of this type of legume are not similar to the others.[33] No other legume types were considered.

2.2.2. Selecting individuals by consumption

Spot urine samples were matched to corresponding individual FFQ data. From a cohort of 828 individuals, 25 subjects were defined as NP consumers and 37 as HP consumers (none of the other participants from both pulse consumer groups fulfilled any criteria). In order to reduce the potential sources of variability not related to pulse exposure, the number of HP consumers was balanced against NP consumers (HP = 25, NP = 25). Finally, dietary data, anthropometry, biochemical parameters, health status and medication were explored with a view to discarding any variability unrelated to pulse consumption.

2.3. Metabolomics analysis

138 2.3.1. Urine sample analysis and data processing

139 Morning fasting spot urine samples were collected, aliquoted, encoded and
140 frozen at -80 ºC until were use. Sample preparation was based on the
141 methodology previously published.[19] The ^{1}H-NMR urinary spectra were
142 acquired using a Varian-Inova-500 MHz NMR spectrometer with presaturation
143 of the water resonance using a NOESYPRESAT pulse sequence. During the
144 acquisition, the internal temperature was kept constant at 298 K. An exponential
145 window function was applied to the free induction decay (FID) with a line-
146 broadening factor of 0.3 Hz prior to Fourier transformation. For each sample, a
147 total of 128 scans were collected into 32 K data points with a spectral width of
148 14 ppm at 300 K, an acquisition time of 3.2 s and a relaxation delay of 3 s.

149 ^{1}H-NMR spectra were phased, baseline-corrected and calibrated (TSP, 0.0
150 ppm) using TopSpin software (version 3.0, Bruker, BioSpin, Germany). After
151 baseline correction, original spectral data were bucketed in intelligent bucketing
152 domains of 0.005 ppm with ACD/NMR Processor 12.0 software (Advanced
153 Chemistry Development, Toronto, Canada). The water signal and noise regions
154 above 9.5 ppm and below 0.5 ppm were excluded from the analysis.

155 Data were submitted to MetaboAnalyst 3.0 for interquartile range filtering and
156 normalization by the sum of the intensities of the spectra.[34]

157

158 2.3.2. Statistical analysis

159 The NMR data set was log-transformed, Pareto-scaled and posteriorly analysed
160 in a multivariate approach using SIMCA-P+13.0 software (Umetrics, Umeå,

Sweden). Interindividual variation may confuse the effects of intervention, particularly in multivariate data of high dimensionality. Therefore, partial least squares discriminant analysis with orthogonal signal correction (OSC-PLS-DA) was used to explore the differences in metabolomes among the pulse consumption.[35] OSC filtration was used to reduce the variability not associated with dietary classification, as has been done in other published nutrimetabolomic studies.[19,31,36] The quality of the models was evaluated by the proportion of the variance of the response variable that is explained by the model (R^2Y) and the predictive ability (Q^2) parameters.[35] Validation of the models and the evaluation of the degree of overfitting were carried out using a permutation test (n = 200), and the correlation coefficient between the original Y and the permuted Y plotted against the cumulative R^2 and Q^2 was calculated. Those NMR signals with variable importance for projection (VIP) values ≥1 in the component of the OSC-PLS-DA model were selected as being relevant for explaining the differences in metabolic profiles. These variables were further studied through the univariate Student's t-test among HP and NP consumers to assess the statistical significances. Multiple tests were controlled by the false discovery rate (FDR). Statistical significance was considered at an FDR-adjusted p-value <0.05. Then, Cliff's delta was chosen for estimation of the effect size[37] and calculated for each feature.

2.3.3. Metabolite identification

Metabolite identification was performed using the Chenomx NMR Suite Professional Software package (version 8.1; Chenomx Inc., Edmonton,

185 Canada) and by comparing NMR spectral data to those available in databases
186 such as the Human Metabolome Database (http://www.hmdb.ca), the Biological
187 Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu) and the Madison
188 Metabolomics Consortium Database (www.mmcd.nmrfam.wisc.edu), along with
189 the existing NMR-based metabolomics literature. Further, a Pearson's
190 correlation test and clustering analysis with Pearson distance and Ward's
191 minimum variance using PermutMatrix 1.9.3.0 software[38] were applied in order
192 to identify the signals corresponding to the same metabolite.

193

194 2.4. Study of combined urinary biomarker model

195 The interaction between gender and the resulting metabolites was evaluated by
196 a logistic regression for discarding any effect on the biomarkers. Then, these
197 metabolites were submitted to a stepwise logistic regression analysis (IBM
198 SPSS Statistics 20 software, SPSS, Inc., Chicago, IL, USA) to evaluate whether
199 the combination of more than one biomarker improves the discrimination[20] of
200 pulse consumption. The models were constructed through a dichotomous
201 variable of pulse consumption as dependent variable and identified metabolites
202 as independent variables, with a p-value of <0.05 as a condition required for
203 entering and remaining in the model. For validation of models, the analysis with
204 a training set of 2/3 of the samples (removing 1/3 of the individuals as the
205 validation set) was permuted 20 times. Spearman's rank correlation coefficient
206 was used to assess correlations between the combined models and pulse
207 consumption.

