

Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses

Etienne A. Thévenot,^{*,†,¶} Aurélie Roux,^{‡,¶} Ying Xu,[‡] Eric Ezan,[‡] and Christophe Junot^{*,‡}

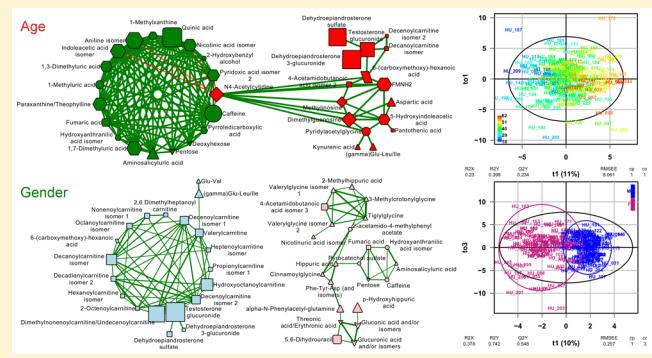
[†]CEA, LIST, Laboratory for Data Analysis and Smart Systems, MetaboHUB Paris, F-91191 Gif-sur-Yvette, France

[‡]Laboratoire d'Etude du Métabolisme des Médicaments, DSV/iBiTec-S/SPI, MetaboHUB Paris, CEA-Saclay, Gif-Sur-Yvette, France

Supporting Information

ABSTRACT: Urine metabolomics is widely used for biomarker research in the fields of medicine and toxicology. As a consequence, characterization of the variations of the urine metabolome under basal conditions becomes critical in order to avoid confounding effects in cohort studies. Such physiological information is however very scarce in the literature and in metabolomics databases so far. Here we studied the influence of age, body mass index (BMI), and gender on metabolite concentrations in a large cohort of 183 adults by using liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS). We implemented a comprehensive statistical workflow for univariate hypothesis testing and modeling by orthogonal partial least-squares (OPLS), which we made available to the metabolomics community within the online Workflow4Metabolomics.org resource. We found 108 urine metabolites displaying concentration variations with either age, BMI, or gender, by integrating the results from univariate *p*-values and multivariate variable importance in projection (VIP). Several metabolite clusters were further evidenced by correlation analysis, and they allowed stratification of the cohort. In conclusion, our study highlights the impact of gender and age on the urinary metabolome, and thus it indicates that these factors should be taken into account for the design of metabolomics studies.

KEYWORDS: urine, metabolomics, high-resolution mass spectrometry, cohort, age, BMI, gender, orthogonal partial least-squares, statistics, workflow



INTRODUCTION

Scientific examination of urine has been used since antiquity for diagnostic purposes.¹ Compared to other biofluids, it can be collected noninvasively, in large amounts, and by repeat sampling. The preparation of urine samples for analysis is simple, and the concentration of many metabolites is amplified by bladder storage. As the primary route of elimination of water-soluble waste compounds, urine is a window to biochemical changes² and a medium of choice for biomarker discovery by using metabolomics based approaches.³

Analysis of urine metabolites (small molecules <1.5 kDa) has been performed using nuclear magnetic resonance (NMR)^{4–7} and mass spectrometry coupled to either gas (GC-MS)^{8,9} or liquid chromatography (LC-MS).^{10–18} Reference protocols have been published for NMR,¹⁹ GC-MS,²⁰ and LC-MS²¹ analyses.

Stability of urine metabolites has been shown by multivariate visualization of sample patterns after LC-MS acquisition for up to 48 h at 4 °C, up to one month at –20 °C, and after nine freeze–thaw cycles.¹² Detailed follow-up of 280 metabolites further

confirmed that 93% of them did not show significant concentration variations up to 72 h after collection when kept at 4 °C.¹⁸ To correct for random intersubject variations of diuresis, various normalization methods have been described, including the traditional use of urinary creatinine concentration, osmolality,¹³ the totally useful MS signal,¹³ and specific gravity,^{22,23} but also a combination of creatinine concentration and MS signals's normalization,²⁴ and the determination of the total concentration of chemically labeled metabolites by using LC-UV.²⁵

Comprehensive annotation and identification of human urinary metabolites have been reported by a few recent studies, such as that of Zhang et al. which led to the detection of approximatively 970 metabolite signals using hydrophilic interaction liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS),²⁶ the work of Roux et al. in

Received: April 27, 2015

which 384 metabolites were annotated by LC-HRMS/MS,¹⁶ or the work of Bouatra et al. which reports the detection of 445 metabolites by a combination of NMR, GC-MS, DFI/LC-MS/MS, and ICP-MS.¹⁷ The human urine metabolome database UMDB currently contains about 3,100 annotated metabolite species.¹⁷

In contrast, variations of metabolite concentrations in urine due to physiological effects (such as age, body mass index BMI, or gender) have received little attention so far. Such physiological differences have however to be carefully taken into account in biomarker studies to avoid potential confounding effects between cases and controls.^{27,28} The few data available in urine come from NMR studies only, and they are very scarce in metabolomics databases (such as HMDB) to date. In a pioneer study in rats, Bolland et al. investigated the variation of urine composition due to various physiological effects including age, gender, diet, species, and hormonal status.⁴ In humans, 17 metabolites were reported to vary with either age, gender, diet, or circadian rhythms in two publications, when analyzing groups of 60 and 24 volunteers, respectively.^{5,7} An untargeted metabolomics analysis of a large cohort by using a sensitive LC-HRMS based technology would therefore be very useful to provide a complementary and comprehensive description of the physiological variations of the urine metabolome.

To detect variations in metabolite concentrations between conditions of interest, robust statistical methods are critical.²⁹ Both univariate hypothesis testing and multivariate orthogonal partial least-squares (OPLS) modeling approaches are increasingly gaining popularity.^{30–33} On the one hand, univariate approaches, where each variable is independently investigated by hypothesis testing³⁴ have been extensively used, and several methods have been developed to correct the inflation of the false positive rate due to the repetition of tests.³⁵ Univariate strategies provide a direct measure of significance, the *p*-value, and nonparametric tests are useful in metabolomics to handle data with few samples, non-Gaussian distributions, or outlier intensities.³⁶

On the other hand, chemometrics methods, which provide a single model for the whole data set, have been developed.³⁷ In particular, partial least-squares, which is a latent variable regression method based on covariance between the predictors and the response, has been shown to efficiently handle data sets with multicollinear predictors, as in the case of spectrometry measurements.³⁸ More recently, Trygg and Wold introduced the orthogonal projection to latent structures (OPLS) algorithm to model separately the variations of the predictors correlated and orthogonal to the response.³⁹ OPLS has a similar predictive capacity compared to PLS and improves the interpretation of the predictive components and of the systematic variation.⁴⁰ In particular, OPLS modeling of single responses only requires one predictive component. Diagnostics such as the Q2Y metrics and permutation testing are of high importance to avoid overfitting and assess the statistical significance of the model. The variable importance in projection (VIP), which reflects both the loading weights for each component and the variability of the response explained by this component,^{40,41} is often used for feature selection.^{39,40} OPLS is available in the SIMCA-P commercial software (Umetrics, Umeå, Sweden).⁴² In addition, the kernel-based version of OPLS⁴³ is available in the open-source R statistical environment,⁴⁴ and a single implementation of the linear algorithm in R has been described very recently⁴⁵ but still lacks major functionalities for computation (handling of missing values), diagnostics (R² and Q², permutation testing), and

feature selection (VIP). An implementation of the full OPLS approach in R would therefore be of high interest, since it would facilitate direct comparison of results and integration with other data mining methods within a comprehensive statistical workflow.

In a previous study, we described the structural annotation of urine metabolites by nontargeted liquid chromatography coupled to high resolution mass spectrometry (LC-MS) analysis of a large adult cohort consisting of employees working at the CEA research institute.¹⁶ The objective of the present work is to characterize the physiological variability with age, body mass index (BMI), and gender of these metabolites. By implementing a statistical workflow including univariate hypothesis testing and OPLS modeling, we could compare and integrate the results from both approaches and identify 108 metabolites showing significant concentration variations with age, BMI, or gender. We further characterized several clusters of correlated metabolites which provide a stratification of the cohort.

■ EXPERIMENTAL SECTION

Cohort

Urine samples were collected from a cohort of 189 human adults during their routine examination as employees working at the CEA research center of Saclay. Sample collection was performed with informed consent of the subjects, in accordance with the 1964 Helsinki declaration and its later amendments. Five volunteers with high BMI values (from 32.9 to 41.3 kg/m²) were not included since these values were considered extreme (as confirmed by significant *p*-values with the Grubbs outlier test) and representative of too few subjects to provide robust statistical information. An additional sample was further discarded after normalization due to outlier values in the negative ionization mode, leading to 183 individual samples under study.

Sample Preparation, UPLC–MS Analysis, and Metabolite Identification

The experimental protocol (chemicals, biological material, sample preparation, liquid chromatography, mass spectrometry conditions, compound identification) has been described previously (“439020” study).¹⁶ Briefly, an Accela liquid chromatographic system (Thermo Fisher Scientific, Villebon-sur-Yvette, France) was coupled to an LTQ-Orbitrap Discovery (Thermo Fisher Scientific, Villebon-sur-Yvette, France) fitted with an electrospray (ESI) source. The ultrahigh performance liquid chromatographic (UHPLC) separation was performed on a Hypersil GOLD C18 1.9 μm, 2.1 mm × 150 mm column (Thermo Fisher Scientific, Villebon-sur-Yvette, France). The flow rate was 0.5 mL/min with mobile phases A (100% water) and B (100% acetonitrile), both containing 0.1% formic acid. The gradient consisted of an isocratic step of 2 min at 100% phase A, followed by a linear gradient from 0% to 100% of phase B for the next 11 min, before returning to 100% A in 0.5 min. The column was then allowed to equilibrate for 3.5 min. The mass resolution of the spectrometer was 30,000 ($m/\Delta m$ at 400 u) and the mass accuracy was within a 5 ppm range. All samples were analyzed in three batches (one for the positive mode *pos* and two for the negative mode: *ne1* and *ne2*). The ESI source was cleaned between *pos* and *ne1*, and between *ne1* and *ne2*. The same quality control (QC) sample for all batches (mix of urines from a subset of the cohort) was injected regularly throughout the run after every ten samples of interest. Peak detection, grouping, and retention time correction with the XCMS software⁴⁶ resulted in two peak tables (one for each ionization mode). Annotation was

performed by using our in-house ESI-mass spectra database, as well as the HMDB,⁴⁷ KEGG,⁴⁸ and METLIN⁴⁹ public databases, and identities were confirmed with additional MS/MS experiments when required.

