

UNTARGETED SERUM METABOLIC PROFILING BY COMPREHENSIVE TWO- DIMENSIONAL GAS CHROMATOGRAPHY—HIGH-RESOLUTION TIME-OF-FLIGHT MASS SPECTROMETRY

Nicolas Di Giovanni, Marie-Alice Meuwis, Edouard Louis and Jean-Francois Focant

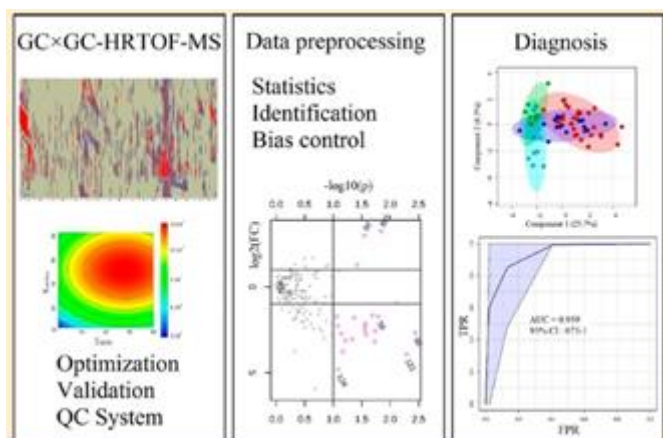
Department of Chemistry, Organic and Biological Analytical Chemistry Group, Quartier Agora, University of Liege

GIGA institute, Translational Gastroenterology and CHU de Liege, Hepato-Gastroenterology and Digestive Oncology, University of Liege

KEYWORDS: GC×GC, comprehensive gas chromatography, untargeted metabolomics, QC system, Crohn's disease, inflammatory bowel disease

Abstract

While many laboratories take appropriate care, there are still cases where the performances of untargeted profiling methods suffer from a lack of design, control, and articulation of the various steps involved. This is particularly harmful to modern comprehensive analytical instrumentations that otherwise provide an unprecedented coverage of complex matrices. In this work, we present a global analytical workflow based on comprehensive two-dimensional gas chromatography coupled to high-resolution time-of-flight mass spectrometry. It was optimized for sample preparation and chromatographic separation and validated on in-house quality control (QC) and NIST SRM 1950 samples. It also includes a QC procedure, a multiapproach data (pre)processing workflow, and an original bias control procedure. Compounds of interest were identified using mass, retention, and biological information. As a proof of concept, 35 serum samples representing three subgroups of Crohn's disease (with high, low, and quiescent endoscopic activity) were analyzed along with 33 healthy controls. This led to the selection of 33 unique candidate biomarkers able to classify the Crohn's disease and healthy samples with an orthogonal partial least-squares discriminant analysis Q2 of 0.48 and a receiver-operating- characteristic area under the curve of 0.85 (100% sensitivity and 82% specificity in cross validation). Fifteen of these 33 candidates were reliably annotated (Metabolomics Standards Initiative level 2).



Introduction

Since its debut with Pauling in 1971¹ and the introduction of the specific term in 1998,² metabolomics, the comprehensive study of the low-molecular weight metabolites of a biological sample, has considerably evolved to become a widespread discipline commonly employed to provide phenotypic insights in applications such as clinical medicine.³ Next to nuclear magnetic resonance (NMR), mass spectrometry (MS) appeared as one of the most powerful analytical platforms in the field, especially when combined with separation techniques such as liquid or gas chromatography.⁴ Comprehensive two-dimensional gas chromatography (GC×GC) has soon demonstrated high potential for the elucidation of biological samples.^{5–7} It has been subject to major enhancements and numerous studies during the last two decades,⁸ leading the way to a mature technique.^{9,10} Further coupling with high-speed MS analyzers such as time-of-flight (TOF) MS allowed higher resolution, sensitivity, and mass spectral quality.¹¹ In comparison to single-dimensional (1D) GC–MS, this resulted in an increased number of detected peaks,^{9,12} up to 20 times-enhanced detection limits¹³ and more accurate and reliable results (less false positives in biomarker research). Moreover, new high-resolution (HR) TOF-MS instruments not only strengthen compound annotation but also further improve the quantitation based on distinctive masses as well as the selectivity through deconvolution, which leads to an increase in the number of unique metabolites detected and a facilitated search for targeted compounds.¹⁴ In addition, many key aspects have been extensively investigated, such as quality assurance/quality control (QA/QC),^{15–17} sample preparation,¹⁸ separation, and detection,¹⁹ or identification.²⁰ This is also the case for data (pre)processing,^{21–23} knowing that single GC–MS protocols are, to a large extent, transferable to GC×GC. In addition, specific resources and protocols have been developed and perfected over the years, such as software (GC Image (LLC, Lincoln, Nebraska, U.S.A.), ChromaTOF (LECO Corporation, Saint Joseph, Michigan, U.S.A.), ChromSquare (Shimadzu, Kyoto, Japan), and GC×GC-Analyzer (MsMetrix, Utrecht, Netherlands)), methods (PARAFAC²⁴) and platforms (the more general MetaboAnalyst²⁵ and MetPP²⁶).

Nevertheless, the technique is undermined by various issues: first, the general limitations of GC–MS, such as sample preparation,^{18,27} compound annotation,²⁸ and thermal decomposition,^{29,30} which are still not fully addressed, and second, difficulties common to comprehensive technologies regarding the analytical stability and handling of large data sets, difficulties that make the detection of the significant variables more challenging and less reliable and limit their use in routine laboratories.³¹ Also, in the larger picture, while many studies take appropriate care, there are still cases of untargeted analyses suffering from a lack of articulation. Indeed, when the various steps are considered separately, it likely leaves weak links in the chain that limit the overall quality of the global process and contribute to maintaining the gap with the clinic.

This work aims to perform global profiling of serum samples with GC×GC–HRTOF-MS in a quantitative and qualitative manner, that is, through a coordinated workflow in which all

steps are carefully designed and controlled, based on previous studies as well as general knowledge, and take advantage of using multiple approaches. As a proof of concept, we applied the method to serum samples and looked for potential biomarkers of Crohn's disease endoscopic activity. Crohn's disease (CD) is an inflammatory bowel disease (IBD) characterized by chronic inflammation of any part of the gastrointestinal tract with remission periods. Its negative impact on the quality of life of patients³² is amplified by a high and increasing prevalence.³³ Moreover, current treatments are costly and may present severe side effects.³⁴ Its diagnosis still largely relies on clinical evaluation, namely, invasive colonoscopy and histological evaluation of biopsy together with clinical and endoscopic activity scores. Therefore, the use in clinics of noninvasive companion biomarkers able to indicate patients with endoscopic activity would be useful. Besides, the altered metabolites correlated to endoscopic activity could contribute to a better understanding of the pathophysiology of the disease. To this date, several metabolic profiling studies have been conducted. Most often, CD has been analyzed along with ulcerative colitis (UC) to be differentiated from healthy controls.^{35–41} However, mouse models, either with interleukin-10 deficiency^{42,43} or with induced colitis,^{44,45} have also been investigated. The matrices of choice have been fecal extracts,^{35, 38, 41, 46 – 49} serum,^{37,39,44,45,47,49,50} urine,^{36,40,42,44,50} and colonic mucosal tissues.^{51–53} 1H NMR has been the most frequently applied analytical platform^{35–40,42,44,46,50} with its variant 1H MRS^{51–53} before GC–MS^{43,47,48,54} and LC–MS.^{45,49} GC×GC, the instrumentation used in this work on human serum, is obviously able to detect the chemical classes observed in GC–MS that include most of the ones observed in NMR, with individual metabolite differences.^{55,56} Those classes are mainly amino acids, organic/carboxylic acids, fatty acids, sugars, and sugar alcohols.^{55,57} LC–MS, depending on the ionization technique (usually electrospray ionization (ESI)) and the acquisition mode (positive, negative, or both), is able to detect more lipidic compounds,⁵⁸ such as (lyso)phosphatidylcholines (PCs and LPCs),^{45,59} di- and triglycerides (TG), fatty acids (FA),⁶⁰ and phospholipids^{59,61,62} or carnitines and nucleotides.⁴⁹

Material and methods

SAMPLES

Patient recruitment and serum sample intakes were performed at the university hospital of Liege, Belgium. The project protocol received full approval of the institutional ethical reviewing board, and each patient provided a signed informed consent (no. BE70721423133). For all samples, the sample intake, processing, and storage procedures were standardized and followed our biobanking guidelines developed from proteomics studies and utilized also for clinical trials (Supporting Information, Section S1A). The CD patients (n = 35) were classified into three groups of endoscopic activity according to the related Crohn's disease endoscopic index score (CDEIS) as follows: quiescent patients (n = 9) with CDEIS = 0, low- endoscopic activity patients (n = 14) with 0 < CDEIS < 3, and high-endoscopic activity patients (n = 12) with CDEIS > 3. The controls (n = 33), qualified as "healthy" in the sense that they were negative for IBD or any other known cancer at the time of sample intake, were matched for gender and age. The clinical data of the patients are summarized in the Supporting Information (Section S1B). The internal QC samples consisted of 30 µL aliquots, made through one

freeze–thaw cycle, 63 of human serum samples distinct from the study samples and were collected using the same procedure.

CHEMICALS

Hexane, chloroform, methanol absolute, pyridine, methoxyamine hydrochloride (MeOX), N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) + 1% 2,2,2-trifluoro-N-methyl-N-(trimethylsilyl)-acetamide, chlorotrimethylsilane (TMCS), N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA), methyl chloroformate, alkane solutions (C8–C20, ~40 mg/L in hexane), and standards glycine-2,2-d₂, succinic acid-2,2,3,3-d₄, fumaric acid-2,3-d₂, and 4,4'-dibromooctafluorobiphenyl were all purchased from Sigma (Sigma-Aldrich Co. LLC., St. Louis, Missouri, U.S.A.). "Hole-shaped" vials (microliter caps DN11, 0.9 mL) were purchased from VWR (VWR International BVBA, Leuven, Belgium), and glass wool liners (Sky Liner precision and single taper) and capillary chromatography columns (Restek) were from Interscience (Louvain-la-Neuve, Belgium). The NIST-certified reference material SRM 195064 was purchased from the National Institute of Standards and Technology (Rockville, Maryland, U.S.A.) and stored at –80 °C upon arrival. Fresh solutions of MeOX (30 mg/mL in pyridine) and standards (100 µg/mL either in Milli-Q water (Q-Gard 1, Merck KGaA, Darmstadt, Germany) or in hexane for pre- and post-extraction standards, respectively) were prepared every week.

SAMPLE PREPARATION

All samples were stored at –80 °C until use. After thawing at 4 °C for 1 h, 30 µL of each sample was put in hole-shaped vials to which 3 µL of glycine-2,2-d₂ and succinic acid-2,2,3,3-d₄ were added as isotopically labeled internal standards. Metabolites were extracted (see optimization step for details) before cooling at –20 °C for 10 min and centrifugation at 9300 RCF for 10 min. Fumaric acid-2,3-d₂ (3 µL) was then added before the samples were dried under a gentle N₂ flow while heated at 40 °C. They were stored again at –80 °C until use. Before injection, the metabolites were made volatile with a MeOX and MSTFA two-step derivatization protocol (see optimization step for details). The medium was homogenized after each step by 1 min of vortexing. It was recovered, and 3 µL of 4,4'-dibromooctafluorobiphenyl was finally added. All steps were performed manually. For security precautions, see the Supporting Information, Section S2.

GC×GC–HRTOF-MS

The instrumentation consisted of a JMS-T100GCVV GC×GC–TOF-MS (Jeol Ltd., Tokyo, Japan) using an Agilent 6890 GC (Santa Clara, California, U.S.A.) coupled to a CTC autosampler (CTC Analytics AG, Zwingen, Switzerland). We used a set of capillary columns of 30 and 1 m, both 0.25 mm in i.d. and 0.25 µm in film thickness, to optimize the overall flow and compromise between peak capacity and sample loadability.¹³ A guard column (2 m, 0.25 mm) was placed after the injector, and 10 cm was cut after each step of routine maintenance of the instrument. After

which, the system was reconditioned with five injections of a QC sample. SilTite μ - Unions (SGE Analytical Science) were used for the connections. The peaks were modulated through a ZX1 loop cryogenic modulator using liquid nitrogen (Zoex Corporation, Houston, U.S.A.). Helium served as a carrier gas in a constant flow mode. The GC oven was held at the initial temperature for 1 min (purge time was 1 min at 50 mL/min) and at the final temperature of 300 °C for 5 min. The ion source and transfer line temperatures were set to 250 °C. Ions were generated by electron ionization (EI) in the positive mode at 70 eV and 300 μ A. The data were acquired at 25 Hz at m/z 50–700.