208 The global performance of the models was evaluated by receiver operating
209 characteristic (ROC) curve and estimation of the area under the curve (AUC)
210 values. The optimum cut-off for sensitivity and specificity of the biomarkers was
211 determined as the minimum distance to the top-left corner.[39]

212

213 ## 3. Results

214 A flow chart of the participants allocated in the present study is presented in the
215 Supplementary Information (Figure S1). Anthropometric measurements and
216 biochemical analyses were performed using standardized methods.[28] HP
217 consumers showed a pulse consumption of 38.45 ± 14.68 g/day, while NP
218 consumers reported a consumption of 3.75 ± 3.95 g/day (mean ± SD). The
219 characteristics of participants classified by pulse consumption (Table S2) are
220 presented in the Supplementary Information. The stratified populations were not
221 different in terms of disease (type 2 diabetes mellitus or cardiovascular risk
222 factors), medications or biochemical parameters, among other data. Subjects
223 who were HP consumers showed higher amounts of both dietary fibre ($p <$
224 0.01) and polyunsaturated fatty acid ($p < 0.05$) intakes as a consequence of
225 legume macronutrient composition.[6,40] No significances other than pulses were
226 found with regard to food intake.

227

228 ### 3.1. Selection of significant biomarkers related to pulse consumption

229 For the analysis of the features belonging to pulse consumption in the urinary
230 metabolome of the HP and NP consumers, an orthogonal signal correction was
231 applied before PLS-DA analysis. The OSC-PLS-DA analysis of the two groups

232 resulted in a latent variable model with R^2Y and Q^2 values of 0.954 and 0.809,
233 respectively, indicating that the model was able to classify each subject in the
234 correct consumption group. The corresponding permutation tests showed
235 negative Q^2 intercepts with a value of -0.164, implicating validation of the
236 model.[35] With the purpose of selecting the most discriminative urinary markers
237 of consumption, only the statistically significant variables coming from both
238 multi- and univariate analyses simultaneously were considered.

239

240 3.2. Identified biomarkers of habitual pulse consumption

241 A total of 16 compounds were identified as discriminant metabolites of pulse
242 consumption. Metabolites and chemical shifts identified corresponding to
243 statistical analyses are presented in Table 1. The total number of metabolites
244 related to pulse consumption was divided into categories as follows: (i) choline
245 metabolism: choline, dimethylglycine, trimethylamine-N-oxide (TMAO) and
246 dimethylamine; (ii) protein-related compounds: 3-methylhistidine,
247 methylguanidine, phenylalanine, glutamine and n-acetylglutamine; and (iii)
248 energy metabolism: glucose, leucine, isovalerylglycine, and isobutyric,
249 acetoacetic, citric and cis-aconitic acids.

250

251 3.3. Combined urinary biomarker approach

252 Logistic regression analysis revealed that there was no significant interaction
253 between gender and the metabolites ($p>0.05$; all) shown in Table S3
254 (Supplementary Information). To study the improvement of the discrimination

between groups (HP and NP consumers), a conditional stepwise variable selection method, through a binary logistic regression analysis, was used on a combination of more than one discriminant metabolite. Table S4 (Supplementary Information) shows the resulting metabolites included in all 20 permuted models and the contribution to the model. Three metabolites were included in the fitted model according to the maximum AUC, which contained two protein-related metabolites (glutamine and 3-methylhistidine) and one choline-related metabolite (dimethylamine). These three metabolites correlated individually with the pulse consumption. However, the combined model exhibited the strongest correlation ($r=0.73$, $p<0.01$) with the pulse exposure, as shown in Table S5 (Supplementary Information).

The ROC curve analysis was used to evaluate the combined metabolite model and their metabolites using both training and validation sets separately. The highest AUC was for the combined metabolite model for both training (AUC = 95.6%) and validation (AUC = 94.4%) sets, including glutamine, 3-methylhistidine and dimethylamine followed by the individual metabolites 3-methylhistidine (AUC = 82.4%), glutamine (AUC = 81.6%) and dimethylamine (AUC = 75.0%), as shown in Figure 1. The equations generated from the logistic regression and the AUCs from the models with their sensitivity and specificity are shown in Table 2.

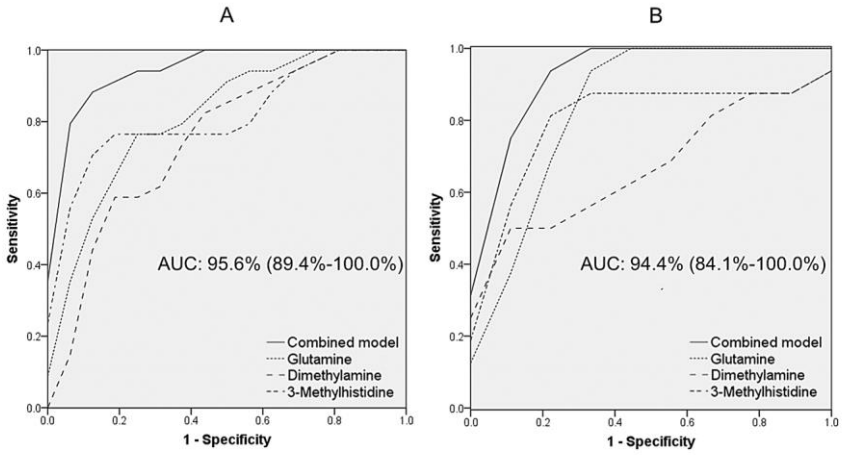


Figure 1. Receiver operating characteristic (ROC) curves of combined model (continuous line) with the area under the ROC curve and of included individual metabolites (discontinuous lines) in the training (A) and validation (B) sets.