In this study, integration of each annotated peak was performed with the Quan Browser module from the Xcalibur software (Thermo Fisher Scientific) after visual checking of the retention time limits, in order to refine signal quantitation and avoid any bias that may have occurred during the initial XCMS procedure. The few missing values in the *pos* and *neg* peak tables (less than 2%) were replaced by half the minimum of the nonmissing values (results without replacement of missing values were similar).

Signal Drift Correction and Batch Effect Removal

Within each peak table, intensities were corrected for signal drift and batch effect by fitting a locally quadratic (*loess*) regression model to the QC values.^{50,51} The α parameter controlling the smoothing⁵² was set to 1 to avoid overfitting. Once the peak tables were normalized, metabolites with a coefficient of variation (CV) of their QC values >30% (which was the case for 10% and 4% of the features of the *pos* and *neg* tables, respectively) were filtered out.

Normalization and Quality Control

Each urine profile was normalized to the osmolality of the sample.¹³ The two peak tables were subsequently log10 transformed. One sample with both an extreme intensity distribution quantile⁵³ ($p < 10^{-15}$) and an extreme Hotelling's T2 distance in the PCA score plot⁵⁴ ($p < 10^{-15}$) in the negative ionization mode, was discarded. The *pos* and *neg* peak tables in addition to the tables for sample and metabolite metadata are provided as Supporting Information file 2.

Univariate Statistical Analysis

Correlations with age and BMI covariates were assessed with the Spearman rank correlation test, and the Wilcoxon (Mann–Whitney) test was used for the comparison of gender distributions. The *p*-values were adjusted for multiple comparison by controlling the false discovery rate (proportion of false positives among the metabolites called significant) at a 5% threshold.⁵⁵

Multivariate Modeling

The *ropls* R package implements the PCA, PLS(-DA) and OPLS(-DA) approaches with the original, NIPALS-based, versions of the algorithms.^{38,39} It includes the R2 and Q2 quality metrics,^{42,55} the permutation diagnostics,⁵⁶ the computation of the VIP values,⁵⁷ the score and orthogonal distances to detect outliers,⁵⁸ as well as many graphics (scores, loadings, predictions, diagnostics, outliers, etc.). The *roppls* package is available from the Bioconductor repository.⁵⁹

The number of orthogonal components of the OPLS models for the age, BMI, and gender responses, as assessed by the Q2Y metrics, was 1, 1, and 3, respectively. The variables were standardized (i.e., mean-centered and unit-variance scaled) prior to model building as it resulted in the highest Q2Y (results with Pareto scaling were similar).

Correlation Networks and Hierarchical Clustering

The Spearman correlation (cor_{spe}) coefficient was used as the measure of similarity between metabolite profiles among individuals. Only edges corresponding to an absolute correlation >0.5 are shown. Hierarchical clustering of individuals and

metabolites were performed using $1-cor_{spe}$ as the dissimilarity, and the Ward's linkage method.

Software

All computational modules (for data normalization, quality control, univariate testing and multivariate modeling) were written in R (version 3.0.2)⁴⁴ and integrated into the Workflow4Metabolomics.org online resource for computational metabolomics.⁶⁰

Correlation networks were visualized with the Cytoscape software⁶¹ (version 3.2.1). When required, image editing was performed with the GIMP software (www.gimp.org; version 2.8.10).

RESULTS

Data Description and Normalization

Urine samples were collected from a cohort of 183 human adults (Figure 1): 100 males (55%) and 83 females (45%), aged 40.9 \pm

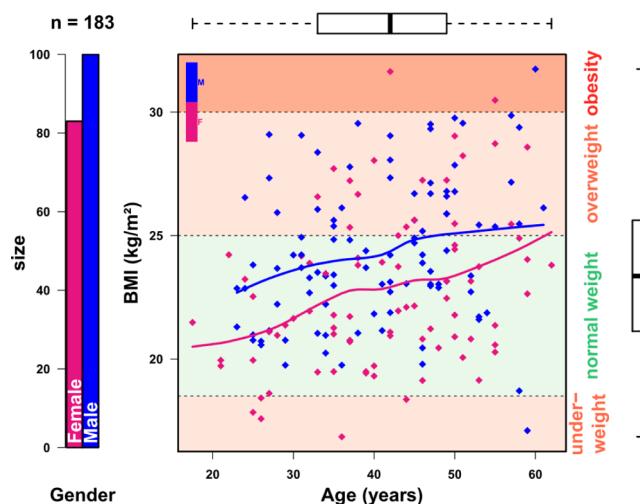


Figure 1. Cohort description. Left: Number of females and males in the cohort. Right: BMI as a function of age among males and females. Horizontal dotted lines indicate underweight (below 18.5 kg/m²), normal weight (18.5–25 kg/m²), overweight (25–30 kg/m²), and obesity (above 30 kg/m²). The *loess* trend for each gender category is shown. The distributions of age and BMI values are displayed by the neighboring boxplots.

10.3 years (mean \pm standard deviation) with a body mass index (BMI) of 23.6 \pm 3.1 kg/m². BMI was found significantly higher in males ($p = 2 \times 10^{-4}$) and positively correlated with age ($p = 9 \times 10^{-5}$; Figure 1).

Sample analysis by LC-HRMS, as described previously,¹⁶ had been performed in three runs (one, *pos*, and two, *ne1* and *ne2*, batches for the positive and negative ionization modes, respectively), and a single pooled sample had been used as quality control (QC; Figure S1). Preprocessing with XCMS had resulted in two peak tables. Annotation had been achieved by in-house and public database matching and MS/MS experiments.¹⁶

This work led to the identification of 258 metabolites in human urine samples¹⁶ either formally (when at least two physicochemical parameters, such as chromatographic retention time and MS/MS spectrum, matched those of our spectral library of reference compounds) or putatively (based on information from databases and the interpretation of MS and MS/MS spectra), corresponding to levels 1 and 2 from the metabolomics standard initiative⁶² (MSI). Among them, 226 were found to be

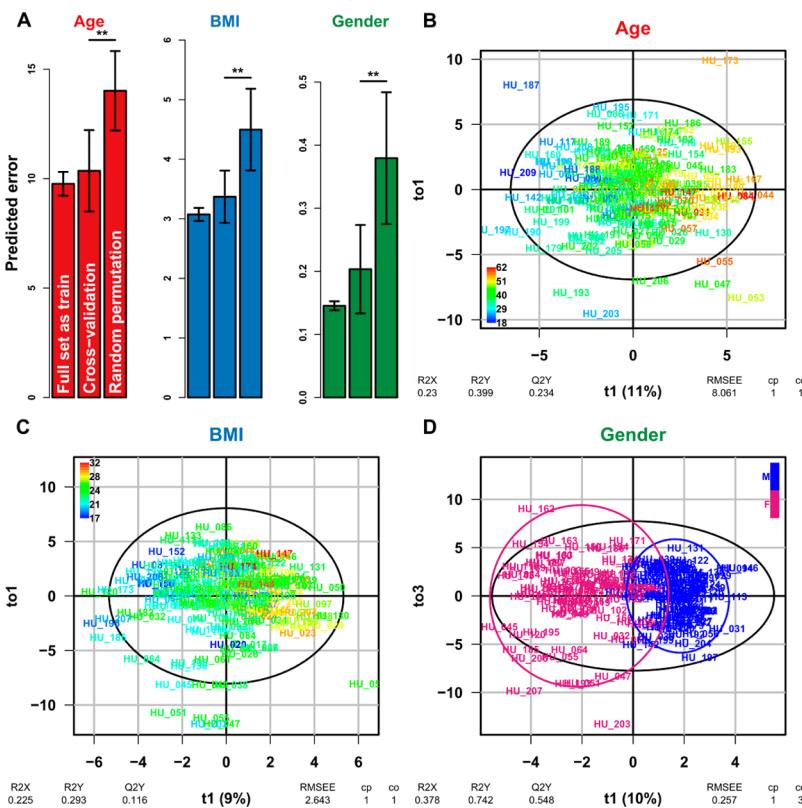


Figure 2. Multivariate modeling of physiological variations using OPLS. Models with one predictive (t1) and one to three orthogonal components (to) were built with the metabolites from the positive data set as predictors and age, BMI, or gender as the response. (A) The quality of each model was evaluated by computing either the root mean squared errors (age and BMI regression) or the mean of the false positive and negative rates (gender classification) after splitting the data set into 7 segments: predictions were computed either (i) by using the same subset for training and testing, (ii) by testing the model on the left out segment, or (iii) by randomly permuting the response values during the training before testing the model on the left out segment. The asterisks indicate the *p* value interval associated with the hypothesis testing of an equality of error means with or without random permutation (**: *p* < 0.01; ***: *p* < 0.001). (B–D) Score plot of each model in the plane of the first predictive (t1) and the last orthogonal (to) components. The percentage of response variance explained by the predictor component only (t1) is indicated in parentheses. R²X (respectively R²Y): percentage of predictor (respectively response) variance explained by the full model. Q²Y: predictive performance of the model estimated by cross-validation. For the classification model (gender), the ellipses corresponding to 95% of the multivariate normal distributions with the samples covariances for each class are shown.

reliable for statistical and biological interpretation (i.e., no issue of isomers), including 195 metabolites which were detected in QC samples. Finally, 25 were removed due to detection sensitivity issues, resulting in a total of 170 metabolites for biological interpretation.

The objective of the present study was to characterize, among these 170 previously identified or annotated metabolites, those that vary with either the age, BMI, or gender physiological factors. To avoid any analytical bias and optimize the quality of the data for subsequent statistical analysis, the two peak tables were processed as follows. First, instrumental drift (within each batch) and offset differences between batches *ne1* and *ne2* observed in the two peak tables were adjusted after fitting a regression model to the QC values (Figure S2).^{50,51} This resulted in 90% (*pos*) and 96% (*neg*) of the metabolites meeting the 30% quality control criterion for the coefficient of variation of QC intensities. Second, to correct for the random variability of solute concentrations in urine (e.g., due to the amount of water consumption), we used the osmolality measurement of each sample and normalized the corresponding metabolite profile.¹³ The data were then log10 transformed in order to stabilize the variance. Finally, the quality of the sample profiles was controlled by computing the Hotelling's T₂ statistics⁵⁴ and the Z-scores of the distribution quantiles,⁵³ and one sample with significant *p*-

values for both statistics in the negative ion mode was discarded. The two processed peak tables, in addition to the associated sample and metabolite metadata, are provided in Supporting Information file 2.