METHOD OPTIMIZATION

Two extraction solvents (methanol (3:1)²³ and methanol/ chloroform/water (2.5:1:1:1)⁶⁵) and four silylation agents (MSTFA, BSTFA, MTBSTFA, and methyl chloroformate⁶⁶) were tested¹⁸ in triplicate. A Box–Behnken design of experiment (DoE, Unscrambler X, Camo, Oslo, Norway) helped determine the most appropriate⁶⁷ volumes (for the derivatization reagents: 5–30 μ L for 30 μ L of serum), temperatures (20–100 °C), and durations (0.5–4 h). Details are provided in the Supporting Information, Section S3. Regarding separation and detection,⁶⁸ we evaluated multiple stationary phases (normal Rxi-5/Rxi-17, inverse Rxi-17/Rxi-5, and alternative sets Rtx-200/Rxi-5⁶⁹), flows (0.5, 1, and 1.5 mL/min), initial temperatures (40, 50, and 90 °C), temperature ramps (3, 5, 7, and 10 °C/min), modulation periods (PM = 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, and 8 s), hot jet durations (at first, a fifth of the modulation period then 500, 700, and 900 ms for 3.5 s of PM), injection temperatures (200, 250, and 280 °C), liner designs (“single taper gooseneck” and “precision” 4 mm Sky liners with glass wool), and injection volumes and splits (1 and 2 μ L in the splitless mode and 5 μ L with split ratios of 5:1 and 10:1).

METHOD VALIDATION

The method was validated on internal QC samples and a certified reference material (NIST SRM 1950). Six QC samples were injected consecutively (intra-batch), and five of them were injected over five days (inter-batch). We monitored 19 metabolites representative of the matrix regarding the chemical classes, retention times in both dimensions (1t_R and 2t_R), and peak volumes (see the Supporting Information, Section S4). Four SRM 1950 samples were injected consecutively (intra- batch), and five were injected over five days (inter-batch). Twenty-eight metabolites (16 amino acids, 3 clinical markers, and 9 fatty acids, Supporting Information, Section S5) were properly annotated and followed.

QC SYSTEM

To define the control state of the system, nine QC samples were injected over two days after maintenance of the instrument.

PROOF OF CONCEPT

We analyzed 35 serum samples of three subgroups of Crohn's disease patients: high endoscopic activity ($n = 12$), low endoscopic activity ($n = 14$), and quiescence ($n = 9$), along with 33 healthy controls, 15 QC samples (one every five samples), and 4 SRM 1950 samples. Three process blanks¹⁷ were injected before each QC. All analytical steps were randomized. Six samples, a single one and a batch of 5, had to be reinjected after rejection by the QC system.

DATA PROCESSING

First, the raw data acquired were submitted to centroidization (Data Manager software, JEOL Ltd., Japan; .CDF and .GCI data formats). GC Image 2.5 software (LLC, Lincoln, Nebraska, U.S.A.) automatically performed process blank subtraction and baseline correction. In the proof-of-concept study where multiple chromatograms were to be processed together, they were aligned and created a composite template that included all detected features. The template was manually verified for artifacts, degradation zones, coelutions, and multiple peaks of single compounds (amino acids, sugars, and wraparounds that were removed, split, and merged). Then, it was applied to each individual chromatogram to generate consistent peak lists examined for transcription errors. Only metabolites present in at least half the samples of any class were kept. The remaining missing values were replaced by the lowest signal measured for the metabolite divided by two. The peak volumes were normalized to the internal standard giving the lowest relative standard deviation (RSD) in the QC samples.

Outliers were sought by calculating robust z scores, according to the median absolute deviation of the variable in all concerned samples. For z scores of >50 , the outlier was replaced by the closest value divided or multiplied by two (maximum or minimum outlier). Finally, the peak volumes were corrected for the analytical variation measured in the QC samples through a locally estimated scatterplot smoothing (LOESS) procedure. Only the metabolites with a corrected analytical variation (RSD) lower than 30% were kept with the idea of using "fewer but valid signals, rather than more inaccurate and noisy" ones.³¹ The statistical tests and methods applied in the following steps (scaling evaluation, biomarker research, bias control, and performance assessment) were conducted with MetaboAnalyst web resource,⁷⁶ Excel add-in Tanagra⁷⁷ and RapidMiner,⁷⁸ and R programs.⁷⁹

ANNOTATION

Relying only on a full mass spectrum comparison provides limited confidence. Therefore, the most plausible results obtained this way (NIST 14 and Wiley 9 libraries, unit mass resolution) were further tested by comparing the theoretical (NIST14, Wiley 9 libraries, ChemSpider, and NIST Webbook web resources) and experimental linear retention indices (LRIs).¹¹ The latter were measured with a standard mix of n -alkanes (C8–C20) added to three of the study samples. As a third orthogonal property,²⁸ the mass accuracy was calculated for the highest selective fragments available, usually molecular $(M)^{+ \cdot}$ or $(M-CH_3)^{+ \cdot}$ radical ions. To do so, six study samples (two with low endoscopic activity, two with high endoscopic activity, and two healthy controls) were chosen for their high concentrations in the potential biomarkers. Their biological replicates were analyzed, according to the previously optimized sample preparation and

separation conditions on Pegasus GC-HRT 4D (LECO Corporation, Saint Joseph, Michigan, U.S.A.; resolution of 25,000). The secondary oven's temperature was 15 °C higher than that of the main one. Acceptance thresholds were set for the three types of information: match factor of >70023 or >600 with a probability of >50%, $\Delta\text{LRI} <^{25,80}$ and mass error < 1 ppm for any specific fragment. We finally assessed the biological plausibility of the annotated candidates with the Human Metabolome Database (HMDB81) and the Kyoto Encyclopedia of Genes and Genomes (KEGG82) as well as by conducting a manual literature review. Further interpretation and exploration of these results are beyond the scope of the present paper. Ideally, they should include a targeted confirmation of the variations observed, conducted on an independent set of patients. A powerful approach would be to integrate other omics data obtained on these specific biological samples.

RESULTS AND DISCUSSION

METHOD OPTIMIZATION

The parameters and the ranges to optimize were chosen based on previous studies. Due to the untargeted nature of the method, we aimed to maximize three factors: the whole chromatogram signal, the signal for the less reactive amino groups, and the number of resolved metabolites. High-quality chromatography was also sought in order for the following automated steps, particularly preprocessing,⁸³ to be conducted reliably and minimize the need for manual correction. Mainly independent from each other, the sample preparation and separation were optimized separately but successively in order to maintain the articulation of the analytical process. In the experiments conducted, methanol (3:1) revealed to be the most efficient extraction solvent, confirming previous results²³ (see the Supporting Information, Section S6A, for the detailed results). MSTFA was the most efficient derivatization agent, and its volume was the most influential parameter in the DoE. The global derivatization used 15 μL of MeOX and 10 μL of MSTFA with both reactions performed at 40 °C for 1 h. The stability of the response near the optimal values of temperature and duration increased the robustness of the process. Again in accordance with previously published results,^{27,84} the parameters interacted and compromised between the completion of the derivatization,⁷⁵ solubilization¹⁸ and dilution of the dried metabolites, production of artifacts,¹⁸ and hydrolytic degradation⁸⁵ of the adducts,⁷⁵ overall leading to moderate optimum conditions. Despite the potential of reverse column sets to use the chromatographic space in specific applications,^{7,13} the best resolution was achieved with a classical nonpolar/semipolar set (Supporting Information, Section S6B). The (linear) temperature ramp was found to be the dominant separation factor with an optimum value of 3 °C/min. Such a low ramp is quite unusual and, to the best of our knowledge, has never been reported, possibly because it increases the analysis time. Nonetheless, we believe that it takes full advantage of the second dimension of separation. Indeed, producing low elution temperatures at the end of the first column allows the metabolites to spend more time in the second column and to separate better, overcoming the resulting peak broadening. Interestingly, the lower elution temperatures should also help reduce thermal degradation and thus increase the sensitivity. The data acquisition and the temperature ramp were stopped at 240 °C because no metabolite was eluting at higher temperatures. Over 240 °C, a ramp of 15 °C/min

was applied, up to the final temperature of 300 °C. The other optimum parameters were a 1 mL/min flow, starting temperature of 50 °C, modulation period of 3.5 s coupled to a 700 ms hot jet, and 250 °C splitless injection of 1 µL using a “precision” liner.

The importance of optimizing an analytical method intended to perform untargeted analysis of complex matrices is already well documented.¹¹ This was particularly the case here for the derivatization conditions, column set, and temperature ramp. Less influential factors should be considered as well because they can contribute to the high-quality chromatography desired, especially in a derivatized environment. However, since many parameters exhibit similar optimum values across various studies that were confirmed for the main part here, time and efforts can be saved by searching around these values or using them directly and controlling them in that they give satisfactory analytical performances.

METHOD VALIDATION

To determine the parameters to assess and how to do so, the bioanalytical method validation guidance draft published by the Food and Drug Administration (FDA)⁸⁶ was used as a reference. When necessary, it was completed by specific publications. Accuracy, Precision, and Recovery. The 2D peak volume accuracy compared, through z scores, the NIST certified or reference values of the 28 monitored metabolites to the corresponding measured values (corrected with the most chemically similar IS). As a result, the method was found to be limited to semiquantitation due to the lack of the exact response factors (Table 1 and Section S7 of the Supporting information Information). However, for glycine, the only NIST metabolite that we used as an IS, the response factor was calculable and the z scores were <1 in both intra- and inter-batch evaluations (0.5 and 0.9, respectively). Therefore, absolute quantitation seems possible but at the cost of one IS per compound with proper calibration, which would be extremely difficult to achieve in global profiling.¹¹

The precision was assessed through the relative standard deviation in intra- and inter-batch measurements. The chromatographic separation was found to be stable and reproducible regarding the retention times, measured in the QC samples (Supporting Information, Section S4). Regarding the peak volumes, the RSD measured in the NIST samples was under 30% for all 28 metabolites monitored in both intra- and inter-batch evaluations. They were even under 10% for 24 and 18 metabolites. This is comparable to the NIST reported values,⁸⁷ in fact, lower in 13 and 7 cases. It shows the stability of the method that could nevertheless be further improved through the automation of the sample preparation. In the QC samples, the peak volume precision was lower, albeit still good, with 14 and 11 metabolites (out of 16 since the IS had no variation after normalization), having an RSD < 30%. GC–MS data reported previously⁸⁸ found RSD values lower than 20% for all the 32 representative compounds tested and lower than 10% for most of them. The modulation system and the second chromatographic column therefore seem to be a source of additional variations, hence the need to correct and monitor the peak volumes, here through the use of a QC system and LOESS (partial) correction. It is particularly true for multiple batches since the precision was lower in inter-batch measurements. Homogenization issues due to manual aliquoting could explain the higher RSD in the QC samples compared to the NIST SRM, although the metabolites monitored were not the same. Such issues have been

reported.⁸⁸ It is of primary importance since the QC samples constitute the core of the global quality system on which the method relies. In addition, they should also be aliquots of all the study samples, which is a minor limitation of this study.

Regarding the recovery, the differences between the response factors of the standards⁷³ prevented the direct comparison of their signals, especially between the non-derivatized injection standard and the others, which is why negative recoveries were obtained in this case. If absolute values cannot be interpreted, their reproducibility is still meaningful.⁸⁹ We found a maximum RSD of 15.6% for three different batches of SRM 1950 and QC samples (Supporting Information, Section S8), indicating the robustness of the optimized sample preparation.