4. Discussion

In this study, we present a panel of different urinary metabolites related to habitual pulse exposure using a ^{1}H-NMR-based untargeted nutrimetabolomic approach in a free-living population. In addition, high correlations were found when the exposure was assessed as a continuous variable (defined by the combined biomarker panel).

4.1. Characterization of pulse fingerprinting in urine

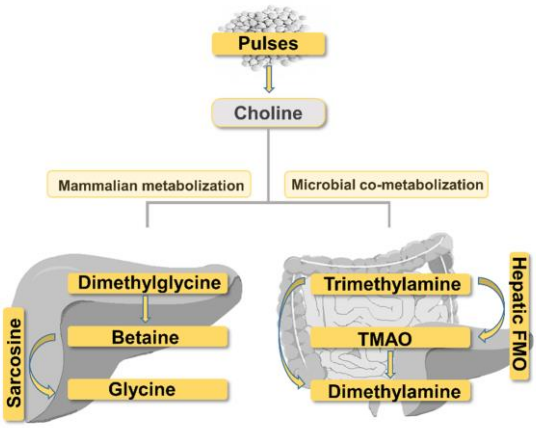
4.1.1. Pulse metabolomic fingerprinting and choline metabolism

Several compounds found in the spot urine of pulse consumers are related to choline. Thus pulses, as a rich source of choline,[41] may be the precursor of

291 additional metabolites that are susceptible to microbial degradation generating
292 new compounds.[42] Therefore, the increase of several intermediates of choline
293 metabolism, such as choline itself, TMAO and dimethylamine, appears to be a
294 consequence of the microbial activity in HP consumers. In relation to this, De
295 Filippis and co-workers found an inverse correlation between urinary TMAO and
296 vegetarian diets compared with omnivore ones. However, they suggest different
297 food sources of carnitine and choline such as eggs, beef, pork and fish.[43]
298 Hence, legumes from vegetarian diets should be proposed as a food choline
299 source. The increase of dimethylamine, which is also a downstream product of
300 choline, supports the microbial degradation of TMAO from choline. Furthermore,
301 TMAO was identified as a major source of urinary dimethylamine in humans,[44]
302 directly related to gut microbiota metabolism.[45] On the other hand, the increase
303 of urinary dimethylglycine may also come from the choline contained in pulses.
304 The enzymes choline dehydrogenase, betaine aldehyde dehydrogenase and
305 betaine homocysteine methyltransferase lead to dimethylglycine from choline.[46]
306 Therefore, the results of the present study suggest a possible impact on urinary
307 metabolome by choline from pulses that is degraded via both (i) mammalian
308 pathways in which choline is converted to dimethylglycine through betaine, and
309 (ii) microbial metabolism in which choline is degraded to trimethylamine, TMAO
310 and dimethylamine. For this reason, we propose dimethylamine and
311 dimethylglycine in spot urine as potential candidates for biomarkers of pulse
312 consumption. Nevertheless, these choline-related metabolites need to be
313 further explored in controlled studies confirming that they are food intake
314 biomarkers instead of reflecting metabolic differences due to the pulse

315 consumption. Figure 2 shows both proposed pathways for downstream
316 products of choline.

317



318

319 **Figure 2.** Proposed pathways for choline degradation from pulses including
320 significant metabolites in HP consumers in the present study. Image courtesy of
321 Francisco Madrid-Gambin. Copyright 2016.

322

323 4.1.2. Pulse metabolomic fingerprinting and protein-related compounds

324 With regard to the increases in glutamine and the acetylated form n-
325 acetylglutamine, several explanations may be proposed. Glutamine and n-
326 acetylglutamine could come from dietary sources since glutamine is found in
327 high-protein foods, such as pulses.[47] Another explanation could be the
328 alteration of urinary levels previously shown in this type of population,[48] affected
329 by pulse consumption. There was a higher excretion of 3-methylhistidine in HP
330 consumers. This metabolite is a biomarker of meat and fish consumption,[49]
331 denoting a potential role as a biomarker of consumption. Interestingly, all food
332 sources of this metabolite are also protein sources, including pulses as a

vegetable source, as highlighted in the present study. However, 3-methylhistidine is also a muscle protein breakdown that is sensitive to gender and age.[50] Methylguanidine is derived from protein catabolism and from the breakdown of creatinine,[51,52] therefore it may be related to protein from pulses.

4.1.3. Pulse metabolomic fingerprinting and energy metabolism

The signals of several usual metabolites were altered between the two groups. However, the definition as food intake biomarkers is controversial. Instead, they probably reflect metabolic differences associated with being a low and high consumer, based on the study design. Most of the biomarkers found in the present study are metabolites related to energy metabolism. The lower excretion of acetoacetic acid, glucose and tricarboxylic acid (TCA) cycle intermediates (citric and cis-aconitic acids) appears to involve a different energy modulation according to the pulse consumption. This fact is in part reinforced by changes in BCAAs and subproducts, which are involved in energy metabolism. For example, isobutyric acid is a short-chain fatty acid that is a product of BCAA catabolism of valine, which is a glucogenic BCAA metabolized via the methylmalonyl-CoA in the TCA cycle.[53] On the other hand, acetoacetic acid is a ketone body produced in the human liver for fatty acid breakdown,[54] which serves as a source of energy when normal glycolysis is altered. Interestingly, acetoacetic acid was shown to be increased in diabetes mellitus.[55] Hence, we hypothesize that gluconeogenesis may be diminished in pulse consumers, supported by the urinary reduction of acetoacetic and isobutyric acids (lower fatty acid catabolism), and the reduction of TCA cycle intermediates and urinary

glucose (better use of glucose). Furthermore, it was observed that pulse consumption has a glucose-lowering role in diabetes mellitus,[56,57] thereby explaining the lower plasma glucose concentration and lower urinary excretion. Figure 3 shows the resulting endogenous metabolites connected to the TCA cycle. Nevertheless, the small sample size that resulted after the stratification of the population leads to only exploratory results that should be confirmed.