Implementation of a Comprehensive Workflow for Univariate Hypothesis Testing and OPLS Modeling

In order to identify the metabolites showing physiological variations, we implemented hypothesis testing and orthogonal partial least-squares (OPLS) modeling. These statistical approaches are often used in metabolomics for biomarker discovery⁶³ and may be complementary.⁶⁴

First, for each variable, pFDR values (i.e., *p*-values corrected for multiple testing) from the nonparametric hypothesis testing of the correlation with age or BMI, or the median differences between genders, were computed. A total of 124 metabolites were significant with this univariate approach. Except for pyridoxic acid isomer 1, whose correlations with age were significant in both modes but with opposite signs (this metabolite was therefore discarded as an outlier), results were in good agreement for the metabolites detected in both modes, with an average agreement for the three physiological conditions of 81% in statistical significance and of 96% agreement in trend (sign of

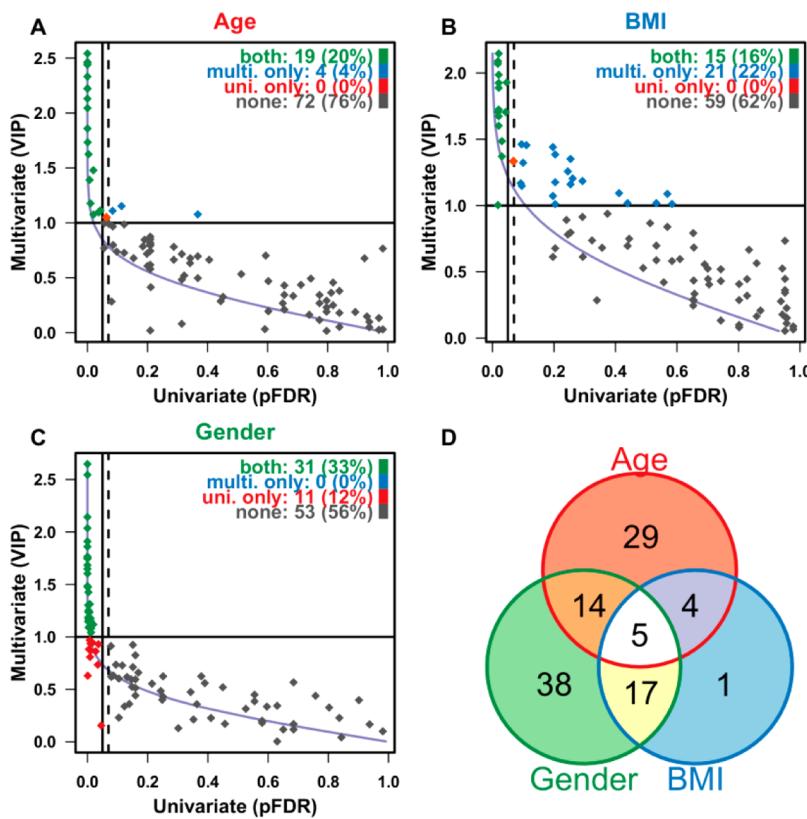


Figure 3. Selection of variable metabolites by combined univariate and multivariate approaches. (A–C) Comparison between the univariate and multivariate variable selection for the data set from the positive ionization mode. The metabolites are colored according to their selection by one of the approaches or both. The $p\text{FDR} = 0.05$ (respectively $\text{VIP} = 1$) thresholds are displayed as a vertical (respectively horizontal) line. The dashed vertical line indicates the $p\text{FDR} = 0.07$ threshold used for the combination of univariate and multivariate results. The metabolites added to the univariate selection (i.e., those in the $0.05 \leq p\text{FDR} < 0.07$ and $\text{VIP} > 1$ region) are colored in orange. The relationship between VIP and the noncorrected p -values is shown as a purple curve. (D) Venn plot comparing the numbers of metabolites showing variations with age, BMI, or gender.

the correlation with age or BMI, or of the difference between gender medians).

Second, multivariate models were built by using orthogonal partial least-squares (OPLS).^{39,40} To this end, we wrote an R package named *ropls* for multivariate modeling which we used to model each response (Figure 2 and Figure S3).

OPLS modeling of all responses was found significant (except the BMI modeling with the *neg* data set) when compared with models built on random permutation of the labels (Figure 2A and Figure S3A). The predictive ability (estimated by the Q_2Y metrics) was relatively high for the gender models (Figure 2B–D and Figure S3B–D) and was lower (but still significant, Figure S4) for the age and BMI models.

Statistical Selection of Metabolites with Physiological Variations

Univariate ($p\text{FDR}$) and multivariate (VIP) metrics were first compared graphically (Figure 3A–C and Figure S5). The $p\text{FDR} < 0.05$ and $\text{VIP} > 1$ rules were in good agreement, with an average value for both *pos* and *neg* data sets of 96%, 81%, and 89%, for the age, BMI, and gender response respectively (agreement is computed as the mean of the number of metabolites for which both rules result in acceptance or rejection, over the total number of metabolites). On the BMI plot (Figure 3B), a shift to the right was observed, indicating a lower Spearman correlation of the metabolite profiles with this response.

As a clear pattern was observed in the ($p\text{FDR}$, VIP) cloud plots (Figure 3A–C and Figure S5), we studied the relationship between the two metrics. In fact, in this specific case of OPLS

standardized modeling of a single response, we could demonstrate a mathematical relationship between VIP and noncorrected p -values from a Pearson correlation test (see Figure S6 and accompanying text in Supporting Information file 1) which is shown as a purple curve in Figure 3A–C. In our plots, the observed deviations of the points from the curve therefore resulted solely from the fact that we used the nonparametric (Spearman) version of the correlation test and that our p -values were corrected for multiple testing (resulting in a shift to the right of the point cloud). As a result, the VIP values did not provide information about interactions between variables, but they could nevertheless be used to highlight the few metabolites with a strong Pearson correlation which would not have been detected due to the lower power of our nonparametric univariate test. We thus defined a methodology combining our $p\text{FDR}$ and VIP values as follows (Figure 3A–C): the nonparametric univariate $p\text{FDR} < 0.05$ was kept as a reference (since many metabolite profiles deviate from the normal distribution, despite the \log_{10} transformation, the nonparametric correlation coefficient was more robust), and the metabolites with $\text{VIP} > 1$ up to $p\text{FDR} < 0.07$ were added to the selection (we observed for higher $p\text{FDR}$ values that $\text{VIP} > 1$ could misleadingly result from non-normal metabolite profiles). This resulted in the addition of four metabolite variations to our univariate list, including the previously not selected 3-hydroxyphenylacetic acid.

Table 1. Identified Metabolites with Significant Correlation with Age or BMI, or Median Difference between Genders^a

Molecule	Annot. level	Chemical class	HMDB	Mode	Age (log2 ratio)	Age: BMI (sig.)	BMI (log2 ratio)	BMI: Gender (sig.)	Gender (log2 ratio)	Age: Gender (sig.)	Age: BMI: Gender (sig.)	Cluster group
Aniline isomer	2	AroHoM		pos	8.9							2
Indoleacetic acid isomer	2	AroHeP		pos	7.6							2
Pyridoxic acid isomer 2	2	AroHeM		pos	5.6	*						2
Nicotinic acid isomer	2	AroHeM		pos	5.3							2
Caffeine	1	Xenobi	HMDB01847	pos	5.2				-0.95			2
Quinic acid	1	AliHoM	HMDB03072	neg	4.4							2
Aminosalicyluric acid	2	AA-pep		pos	3.6				-0.96	*		2
				neg	2.3							2
Adipoylcarnitine	2	AcyCar		pos	3.5				1.1			1
1,7-Dimethyluric acid	1	AroHeP	HMDB11103	pos	3.3							2
				neg	2.3							2
1,3-Dimethyluric acid	1	AroHeP	HMDB01857	pos	3.1		2					2
				neg	2.8		1.7	*				2
Paraxanthine/Theophylline	2	AroHeP		pos	3							2
1-Methylxanthine	1	AroHeP	HMDB10738	neg	1.9							2
Fumaric acid	1	Lipids	HMDB00134	neg	1.8				-0.35			2
Pyrroledicarboxylic acid	2	AroHeM		neg	1.8							2
1-Methyluric acid	1	Xenobi	HMDB03099	neg	1.6							2
Hydroxyanthranilic acid isomer	2	AroHeM		pos	1.6				-0.46	*		2
2-acetamido-4-methylphenyl acetate	2	AroHoM		pos	1.6				-0.62			3
				neg			-1.6		-0.51			3
Acetylphenylalanine	1	AA-pep	HMDB00512	pos	1.6				-0.8			3
				neg	1.2				-0.71			3
Methoxysalicylic acid isomer	2	AroHoM		neg	1.4							2
Tetrahydrohippuric acid	2	AcyGly		pos	1.3							3
Normetanephrine isomer	2	AroHoM		pos	1.3							3
2-Hydroxybenzyl alcohol	1	AroHoM	HMDB59709	neg	1.3							2
Threo-3-Phenylserine	1	AA-pep	HMDB02184	pos	1.3							3
N-Acetyltryptophan	1	AA-pep	HMDB13713	neg	1.2							3
Deoxyhexose	1	Carboh	HMDB00849	neg	0.94							2
Pentose	2	Carboh		neg	0.73				-0.15			2
Dehydroepiandrosterone sulfate	1	Steroi	HMDB01032	neg	-7	*			2.6			1
Testosterone glucuronide	2	Steroi	HMDB03193	neg	-3.1				1.5	*		1
Dehydroepiandrosterone 3-glucuronide	1	Steroi	HMDB10348	neg	-2.9				0.5			1
FMNH2	2	AroHeP	HMDB01142	neg	-2.6	*						1
N4-Acetylcytidine	1	Nucleo	HMDB05923	neg	-2.2	*						1
6-(carboxymethoxy)-hexanoic acid	2	Organi		neg	-1.8				0.26			1
Tryptamine	1	AroHeP	HMDB00303	pos	-1.8	*						1
4-Hydroxybenzoic acid	1	AroHoM	HMDB00500	neg	-1.4		-1					3
Methylinosine	2	Nucleo		neg	-1.3				*	*		1
Indoleacetyl glutamine	2	AA-pep	HMDB13240	pos	-1.2	*						1
Decanoylcarnitine isomer	2	AcyCar		pos	-1.1				1			1
Decenoylcarnitine isomer 2	2	AcyCar		pos	-1.1		1.6		1.3			1
Tryptophan	1	AA-pep	HMDB00929	neg	-0.91							3
Dimethylguanosine	2	Nucleo	HMDB04824	neg	-0.91				*	*		1
Hydroxysuberic acid isomer 2	2	Organi		neg	-0.87							3
Aspartic acid	1	AA-pep	HMDB00191	neg	-0.84		-0.88		*			3
Pyridylacetylglycine	2	Xenobi	HMDB59723	neg	-0.8	*						1
N-Acetyl-aspartic acid	1	AA-pep	HMDB00812	neg	-0.75		-0.73		-0.26			3
Kynurenic acid	1	AA-pep	HMDB00715	neg	-0.63							1
Heptylmalonic acid	2	Lipids	HMDB59719	neg	-0.62							3
4-Acetamidobutanoic acid isomer 2	2	Lipids		neg	-0.57	*						1
Propionylcarnitine isomer 1	2	AcyCar		pos			2.9		1.3			1
Dimethylnonenoylcarnitine/Undecenoylcarnitine	2	AcyCar		pos			2.1		1.9			1
Octanoylcarnitine isomer	2	AcyCar		pos			1.2		0.76			1
Decadienylcarnitine isomer 2	2	AcyCar		pos			1.2		0.72			1
2-Octenoylcarnitine	2	Lipids	HMDB13324	pos			1.1		0.35			1
Heptenoylcarnitine isomer	2	AcyCar		pos			0.97		0.41			1
Valerylcarnitine	2	AcyCar	HMDB13128	pos			0.92		0.45			1
p-Anisic acid	1	AroHoM	HMDB01101	neg			-4.1		-3.4			3
Mevalonic acid isomer 1	2	Organi		pos			-2.5		-1.2			3
4-Acetamidobutanoic acid isomer 3	2	Lipids		pos			-2		-0.67			3
				neg	***		-1.7		-0.59			3
N-Acetyltryptophan isomer 3	2	AA-pep		pos			-1.6					3
				neg	-1.2		-1.6	*				3
Cinnamoylglycine	1	AcyGly	HMDB11621	neg			-1.3		-0.55			3
3-Hydroxyphenylacetic acid	1	AroHoM	HMDB00440	neg			-0.99					3
Phe-Tyr-Asp (and isomers)	2	AA-pep		neg			-0.96		-0.25			3
Tiglylglycine	1	AcyGly	HMDB00959	pos		*	-0.88		-0.46			3
3-Methylcrotonylglycine	1	AcyGly	HMDB00459	pos		*	-0.86		-0.45			3
				neg	-0.85	*	-1.2		-0.23			3