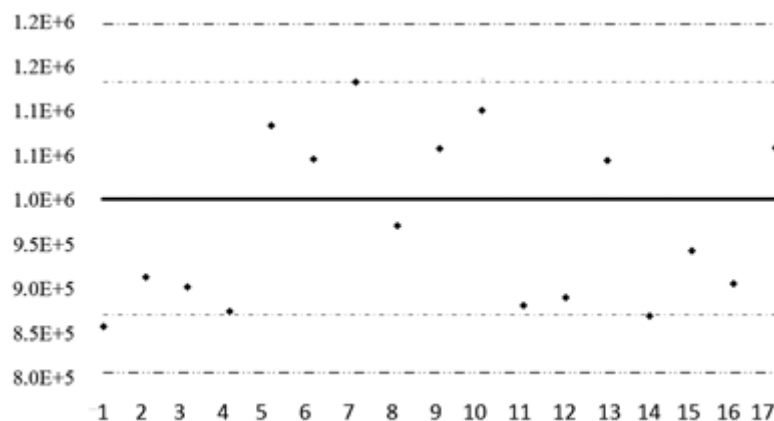
Linearity, Range, and Sensitivity. These parameters were evaluated using glycine-2,2-d₂, succinic acid-2,2,3,3-d₄, and fumaric acid-2,3-d₂ added to QC samples at five different concentrations (process blank¹⁷ and 0.05, 0.25, 1.67, and 10 µg/ mL), each in six replicates. Selective m/z values representative of the IS (deuterated masses) were chosen and used in different types of regressions of increased complexity (unweighted and \sqrt{S} , S, S₂, S₃, and S₄ weights) in order to take the uncertainty of measure into account. To assess these regressions, we avoided the R² coefficient and preferred the sum of relative errors, the residual plots,⁹⁰ the range of linearity, and the sensitivity. We found that using a weighted regression was more efficient for succinic and fumaric acids (S₂ weights for both). However, the regression parameters of the second were only slightly changed (from $y = 43,158x + 6964$ to $y = 44,021x + 9$; see the Supporting Information, Section S9, for details). Glycine and fumaric acid were reproducibly linear along the whole range tested. Succinic acid showed distinct behaviors below and above 0.25 ng/µL. The one below was used to assess the sensitivity. The limits of detection and quantification (LOD, LOQ) were calculated using the regression curves,⁹¹ process blank¹⁷ samples,^{92–94} and the relative standard deviation (RSD) of real injections⁹⁵ (Supporting Information, Section S10A). As a result, fumaric and succinic acids had consistently low LODs (approximately 2 and 5 pg on the column, respectively) and LOQs (6–9 and 16 pg on the column; Supporting Information, Section S10B, for details). Glycine, due to its higher blank signal, had a respective LOD and LOQ of 64 and 90 pg on the column. This was in good agreement with previous GC×GC studies,^{80,95} especially the one from Koek¹³ who used an S/N method (extrapolated S/N = 3) and various chromatographic columns settings and obtained LODs ranging from 0.5 to 4 pg on the column for fumaric and succinic acids and from 70 to 280 pg for glutamine (glycine was not assessed). Our results were also compared advantageously with GC–MS where a method optimized for extraction and derivatization reported an LOD of approximately 1–5 pmol on the column⁸⁸ (100–500 pg injected on the column for an average molecular mass of 100 g/mol and 1 µL injection). In addition, we manually controlled the selectivity through the chromatographic and mass spectral resolutions of the metabolites, which was facilitated by the 2D chromatogram visualization offered by GC×GC. The stability of the derivatized compounds was maximized by limiting the storage time before injection to 2 h at –20 °C.^{18,27} Overall, despite the impossibility to reach optimal conditions for all metabolites,¹¹ the method was found to be fit-for-purpose. Indeed, according to the results obtained for the metabolites assessed, it should allow the untargeted sensitive detection and semiquantitation (linear behavior) of a large range of metabolites of various chemical classes and concentrations in a stable, therefore reproducible, manner. To do so, the ISs have to be representative of the many metabolites present in the samples. Therefore, the method presented could be further

improved by using a larger pool of (carefully selected) ISs. We would recommend to inject at least three ISs per major chemical class and one per minor one. One IS per intensity level (high, medium, or low) and one per retention window (1tR divided into three equal parts). One could also find it useful to implement some recently developed procedures that combine various ISs (CCMN96 and NOMIS97). Absolute quantification, which would require one IS per compound, seems out of reach for global profiling in the actual state of the field. Finally, the method developed for the untargeted analysis of 30 μ L of serum samples should be adaptable to other matrices common in metabolomics, such as plasma and urine.^{37,50} QC System. A QC system repeats QC sample injections to monitor the stability of the performances over time, providing assurance of the quality¹⁶ of the data. It can also be used to increase it by correcting for the analytical variations through an LOESS procedure.²³ Here, control charts were constructed for the 19 representative compounds defined at the validation step. Based on the validation results as well as on previous studies,^{80,57} the acceptable variations (around the mean value) for the retention times in both dimensions and normalized peak volumes were set to 2 times the modulation period, 5% of the PM, and 30% of the mean. The warning and action limits were drawn at 2 and 3 times these values. Finally, acceptance and rejection criteria were defined for individual injections and batches (Supporting Information, Section S11A). The resulting charts confirmed that the retention times and therefore the chromatographic separation were very stable. Indeed, 18 representative metabolites were well inside the warning limits. The only exception was the cysteine 1tR. Regarding the peak volumes, 11 and 1 metabolites (out of 16 since the 3 ISs were completely corrected by the normalization) had all their nine values inside the warning and action limits. For the four metabolites remaining, which violated at least once the action limits, the control charts were reconstructed with warning and action thresholds set to 2 and 3 times their measured standard deviations.⁹⁸ This way, the system was considered under control as long as these metabolites behaved within their usual ranges of values, even if it was over the acceptable 30% threshold. From a methodological point of view, the results obtained with nine injections over two days seemed to define well enough the control state of the analytical system. Anyway, for studies analyzing large cohorts over long periods of time, it would be advisable to inject multiple batches (at least five) of multiple (at least three) daily samples distributed over at least two weeks. Performing maintenance actions in that time frame would allow to evaluate their influence on the system. A representative example of a QC chart is provided in Figure 1.

Table 1. *Intra- and Inter-Batch Accuracy and Precision Values for Selected Metabolites Assessed in the NIST SRM 1950*

Metabolites	Mean (μmol/L)	accuracy		NIST	precision	
		Intra-batch	Inter-batch		Intra-batch	Inter-batch
		Z score			RSD %	
phenylalanine	51.0	11.3	13.1	14	4.2	10.7
methionine	22.3	7.7	6.4	8	14.8	20.2
glycine	245.0	0.5	0.9	7	4.1	4.2
tyrosine	57.3	25.0	26.3	5	4.7	11.2
histidine	72.6	17.8	16.2	5	9.0	7.7
glutamic acid	67.0	1.0	1.2	27	3.0	9.2
uric acid	254.0	27.9	26.2	2	2.8	16.2
dodecanoic acid	9.5	13.6	12.0	6	3.7	3.6
myristic acid	80.1	4.1	3.9	21	3.4	4.3
palmitic acid	2364.0	29.2	28.9	3	7.3	7.7
heptadecanoic acid	17.6	23.9	23.3	4	9.4	7.8
oleic (Z)-acid	1614.0	9.6	9.5	10	3.4	9.9

Figure 1. Control chart of the peak volume (normalized raw signal against time of analysis in days) of glutamic acid. All values lie inside the action limits (bold dashed lines, $\mu \pm 3\sigma/\sqrt{n}$). The ones lying outside the warning limits (light dashed lines, $\mu \pm 2\sigma/\sqrt{n}$), that is, days 1, 7, and 14, could indicate some analytical instability, which would have to be confirmed on the other monitored metabolites.



PROOF-OF-CONCEPT STUDY

To further evaluate the analytical platform developed as well as to design and assess proper workflows for data (pre)processing, statistical treatments, bias control, and performance assessment (summarized in Figure 2), a real-life biomarker research study was conducted. It aimed to separate the Crohn's disease samples put together regardless of the endoscopic activity ($n = 35$) from the healthy controls ($n = 33$). Further, the possibility to separate the three subgroups of Crohn's disease patients from each other was evaluated. Raw Data Processing. A total of 922 unique peaks were consistently detected across all the study chromatograms of which 524 were kept after manual control for their chromatographic and mass spectral specificity. This is in good agreement with the 517 and 566 peaks obtained by Almstetter⁹⁹ and Castillo,⁸⁰ respectively, and empirically confirms the high peak capacity of this analytical platform. Further preprocessing included an LOESS (partial) correction and selection (Figure 2), which effect is shown in the Supporting Information, Section S11B. It led to a final data set of 183 metabolites having an RSD of $<30\%$, almost twice as many as with 1D GC-MS,⁵⁷ despite the strict quality criteria applied. This demonstrates the added value of comprehensive GC \times GC in untargeted applications. The exact chemical nature of the additionally resolved and detected metabolites was not investigated. Most probably, they belonged to the classes usually seen in GC-MS, that is, mainly amino acids, organic/carboxylic acids, fatty acids, sugars, and sugar alcohols.^{55,57} The selection made not only aimed to guarantee the quality of the data but also reduced the size of the data set and made it easier to handle at the following steps. Data Scaling. We tested the effect of auto, pareto, level, range, and vast scaling on the structure of the data-natural trends and outliers-and on the biomarker research,^{21,100} here to separate the four groups of samples at once. To do so, we applied unsupervised hierarchical clustering analysis (HCA) and principal component analysis (PCA) along with the 95% confidence ellipse.¹⁰¹ We also selected various, differently operating, supervised methods: univariate ones: analysis of variance (ANOVA), Fisher ratios, and Stepdisc feature selection and multivariate ones: partial least squares discriminant analysis (PLS-DA) and random forests (RF). We found that level scaling was very sensitive to minor differences in samples. It tended to artificially create outliers that did not seem to exist in the raw data and was left out (see the Supporting Information, Section S12, for related figures). The other four methods produced similar PCA and PLS-DA plots as well as RF results. All had a positive effect on biomarker research in comparison to the raw data and led to the same selection of metabolites with similar significance values. Auto and range were slightly better in prediction with the highest predictive Q² (0.28 and 0.31, respectively, on the whole dataset), in accordance with results from van den Berg et al.²¹ However pareto was chosen because it maintained more of the initial structure (PCA and HCA) and was reported to be more stable than these two methods.²¹ In summary, at the exception of level scaling, the methods tested here, based on mean centering, acted positively and similarly on the data structure and biomarker research, confirming that data scaling can affect the final results²¹ and should be considered in method development. Overall, the proposed raw data processing workflow maintained the integrity of the data while preparing them for the following statistical treatment. Biomarker Research. In recent years, several protocols of high quality have been published, which aimed to standardize the main steps of a metabolomics analysis, including data preprocessing and processing.^{23,73,101} Indeed, standardization seems to be the next step in the field.¹⁷ The protocols for biomarker research voluntarily limit the process to several successful methods, such as RF, ANOVA, receiver operating characteristic (ROC) curves, and bootstrap validation. Nevertheless, other methods exist, and even the advocates of standardization

recognize that no workflow is appropriate in all situations.¹⁰¹ For this reason, in the experimental procedure presented here, the data set was submitted to multiple tests characterized by different ways of operating algorithms and assumptions,¹⁰¹ such as correlation versus hypothesis testing versus modelization versus diagnosis and parametric versus non parametric methods. By doing so, we intended to obtain a comprehensive and robust process applicable on data sets of various structures and sizes, including the ones characterized by low subject-to-variable ratios. However, in order to not complicate it excessively, a prior selection of the methods used was made based on general statistical knowledge and literature review. Practically, the first step consisted of the application of univariate testing (Fisher ratios (FR), correlations, and ANOVA; the exact test depending on the normality of the distribution), multivariate feature selection (Stepdisc, Runs, and Relief filterings), and multivariate modelization (orthogonal partial least-squares discriminant analysis (OPLS-DA), PLS-DA, RF, Naive Bayes classification, support vector machine (SVM), and neural networks (NN)). Individual fold changes (FC), two- tailed statistical power (through Cohen's $d^{102,103}$), and ROC curves were also determined. Further details are given in the Supporting Information, Section S13A. Second, multiple testing and overfitting issues were taken into account through Bonferroni correction for univariate analysis¹⁰⁴ and resampling error rates (bootstrapping, cross, and leave-one-out validations) for multivariate analysis, respectively.¹⁰¹ Third, the metabolites were ranked according to their significance based on p values, order of selection, weight, or variable importance in the projection (VIP) score. The most consistently significant ones were selected as candidate biomarkers (CB; Supporting Information, Section S13B).^{101,105} Fourth, to reduce the chance of missing any discriminant compound,¹⁰⁴ the candidates were removed from the data set and PLS-DA, OPLS-DA, and RF were applied to it. In the same perspective, we compared the supervised plots drawn on the data set before and after the selection and removal of the candidates. Again, the procedure is summarized in Figure 2. Its application to discrimination between the Crohn's disease samples and healthy controls gave slight differences between the techniques, especially for Bayes classifier, neural networks,¹⁰⁶ and SVM.²² Apart from that, the same candidates were consistently highlighted as the most significant ones ($n = 22$) (Table 2 and Section S14 in the Supporting Information for detailed results). The discrimination between the three Crohn's disease subgroups, characterized by different endoscopic activities, suffered from the nonapplicability of ROC curves, OPLS-DA, SVM, FC, and statistical power (number of classes > 2) as well as from the difficulty to interpret the Bayes classifier and neural network in terms of variable importance in such a case (Section S15). In addition, the statistical values obtained were lower than for the comparison between the CD and healthy control (HC) samples. Three reasons could contribute to explaining this: first, the logically higher biological proximity of the samples, second, the limits of the analytical method with the metabolites detected with this technique in this matrix possibly not being able to fully discriminate, and third, the difficulty for the statistical methods and models to differentiate three groups at once. It could justify the restratification of the data set to test new binary and therefore simpler hypotheses of work. This would also increase the number of possible methods available and facilitate the interpretation in terms of explanatory variables.¹⁰¹ Despite this issue, 29 candidates were consistently highlighted. Twelve of these were common to the previous separation and could be of particular interest in biological interpretation if confirmed on another cohort of patients. In both cases, the PLS-DA/OPLS- DA/RF posterior tests did not highlight any new potential candidate. In addition, the comparison of the supervised plots before