The role of other findings such as increases of leucine and phenylalanine in pulse consumers is unclear. On the one hand, these habitual urinary compounds could be increased as a consequence of pulses being the source. However, another explanation of these findings could support the hypothesis above. Leucine, which is an acetoacetic acid precursor, may modulate glucose metabolism through oxidation, as well as insulin signalling and release. In addition, stimulation of glucose recycling via the glucose-alanine cycle by leucine may inhibit protein breakdown.[58,59] However, alterations in urinary leucine have also been proposed for the prediction of diabetes mellitus, probably related to the perturbed energy metabolism.[55] The origin of increased phenylalanine is also uncertain. This ketogenic amino acid can stimulate insulin and glucagon concentration, enhancing glucose homeostasis,[60] and is also altered in an insulin-resistant state and obesity.[61] Overall, the consumption of pulses seems to affect the energy metabolism in the studied population.

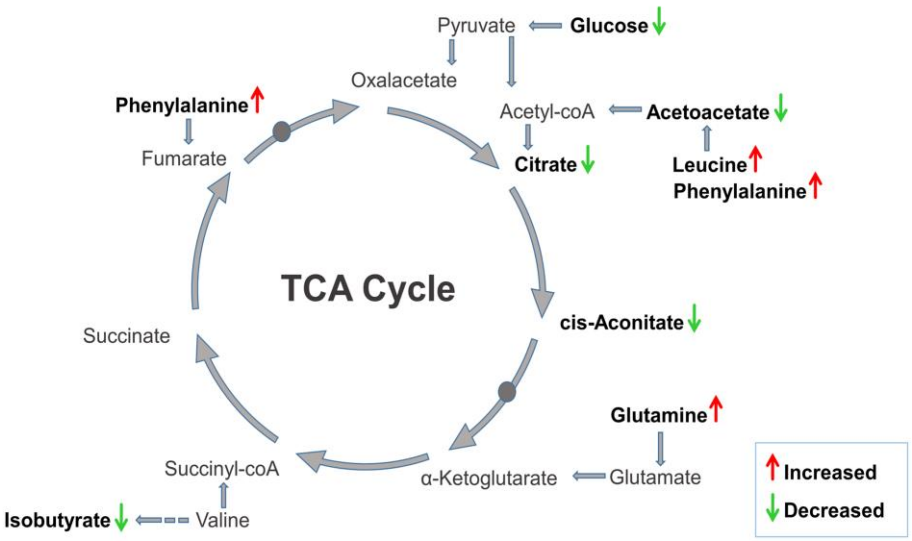


Figure 3. Modified metabolites found in HP consumers connected to energy metabolism.

4.2. New biomarker panel to characterize habitual pulse consumption

To delimit the prediction of habitual pulse intake, comprising lentils, chickpeas and beans, a combination of more than one discriminatory metabolite had to be studied. The combination of three metabolites enhanced considerably the AUC and the confidence interval of the model in comparison with individual metabolites, as shown in Table 2. The developed model indicated that glutamine, 3-methylhistidine and dimethylamine were the strongest candidates for exposure biomarkers. It is important to note that the role of the component coming from choline metabolism suggests the importance of this metabolite as a biomarker of intake. Interestingly, metabolites displaying changes in energy metabolism were scarcely considered by the stepwise logistic regression. None of the other metabolites entered the model, probably as a result of collinearity in the evidence provided by these compounds, which may originate from the same

metabolic pathways, giving similar biological or dietary information.[36] Instead, two metabolites related to protein coming from pulses and one connected to microbiota choline degradation were established in the combined metabolite model, giving complementary information, showing a better discrimination (AUC > 90% in both training and validation sets) than each metabolite individually (AUC < 90% in all cases), and reinforcing the improved capacity of biomarker patterns to distinguish between different dietary exposures.

5. Conclusions

We applied an untargeted ^{1}H-NMR-based metabolomic strategy to distinguish the urinary metabolome of habitual pulse consumption in a free-living population. Stepwise logistic regression analysis exhibited a useful approach to designing a combined urinary biomarker model taking into consideration the different characteristics of pulses. With regard to food metabolome, this study points to a central role of choline contained in pulses and breakdown products such as dimethylglycine, TMAO and dimethylamine. Protein-related compounds such as glutamine, 3-methylhistidine and methylguanidine were also increased in the urine of HP consumers. The combined metabolite model indicated that dimethylamine, 3-methylhistidine and glutamine were the strongest candidates for exposure prediction. In relation to energy metabolism, numerous compounds connected to the TCA cycle, including BCAAs and acetoacetic acid, were modified, denoting a substantial impact on energy metabolism modulation and on urinary glucose in this population. However, since the status of type 2 diabetes mellitus or three or more major cardiovascular risk factors in the

studied population could have a distinctive energy modulation, properly controlled interventions could confirm the findings observed in this cross-sectional study.

6. Supporting Information

Table S1 – Criteria for stratifying participants by frequency of consumption.

Table S2 – Characteristics of the study population according to pulse consumption.