Table 1. continued

Molecule	Annot. level	Chemical class	HMDB	Mode	Age (log2 ratio)	Age: BMI (sig.)	BMI (log2 ratio)	BMI: Gender (sig.)	Gender (log2 ratio)	Age: Gender (sig.)	Age: BMI: Gender (sig.)	Cluster group
Valerylglycine isomer 2	2	AcyGly		pos		*	-0.71	*	-0.22			3
							-0.76	*	-0.24			3
alpha-N-Phenylacetyl-glutamine	1	AA-pep	HMDB06344	pos			-0.61	*	-0.36			3
							-0.83		-0.36			3
2-Aminoadipic acid	1	AA-pep	HMDB00510	neg			-0.61		-0.26			3
Hydroxyoctanoylcarnitine	2	AcyCar		pos					1.3			1
Decenoylcarnitine isomer 1	2	AcyCar		pos					1.2			1
Hypoxanthine	1	AroHeP	HMDB00157	pos					0.77			3
(gamma)Glu-Leu/Ile	2	AA-pep		pos					0.66			1
							-0.93		0.42			1
Nonenoylcarnitine isomer 1	2	AcyCar		pos					0.58			1
Niacinamide	2	AroHeM	HMDB01406	pos					0.52			1
2,6 Dimethylheptanoyl carnitine	2	AcyCar	HMDB06320	pos					0.51			1
Asp-Leu/Ile isomer 1	2	AA-pep		neg					0.43			1
Proline	1	AA-pep		pos					0.4			1
3-Methyl-2-oxovaleric acid	1	Lipids	HMDB00491	neg					0.37			3
Hexanoylcarnitine isomer	2	AcyCar		pos					0.35			1
Glu-Val	2	AA-pep	HMDB59717	neg					0.33			1
Asp-Leu/Ile isomer 2	2	AA-pep		neg					0.32			1
Pyridoxamine isomer 1	2	AroHeM		pos					0.27			2
Acetaminophen glucuronide	1	Carboh	HMDB10316	neg					-5.1			3
Creatine	1	AA-pep	HMDB00064	pos					-1.5			3
Mevalonic acid isomer 2	2	Organic		pos					-1.3			3
Monoethyl phthalate	2	Xenobi	HMDB02120	neg					-1.1			3
Oxoglutaric acid	2	Organic	HMDB00208	neg					-0.84	***		3
Methylglutaryl carnitine isomer 2	2	AcyCar		pos					-0.77			3
Malic acid	1	Organic	HMDB00156	neg					-0.69	*		3
Pantothenic acid	1	AliAcy	HMDB00210	pos					-0.65			3
							-0.79	-0.98	-0.55			3
2-Methylhippuric acid	1	AcyGly	HMDB11723	pos					-0.64			3
							neg		-0.61			3
p-Hydroxyhippuric acid	2	AcyGly	HMDB13678	pos					-0.62			3
							neg	-0.66	-0.58			3
3,5-dihydroxybenzoic acid/3,4-dihydroxybenzoic acid	2	AroHoM		neg					-0.6			2
Gluconic acid and/or isomers	2	Carboh		neg					-0.53			3
3-Hydroxybenzyl alcohol	1	AroHoM	HMDB59712	neg					-0.52			3
5,6-Dihydrouracil	1	AliHeM	HMDB00076	pos					-0.51			3
Valerylglycine isomer 1	2	AcyGly		pos					-0.49			3
							neg		-0.26			3
Glucuronic acid and/or isomers	2	Carboh		neg					-0.48			3
Trimethylamine N-oxide	1	AliAcy	HMDB00925	pos					-0.47			3
3,4-Dihydroxybenzenoic acid	1	AroHoM	HMDB01336	neg					-0.4			3
Hydroxybenzyl alcohol isomer	2	AroHeM		neg					-0.38			2
Citric acid	1	Organic	HMDB00094	neg					-0.38			3
Hippuric acid	1	AcyGly	HMDB00714	pos					-0.36			3
N2-Acetylaminoadipic acid	2	AA-pep		neg					-0.35			3
							neg		-0.36			1
5-Hydroxyindoleacetic acid	1	AroHeP	HMDB00763	pos					-0.34	*		3
							neg	-1.1	-0.22	*		3
Threonic acid/Erythronic acid	2	Carboh		neg					-0.32	*		3
Pyruvic acid	1	Organic	HMDB00243	neg					-0.28	**		3
Pyrocatechol sulfate	2	AroHoM	HMDB59724	neg					-0.28			2
Glyceric acid	1	Carboh	HMDB00139	neg					*	-0.27		3
Nicotinuric acid isomer	2	AA-pep		neg					-0.21			3

^aMetabolites were sorted by decreasing log2 ratios of the predicted intensities by a linear model at the maximal compared to the minimal value of the factor (for age and BMI) and of the male over female medians (for gender). The log2 ratios are given in green/red (resp. light blue/pink for gender) when positive/negative. Positive and negative ionization modes of the same molecule were grouped. The annotation level according to MSI criteria, chemical class, HMDB ID (when available), and ionization mode (*pos* or *neg*) is indicated. Interactions between the factors are displayed only when at least one of the interacting factors is significant. The interaction is color-encoded according to the (non-adjusted) *p*-value: ***, red, *p* < 0.001, **, yellow, *p* < 0.01; *, green, *p* < 0.05. The partition of the metabolites by hierarchical clustering into three groups according to their profile among individuals is indicated (see Figure 6). Abbreviations: (AA-pep) Amino Acids, Peptides, and Analogues; (AcyCar) Acyl Carnitines; (AcyGly) Acyl Glycines; (AliAcy) Aliphatic Acyclic Compounds; (AliHeM) Aliphatic Heteromonocyclic Compounds; (AliHoM) Aliphatic Homomonocyclic Compounds; (AroHoM) Aromatic Heteromonocyclic Compounds; (AroHeP) Aromatic Heteropolycyclic Compounds; (AroHoM) Aromatic Homomonocyclic Compounds; (Carboh) Carbohydrates and Carbohydrate Conjugates; (Lipids) Lipids; (Nucleo) Nucleosides, Nucleotides, and Analogue; (Organic) Organic Acids and Derivatives; (Steroi) Steroids and Steroid Derivatives; (Xenobi) Xenobiotics.

Chemical and Biological Characterization of Selected Metabolites

A total of 108 metabolites out of 170 were shown to significantly vary with at least one of the physiological factors studied: 52 with

age, 27 with BMI, and 74 between genders (Figure 3D and Table 1). Annotations provided by our previous study¹⁶ indicated that 41% of these metabolites were identified at level 1 according to the metabolomics standard initiative (MSI),⁶² whereas the others

remained putatively annotated (MSI level 2). Amplitudes of physiological variations, as measured at the extremities of the regression interpolation (Table 1), were higher for age (129 and 467 ratios, for the steepest decrease and increase, respectively), than for BMI (17 and 8) and gender (33 and 6). For one-third of the metabolites, variations occurred with two or three of the factors (Figure 3D). Five metabolites exhibited physiological variations with all three factors: 2-acetamido-4-methylphenyl acetate, decenoylcarnitine isomer 2 (Figure 4G), N-acetylaspartic acid, 3-methylcrotonylglycine, and pantothenic acid. To further investigate interactions between factors, we performed analysis of covariance. Although *p*-values after correction for multiple testing were all above the FDR = 0.05 threshold, significant values were obtained before correction and are reported in Table 1 as an indication of putative interaction. The chemical classes of the metabolites were also detailed, by using the “super class” taxonomy of the UMDB human urinary metabolome database,¹⁷ supplemented with the acylcarnitine, acylglycine, steroid and steroid derivative, and xenobiotic categories (Table 1). Amino acids, peptides, and analogues were overrepresented in our selection by a 3.4 factor compared to the HMDB content, whereas aromatic acids were underrepresented by an average 1.7 factor (Figure 4A). For each physiological response, the chemical distribution of significant metabolites is summarized in Figure 4A. Finally, clustering of significant metabolites was investigated by analyzing the correlation networks of their profiles among individuals (Figure 5).