and after the removal of the candidates showed a large degradation of the separation, which supported the efficiency of the selection process. Based on these results, we consider that, for the global profiling of complex samples, the most appropriate methods are correlations (both Pearson/Spearman and Kendall), FC, Welch, or Kruskal-Wallis ANOVA or Fisher ratios (all working on a F distribution), Stepdisc as feature selection, PLS-DA and RF as modeling methods, ROC curves, univariate statistical power, and boxplots. Here, these methods managed well the small sample sizes (minimal $n = 9$), the noise inherent to biological variables, and, for the multivariate ones, the many variables encountered. If using all of them takes time and demands efforts, it gives a more comprehensive view of the data set. At least two or three should be used to be able to cope with any malfunctioning. Regarding the other techniques, they do not seem absolutely necessary because they are not always applicable (neither are ROC curves and statistical power); they can be more difficult to interpret (weights) and/or they tend to be redundant. Indeed, here, the same candidates were selected whether they were used or not (Table S15). Nonetheless, these methods can be useful in specific cases if satisfying results cannot be found otherwise. In general, the consensus metabolites are the most reliable and should be selected first as candidates. The significant compounds highlighted by only one or a few methods should be further investigated by looking to the raw data (FC, boxplots, presence of atypical values that could alter the parametric methods, and influence of zero values or ties). Looking for possibly missed interesting metabolites is also useful, particularly since the number and significance of the candidates depend on the biological proximity of the groups as well as on the coverage of the metabolome allowed by the analytical instrumentation. Finally, since it is difficult to define precisely what is a relevant feature,¹⁰¹ one could find it desirable to adapt the selection thresholds. For example, it seems acceptable in an exploratory perspective to overlook the Bonferroni correction, provided that it is mentioned. It is however not advisable to lower them too much. Bias Control. Despite careful experimental design and selection of the samples, confounding factors can have a non-negligible¹⁰⁵ impact to the point that they could be the “most important threat to validity”.¹⁰⁷ The following procedure was designed to avoid spurious results^{104,16} by assessing the extent to which the candidate biomarkers truly reflected the classes of interest and not the bias factors. All metadata available were considered: analytical: injection order, hemolysis of the samples, sample extraction, and drying, and clinical: age, gender, body mass index (BMI), disease location, tobacco and alcohol consumptions, medication for hypothyroidism, gastroesophageal reflux and ulcers, antitumor necrosis factor α (anti-TNF α), and immunosuppression (Section S1). The statistical methods employed were chosen to deal with the various situations met. First, the repartition of the factors between the biological classes studied was investigated:¹⁰⁴ for continuous variables, through Pearson’s r or Spearman’s ρ , according to normality, as well as Kendall’s τ , Welch ANOVA, and Kruskal-Wallis test and, for categorical ones, through, Goodman-Kruskal’s τ , and Theil’s U . As a complement, the means, medians, and percentages were compared between the classes, and boxplots and scatterplots were drawn. When a factor was not adequately matched between the groups, we aimed to determine which candidates it could affect. The second step therefore evaluated the direct relationship between the factor and the candidates. Through Pearson’s r or Spearman’s ρ , according to normality, as well as Kendall’s τ , Welch ANOVA, and Kruskal-Wallis test. Third, we looked for the effect of the factor on the candidates’ ability to discriminate, the very point of the study. It was done through partial correlations and the equilibration of the classes for the confounding factor. To do the latter, we determined the samples most responsible for the

imbalance and we temporarily removed them. After which, the new repartition was tested again for imbalance. Then, correlations (Pearson/Spearman and Kendall), Welch ANOVA and Kruskal-Wallis tests, individual fold changes, and ROC curves were compared before and after the removal, that is, with and without the imbalance. As a visualization tool, PCA¹⁰¹ and HCA plots were constructed on the candidates with each point tagged according to the bias factor. For categorical bias factors having a sufficiently low number of subgroups, boxplots and univariate ROC curves were drawn according to those subgroups and compared. The global protocol is summarized in Figure 2. Regarding the separation between the CD samples and the HC, the anti-TNF α and immunosuppressive medications had significant imbalance (p values of <0.01 and <0.05 , respectively, see the Supporting Information, Section S16). Nine and twelve candidates were found to be linked to the two factors at various degrees, but only three of them were biased to such an extent in that their residual ability was too poor to still be significant. They were therefore removed from the set of biomarkers. In the separation between the CD subgroups, unequal repartitions were observed for the anti-TNF α (p value of <0.01) and immunosuppressive medications as well as for the age and BMI of the patients (p values < 0.05 , Supporting Information, Section S17). Five, four, seven, and three candidates were correlated to these factors. The control of the residual ability to separate relied only on the equilibration because of the unavailability of partial correlations for $K > 2$. It led to the removal of five candidates. Two others were biased but still discriminant. They were therefore not taken into account in the separation performance assessment but kept as relevant metabolites. To conclude, the procedure proposed dealt well with both continuous and categorical factors, both normal and non-normal distributions, and small sample sizes. It appeared to be efficient as well as suitable for the bias factors and the data structure that we were confronted with. It remained quite simple despite the complexity inherent to such a process.¹⁰⁷ It should therefore be applicable and beneficial to other studies willing to control for potential biases. Three more things should be noted regarding the practical use of this procedure. First, it aims less to reject candidate biomarkers, even if it can be used to do so, than to evaluate and document their reliability. This is of primary importance, for example, for meta-analysis. Second, the process is sensitive to the sample size. Large cohorts are helpful to smooth the data relative to potential bias factors as well as to keep enough patients after the equilibration to maintain statistical significance. Third, the procedure can only control the presence of potential known biases, not correct for them. Therefore, it does not replace in any way an appropriate experimental design where the samples are matched as properly as possible, given the usual clinical limitations.

Classification and Separation Performances of the Candidate Biomarkers. The goal here was to see if the statistically significant associations measured could lead to biomarkers able to discriminate.¹⁶ First, the unbiased candidates highlighted for both separations ($n = 19$ and $n = 22$) were sequentially grouped, starting with the most discriminant ones, into metamarkers (Supporting Information, Section S18A). Second, their performances were assessed, again with multiple methods: 101 separation space visualization: HCA, PCA, PLS-DA, and OPLS-DA plots; PLS-DA and OPLS-DA classification models; linear discriminant analysis (LDA): Hotelling's T^2 and Wilks's λ through Rao transformation, the most appropriate for small sample sizes;¹⁰⁸ multivariate ROC curves using several algorithms; multivariate statistical power; and mean error rates for LDA, Bayes classifier, PLS-DA, neural network, and SVM models. When applicable (classification models, error rate models, and ROC curves), permutation testing and resampling (including bootstrapping)¹⁰¹ and test validations were also evaluated.¹⁰⁴ For the latter, training

and test sets were constructed in a 2:1 ratio,¹⁰⁴ and they were matched and tested for biases (Supporting Information, Section S18B). Third, to further assess the relative significance of the candidates, we considered their respective order of importance in the performance models. A summary of this protocol can be found in Figure 2. The results showed that the CD samples were effectively and consistently separated from the HC. The diagnosis ability (ROC) was also good^{104,109} with the most efficient model able to diagnose with 100% sensitivity and 82% specificity (Figure 3 and Section S19 in the Supporting Information for the detailed results). For a given method, the performances were similar regardless of the number of candidates included in the meta-marker. This means that the most significant candidates, which formed the core of all the biomarkers tested, produced most of the separation and therefore should be used in classification.¹⁰⁴ In the meantime, the three CD subgroups could not be completely separated, and the performances were logically lower (Figure 4 and Section S20 in the Supporting Information). Among the three possible reasons for this, already exposed in Biomarker Research, the most influential here is the difficulty to discriminate more than two classes with only one frontier of separation,¹⁰¹ regardless of the sophistication of the algorithm. Indeed, it makes it harder for the models to deal with the inherent noise of the variables and to distinguish it from the biologically interesting information. It thus decreases the performances and makes it necessary to employ more variables to achieve sufficient separation. Another issue already mentioned is the nonapplicability of ROC curves, OPLS-DA, SVM, and statistical power for more than two classes. In both cases, the various models converged,¹⁰¹ and the candidates selected as the most significant were found to be the most influential in the performance models. In addition, when the QC samples were added to the plots, they were grouped at the center, close to the healthy controls. This indicated that the analytical variability was much lower than the biological one and confirmed the stability of the designed analytical process.¹⁶ From a methodological point of view, using methods that operate differently allows a comprehensive assessment of the candidates, that is, a broad perspective on their separation, classification, and potential diagnosis capabilities. We particularly recommend to employ an OPLS-DA model, LDA model, error rates for PLS-DA, ROC curves (RF algorithm), and PCA, HCA, and OPLS-DA plots. Resampling and test validations as well as permutation testing should be used because they evaluate more accurately the generalization of the models to new samples. They also allow the calculation of confidence intervals, both features that are particularly important with low sample sizes. The order of importance of the candidates in the classification models (particularly PLS-DA, OPLS-DA, and RF) is a useful evaluation of their relative effectiveness. Adding the QC samples to the unsupervised PCA and HCA plots is a useful posterior control of the analytical quality of the data. To finish, it is important to stress that, regardless of their usefulness in prediction, all candidate biomarkers and, more largely, all significantly altered metabolites have a potential biological interest and should be kept, provided that they are reliably identified or at least annotated. Annotation of the Candidate Biomarkers. Of the 33 unique, unbiased candidates, 10 met the three defined mass criteria and 7 met two of those (Table 3; all metabolites are given in the Supporting Information, Section S21). Two of them, terephthalic acid and threonolactone, were not reported to have any biological function and were probable artifacts (see the Supporting Information, Section S22). Terephthalic acid is a bulk chemical used to produce polyethylene terephthalate (PET). It was likely introduced by a plastic material used in the analytical process, such as gloves, despite the subtraction of the process blank in the data preprocessing. Threonolactone is a biological artifact produced by the autoxidation of L-ascorbic

acid in methanol,¹¹⁰ the solvent used here for protein extraction. For this reason, it has been found to have poor chemical stability in urine.¹¹¹ By combining three kinds of information, putative (Metabolomics Standards Initiative (MSI) level 2) annotations^{112,113,23} were achieved for approximately half of the candidates. Such a typical fraction²⁸ is quite satisfactory given the complexity of the metabolome. On the other hand, it means that half of the significant metabolites are lost for any clinical or biological use.¹⁰⁹ Increasing the resolution from 5000 to 25,000 (mass error from 3–100 ppm to 1 ppm or less) and using linear retention indices improved the confidence of the annotation. It also compensated for the lack of specificity of the low-resolution mass spectra due to the derivatization masses (m/z 73 and 147). In addition, LRIs were confirmed¹¹⁴ to be particularly useful for sugars because of their unselective mass spectra and the absence of the molecular ion in electron ionization (EI). Using a pool of alkanes limited to C20 was found to be sufficient, given that the indices were extrapolated over C20 (linear temperature ramp). Importantly, any method willing to achieve non equivocal, level 1 identification, should compare the mass and retention information measured to the ones of the corresponding standards analyzed on the same platform under the same conditions.^{112,113,23} The literature review that included various biological matrices and various analytical platforms showed that, among the annotated candidates, several had previously been associated to and are therefore likely to play a role in CD, IBD, and, more largely, gastrointestinal inflammation processes (Supporting Information, Section S23). It also showed that there is little consensus about the candidate biomarkers, sometimes even for similar matrices studied with similar instrumentations. In agreement with Barding,⁵⁵ the candidates highlighted with NMR (and obviously GC-MS) belong to chemical classes observable with GC×GC. Therefore, while this is not the goal of this untargeted study, this instrumentation is capable of targeting and challenging most of the significant metabolites found in other studies with the exception of the more lipidic fractions observed with LC–MS. In addition, due to its increased sensitivity and resolution, it is able to detect more metabolites, that is, more potential candidates, including new ones.