Table S3 – Interaction between gender and the metabolites found in the present study.

Table S4 – Permuted models used in training/validation sets with the resulting metabolites.

Table S5 – Correlations between legume consumption and the combined model for prediction of legume exposure and considered individual metabolites.

Figure S1 – Flow chart of subjects from the PREDIMED subcohort included in the study.

7. Conflict of interest disclosure

The authors declare no competing financial interest.

8. Acknowledgements

This study is supported by Spanish National Grants from the Ministry of Economy and Competitiveness (MINECO), and co-funded by FEDER (Fondo Europeo de Desarrollo Regional): AGL2009-13906-C02-01, JPI HDHL FOODBALL Project (PCIN-2014-133-MINECO Spain), and the award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR. We also thank the EU Joint Programming Initiative "A Healthy Diet for a Healthy Life" on Biomarkers BioNHFOODBALL. F.M-G. acknowledges the APIF PhD fellowship (University of Barcelona). M.U-S. would like to thank the "Ramón y Cajal" programme (RYC-2011-09677) from MINECO and the Fondo Social Europeo. EAA would like to thank CONACYT (Mexico) for the PhD fellowship.

Abbreviations

AUC, area under the curve; FFQ, food frequency questionnaire; FID, free induction decay; HP, habitual pulses; ISRCTN, International Standard Randomized Controlled Trial Number; KOD, potassium deuteroxide; MD, Mediterranean diet; NMR, nuclear magnetic resonance; NP, non-pulses; OSC-PLS-DA, partial least-squares discriminant analysis with orthogonal signal correction; ROC, receiver operating characteristic; TCA, tricarboxylic acid; TMAO, trimethylamine-N-oxide; TSP, 3-(trimethylsilyl)-proprionate-2,2,3,3-d_4; VIP, variable importance projection.

References

(1) Bergouignan, A.; Momken, I.; Schoeller, D. A.; Simon, C.; Blanc, S. Metabolic fate of saturated and monounsaturated dietary fats: the Mediterranean diet revisited from epidemiological evidence to cellular mechanisms. *Prog. Lipid Res.* **2009**, *48* (3–4), 128–147.

(2) Salas-Salvadó, J.; Bulló, M.; Babio, N.; Martínez-González, M. Á.; Ibarrola-Jurado, N.; Basora, J.; Estruch, R.; Covas, M. I.; Corella, D.; Arós, F.; et al. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care* **2011**, *34* (1), 14–19.

(3) Marlow, G.; Ellett, S.; Ferguson, I. R.; Zhu, S.; Karunasinghe, N.; Jesuthasan, A. C.; Han, D. Y.; Fraser, A. G.; Ferguson, L. R. Transcriptomics to study the effect of a Mediterranean-inspired diet on inflammation in Crohn's disease patients. *Hum. Genomics* **2013**, *7* (1), 24.

(4) Lopez-Legarrea, P.; Fuller, N. R.; Zulet, M. A.; Martinez, J. A.; Caterson, I. D. The influence of Mediterranean, carbohydrate and high protein diets on gut microbiota composition in the treatment of obesity and associated inflammatory state. *Asia Pac. J. Clin. Nutr.* **2014**, *23* (3), 360–368.

(5) Estruch, R.; Ros, E.; Salas-Salvadó, J.; Covas, M.-I.; Corella, D.; Arós, F.; Gómez-Gracia, E.; Ruiz-Gutiérrez, V.; Fiol, M.; Lapetra, J.; et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N. Engl. J. Med.* **2013**, *368* (14), 1279–1290.

(6) Bouchenak, M.; Lamri-Senhadji, M. Nutritional quality of legumes, and their role in cardiometabolic risk prevention: a review. *J. Med. Food* **2013**, *16* (3), 1–14.

(7) Roy, F.; Boye, J. I.; Simpson, B. K. Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. *Food Res. Int.* **2010**, *43* (2), 432–442.

(8) Campos-Vega, R.; Loarca-Piña, G.; Oomah, B. D. Minor components of pulses and their potential impact on human health. *Food Res. Int.* **2010**, *43* (2), 461–482.

(9) Sofi, F.; Abbate, R.; Gensini, G. F.; Casini, A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. *Am. J. Clin. Nutr.* **2010**, *92* (5), 1189–1196.

(10) Dilis, V.; Trichopoulou, A. Nutritional and health properties of pulses. *Mediterranean Journal of Nutrition and Metabolism*. IOS Press January 1, 2009, pp 149–157.

(11) Ramalingam, A.; Kudapa, H.; Pazhamala, L. T.; Weckwerth, W.; Varshney, R. K. Proteomics and Metabolomics: Two Emerging Areas for Legume Improvement. *Front. Plant Sci.* **2015**, *6* (December), 1116.

(12) Faris, M. A. I. E.; Takruri, H. R.; Issa, A. Y. Role of lentils (Lens culinaris L.) in human health and nutrition: A review. *Mediterranean Journal of Nutrition and Metabolism*. IOS Press January 1, 2013, pp 3–16.

(13) Bach-Faig, A.; Berry, E. M.; Lairon, D.; Reguant, J.; Trichopoulou, A.; Dernini, S.; Medina, F. X.; Battino, M.; Belahsen, R.; Miranda, G.; et al. Mediterranean diet pyramid today. Science and cultural updates. *Public Health Nutr.* **2011**, *14* (12A), 2274–2284.

(14) U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015–2020 Dietary Guidelines for Americans http://health.gov/dietaryguidelines/2015/ (accessed May 9, 2016).