Metabolites with age-correlated concentrations could be divided into two clusters (Figure 5): an age-increasing concentration group contained caffeine (Figure 4B) and its metabolites (including paraxanthine/theophylline, 1-methylxanthine, 1,3-dimethyluric acid, and 1,7-dimethyluric acid), in addition to aromatic acids (e.g., indoleacetic acid isomer, Figure 4C) and carbohydrates (deoxyhexose, pentose). Some of these metabolites had higher concentrations in females (e.g., caffeine). The second cluster, consisting of metabolites with negative age-correlations, comprised steroids (testosterone glucuronide, dehydroepiandrosterone sulfate) and some acyl carnitines (with higher concentrations in males) but also nucleosides (N4-acetylcytidine, methylinosine, dimethylguanosine) and a few amino acids and peptides (aspartic acid; Figure 4E).

Apart from 3-hydroxyphenylacetic acid, metabolites with BMI-correlated concentrations were also significant for age-correlation or difference between genders (Table 1 and Figure 4E–G). Except for decenoylcarnitine isomer 2 (Figure 4G), correlations with age and BMI were of the same sign. A group of 8 correlated acyl carnitines showed increased concentrations with BMI (Figure 4G). Conversely, BMI-decreasing metabolites included 4-acetamidobutanoic acid isomer 3 (Figure 4F) and a group of 5 acyl glycines (Table 1).

Metabolites with difference in median concentrations between genders included lipids and amino-acids/peptides (which were overrepresented compared to the HMDB content; Figure 4A) and could be split into two clusters (Figure 5). First, metabolites with higher concentrations in males included the three steroids previously shown to be associated with age, in addition to a large group of 15 acyl carnitines and several dipeptides (Figure 5 and Table 1). Metabolites with higher concentrations in females included some metabolites previously shown to increase with age, such as caffeine (Figure 4B), and derivatives of amino acids (*N*-acetylphenylalanine) but also organic acids with gender variations only (malic acid, Figure 4H, oxoglutaric acid; Table 1).

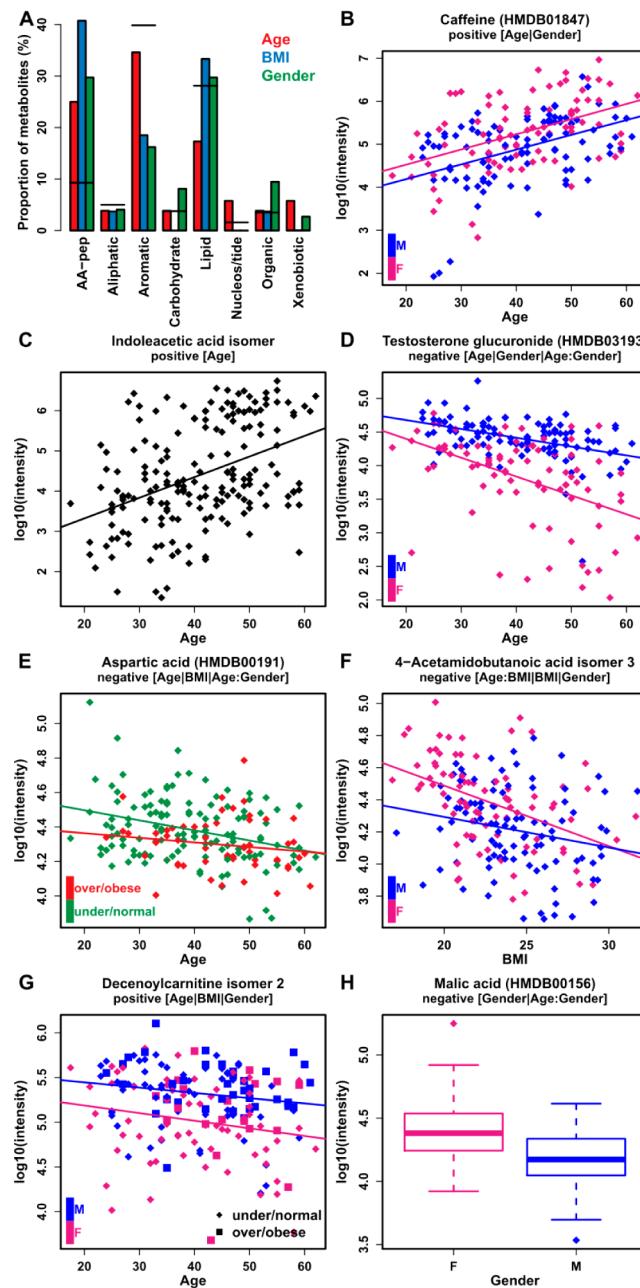


Figure 4. Chemical classes and signal intensity variations of the selected metabolites with age, BMI, or gender. (A) The distribution of the chemical classes of the selected metabolites is plotted as bars of the corresponding factor color. The horizontal segments indicate the corresponding values for the whole HMDB database. (B–H) Intensities among individuals for a selection of representative metabolites. The type of graphic depends on the significant factor(s) and interaction(s) between factors. The scatter plot of the metabolite significant with age only is in black (C). When age and BMI are both significant (E), BMI values are colored in green (respectively red) if they are below (respectively above) 25. In the case of three significant factors (G), BMI values are displayed as diamonds (respectively squares) if they are below (respectively above) 25. Significant interactions result in crossing regression lines. Graphics for all the significant metabolites listed in Table 1 are available in Supporting Information file 3.

Second, a group of 8 acyl-glycines had higher concentrations in females (hippuric acid, 2-methylhippuric acid; Figure 5), some of them also decreasing with BMI (Table 1).

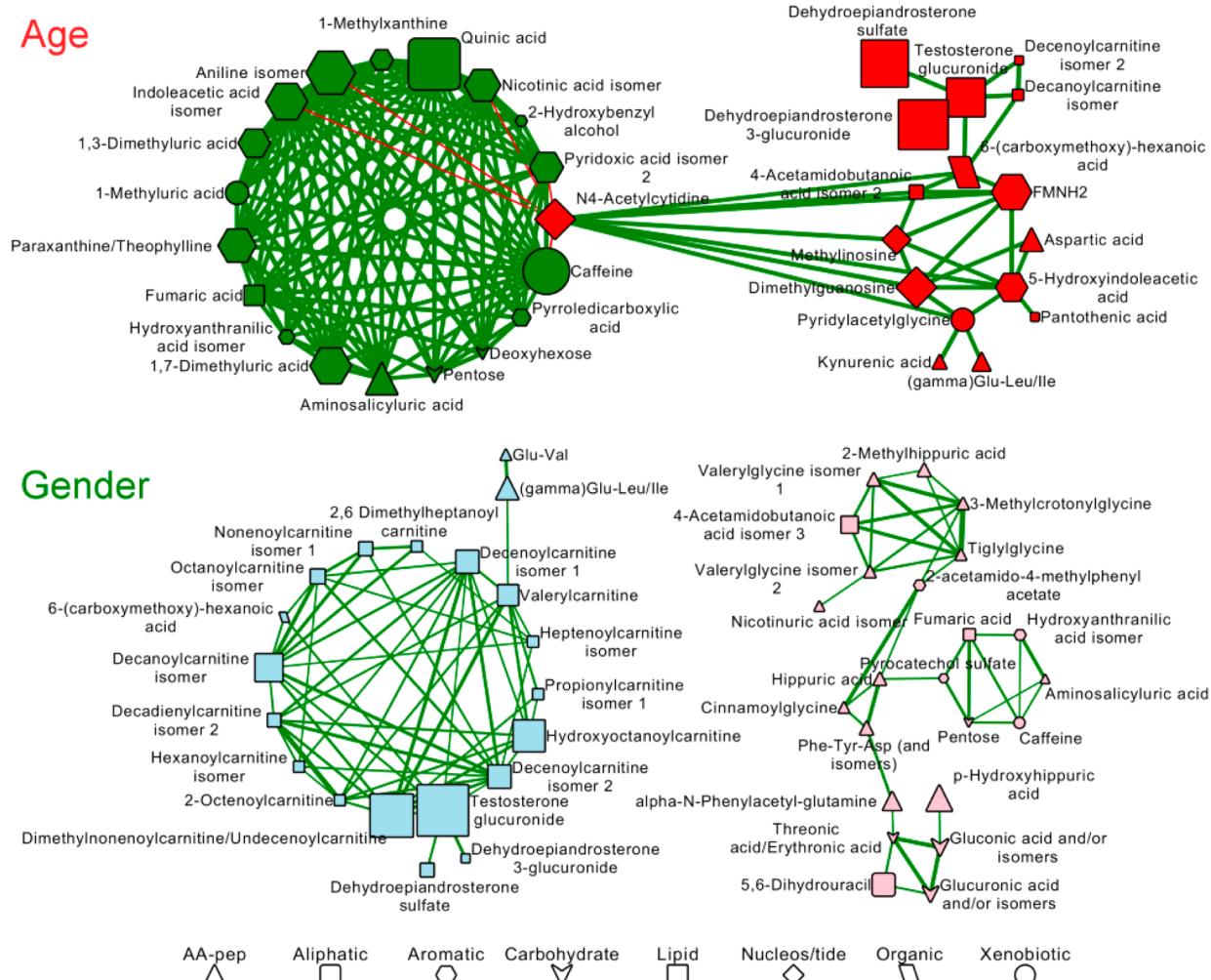


Figure 5. Correlation networks of the metabolites with significant variations with age and gender. Edges represent correlations with absolute value >0.5 . Groups of 3 metabolites or less are not shown. The green (respectively red) edge color corresponds to a positive (respectively negative) correlation. The edge width is proportional to the value of this correlation (scale ranges from 1 to 3). Node size is proportional to $-\log_{10}(p\text{FDR})$, where pFDR is the p -value of the metabolite after correction for multiple tests (scale ranges from 1 to 5). Node shape corresponds to the chemical class of the metabolite, as defined in Table 1. Color conventions are identical to those in Table 1: For the age response, positive (respectively negative) correlation of the metabolite with the factor results in a green (respectively red) node color; for the gender factor, the light blue (respectively pink) node color indicates a higher concentration in males (respectively females).