Figure 2. Summary of the global data processing workflow from raw data to verified and annotated candidate biomarkers.

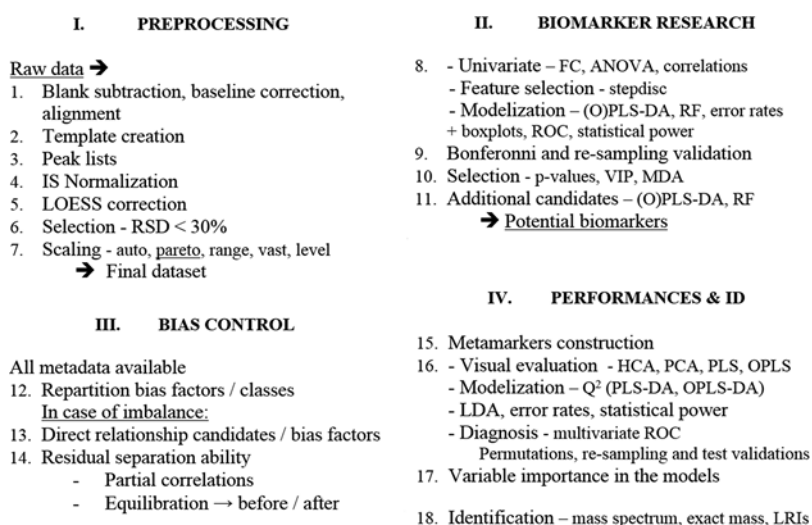


Table 2. Main Statistical Values for the Candidate Biomarkers Able to Discriminate the CD Samples from the Healthy Controls^a

candidate	normality	FC	Pearson/ Spearman	Kendall	Welch/ K-W	FR value	p	Stepdisc	NN abs diff.	SVM weight	PLS-DA VIP	RF, MDA	ROC AUC	stat. power
1	0.00	4.6	0.00	0.00	0.00	0.00		1	0.67	1.33	2.8	0.0151	0.87	0.93
2	0.00	1.6	0.00	0.00	0.00	0.00			0.43	0.73	2.3	0.0104	0.79	0.75
3	0.00	4.7	0.00	0.00	0.00	0.00		2	0.62	1.51	2.3	0.0072	0.78	0.74
4	0.00	1.6	0.00	0.00	0.00	0.00		7	0.60	2.65	2.2	0.0052	0.76	0.67
5	0.00	1.7	0.00	0.00	0.00	0.00			0.50	0.42	2.5	0.0060	0.81	0.84
6	0.00	2.3	0.00	0.00	0.00	0.00		17	0.56	3.42	2.1	0.0030	0.70	0.62
7	0.02	1.5	0.00	0.00	0.00	0.00		3	0.37	2.46	2.2	0.0045	0.75	0.06
8	0.00	2.8	0.00	0.00	0.00	0.00			0.11	0.86	2.1	0.0127	0.84	0.62
9	0.00	1.7	0.00	0.00	0.00	0.00			0.18	1.61	1.9	0.0040	0.73	0.54
10	0.00	1.6	0.00	0.00	0.00	0.00			0.44	0.52	2.1	0.0018	0.74	0.67
11	0.00	1.3	0.01	0.00	0.01	0.01			0.43	2.66	1.6	0.0052	0.75	0.39
12	0.00	1.4	0.01	0.00	0.01	0.01		8	0.76	3.47	1.7	0.0010	0.69	0.41
13	0.00	1.5	0.05	0.00	0.05	0.05			0.52	2.79	1.4	0.0042	0.70	0.26
14	0.00	1.3	0.05	0.00	0.05	0.05			0.44	2.83	1.3	0.0012	0.68	0.26
15	0.06	1.4	0.01	0.00	0.01	0.00			0.22	0.38	1.8	0.0011	0.69	0.03
16	0.00	1.3	0.03	0.01	0.03	0.03		5	0.61	3.95	1.9	0.0001	0.65	0.31
17	0.82	1.1	0.03	0.01	0.03	0.03			0.28	1.30	1.5	0.0000	0.66	0.30
18	0.03	1.2	0.02	0.01	0.02	0.02			0.48	1.17	1.5	0.0014	0.66	0.34
19	0.00	1.9	0.08	0.00	0.10	0.08			0.21	1.08	1.2	0.0020	0.75	0.21
20	0.00	1.3	0.30	0.00	0.29	0.30			0.19	1.83	0.9	0.0103	0.80	0.10
21	0.00	1.6	0.03	0.01	0.03	0.03			0.42	0.24	1.5	0.0003	0.66	0.32
22	0.00	1.3	0.01	0.01	0.01	0.01			0.25	2.39	1.7	0.0008	0.66	0.41

^a Univariate tests : FC, Pearson, Spearman, and Kendall correlations, Welch and Kruskal-Wallis (K-W) tests, FR, and statistical power. Multivariate methods: Stepdisc feature selection, NN (absolute distance between the groups), PLS-DA, and RF (mean decrease accuracy). Diagnosis: ROC (area under the curve, AUC). The values meeting the statistical thresholds set for significance are in bold.

Figure 3. HCA and PCA plots (six candidates used), OPLS-DA plot (12 candidates), and ROC curve (all 19 candidates, RF algorithm) for the separation between the CD samples (red) and the HC (green). In the ROC curve, the blue line pictures the mean value of the 100 cross validation classifications, the blue surface represents the corresponding confidence intervals at 95%, and the red line pictures the test classification.

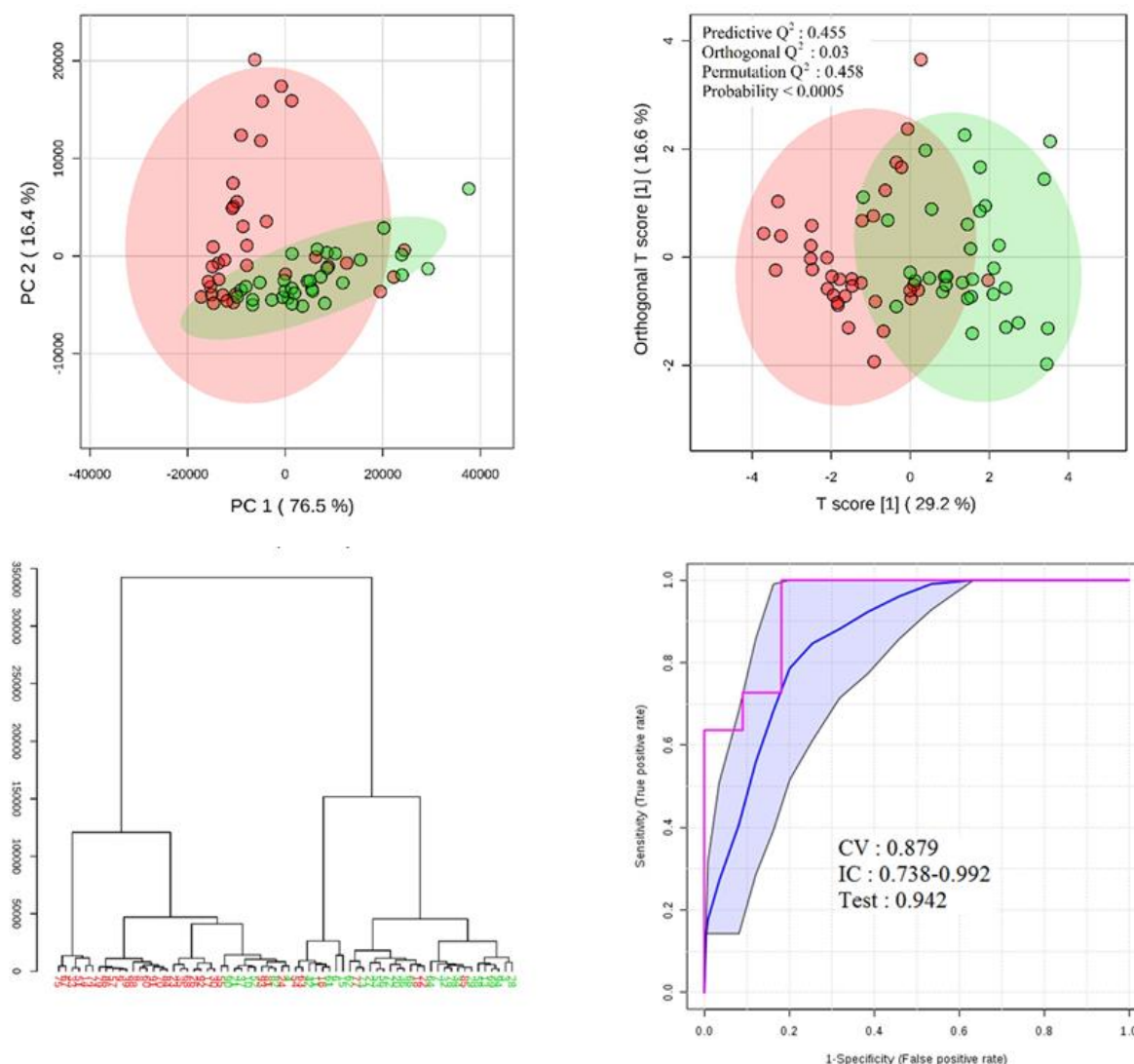


Figure 4. 3D PCA plot (top, six candidates used; PC1: 80.2%, PC2: 8.7%, and PC3: 5.5%) and PLS-DA plot (bottom, 22 candidates used; PC1: 30.5%, PC2: 19.8%, PC3: 9.3%) for the separation between the three CD groups. High-endoscopic activity samples are in red; low- endoscopic activity samples are in green and quiescent samples in blue.

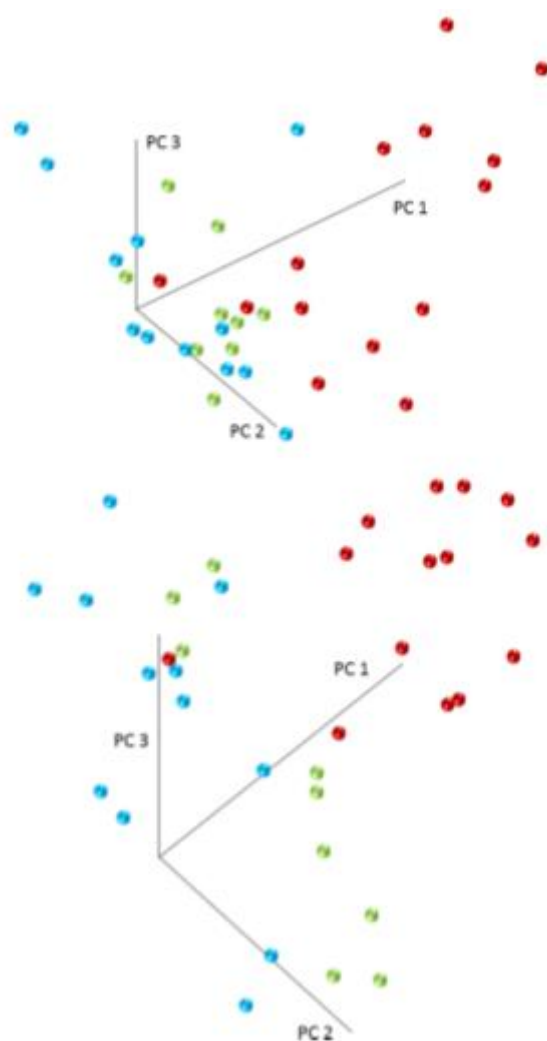


Table 3. Annotation of the Candidate Biomarkers^a

candidate	ID	match	probability	ΔRT	Δ (ppm)	criteria met
2	erythronic acid/threonic acid	902	49	57	0.1	2
7	aspartic acid	758	82	19	0.2	3
8	threonolactone*	808	57	40	0.6	2
9	glutamic acid	853	78	5	0	3
10	xylose	757	6	28	0.7	2
13	methionine	678	86	0	0.1	3
14	2-hydroxybutyric acid	846	27	1	0.1	3
17	1,5-anhydroglucitol	912	66	11	0.4	3
18	citric acid	872	95	12	0.2	3
19	galactose	739	11	10	34	2
23	erythrose/threose	572	12	24	0.4	2
26	capric acid	795	93	0	0.3	3
31	erythritol/threitol	823	23	16		2
33	myristic acid	879	96	14	0.1	3
34	glucose	899	15	7	0.6	3
36	lauric acid	806	90	8	0.1	3
37	terephthalic acid*	644	0	10	0.2	2

*Candidates highlighted in the CD vs HC separation (top) and candidates highlighted in the three CD subgroups separation (bottom). The values that meet the significance thresholds are in bold. *Threonolactone and terephthalic acid are probable artifacts.