(15) Britten, P.; Marcoe, K.; Yamini, S.; Davis, C. Development of food intake patterns for the MyPyramid Food Guidance System. *J. Nutr. Educ. Behav.* **2006**, *38* (6 Suppl), S78-92.

(16) Mithril, C.; Dragsted, L. O.; Meyer, C.; Tetens, I.; Biltoft-Jensen, A.; Astrup, A. Dietary composition and nutrient content of the New Nordic Diet. *Public Health Nutr.* **2013**, *16* (5), 777–785.

(17) MacWilliam, S.; Wismer, M.; Kulshreshtha, S. Life cycle and economic assessment of Western Canadian pulse systems: The inclusion of pulses in crop rotations. *Agric. Syst.* **2014**, *123*, 43–53.

(18) O'Gorman, A.; Gibbons, H.; Brennan, L. Metabolomics in the Identification of Biomarkers of Dietary Intake. *Comput. Struct. Biotechnol. J.* **2013**, *4* (5), 1–7.

(19) Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Lupianez-Barbero, A.; Estruch, R.; Corella, D.; Fitó, M.; Arós, F.; Ruiz-Canela, M.; Salas-Salvadó, J.; et al. Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *J. Proteome Res.* **2015**, *14* (1), 531–540.

(20) Garcia-Aloy, M.; Llorach, R.; Urpi-Sarda, M.; Tulipani, S.; Estruch, R.; Martínez-González, M. a.; Corella, D.; Fitó, M.; Ros, E.; Salas-Salvadó, J.; et al. Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: The predimed study. *J. Proteome Res.* **2014**, *13* (7), 3476–3483.

(21) Perera, T.; Young, M. R.; Zhang, Z.; Murphy, G.; Colburn, N. H.; Lanza, E.; Hartman, T. J.; Cross, A. J.; Bobe, G. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol. Nutr. Food Res.* **2015**, *59* (4), 795–806.

(22) Bonetti, A.; Marotti, I.; Dinelli, G. Urinary excretion of kaempferol from common beans (Phaseolus vulgaris L.) in humans. *Int. J. Food Sci. Nutr.* **2007**, *58* (4), 261–269.

(23) Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; van der Hooft, J. J. J.; Wishart, D. S. The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.*

2014, *99* (6), 1286–1308.

(24) McNiven, E. M. S.; German, J. B.; Slupsky, C. M. Analytical metabolomics: nutritional opportunities for personalized health. *J. Nutr. Biochem.* **2011**, *22* (11), 995–1002.

(25) Souza, R. G. M.; Gomes, A. C.; Naves, M. M. V; Mota, J. F. Nuts and legume seeds for cardiovascular risk reduction: scientific evidence and mechanisms of action. *Nutr. Rev.* **2015**, *73* (6), 335–347.

(26) Wilson, C. Nutrition: Consumption of legumes might be beneficial in type 2 diabetes mellitus. *Nat. Rev. Endocrinol.* **2013**, *9* (1), 3.

(27) Martinez-Gonzalez, M. A.; Corella, D.; Salas-Salvado, J.; Ros, E.; Covas, M. I.; Fiol, M.; Warnberg, J.; Aros, F.; Ruiz-Gutierrez, V.; Lamuela-Raventos, R. M.; et al. Cohort Profile: Design and methods of the PREDIMED study. *Int. J. Epidemiol.* **2012**, *41* (2), 377–385.

(28) Estruch, R.; Martinez-Gonzalez, M. A.; Corella, D.; Salas-Salvado, J.; Ruiz-Gutierrez, V.; Covas, M. I. Effects of a Mediterranean-Style Diet on Cardiovascular Risk Factors. *Ann. Intern. Med.* **2006**, *145* (1), 1–11.

(29) Fernández-Ballart, J. D.; Piñol, J. L.; Zazpe, I.; Corella, D.; Carrasco, P.; Toledo, E.; Perez-Bauer, M.; Martínez-González, M. A.; Salas-Salvadó, J.; Martín-Moreno, J. M. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. *Br. J. Nutr.* **2010**, *103*, 1808–1816.

(30) Khan, N.; Khymenets, O.; Urpí-Sardà, M.; Tulipani, S.; Garcia-Aloy, M.; Monagas, M.; Mora-Cubillos, X.; Llorach, R.; Andres-Lacueva, C. Cocoa polyphenols and inflammatory markers of cardiovascular disease. *Nutrients* **2014**, *6* (2), 844–880.

(31) Pujos-Guillot, E.; Hubert, J.; Martin, J. F.; Lyan, B.; Quintana, M.; Claude, S.; Chabanas, B.; Rothwell, J. a.; Bennetau-Pelissero, C.; Scalbert, A.; et al. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: Citrus fruit as a case study. *J. Proteome Res.* **2013**, *12* (4), 1645–1659.

(32) Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Khymenets, O.; Bullo, M.; Corella, D.; Fito, M.; Martinez-Gonzalez, M. A.; Estruch, R.; Andres-Lacueva, C. An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics* **2015**, *11* (4), 797–806.

(33) Food and Agriculture Organization. Definition and classification of commodities, 1994. http://www.fao.org/ES/faodef/fdef04e.htm (Accessed May 2016).

(34) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res.* **2015**, *43* (W1), W251–W257.