Stratification of the Cohort

Finally, we investigated the clustering of the cohort individuals according to their metabolite fingerprints. The *pos* and *neg* data sets were concatenated after standardization of the variables, and hierarchical clustering of the samples (and metabolites) was computed (Figure 6A). Seven (respectively 3) groups for individuals (respectively metabolites) named A–G (respectively 1–3) could be distinguished. Metabolite cluster 1 mainly comprised age-concentration decreasing metabolites such as nucleosides, and metabolites with higher concentrations in males (some of them also increasing with BMI) such as acyl-carnitines and steroid hormones. Cluster 2 included some of the age-concentration increasing metabolites with some of them having higher concentration in females described previously (Table 1 and Figure 5): metabolites originating from food (caffeine and its metabolites, quinic acid), some aromatic compounds, and carbohydrates. Metabolites from cluster 3 had concentrations either increasing with age (e.g., some amino acid derivatives such as their acetylated forms) or decreasing with BMI (e.g., acyl-glycines), and they were generally more concentrated in females.

Compared to group A, group D profiles displayed lower concentrations of cluster 2 metabolites and a moderate increase of concentrations of cluster 1. As suggested by the above remarks, group D individuals were thus expected to be younger than group A, which was indeed the case when checking the physiological covariate of these individuals (Figure 6B). Similarly, further increase of cluster 1 concentrations from group D to group G was expected to reflect a marked increase of the proportion of males in the latter group, which was again confirmed by the metadata.

DISCUSSION

Urine metabolome variations with age, BMI, and gender were studied by untargeted LC-HRMS analysis of samples from a large adult cohort (183 volunteers, 170 identified metabolites). This is the first time that such a large-scale study is conducted in human urine.

Urine profiles were first normalized to the osmolality of the sample since (i) interindividual variations of urine concentration were expected and (ii) diuresis was not a factor of interest in this physiological study. Then, to select the metabolites with

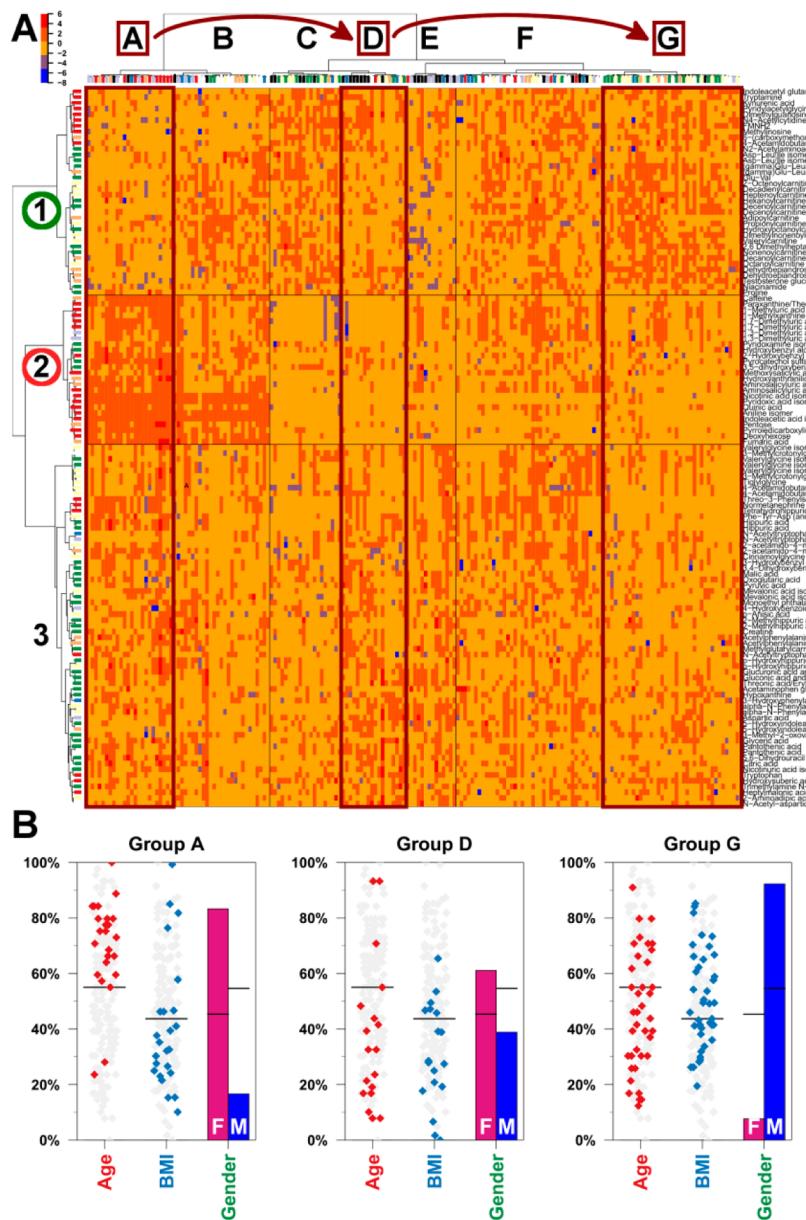


Figure 6. Cohort profiling. (A) Clustering of metabolites (rows) and individual urine samples (columns). Intensities are color encoded from -7.4 (purple) to 4.6 (red) in the logarithmic scale. Borders resulting from the cutting of the metabolite (respectively sample) dendograms into $k = 3$ (respectively $k = 7$) groups are drawn as black lines. The metabolite groups are reported in Table 1. For each metabolite, the significance of the physiological variation from Table 1 is reported on the left vertical side bar with the same color encoding as in Figure 3D (for instance, yellow indicates a significant correlation with BMI and a gender difference between medians). The top horizontal side bar reflects the actual age, BMI, and gender of the individual as a $2 \times 2 \times 2$ level category (below/above median for age and BMI, and female/male for gender) with the same color code as Figure 3D (in addition to black encoding the $0 \times 0 \times 0$ level); hence, individuals with a black label are females below age and BMI medians, whereas green labels encode males below age and BMI medians. (B) Clinical metadata (age, BMI, gender) of individuals from groups A, D, and G. The cohort median (or proportion for gender) is indicated with a horizontal segment.

physiological variations, we used both univariate hypothesis testing and multivariate modeling by orthogonal partial least-squares (OPLS).³⁹ OPLS is a modification of the partial least-squares algorithm which facilitates model interpretation by allowing separate analysis of the variations correlated or orthogonal to the factor of interest. Univariate and PLS (or OPLS) approaches are becoming popular in biomarker studies: In a recent review covering LC-MS applications in urine from the past two years, Zhang and Watson reported that both methods were used in the majority of the articles.³ It has been suggested theoretically and by a few simulation studies that univariate and

multivariate approaches should be complementary for feature selection.^{64,65} To allow direct comparison and integration of hypothesis testing and OPLS results, we implemented the full functionalities of OPLS in an R package named *ropls*. We further integrated this module into the Workflow4Metabolomics.org online resource for computational metabolomics⁶⁰ (W4M), which provides a user-friendly, web-based environment for data preprocessing, statistical analysis, and annotation.

Comparison of the metabolites selected with the univariate significance threshold ($FDR < 5\%$) and the multivariate $VIP > 1$ rule revealed a strong agreement, ranging from 81% to 96%. In

fact, we could demonstrate, in this specific case of a single predictive component OPLS model with standardized variables, a mathematical relationship between the *p*-values from the Pearson correlation test and the VIPs. This is the first time that univariate *p*-values and multivariate VIPs are compared at both the graphical and theoretical levels. We integrated our univariate and OPLS results by selecting all metabolites with FDR < 5%, and up to 7% in the case of VIP > 1. The latter rule resulted in the addition of four metabolites, including a new one (3-hydroxyphenylacetic acid) with a clear 2-fold decrease in concentrations with BMI. To fully exploit the multivariate structure of the data sets, PLS modeling with more than one component will be of interest in future studies. Alternatively, metrics others than VIP (e.g., the regression coefficients) could be used (see Mehmmood et al, 2012, for a review).⁴¹ In addition, to facilitate the comparison between feature selection methods, the statistical significance of the chosen metrics could be further evaluated, as in the recently described permutation approach.⁶⁶

Application of our statistical workflow to the urine sample profiles resulted in the identification of 108 metabolites whose concentrations varied with either age, BMI, or gender of the individuals. As explained below, the metabolite concentration trends already found in previous studies validate our findings, whereas many concentration variations are reported for the first time. We observed three clusters of metabolites which could be used to stratify the cohort.

The first cluster contained metabolites with concentrations decreasing with age, or increasing with BMI, and higher in males. It included male steroid hormones and their metabolites (testosterone glucuronide, dehydroepiandrosterone sulfate, and 3-glucuronide), nucleosides, and a large group of C0–C10 acylcarnitines. Higher concentrations of testosterone in men and decline with age are well-known, as is the case with DHEA sulfate.⁶⁷ Higher concentrations of acylcarnitines in male,⁶⁸ and positive correlation with age (except for decanoylcarnitine)⁶⁹ and BMI⁷⁰ have been reported in blood and serum recently. Concentrations of acylcarnitines were also increased in urines of male volunteers submitted to moderate weight gain using a lipid-enriched overfeeding protocol.⁷¹

The second cluster, with concentrations increasing with age and higher in females, comprised xenobiotics (caffeine and its metabolites,⁷² quinic acid), fumaric acid, acetylphenylalanine, carbohydrates (deoxyhexose, pentose), and aromatic compounds. The higher concentrations of caffeine and its metabolites suggest an increased consumption and/or a decreased metabolic activity with age, and in females. Such a trend has previously been described in plasma metabolomic studies.^{67,73}

The third and largest cluster contained a majority of metabolites with concentrations higher in females, including age-increasing or BMI-decreasing variations. Amino acid derivatives such as acetylated amino acids were abundant, in addition to a group of acyl-glycines, oxoglutaric acid, citric acid, and acetaminophen glucuronide. Higher concentrations of citric acid in females have been described in serum both in the rat⁴ and in human.⁷ However, a recent LC-MS/MS study of acyl-glycines quantification did not report any correlation with BMI or gender.⁷⁴ The small number of volunteers (20) in the study by Stanislaus et al. may have been too low to evidence such variations.⁷⁴

A single metabolite was found with significant variations for BMI only (3-hydroxyphenylacetic acid). This compound is a phenolic acid known to be a rutin metabolite, and its relationship with BMI remains to be elucidated. The small number of

individuals with extreme BMI values in our cohort probably limits the number of significant variations (and the predictive ability of the multivariate models). A further study is therefore required to refine the characterization of metabolite variations with BMI and compare the results with those published recently in blood.⁷⁰ Moreover, volunteers of our study were enrolled without exclusion criteria. As a result, the cohort probably contains a small proportion of individuals treated for diseases. We therefore used a large number of samples, as well as nonparametric tests, to minimize the risk of false positive variations due to pathologies. It will be of interest to replicate these results in a healthy cohort, where additional metadata (such as the smoking status)⁷³ will be recorded.