CONCLUSIONS

This work showed that it is not straightforward to perform exploratory hypothesis-generating analysis on complex biological samples in a semiquantitative and reliable manner, particularly with a comprehensive analytical platform such as GC×GC–MS despite its power in terms of peak capacity. Indeed, if the global approach proposed was validated and successfully applied to separate Crohn's disease samples from negative, "healthy" controls, it required careful optimization, the control and articulation of the various steps, the use of multiple statistical treatments and the evaluation of the potential biases that were not compensated in the experimental design. Regarding the results obtained, this study is limited to a proof of concept. Therefore, the candidates highlighted, despite their analytical and statistical reliability, cannot be seen as true biomarkers. To achieve that, the first step would be to perform a targeted quantitative analysis of the samples. Then, assuming that the variations observed would be confirmed, these compounds should be validated on a second independent set of samples to see if they translate to a different group of patients.

ASSOCIATED CONTENT

SUPPORTING INFORMATION

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00535>.

Sample collection and clinical and analytical metadata for the Crohn's disease study (Section S1): clinical metadata for the samples of the proof of concept study (Table S1); security precautions for the sample preparation (Section S2); Box–Behnken design of experiment (Section S3); list of the metabolites monitored in the internal QC samples, intra- and inter-batch variations (Section S4); list of the 19 representative metabolites monitored in the internal QC samples (Table S2); list of the metabolites assessed in the NIST SRM 1950 (Section S5): extracts from the NIST SRM 1950 (metabolites in frozen human plasma) certificate (Figure S1); method optimization, sample preparation and separation and detection (Section S-6): DoE response surfaces for MSTFA and MeOX volumes and temperatures (Figure S2), sample preparation optimization (Table S3), examples of chromatograms produced by the three sets of columns (Figure S3), separation parameters for various flow rates and temperature ramps (Figure S4), chromatograms obtained at 3 and 5 °C/min (Figure S5), run times for the four temperature ramps evaluated (Table S4), effect of the initial temperature on the early eluting compounds (Figure S6), effect of the hot jet duration on the resolution in the overloaded zones of urea and glucose (Figure S7), influence of the injection temperature on the normalized peak volume (Table S5), impact of the injection volume on the chromatographic resolution in an area at risk of overloading (urea) (Figure S8), effect of the liner design on the global chromatogram (Figure S9), and summary of the optimized analytical conditions (Figure S10); accuracy and precision assessment in the NIST SRM 1950 (Section S7): accuracy and precision assessment in NIST SRM 1950 (Table S6); recovery assessment in the NIST SRM 1950 and the internal QC samples (Section S8): recovery assessment in NIST SRM 1950 and internal QC (Table S7); regression methods and determination of the best fit (Section S9): determination of the most efficient regression for succinic acid (Table S8), determination of the most efficient regression for fumaric acid (Table S9), determination of the most efficient regression for glycine (Table S10), residual plots for the three IS using unweighted and weighted regressions (Figure S11), and unweighted and weighted linear regression for succinic acid, fumaric acid, and glycine (Figure S12); sensitivity assessment, methods and calculation of LOD and LOQ (Section S10): LOD, LOQ, and sensitivity assessment using three different methods (Table S11); QC system, acceptance/rejection criteria and effect of the QC correction (Section S11): effect of the LOESS (partial) correction on the QC sample signals for two metabolites of the final dataset (Figure S13) and QC and study samples signals before and after the LOESS (partial) correction (Figure S14); data scaling (Section S12): HCA plots for the various scaling methods (Figure S15), PCA 3D plots for the various scaling methods (Figure S16), PCA 2D plots for level scaling with all samples and without the two alleged outliers (Figure S17), PLS-DA 3D plots for the various scaling methods (Figure S18), effect of scaling on random forests' (RF) ability to highlight the most significant metabolites (Table S12), PLS-DA Q² and R² for the various scaling methods (Table S13), and percentages of variation explained by the first two axes in PCA and PLS-DA

plots (Table S14); statistics for biomarker research, statistical treatment details and significance (Section S13); data control and selection process for the CD samples and the healthy controls (Section S14): OPLS-DA plots for the separation between CD and HC samples, constructed on all the 183 metabolites taken in the final dataset (Figure S19), global significance score obtained through the aggregation of the individual statistical significance values (Table S15), candidates meeting the significance threshold of 0.05 after Bonferroni correction for multiple testing ($0.05/183$) for four statistical tests: Pearson/Spearman, Kendall correlations, Welch/Kruskal-Wallis ANOVA, and FR (Table S16), additional biomarker research through PLS-DA, OPLS-DA, and RF (Table S17), ranking of the potential additional candidate biomarkers in the initial selection process (Table S18), OPLS-DA plots constructed on all 183 metabolites of the final dataset and on the 161 metabolites left after the selection of the most significant ones as candidate biomarkers (Figure S20), and examples of boxplots for some of the candidate biomarkers and for some of the potential additional ones (Figure S21); data control and selection process the three subgroups of CD samples (Section S15): sPLS-DA plots for the separation between CD subgroups samples (Figure S22), main statistical values for the candidate biomarkers selected to separate the three CD subgroups (Table S19), sPLS-DA plots constructed on the 183 metabolites of the final dataset and on the 154 metabolites left after selection of the most significant ones as candidate biomarkers (Figure S23), and examples of boxplots for some of the candidate biomarkers (Figure S24); bias control for the CD candidates, repartition of the factors between the biological classes, relationships between the candidate biomarkers and the bias factors, and residual separation ability (Section S16): repartition of the bias factors between the CD samples and the healthy controls given by the numbers of samples, percentages, means, and medians (Table S20), p values for the potential imbalances (Table S21), p values for the relationships between the candidate biomarkers and both anti-TNF α and immuno-suppression medications (Table S22), samples removed for the anti-TNF α equilibration and summary of the samples left (Table S23), samples removed for the immunosuppressor equilibration and summary of the samples left (Table S24), effect of the factors on the separation ability measured through the residual separation capability and on the capability variation (Table S25), and comparison of the separation ability before and after the equilibration for the candidates potentially affected (Table S26); bias control for the three CD subgroups candidates, repartition of the factors between the biological classes, relationships between the candidate biomarkers and the bias factors, and residual separation ability (Section S17): repartition of the bias factors considered between the three CD groups given by the numbers of samples, percentages, means and medians (Table S27), p values for the potential imbalances (Table S28), HCA and PCA plots constructed on all 183 metabolites and the selected candidates ($n = 29$) to see the influence of disease location (Figure S25), p values for the relationships between the candidate biomarkers and the four potential bias factors (Table S29), scatterplots of the candidates' detected signal against age with the corresponding linear regression (Figure S26), scatterplots of the candidates' detected signal against BMI with the corresponding linear regression (Figure S27), samples removed for the BMI equilibration, summary of the samples left, and p values for the repartition after the removal (Table S30), samples removed for the anti-TNF α equilibration, summary of the samples left, and p values for the repartition after the removal (Table S31), samples removed for the immunosuppressor equilibration, summary of the samples left, and p values for the repartition after the removal (Table S32), samples removed for the age equilibration, summary of the samples left, and p

values for the repartition after the removal (Table S33), effect of the factors on the separation ability measured through the residual separation capability (Table S34), and comparison of the separation ability (p values) before and after the equilibration for the candidates potentially affected (Table S35); data processing, metamarkers and testing of the potential biases between the training and test sets (Section S18): distribution of the samples and corresponding p values between training and test sets for the various potential bias factors (Table S36); separation between healthy controls and Crohn's disease samples, performances and selection order in the models (Section S19): discrimination performances achieved with the three different metamarkers using OPLS-DA, PLS-DA, LDA, sPLS-DA, SVM, and RF models (Table S37), ROC curves for an increasing number of candidates and multivariate statistical power for all the candidates at FDR 0.01 (Figure S28), HCA and PCA plots constructed on the most significant candidate biomarkers (Figure S29), PCA and OPLS plots constructed on the three different metamarkers (Figure S30), ranking or selection order of the highlighted candidate biomarkers in neural network, Bayes classifier, PLS-DA, OPLS-DA, SVM, and RF models (Table S38); separation between the three Crohn's disease groups, performances and selection order in the models (Section S20): discrimination performances achieved with the three different metamarkers using PLS-DA, LDA, sPLS-DA, and RF models (Table S39), HCA and PLS plots for the separation between the three CD groups, statistical power for the 15 most significant candidates at FDR 0.1 (Figure S31), HCA and PCA plots constructed on all the candidate biomarkers (Figure S32), and ranking or selection order of the highlighted candidate biomarkers in NN, Bayes classifier, PLS-DA, and RF models (Table S40); annotation of the candidate biomarkers (Section 21): annotation of the candidate biomarkers (Table S41); biological functions of the candidate biomarkers (Section S22): biological functions of the candidate biomarkers (Table S42); and the literature review (Section S23): known and unknown metabolites in IBD through the literature review (Table S43) and literature review for the selected and annotated potential biomarkers (Table S44).

AUTHOR INFORMATION

CORRESPONDING AUTHOR

*E-mail: jf.focant@uliege.be. Phone: +32 366 35 31. Fax: +32 4 366 43 87.

ORCID Nicolas Di Giovanni: 0000-0003-2768-6834

AUTHOR CONTRIBUTIONS

N.D.G. designed the experiments, conducted them, and wrote the manuscript under the supervision of J.-F.F. M.-A.M. and E.L. provided the samples, clinical metadata, and expertise and helped with manuscript writing. All authors have given approval to the final version of the manuscript.

FUNDING

This work was funded by the chemistry department of the University of Liege.

NOTES

The authors declare no competing financial interest.

The mass spectrometry data have been deposited to the HARVARD DATAVERSE repository: 10.7910/DVN/E95B05.

ACKNOWLEDGMENTS

We thank Jeol (Europe) BV for instrumental support, Restek for GC consumables, and P.-H. Stefanuto for lab support and advice during manuscript writing.

References

1. Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography. *Proc. Natl. Acad. Sci. U. S. A.* 1971, 68, 2374–2376.
2. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. Systematic Functional Analysis of the Yeast Genome. *Trends Biotechnol.* 1998, 16, 373–378.
3. Zhang, A.; Sun, H.; Wang, X. Serum Metabolomics as a Novel Diagnostic Approach for Disease: A Systematic Review. *Anal. Bioanal. Chem.* 2012, 404, 1239–1245.
4. Bedair, M.; Sumner, L. W. Current and Emerging Mass- Spectrometry Technologies for Metabolomics. *TrAC, Trends Anal. Chem.* 2008, 27, 238–250.
5. Li, X.; Xu, Z.; Lu, X.; Yang, X.; Yin, P.; Kong, H.; Yu, Y.; Xu, G. Comprehensive Two-Dimensional Gas Chromatography/Time-of- Flight Mass Spectrometry for Metabonomics: Biomarker Discovery for Diabetes Mellitus. *Anal. Chim. Acta* 2009, 633, 257–262.
6. Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. Comprehensive Two-Dimensional Gas Chromatog- raphy Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Anal. Chem.* 2006, 78, 2700–2709.
7. Kouremenos, K. A.; Pitt, J.; Marriott, P. J. Metabolic Profiling of Infant Urine Using Comprehensive Two-Dimensional Gas Chromatography: Application to the Diagnosis of Organic Acidurias and Biomarker Discovery. *J. Chromatogr. A* 2010, 1217, 104–111.
8. Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. Fisher Ratio Method Applied to Third-Order Separation Data to Identify Significant Chemical Components of Metabolite Extracts. *Anal. Chem.* 2006, 78, 5068–5075.
9. Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. Comprehensive Two-Dimensional Gas Chromatography–Time-of-Flight Mass Spectrometry (GC × GC- TOF) for High Resolution Metabolomics: Biomarker Discovery on Spleen Tissue Extracts of Obese NZO Compared to Lean C57BL/6 Mice. *Metabolomics* 2005, 1, 65–73.
10. Almstetter, M. F.; Oefner, P. J.; Dettmer, K. Comprehensive Two-Dimensional Gas Chromatography in Metabolomics. *Anal. Bioanal. Chem.* 2012, 402, 1993–2013.