(35) Llorach-Asunción, R.; Jauregui, O.; Urpi-Sarda, M.; Andres-Lacueva, C. Methodological aspects for metabolome visualization and characterization: a metabolomic evaluation of the 24 h evolution of human urine after cocoa powder consumption. *J. Pharm. Biomed. Anal.* **2010**, *51* (2), 373–381.

(36) Garcia-Aloy, M.; Llorach, R.; Urpi-Sarda, M.; Jáuregui, O.; Corella, D.; Ruiz-Canela, M.; Salas-Salvadó, J.; Fitó, M.; Ros, E.; Estruch, R.; et al. A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED study. *Mol. Nutr. Food Res.* **2015**, *59* (2), 212–220.

(37) Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. **1993**, *114* (3), 494–509.

(38) Caraux, G.; Pinloche, S. PermutMatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* **2005**, *21* (7), 1280–1281.

(39) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **2013**, *9* (2), 280–299.

(40) De Almeida Costa, G. E.; Da Silva Queiroz-Monici, K.; Pissini Machado Reis, S. M.; De Oliveira, A. C. Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. *Food Chem.* **2006**, *94* (3), 327–330.

(41) Lewis, E. D.; Kosik, S. J.; Zhao, Y.-Y.; Jacobs, R. L.; Curtis, J. M.; Field, C. J. Total choline and choline-containing moieties of commercially available pulses. *Plant Foods Hum. Nutr.* **2014**, *69* (2), 115–121.

(42) Tremaroli, V.; Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **2012**, *489* (7415), 242–249.

(43) De Filippis, F.; Pellegrini, N.; Vannini, L.; Jeffery, I. B.; La Storia, A.; Laghi, L.; Serrazanetti, D. I.; Di Cagno, R.; Ferrocino, I.; Lazzi, C.; et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut* **2015**, *65* (11), 1812–1821.

(44) Zhang, A. Q.; Mitchell, S. C.; Ayesh, R.; Smith, R. L. Dimethylamine formation in man. *Biochem. Pharmacol.* **1993**, *45* (11), 2185–2188.

(45) Dumas, M.-E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C.; et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (33), 12511–12516.

(46) Friesen, R. W.; Novak, E. M.; Hasman, D.; Innis, S. M. Relationship of Dimethylglycine, Choline, and Betaine with Oxoproline in Plasma of Pregnant Women and Their Newborn Infants. *J. Nutr.* **2007**, *137* (12), 2641–2646.

(47) Marinangeli, C. P. F.; Jones, P. J. H. Pulse grain consumption and obesity: effects on energy expenditure, substrate oxidation, body composition, fat deposition and satiety. *Br. J. Nutr.* **2012**, *108* (S1), S46–S51.

(48) Roberts, L. D.; Koulman, A.; Griffin, J. L. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol.* **2014**, *2* (1), 65–75.

(49) Brennan, L.; Gibbons, H.; O'Gorman, A. An Overview of the Role of Metabolomics in the Identification of Dietary Biomarkers. *Curr. Nutr. Rep.* **2015**, *4* (4), 304–312.

(50) Aranibar, N.; Vassallo, J. D.; Rathmacher, J.; Stryker, S.; Zhang, Y.; Dai, J.; Janovitz, E. B.; Robertson, D.; Reily, M.; Lowe-Krentz, L.; et al. Identification of 1- and 3-methylhistidine as biomarkers of skeletal muscle toxicity by nuclear magnetic resonance-based metabolic profiling. *Anal. Biochem.* **2011**, *410* (1), 84–91.

(51) Yokozawa, T.; Fujitsuka, N.; Oura, H. Studies on the precursor of methylguanidine in rats with renal failure. *Nephron* **1991**, *58* (1), 90–94.

(52) Ando, A.; Orita, Y.; Nakata, K.; Tsubakihara, Y.; Takamitsu, Y.; Ueda, N.; Yanase, M.; Abe, H. Effect of low protein diet and surplus of essential amino acids on the serum concentration and the urinary excretion of methylguanidine and guanidinosuccinic acid in chronic renal failure. *Nephron* **1979**, *24* (4), 161–169.

(53) Hutson, S. M.; Sweatt, A. J.; Lanoue, K. F. Branched-chain amino acid metabolism: implications for establishing safe intakes. *J Nutr* **2005**, *135* (6 Suppl), 1557S–64S.

(54) Stern, J. R. Enzymes of acetoacetate formation and breakdown. *Methods Enzymol.* **1955**, *1*, 573–585.

(55) Urpi-Sarda, M.; Almanza-Aguilera, E.; Tulipani, S.; Tinahones, F. J.; Salas-Salvadó, J.; Andres-Lacueva, C. Metabolomics for Biomarkers of Type 2 Diabetes Mellitus: Advances and Nutritional Intervention Trends. *Curr. Cardiovasc. Risk Rep.* **2015**, *9* (3), 12.

(56) Mudryj, A. N.; Yu, N.; Aukema, H. M. Nutritional and health benefits of pulses. *Appl. Physiol. Nutr. Metab. = Physiol. Appl. Nutr. métabolisme* **2014**, *39* (11), 1197–1204.

(57) Singhal, P.; Kaushik, G.; Mathur, P. Antidiabetic potential of commonly consumed legumes: a review. *Crit. Rev. Food Sci. Nutr.* **2014**, *54* (5), 655–672.

(58) Layman, D. K.; Walker, D. A. Potential importance of leucine in treatment of obesity and the metabolic syndrome. *J. Nutr.* **2006**, *136* (1 Suppl), 319S–23S.