Descriptions of physiological metabolite variations in body fluids are scarce in the literature, and analyses are based on targeted technologies, or on small numbers of volunteers. Only in blood/serum have five large-scale studies very recently described the correlation with age,^{67,69,73} BMI,^{70,73} and gender.^{67,68,73}

In some of these studies, around 30% of the detected metabolites were associated with age, which was the most predominant factor among gender, BMI, or ethnicity,^{67,75} whereas 12% of the detected metabolites were associated with BMI in a metabolomics study performed on 947 subjects.⁷⁰ Here, we found that 45 out of the 85 (53%) MSI level 1 metabolites had significant variations with at least one of the factors studied. Since all MSI level 1 metabolites had been identified with our generic in-house spectral database without any indication regarding their physiological variations, our results suggest that the urine medium could be more impacted by physiological factors than plasma.

Interestingly, several urine biomarkers described recently were found to have physiological variations in our study: for instance, hippuric acid,^{32,33,76} kynurenic acid,^{30,33,76} citric acid,^{31,33,77} gluconic acid,^{30,31} pyruvic acid,^{31,33} or α -N-phenylacetylethylglutamine.^{30,77} Although most of these studies used age, BMI, and gender matching cases and controls to avoid biases, our description of physiological variations should further help avoid false positives in future biomarker discovery studies.

In conclusion, we conducted the first comprehensive characterization of urine metabolome physiological variations, and we identified 108 metabolites with concentration variations according to either age, BMI, or gender. Our analytical and statistical framework, including the associated online software tools, therefore provides a unique contribution to the human urinary metabolome annotation and to the robust and high-throughput selection of metabolite biomarkers.

ASSOCIATED CONTENT

S Supporting Information

Text document containing the supplementary figures (Figure S1, data description; Figure S2, signal drift correction and batch effect removal; Figure S3, OPLS models with data obtained in the negative ionization mode; Figure S4, significance of R2Y and Q2Y values estimated by permutation testing; Figure S5, *p*-values vs VIP comparison for the data sets from the negative ionization mode; Figure S6, relationship between VIP from one predictive PLS or OPLS model with standardized variables, and *p*-values from the Pearson correlation test) and the theory section comparing univariate and multivariate selection in the case of single-response, single-predictive PLS models with standardized variables. Spreadsheet with the preprocessed peak tables from

the positive and negative ionization modes, in addition to the sample and variable metadata. Graphics of all the significant metabolites listed in Table 1. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00354.

AUTHOR INFORMATION

Corresponding Authors

*Etienne A. Thévenot: Phone: 33-1-69-08-79-23. Fax: 33-1-69-08-89-65. E-mail: etienne.thevenot@cea.fr.

*Christophe Junot: Phone: 33-1-69-08-43-66. Fax: 33-1-69-08-59-07. E-mail: christophe.junot@cea.fr.

Author Contributions

†E.A.T. and A.R. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Michel Tournier (Occupational Medicine Department of CEA-Saclay) and Xavier Millot and Catherine Bourcier and their staff (Clinical Chemistry laboratory of CEA-Saclay) for providing us with urine samples. We thank Anne-Marie Werner for helpful discussions regarding the manuscript. This work was supported by the French Ministry of Research and National Research Agency as part of the French metabolomics and fluxomics infrastructure (MetaboHUB, ANR-11-INBS-0010 grant).

REFERENCES

- (1) Armstrong, J. A. Urinalysis in Western culture: A brief history. *Kidney Int.* **2006**, *71* (5), 384–387.
- (2) Ryan, D.; Robards, K.; Prenzler, P.; Kendall, M. Recent and potential developments in the analysis of urine: A review. *Anal. Chim. Acta* **2011**, *684* (1–2), 17–29.
- (3) Zhang, T.; Watson, D. G. A short review of applications of liquid chromatography mass spectrometry based metabolomics techniques to the analysis of human urine. *Analyst* **2015**, *140* (9), 2907–2915.
- (4) Bolland, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **2005**, *18* (3), 143–162.
- (5) Slupsky, C. M.; Rankin, K. N.; Wagner, J.; Fu, H.; Chang, D.; Weljie, A. M.; Saude, E. J.; Lix, B.; Adamko, D. J.; Shah, S.; Greiner, R.; Sykes, B. D.; Marrie, T. J. Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles. *Anal. Chem.* **2007**, *79* (18), 6995–7004.
- (6) Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E. High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *Int. J. Epidemiol.* **2008**, *37* (suppl 1), i31–i40.
- (7) Rasmussen, L.; Savorani, F.; Larsen, T.; Dragsted, L.; Astrup, A.; Engelsen, S. Standardization of factors that influence human urine metabolomics. *Metabolomics* **2011**, *7* (1), 71–83.
- (8) Dunn, W. B.; Broadhurst, D.; Ellis, D. I.; Brown, M.; Halsall, A.; O'Hagan, S.; Spasic, I.; Tseng, A.; Kell, D. B. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. *Int. J. Epidemiol.* **2008**, *37* (suppl 1), i23–i30.
- (9) Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y. Development and validation of a gas chromatography/mass spectrometry metabonomic platform for the global profiling of urinary metabolites. *Rapid Commun. Mass Spectrom.* **2008**, *22* (19), 2984–2992.
- (10) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst* **2006**, *131* (10), 1075–1078.
- (11) Gika, H. G.; Theodoridis, G. A.; Wingate, J. E.; Wilson, I. D. Within-Day Reproducibility of an HPLC-MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J. Proteome Res.* **2007**, *6* (8), 3291–3303.
- (12) Gika, H. G.; Theodoridis, G. A.; Wilson, I. D. Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine: Sample stability under different handling and storage conditions for metabolomics studies. *J. Chromatogr., A* **2008**, *1189* (1–2), 314–322.
- (13) Warrack, B. M.; Hnatyshyn, S.; Ott, K.-H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. Normalization strategies for metabonomic analysis of urine samples. *J. Chromatogr. B* **2009**, *877* (5–6), 547–552.
- (14) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. *Anal. Chem.* **2011**, *83* (15), 5864–5872.
- (15) Benton, H. P.; Want, E.; Keun, H. C.; Amberg, A.; Plumb, R. S.; Goldfain-Blanc, F.; Walther, B.; Reily, M. D.; Lindon, J. C.; Holmes, E.; Nicholson, J. K.; Ebbels, T. M. D. Intra- and Interlaboratory Reproducibility of Ultra Performance Liquid Chromatography Time-of-Flight Mass Spectrometry for Urinary Metabolic Profiling. *Anal. Chem.* **2012**, *84* (5), 2424–2432.
- (16) Roux, A.; Xu, Y.; Heilier, J.-F.; Olivier, M.-F.; Ezan, E.; Tabet, J.-C.; Junot, C. Annotation of the Human Adult Urinary Metabolome and Metabolite Identification Using Ultra High Performance Liquid Chromatography Coupled to a Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. *Anal. Chem.* **2012**, *84* (15), 6429–6437.
- (17) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. The Human Urine Metabolome. *PLoS One* **2013**, *8* (9), e73076.
- (18) Roux, A.; Thévenot, E.; Seguin, F.; Olivier, M.-F.; Junot, C. Impact of collection conditions on the metabolite content of human urine samples as analyzed by liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy. *Metabolomics* **2015**, DOI: 10.1007/s11306-014-0764-5.
- (19) Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2* (11), 2692–2703.
- (20) Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **2011**, *6* (10), 1483–1499.
- (21) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **2010**, *5* (6), 1005–1018.
- (22) Jacob, C.; Dervilly-Pinel, G.; Biancotto, G.; Le Bizec, B. Evaluation of specific gravity as normalization strategy for cattle urinary metabolome analysis. *Metabolomics* **2014**, *10* (4), 627–637.
- (23) Edmonds, W. M. B.; Ferrari, P.; Scalbert, A. Normalization to Specific Gravity Prior to Analysis Improves Information Recovery from High Resolution Mass Spectrometry Metabolomic Profiles of Human Urine. *Anal. Chem.* **2014**, *86* (21), 10925–10931.
- (24) Chen, Y.; Shen, G.; Zhang, R.; He, J.; Zhang, Y.; Xu, J.; Yang, W.; Chen, X.; Song, Y.; Abliz, Z. Combination of Injection Volume Calibration by Creatinine and MS Signals' Normalization to Overcome Urine Variability in LC-MS-Based Metabolomics Studies. *Anal. Chem.* **2013**, *85* (16), 7659–7665.
- (25) Wu, Y.; Li, L. Determination of Total Concentration of Chemically Labeled Metabolites as a Means of Metabolome Sample Normalization and Sample Loading Optimization in Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2012**, *84* (24), 10723–10731.