11. Hyötyläinen, T. Analytical Methodologies Utilized in the Search for Chronic Disease Biomarkers. *Bioanalysis* 2010, 2, 919–923.
12. Winnike, J. H.; Wei, X.; Knagge, K. J.; Colman, S. D.; Gregory, S. G.; Zhang, X. Comparison of GC-MS and GC×GC-MS in the Analysis of Human Serum Samples for Biomarker Discovery. *J. Proteome Res.* 2015, 14, 1810–1817.
13. Koek, M. M.; Muilwijk, B.; van Stee, L. L. P.; Hankemeier, T. Higher Mass Loadability in Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry for Improved Analytical Performance in Metabolomics Analysis. *J. Chromatogr. A* 2008, 1186, 420–429.
14. Lebedev, A. T.; Polyakova, O. V.; Mazur, D. M.; Artaev, V. B. The Benefits of High Resolution Mass Spectrometry in Environmental Analysis. *Analyst* 2013, 138, 6946–6953.
15. 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, 1075–1078.
16. Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D. The Importance of Experimental Design and QC Samples in Large-Scale and MS-Driven Untargeted Metabolomic Studies of Humans. *Bioanalysis* 2012, 4, 2249–2264.
17. Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* 2018, 14, 72.
18. Gullberg, J.; Jonsson, P.; Nordström, A.; Sjöström, M.; Moritz, T. Design of Experiments: An Efficient Strategy to Identify Factors Influencing Extraction and Derivatization of Arabidopsis Thaliana Samples in Metabolomic Studies with Gas Chromatography/Mass Spectrometry. *Anal. Biochem.* 2004, 331, 283–295.
19. Banerjee, K.; Patil, S. H.; Dasgupta, S.; Oulkar, D. P.; Patil, S. B.; Savant, R.; Adsule, P. G. Optimization of Separation and Detection Conditions for the Multiresidue Analysis of Pesticides in Grapes by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* 2008, 1190, 350–357.
20. Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. *BMC Bioinf.* 2007, 8, 105.
21. van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. *BMC Genomics* 2006, 7, 142.
22. Madsen, R.; Lundstedt, T.; Trygg, J. Chemometrics in Metabolomics-A Review in Human Disease Diagnosis. *Anal. Chim. Acta* 2010, 659, 23–33.
23. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-Mcintyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; et al. Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* 2011, 6, 1060–1083.
24. Bro, R. PARAFAC. Tutorial and Applications. *Chemom. Intell. Lab. Syst.* 1997, 38, 149–171.
25. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0-Making Metabolomics More Meaningful. *Nucleic Acids Res.* 2015, 43, W251–W257.

26. Wei, X.; Shi, X.; Koo, I.; Kim, S.; Schmidt, R. H.; Arteel, G. E.; Watson, W. H.; McClain, C.; Zhang, X. MetPP: A Computational Platform for Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry-Based Metabolomics. *Bioinformatics* 2013, 29, 1786–1792.
27. Christou, C.; Gika, H. G.; Raikos, N.; Theodoridis, G. GC-MS Analysis of Organic Acids in Human Urine in Clinical Settings: A Study of Derivatization and Other Analytical Parameters. *J. Chromatogr., B* 2014, 964, 195–201.
28. Dunn, W. B.; Erban, A.; Weber, R. J. M.; Creek, D. J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; et al. Mass Appeal: Metabolite Identification in Mass Spectrometry-Focused Untargeted Metabolomics. *Metabolomics* 2013, 9, 44–66.
29. Fang, M.; Ivanisevic, J.; Benton, H. P.; Johnson, C. H.; Patti, G. J.; Hoang, L. T.; Uritboonthai, W.; Kurczy, M. E.; Siuzdak, G. Thermal Degradation of Small Molecules: A Global Metabolomic Investigation. *Anal. Chem.* 2015, 87, 10935–10941.
30. de Castro, M. D. L.; Izquierdo, A. Lyophilization: A Useful Approach to the Automation of Analytical Processes? *J. Autom. Chem.* 1990, 12, 267–279.
31. Fiehn, O. Metabolomics: Prospects and Pitfalls. Answers to Common Questions http://files.pharmtech.com/alfresco_images/pharma/2017/04/04/9b3abb1d-a43b-45e8-a45a-6c9694a15ba9/LEC0315_ebook.pdf (accessed Jul 18, 2019).
32. Kalla, R.; Ventham, N. T.; Satsangi, J.; Arnott, I. D. R. Crohn's Disease. *Br. Med. J.* 2014, 349, g6670.
33. Molodecky, N. A.; Soon, I. S.; Rabi, D. M.; Ghali, W. A.; Ferris, M.; Chernoff, G.; Benchimol, E. I.; Panaccione, R.; Ghosh, S.; Barkema, H. W.; et al. Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology* 2012, 142, 46–54.e42.
34. Louis, E.; Baumgart, D. C.; Ghosh, S.; Gomollon, F.; Hanauer, S.; Hart, A.; Irving, P. What Changes in Inflammatory Bowel Disease Management Can Be Implemented Today? *J. Crohn's Colitis* 2012, 6, S260–S267.
35. Bjerrum, J. T.; Wang, Y.; Hao, F.; Coskun, M.; Ludwig, C.; Günther, U.; Nielsen, O. H. Metabonomics of Human Fecal Extracts Characterize Ulcerative Colitis, Crohn's Disease and Healthy Individuals. *Metabolomics* 2015, 11, 122–133.
36. Stephens, N. S.; Siffledeen, J.; Su, X.; Murdoch, T. B.; Fedorak, R. N.; Slupsky, C. M. Urinary NMR Metabolomic Profiles Discriminate Inflammatory Bowel Disease from Healthy. *J. Crohn's Colitis* 2013, 7, e42–e48.
37. Schicho, R.; Shaykhtudinov, R.; Ngo, J.; Nazyrova, A.; Schneider, C.; Panaccione, R.; Kaplan, G. G.; Vogel, H. J.; Storr, M. Quantitative Metabolomic Profiling of Serum, Plasma, and Urine by ¹H NMR Spectroscopy Discriminates between Patients with Inflammatory Bowel Disease and Healthy Individuals. *J. Proteome Res.* 2012, 11, 3344–3357.
38. Marchesi, J. R.; Holmes, E.; Khan, F.; Kochhar, S.; Scanlan, P.; Shanahan, F.; Wilson, I. D.; Wang, Y. Rapid and Noninvasive Metabonomic Characterization of Inflammatory Bowel Disease. *J. Proteome Res.* 2007, 6, 546–551.
39. Williams, H. R. T.; Willsmore, J. D.; Cox, I. J.; Walker, D. G.; Cobbold, J. F. L.; Taylor-Robinson, S. D.; Orchard, T. R. Serum Metabolic Profiling in Inflammatory Bowel Disease. *Dig. Dis. Sci.* 2012, 57, 2157–2165.

40. Williams, H. R. T.; Cox, I. J.; Walker, D. G.; North, B. V.; Patel, V. M.; Marshall, S. E.; Jewell, D. P.; Ghosh, S.; Thomas, H. J. W.; Teare, J. P.; et al. Characterization of Inflammatory Bowel Disease with Urinary Metabolic Profiling. *Am. J. Gastroenterol.* 2009, 104, 1435–1444.
41. Jansson, J.; Willing, B.; Lucio, M.; Fekete, A.; Dicksved, J.; Halfvarson, J.; Tysk, C.; Schmitt-Kopplin, P. Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease. *PLoS One* 2009, 4, e6386.
42. Murdoch, T. B.; Fu, H.; MacFarlane, S.; Sydora, B. C.; Fedorak, R. N.; Slupsky, C. M. Urinary Metabolic Profiles of Inflammatory Bowel Disease in Interleukin-10 Gene-Deficient Mice. *Anal. Chem.* 2008, 80, 5524–5531.
43. Lin, H.-M.; Edmunds, S. J.; Helsby, N. A.; Ferguson, L. R.; Rowan, D. D. Nontargeted Urinary Metabolite Profiling of a Mouse Model of Crohn's Disease Research Articles. *J. Proteome Res.* 2009, 2045–2057.
44. Schicho, R.; Nazyrova, A.; Shaykhutdinov, R.; Duggan, G.; Vogel, H. J.; Storr, M. Quantitative Metabolomic Profiling of Serum and Urine in DSS-Induced Ulcerative Colitis of Mice by ¹H NMR Spectroscopy. *J. Proteome Res.* 2010, 9, 6265–6273.
45. Chen, C.; Shah, Y. M.; Morimura, K.; Krausz, K. W.; Miyazaki, M.; Richardson, T. A.; Morgan, E. T.; Ntambi, J. M.; Idle, J. R.; Gonzalez, F. J. Metabolomics Reveals That Hepatic Stearoyl-CoA Desaturase 1 Downregulation Exacerbates Inflammation and Acute Colitis. *Cell Metab.* 2008, 7, 135–147.
46. Le Gall, G.; Noor, S. O.; Ridgway, K.; Scovell, L.; Jamieson, C.; Johnson, I. T.; Colquhoun, I. J.; Kemsley, E. K.; Narbad, A. Metabolomics of Fecal Extracts Detects Altered Metabolic Activity of Gut Microbiota in Ulcerative Colitis and Irritable Bowel Syndrome. *J. Proteome Res.* 2011, 10, 4208–4218.
47. Minamoto, Y.; Otoni, C. C.; Steelman, S. M.; Büyükleblebici, O.; Steiner, J. M.; Jergens, A. E.; Suchodolski, J. S. Alteration of the Fecal Microbiota and Serum Metabolite Profiles in Dogs with Idiopathic Inflammatory Bowel Disease. *Gut Microbes* 2015, 6, 33–47.
48. Ponnusamy, K.; Choi, J. N.; Kim, J.; Lee, S.-Y.; Lee, C. H. Microbial Community and Metabolomic Comparison of Irritable Bowel Syndrome Faeces. *J. Med. Microbiol.* 2011, 60, 817–827.
49. Kolho, K.-L.; Pessia, A.; Jaakkola, T.; de Vos, W. M.; Velagapudi, V. Faecal and Serum Metabolomics in Paediatric Inflammatory Bowel Disease. *J. Crohn's Colitis* 2016, 11, 321–334.
50. Dawiskiba, T.; Deja, S.; Mulak, A.; Zabek, A.; Jawien, E.; Pawełka, D.; Banasik, M.; Mastalerz-Migas, A.; Balcerzak, W.; Kaliszewski, K.; et al. Serum and Urine Metabolomic Fingerprinting in Diagnostics of Inflammatory Bowel Diseases. *World J. Gastroenterol.* 2014, 20, 163–174.
51. Balasubramanian, K.; Kumar, S.; Singh, R. R.; Sharma, U.; Ahuja, V.; Makharia, G. K.; Jagannathan, N. R. Metabolism of the Colonic Mucosa in Patients with Inflammatory Bowel Diseases: An in Vitro Proton Magnetic Resonance Spectroscopy Study. *Magn. Reson. Imaging* 2009, 27, 79–86.
52. Varma, S.; Bird, R.; Eskin, M.; Dolenko, B.; Raju, J.; Bezabeh, T. Detection of Inflammatory Bowel Disease by Proton Magnetic Resonance Spectroscopy (¹H MRS) Using an Animal Model. *J. Inflammation* 2007, 4, 24.
53. Sharma, U.; Singh, R. R.; Ahuja, V.; Makharia, G. K.; Jagannathan, N. R. Similarity in the Metabolic Profile in Macroscopically Involved and Un-Involved Colonic Mucosa in Patients with Inflammatory Bowel Disease: An in Vitro Proton (¹H) MR Spectroscopy Study. *Magn. Reson. Imaging* 2010, 28, 1022–1029.