(59) Shearer, J.; Duggan, G.; Weljie, A.; Hittel, D. S.; Wasserman, D. H.; Vogel, H. J. Metabolomic profiling of dietary-induced insulin resistance in

669 the high fat-fed C57BL/6J mouse. *Diabetes, Obes. Metab.* **2008**, *10* (10),
670 950–958.

671 (60) Nuttall, F. Q.; Schweim, K. J.; Gannon, M. C. Effect of orally administered
672 phenylalanine with and without glucose on insulin, glucagon and glucose
673 concentrations. *Horm. Metab. Res. = Horm. und Stoffwechselforsch. =*
674 *Horm. métabolisme* **2006**, *38* (8), 518–523.

675 (61) Adams, S. H. Emerging Perspectives on Essential Amino Acid
676 Metabolism in Obesity and the Insulin-Resistant State. *Adv. Nutr. An Int.*
677 *Rev. J.* **2011**, *2* (6), 445–456.

678

679 **Table 1.** Tentative discriminant metabolites derived from the multi- and
680 univariate analysis of ^{1}H-NMR signal intensities in urine from HP consumers[a]

Source	Metabolite	HP vs NP	δ (multiplicity)	FDR p-value[†]	Cliff's delta[§]
Choline metabolism	Choline	↑	3.19 (s)	3.27×10^{-2}	0.475
	Dimethylglycine	↑	2.93 (s)	3.81×10^{-2}	0.386
	TMAO	↑	3.27 (s)	7.29×10^{-3}	0.485
	Dimethylamine	↑	2.72 (s)	1.05×10^{-2}	0.488
Protein-related compounds	N-acetylglutamine	↑	2.04 (s) 2.08 (m) 2.26 (m) 4.18 (m)	2.55×10^{-2}	0.706
	Glutamine	↑	2.12 (m) 2.46 (m) 3.76 (t)	1.17×10^{-6}	0.814
	Phenylalanine	↑	3.19 (m) 3.98 (dd) 7.32 (d) 7.36 (m) 7.42 (m)	3.21×10^{-2}	0.354
	Methylguanidine	↑	2.83 (s)	3.72×10^{-4}	0.635
	3-Methylhistidine	↑	7.18 (s) 7.92 (s)	1.73×10^{-4}	0.658
Energy metabolism	Citric acid	↓	2.55 (dd) 2.69 (dd)	8.43×10^{-5}	-0.690
	Cis-aconitic acid	↓	5.74 (s) 3.12 (s)	1.11×10^{-3}	-0.629
	Glucose	↓	3.50 (m) 4.66 (d) 5.25 (d)	7.89×10^{-5}	-0.718
	Acetoacetic acid	↓	2.27 (s)	1.95×10^{-2}	-0.408
	Isovalerylglycine	↑	0.92 (d) 2.16 (d) 3.74 (d)	2.83×10^{-4}	0.635
	Leucine	↑	0.94 (t) 1.70 (m) 3.72 (m)	1.18×10^{-3}	0.626
	Isobutyric acid	↓	1.06 (d)	1.29×10^{-2}	-0.446

681 [a]All features have VIP values ≥1.0 in the corresponding OSC-PLS-DA model. [†]P-value
682 of Student's t-test with False Discovery Rate correction. [§]Estimation of the effect size
683 by Cliff's delta with thresholds: |n|<0.330 "small", 0.330>|n|<0.474 "medium" and
684 |n|>0.474 "large". TMAO, trimethylamine-N-oxide. s: singlet, d: doublet, t: triplet, dd:
685 double doublet, m: multiplet.

686

687 **Table 2.** Receiver operating characteristic (ROC) curve parameters of
688 combined models and of individual metabolites in both training and validation
689 sets

	Set[†]	Sensitivity (%)	Specificity (%)	AUC (95% CI)
Combined model	Training	88.2	93.7	95.6 (89.4–100.0)
	Validation	87.5	88.9	94.4 (84.1–100.0)
3-Methylhistidine	Training	76.5	87.5	82.4 (67.7–97.0)
	Validation	87.5	77.8	80.6 (56.2–100.0)
Glutamine	Training	76.5	81.2	81.6 (67.0–96.3)
	Validation	87.5	77.8	84.7 (65.3–100.0)
Dimethylamine	Training	82.4	62.5	75.0 (57.8–92.2)
	Validation	50.0	66.7	68.1 (40.4–100.0)

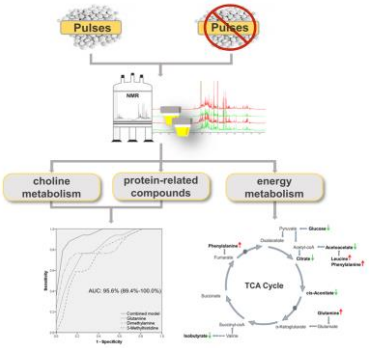
690 AUC: area under the ROC curve. CI: confidence interval. [†]Corresponding to 2/3 of the
691 population for the training and 1/3 for the validation set.

692

Figure 1. Receiver operating characteristic (ROC) curves of combined model (continuous line) with the area under the ROC curve and of included individual metabolites (discontinuous lines) in the training (A) and validation (B) sets.

Figure 2. Proposed pathways for choline degradation from pulses including significant metabolites in HP consumers in the present study.

Figure 3. Modified metabolites found in HP consumers connected to energy metabolism.

703 

704 **For Table of Contents Only**

705 Image courtesy of Francisco Madrid-Gambin. Copyright 2016.

706