- (26) Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G. Evaluation of Coupling Reversed Phase, Aqueous Normal Phase, and Hydrophilic Interaction Liquid Chromatography with Orbitrap Mass Spectrometry for Metabolomic Studies of Human Urine. *Anal. Chem.* **2012**, *84* (4), 1994–2001.
- (27) Kim, K.; Taylor, S.; Ganti, S.; Guo, L.; Osier, M.; Weiss, R. Urine metabolomic analysis identifies potential biomarkers and pathogenic pathways in kidney cancer. *OMICS* **2011**, *15* (5), 293–303.
- (28) Jager, K. J.; Zoccali, C.; MacLeod, A.; Dekker, F. W. Confounding: What it is and how to deal with it. *Kidney Int.* **2007**, *73* (3), 256–260.
- (29) Broadhurst, D.; Kell, D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006**, *2* (4), 171–196.
- (30) Wang, X.; Zhang, A.; Han, Y.; Wang, P.; Sun, H.; Song, G.; Dong, T.; Yuan, Y.; Yuan, X.; Zhang, M.; Xie, N.; Zhang, H.; Dong, H.; Dong, W. Urine Metabolomics Analysis for Biomarker Discovery and Detection of Jaundice Syndrome in Patients With Liver Disease. *Mol. Cell. Proteomics* **2012**, *11* (8), 370–380.
- (31) Chen, T.; Xie, G.; Wang, X.; Fan, J.; Qiu, Y.; Zheng, X.; Qi, X.; Cao, Y.; Su, M.; Wang, X.; Xu, L. X.; Yen, Y.; Liu, P.; Jia, W. Serum and Urine Metabolite Profiling Reveals Potential Biomarkers of Human Hepatocellular Carcinoma. *Mol. Cell. Proteomics* **2011**, DOI: 10.1074/mcp.M110.004945.
- (32) Huang, Z.; Lin, L.; Gao, Y.; Chen, Y.; Yan, X.; Xing, J.; Hang, W. Bladder Cancer Determination Via Two Urinary Metabolites: A Biomarker Pattern Approach. *Mol. Cell. Proteomics* **2011**, DOI: 10.1074/mcp.M111.007922.
- (33) Cheng, Y.; Xie, G.; Chen, T.; Qiu, Y.; Zou, X.; Zheng, M.; Tan, B.; Feng, B.; Dong, T.; He, P.; Zhao, L.; Zhao, A.; Xu, L. X.; Zhang, Y.; Jia, W. Distinct Urinary Metabolic Profile of Human Colorectal Cancer. *J. Proteome Res.* **2012**, *11* (2), 1354–1363.
- (34) van Belle, G.; Fisher, L.; Heagerty, P.; Lumley, T. *Biostatistics—a methodology for the health sciences*; John Wiley & Sons: Hoboken, NJ, 2004.
- (35) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach for multiple testing. *J. R. Stat. Soc. B Met.* **1995**, *57* (1), 289–300.
- (36) Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* **2012**, *2* (4), 775–795.
- (37) Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **2007**, *6* (2), 469–479.
- (38) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130.
- (39) Trygg, J.; Wold, S. Orthogonal projection to latent structures (O-PLS). *J. Chemom.* **2002**, *16* (3), 119–128.
- (40) Pinto, R. C.; Trygg, J.; Gottfries, J. Advantages of orthogonal inspection in chemometrics. *J. Chemom.* **2012**, *26* (6), 231–235.
- (41) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbo, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69.
- (42) Eriksson, I.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and megavariate data analysis. Principles and applications*; Umetrics: Umeå, Sweden, 2001.
- (43) Bylesjo, M.; Rantalainen, M.; Nicholson, J.; Holmes, E.; Trygg, J. K-OPLS package: Kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space. *BMC Bioinformatics* **2008**, *9* (1), 106.
- (44) R Core Development Team (2013). *R: A language and environment for statistical computing*; Foundation for statistical computing: Vienna, Austria, <http://www.r-project.org>.
- (45) Gaude, R.; Chignola, F.; Spiliotopoulos, D.; Spitaleri, A.; Ghitti, M.; Garcia-Manteiga, J.; Mari, S.; Musco, G. muma: An R package for metabolomics univariate and multivariate statistical analysis. *Curr. Metabolomics* **2013**, *1*, 180–189.
- (46) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787.
- (47) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41* (D1), D801–D807.
- (48) Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40* (D1), D109–D114.
- (49) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat. Biotechnol.* **2012**, *30* (9), 826–828.
- (50) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060–1083.
- (51) van der Kloet, F. M.; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *J. Proteome Res.* **2009**, *8* (11), 5132–5141.
- (52) Cleveland, W.; Grosse, E.; Shyu, W. In *Local regression models. In Statistical models*; Chambers, J. M., Hastie, T. J., Eds.; Chapman & Hall: London, U.K., 1997; pp 309–376.
- (53) Alonso, A.; Julia, A.; Beltran, A.; Vinaixa, M.; Diaz, M.; Ibanez, L.; Correig, X.; Marsal, S. AStream: an R package for annotating LC/MS metabolomic data. *Bioinformatics* **2011**, *27* (9), 1339–1340.
- (54) Mason, R.; Tracy, N.; Young, J. A practical approach for interpreting multivariate T2 control chart signals. *J. Qual. Technol.* **1997**, *29* (4), 396–406.
- (55) Tenenhaus, M. *La régression PLS: théorie et pratique*; Technip: Paris, France, 1998.
- (56) Szymanska, E.; Saccenti, E.; Smilde, A.; Westerhuis, J. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **2012**, *8* (1), 3–16.
- (57) Mevik, B.-H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Soft.* **2007**, *18* (2).
- (58) Hubert, M.; Rousseeuw, P.; Vanden Branden, K. ROBPCA: a new approach to robust principal component analysis. *Technometrics* **2005**, *47* (1), 64–79.
- (59) Gentleman, R.; Carey, V.; Bates, D.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. H.; Zhang, J. Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biol.* **2004**, *5*, R80.
- (60) Giacomoni, F.; Le Corguillé, G.; Monsoor, M.; Landi, M.; Pericard, P.; Pétéra, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; Goultquer, S.; Thévenot, E. A.; Caron, C. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* **2015**, *31* (9), 1493–1495.
- (61) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504.
- (62) Sumner, L.; Amberg, A.; Barrett, D.; Beale, M.; Beger, R.; Daykin, C.; Fan, T.; Fiehn, O.; Goodacre, R.; Griffin, J.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A.; Lindon, J.; Marriott, P.; Nicholls, A.; Reily, M.; Thaden, J.; Viant, M. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007**, *3* (3), 211–221.
- (63) Franceschi, P.; Vrhovsek, U.; Mattivi, F.; Wehrens, R. In *Metabolic biomarker identification with few samples. In Chemometrics in practical applications*; Varmuza, K., Ed.; InTech: 2012.
- (64) Braga-Neto, U. Fads and fallacies in the name of small-sample microarray classification - A highlight of misunderstanding and

erroneous usage in the applications of genomic signal processing. *Signal Processing Magazine, IEEE* 2007, 24 (1), 91–99.

(65) Saccenti, E.; Hoefsloot, H.; Smilde, A.; Westerhuis, J.; Hendriks, M. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 2013, 10 (3), 361–374.

(66) Wehrens, R.; Franceschi, P. Thresholding for biomarker selection in multivariate data using Higher Criticism. *Mol. BioSyst.* 2012, 8, 2339–2346.

(67) Lawton, K. A.; Berger, A.; Mitchell, M.; Milgram, K. E.; Evans, A. M.; Guo, L.; Hanson, R. W.; Kalhan, S. C.; Ryals, J. A.; Milburn, M. V. Analysis of the adult human plasma metabolome. *Pharmacogenomics* 2008, 9 (4), 383–397.

(68) Mittelstrass, K.; Ried, J. S.; Yu, Z.; Krumsiek, J.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Polonikov, A.; Peters, A.; Theis, F. J.; Meitinger, T.; Kronenberg, F.; Weidinger, S.; Wichmann, H. E.; Suhre, K.; Wang-Sattler, R.; Adamski, J.; Illig, T. Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers. *PLoS Genet.* 2011, 7 (8), e1002215.

(69) Yu, Z.; Zhai, G.; Singmann, P.; He, Y.; Xu, T.; Prehn, C.; Romisch-Margl, W.; Lattka, E.; Gieger, C.; Soranzo, N.; Heinrich, J.; Standl, M.; Thiering, E.; Mittelstrass, K.; Wichmann, H.-E.; Peters, A.; Suhre, K.; Li, Y.; Adamski, J.; Spector, T. D.; Illig, T.; Wang-Sattler, R. Human serum metabolic profiles are age dependent. *Aging Cell* 2012, 11 (6), 960–967.

(70) Moore, S.; Matthews, C.; Sampson, J.; Stolzenberg-Solomon, R.; Zheng, W.; Cai, Q.; Tan, Y.; Chow, W.-H.; Ji, B.-T.; Liu, D.; Xiao, Q.; Boca, S.; Leitzmann, M.; Yang, G.; Xiang, Y.; Sinha, R.; Shu, X.; Cross, A. Human metabolic correlates of body mass index. *Metabolomics* 2014, 10 (2), 259–269.

(71) Morio, B.; Comte, B.; Martin, J.-F.; Chanseaume, E.; Alligier, M.; Junot, C.; Lyan, B.; Boirie, Y.; Vidal, H.; Laville, M.; Pujos-Guillot, E.; Sébédo, J.-L. Metabolomics reveals differential metabolic adjustments of normal and overweight subjects during overfeeding. *Metabolomics* 2015, DOI: 10.1007/s11306-014-0750-y.

(72) Caubet, M.-S.; Comte, B.; Brazier, J.-L. Determination of urinary ¹³C-caffeine metabolites by liquid chromatography-mass spectrometry: the use of metabolic ratios to assess CYP1A2 activity. *J. Pharm. Biomed. Anal.* 2004, 34 (2), 379–389.

(73) Dunn, W.; Lin, W.; Broadhurst, D.; Begley, P.; Brown, M.; Zelena, E.; Vaughan, A.; Halsall, A.; Harding, N.; Knowles, J.; Francis-McIntyre, S.; Tseng, A.; Ellis, D.; O'Hagan, S.; Aarons, G.; Benjamin, B.; Chew-Graham, S.; Moseley, C.; Potter, P.; Winder, C.; Potts, C.; Thornton, P.; McWhirter, C.; Zubair, M.; Pan, M.; Burns, A.; Cruickshank, J.; Jayson, G.; Purandare, N.; Wu, F.; Finn, J.; Haselden, J.; Nicholls, A.; Wilson, I.; Goodacre, R.; Kell, D. Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics* 2015, No. 1, 11–26.

(74) Stanislaus, A.; Guo, K.; Li, L. Development of an isotope labeling ultra-high performance liquid chromatography mass spectrometric method for quantification of acylglycines in human urine. *Anal. Chim. Acta* 2012, 750, 161–172.

(75) Saito, K.; Maekawa, K.; Pappan, K.; Urata, M.; Ishikawa, M.; Kumagai, Y.; Saito, Y. Differences in metabolite profiles between blood matrices, ages, and sexes among Caucasian individuals and their inter-individual variations. *Metabolomics* 2014, 10 (3), 402–413.

(76) Kloet, F.; Tempels, F.; Ismail, N.; Heijden, R.; Kasper, P.; Rojas-Cherto, M.; Doorn, R.; Spijksma, G.; Koek, M.; Greef, J.; Mäkinen, V.; Forsblom, C.; Holthöfer, H.; Groop, P.; Reijmers, T.; Hankemeier, T. Discovery of early-stage biomarkers for diabetic kidney disease using ms-based metabolomics (FinnDiane study). *Metabolomics* 2012, 8 (1), 109–119.

(77) Posada-Ayala, M.; Zubiri, I.; Martin-Lorenzo, M.; Sanz-Maroto, A.; Molero, D.; Gonzalez-Calero, L.; Fernandez-Fernandez, B.; de la Cuesta, F.; Laborde, C. M.; Barderas, M. G.; Ortiz, A.; Vivanco, F.; Alvarez-Llamas, G. Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int.* 2014, 85 (1), 103–111.