54. Ooi, M.; Nishiumi, S.; Yoshie, T.; Shiomi, Y.; Kohashi, M.; Fukunaga, K.; Nakamura, S.; Matsumoto, T.; Hatano, N.; Shinohara, M.; et al. GC/MS-Based Profiling of Amino Acids and TCA Cycle- Related Molecules in Ulcerative Colitis. *Inflammation Res.* 2011, 60, 831–840.
55. Barding, G. A., Jr.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K. Comparison of GC-MS and NMR for Metabolite Profiling of Rice Subjected to Submergence Stress. *J. Proteome Res.* 2013, 12, 898–909.
56. Bhinderwala, F.; Wase, N.; DiRusso, C.; Powers, R. Combining Mass Spectrometry and NMR Improves Metabolite Detection and Annotation. *J. Proteome Res.* 2018, 17, 4017–4022.
57. Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; et al. The Human Serum Metabolome. *PLoS One* 2011, 6, e16957.
58. Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. LC-MS-Based Metabolomics. *Mol. BioSyst.* 2012, 8, 470–481.
59. Murgia, A.; Hinz, C.; Liggi, S.; Denes, J.; Hall, Z.; West, J.; Santoru, M. L.; Piras, C.; Manis, C.; Usai, P.; et al. Italian Cohort of Patients Affected by Inflammatory Bowel Disease Is Characterised by Variation in Glycerophospholipid, Free Fatty Acids and Amino Acid Levels. *Metabolomics* 2018, 14, 140.
60. Alghamdi, A.; Gerasimidis, K.; Blackburn, G.; Akinci, D.; Edwards, C.; Russell, R. K.; Watson, D. G. Untargeted Metabolomics of Extracts from Faecal Samples Demonstrates Distinct Differences between Paediatric Crohn's Disease Patients and Healthy Controls but No Significant Changes Resulting from Exclusive Enteral Nutrition Treatment. *Metabolites* 2018, 8, 82.
61. Baur, P.; Martin, F.-P.; Gruber, L.; Bosco, N.; Brahmabhatt, V.; Collino, S.; Guy, P.; Montoliu, I.; Rozman, J.; Klingenspor, M.; et al. Metabolic Phenotyping of the Crohn's Disease-like IBD Etiopathology in the TNFΔARE/WT Mouse Model. *J. Proteome Res.* 2011, 10, 5523–5535.
62. Daniluk, U.; Daniluk, J.; Kucharski, R.; Kowalczyk, T.; Pietrowska, K.; Samczuk, P.; Filimoniuk, A.; Kretowski, A.; Lebensztejn, D.; Ciborowski, M. Untargeted Metabolomics and Inflammatory Markers Profiling in Children With Crohn's Disease and Ulcerative Colitis A Preliminary Study. *Inflammatory Bowel Dis.* 2019, 25, 1120–1128.
63. Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma. *Anal. Chem.* 2006, 78, 4307–4318.
64. Phinney, K. W.; Ballihaut, G.; Bedner, M.; Benford, B. S.; Camara, J. E.; Christopher, S. J.; Davis, W. C.; Dodder, N. G.; Eppe, G.; Lang, B. E.; et al. Development of a Standard Reference Material for Metabolomics Research. *Anal. Chem.* 2013, 85, 11732–11738.
65. Nishiumi, S.; Kobayashi, T.; Ikeda, A.; Yoshie, T.; Kibi, M.; Izumi, Y.; Okuno, T.; Hayashi, N.; Kawano, S.; Takenawa, T.; et al. A Novel Serum Metabolomics-Based Diagnostic Approach for Colorectal Cancer. *PLoS One* 2012, 7, No. e40459.
66. Villas-Bôas, S. G.; Smart, K. F.; Sivakumaran, S.; Lane, G. A. Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? *Metabolites* 2011, 1, 3–20.
67. Leardi, R. Experimental Design in Chemistry: A Tutorial. *Anal. Chim. Acta* 2009, 652, 161–172.
- Mostafa, A.; Edwards, M.; Gorecki, T. Optimization Aspects of Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* 2012, 1255, 38–55.
68. Kowalski, J.; Misselwitz, M.; Cochran, J. Investigating GC×GC Separations Using Selective Column Chemistry and Compound Derivatization Pairings for Common Metabolomics Chemical Com-

- pounds. In 4th Multidimensional Chromatography Workshop; Restek Corporation: Toronto, Canada, 2013, 36.
69. Reichenbach, S. E.; Tian, X.; Cordero, C.; Tao, Q. Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* 2012, 1226, 140–148.
 70. Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B., Jr.; Wu, Z.; Fiehn, O. Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GC×GC–HRMS). *Talanta* 2011, 83, 1279–1288.
 71. Noctor, G.; Bergot, G.; Mauve, C.; Thominet, D.; Lelarge-Trouverie, C.; Prioul, J.-L. A Comparative Study of Amino Acid Measurement in Leaf Extracts by Gas Chromatography-Time of Flight- Mass Spectrometry and High Performance Liquid Chromatography with Fluorescence Detection. *Metabolomics* 2007, 3, 161–174.
 72. Fiehn, O. Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr. Protoc. Mol. Biol.* 2016, 114, 30–34.
 73. Hendriks, M. M. W. B.; van Eeuwijk, F. A.; Jellema, R. H.; Westerhuis, J. A.; Reijmers, T. H.; Hoefsloot, H. C. J.; Smilde, A. K. Data-Processing Strategies for Metabolomics Studies. *TrAC, Trends Anal. Chem.* 2011, 30, 1685–1698.
 74. Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O'Hagan, S.; Knowles, J. D.; Halsall, A.; Wilson, I. D.; et al. Development of a Robust and Repeatable UPLC– MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.* 2009, 81, 1357–1364.
 75. Xia, J.; Wishart, D. S. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr. Protoc. Mol. Biol.* 2016, 55, 14–10.
 76. Rakotomalala, R. TANAGRA : Une Plate-Forme d ' Experimentation Pour La Fouille de Donnees. *Rev. Modul.* 2005, 70– 85.
 77. Mierswa, I.; Klinkenberg, R.; Fischer, S.; Ritthoff, O. A Flexible Platform for Knowledge Discovery Experiments: YALE - Yet Another Learning Environment; University of Dortmund: Germany, 2003.
 78. R Core Team. R: A Language and Environment for Statistical Computing. <https://www.r-project.org/> (accessed Jul 18, 2019).
 79. Castillo, S.; Mattila, I.; Miettinen, J.; Oresic, M.; Hyötylainen, T. Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Anal. Chem.* 2011, 83, 3058–3067.
 80. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. HMDB 3.0– The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013, 41, D801–D807.
 81. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 2017, 45, D353–D361.
 82. Bean, H. D.; Hill, J. E.; Dimandja, J.-M. D. Improving the Quality of Biomarker Candidates in Untargeted Metabolomics via Peak Table-
 83. Based Alignment of Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data. *J. Chromatogr. A* 2015, 1394, 111–117.

84. Bekele, E. A.; Annaratone, C. E. P.; Hertog, M. L. A. T. M.; Nicolai, B. M.; Geeraerd, A. H. Multi-Response Optimization of the Extraction and Derivatization Protocol of Selected Polar Metabolites from Apple Fruit Tissue for GC–MS Analysis. *Anal. Chim. Acta* 2014, 824, 42–56.
85. Schummer, C.; Delhomme, O.; Appenzeller, B. M. R.; Wennig, R.; Millet, M. Comparison of MTBSTFA and BSTFA in Derivatization Reactions of Polar Compounds Prior to GC/MS Analysis. *Talanta* 2009, 77, 1473–1482.
86. Food and Drug Administration; Centre for Drug Validation and Research (CDER). Bioanalytical Method Validation. Guidance for Industry. <https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf> (accessed Jul 18, 2019).
87. National Institute of Standards & Technology. Certificate of Analysis. Standard Reference Material 1950. Metabolites in Frozen Human Plasma. <https://www-s.nist.gov/srmors/certificates/1950.pdf> (accessed Jul 18, 2019).
88. Jiye, A.; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklund, S. L.; Moritz, T. Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. *Anal. Chem.* 2005, 77, 8086–8094.
89. Koek, M. M.; Muilwijk, B.; Van Der Werf, M. J.; Hankemeier, T. Microbial Metabolomics with Gas Chromatography/Mass Spectrometry. *Anal. Chem.* 2006, 78, 1272–1281.
90. Kiser, M. M.; Dolan, J. W. Selecting the Best Curve Fit. *LCGC North Am.* 2004, 22, 112–117.
91. Stone, D.; Ellis, J. Calibration and Linear Regression Analysis: A Self-Guided Tutorial (Part 2). <https://sites.chem.utoronto.ca/chemistry/coursenotes/analsci/LinRegr2a.pdf> (accessed Jul 18, 2019).
92. Gustavo Gonzalez, A.; Ángeles Herrador, M. A Practical Guide to Analytical Method Validation, Including Measurement Uncertainty and Accuracy Profiles. *TrAC, Trends Anal. Chem.* 2007, 26, 227–238.
93. Ducauze, C.; Baillet-Guffroy, A.; Bui, T. X. Choix et Validation d'une Methode d'Analyse http://www2.agroparistech.fr/IMG/pdf/Choix_et_Validation_Methode-2.pdf (accessed Jul 18, 2019).
94. Taverniers, I.; De Loose, M.; Van Bockstaele, E. Trends in Quality in the Analytical Laboratory. II. Analytical Method Validation and Quality Assurance. *TrAC, Trends Anal. Chem.* 2004, 23, 535–552.
95. Bressanello, D.; Liberto, E.; Collino, M.; Reichenbach, S. E.; Benetti, E.; Chiazza, F.; Bicchi, C.; Cordero, C. Urinary Metabolic Fingerprinting of Mice with Diet-Induced Metabolic Derangements by Parallel Dual Secondary Column-Dual Detection Two-Dimensional Comprehensive Gas Chromatography. *J. Chromatogr. A* 2014, 1361, 265–276.
96. Redestig, H.; Fukushima, A.; Stenlund, H.; Moritz, T.; Arita, M.; Saito, K.; Kusano, M. Compensation for Systematic Cross-Contribution Improves Normalization of Mass Spectrometry Based Metabolomics Data. *Anal. Chem.* 2009, 81, 7974–7980.
97. Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Oresič, M. Normalization Method for Metabolomics Data Using Optimal Selection of Multiple Internal Standards. *BMC Bioinf.* 2007, 8, 93.
98. Shewhart, W. A. Quality Control Charts. *Bell Syst. Tech. J.* 1926, 5, 593–603.

99. Almstetter, M. F.; Appel, I. J.; Gruber, M. A.; Lottaz, C.; Timischl, B.; Spang, R.; Dettmer, K.; Oefner, P. J. Integrative Normalization and Comparative Analysis for Metabolic Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography - Time-of-Flight Mass Spectrometry. *Anal. Chem.* 2009, 81, 5731–5739.
100. Gromski, P. S.; Xu, Y.; Hollywood, K. A.; Turner, M. L.; Goodacre, R. The Influence of Scaling Metabolomics Data on Model Classification Accuracy. *Metabolomics* 2015, 11, 684–695.
101. Enot, D. P.; Lin, W.; Beckmann, M.; Parker, D.; Overy, D. P.; Draper, J. Preprocessing, Classification Modeling and Feature Selection Using Flow Injection Electrospray Mass Spectrometry Metabolite Fingerprint Data. *Nat. Protoc.* 2008, 3, 446–470.
102. Becker, L. A. Effect Size Calculators. <https://www.uccs.edu/lbecker/> (accessed Jul 18, 2019).
103. Ellis, P. D. Effect Size Calculators. <https://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html> (accessed Jul 18, 2019).
104. Broadhurst, D. I.; Kell, D. B. Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments. *Metabolomics* 2006, 2, 171–196.
105. Dunn, W. B.; Lin, W.; Broadhurst, D.; Begley, P.; Brown, M.; Zelena, E.; Vaughan, A. A.; Halsall, A.; Harding, N.; Knowles, J. D.; et al. Molecular Phenotyping of a UK Population: Defining the Human Serum Metabolome. *Metabolomics* 2015, 11, 9–26.
106. Tu, J. V. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *J. Chronic Dis.* 1996, 49, 1225–1231.
107. Ransohoff, D. F. Rules of Evidence for Cancer Molecular- Marker Discovery and Validation. *Nat. Rev. Cancer* 2004, 4, 309–314.
108. Rakotomalala, R. Analyse Discriminante Lineaire. http://eric.univ-lyon2.fr/~ricco/cours/slides/analyse_discriminante.pdf. (accessed Jul 18, 2019).
109. Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational Biomarker Discovery in Clinical Metabolomics: An Introductory Tutorial. *Metabolomics* 2013, 9, 280–299.
110. Miyake, N.; Otsuke, Y.; Kurata, T. Autoxidation Reaction Mechanism for L-Ascorbic Acid in Methanol without Metal Ion Catalysis. *Biosci., Biotechnol., Biochem.* 1997, 61, 2069–2075.
111. Roux, A.; Thevenot, E. A.; 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, 11, 1095–1105.
112. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed Minimum Reporting Standards for Chemical Analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007, 3, 211–221.
113. Salek, R. M.; Steinbeck, C.; Viant, M. R.; Goodacre, R.; Dunn, W. B. The Role of Reporting Standards for Metabolite Annotation and Identification in Metabolomic Studies. *Gigascience* 2013, 2, 13.
114. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M. Chemical Derivatization and Mass Spectral Libraries in Metabolic Profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* 2005, 56, 219–243.