# Untargeted Serum Metabolic Profiling by GC×GC-HRTOF-MS

**Supporting Information**

Nicolas Di Giovanni[†], Marie-Alice Meuwis[‡], Edouard Louis[‡], Jean-François Focant[*†]

Affiliation

[†]University of Liège, Department of Chemistry, Organic & Biological Analytical Chemistry Group, Quartier Agora, Allée du 6 août, B6c, B-4000 Liège (Sart Tilman), Belgium

[‡] University of Liège, GIGA institute, Translational gastroenterology, Uliège, and CHU de Liège, Hepato-gastroenterology and digestive Oncology, Quartier Hôpital, Avenue de l'hôpital 13, B34-35, B-4000 Liège, Belgium.

[*]To whom correspondence should be addressed:

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

CONTENTS

- Table S-16. Candidates meeting the significance threshold of 0.05 after Bonferroni correction for multiple testing (0.05/183) for four statistical tests: Pearson/Spearman and Kendall correlations, Welch/Kruskal-Wallis ANOVA and FR.
- Table S-17. Additional biomarker research through PLS-DA, OPLS-DA and RF.
- Table S-18. Ranking of the potential additional candidate biomarkers in the initial selection process.
- Figure S-20. 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 S-21. Examples of boxplots for some of the candidate biomarkers and for some of the potential additional ones.

S-15. Data control and selection process the three subgroups of CD samples:

- Figure S-22. sPLS-DA plots for the separation between CD subgroups samples.
- Table S-19. Main statistical values for the candidate biomarkers selected to separate the three CD subgroups.
- Figure S-23. sPLS-DA plots constructed on the183 metabolites of the final dataset and on the 154 metabolites left after selection of the most significant ones as candidate biomarkers.
- Figure S-24. Examples of boxplots for some of the candidate biomarkers.

S-16. Bias control for the CD candidates. A. Repartition of the factors between the biological classes. B. Relationships between the candidate biomarkers and the bias factors. C. Residual separation ability:

- Table S-20. Repartition of the bias factors between the CD samples and the healthy controls given by the numbers of samples, the percentages, means and medians.
- Table S-21. p-values for the potential imbalances.
- Table S-22. p-values for the relationships between the candidate biomarkers and both anti-TNFα and immunosuppression medications.
- Table S-23. Samples removed for the anti-TNFα equilibration and summary of the samples left.

- Table S-24. Samples removed for the immunosuppressor equilibration and summary of the samples left.
- Table S-25. Effect of the factors on the separation ability measured through the residual separation capability and on the capability variation.
- Table S-26. Comparison of the separation ability before and after the equilibration, for the candidates potentially affected.

S-17. Bias control for the three CD subgroups candidates. A. Repartition of the factors between the biological classes. B. Relationships between the candidate biomarkers and the bias factors. C. Residual separation ability:

- Table S-27. Repartition of the bias factors considered between the three CD groups given by the numbers of samples, the percentages, the means and the medians.
- Table S-28. p-values for the potential imbalances.
- Figure S-25. HCA and PCA plots constructed on all 183 metabolites and the selected candidates (n=29) to see the influence of disease location.
- Table S-29. p-values for the relationships between the candidate biomarkers and the four potential bias factors.
- Figure S-26. Scatterplots of the candidates' detected signal against age with the corresponding linear regression.
- Figure S-27. Scatterplots of the candidates' detected signal against BMI with the corresponding linear regression.
- Table S-30. Samples removed for the BMI equilibration, summary of the samples left and p-values for the repartition after the removal.
- Table S-31. Samples removed for the anti-TNFα equilibration, summary of the samples left and p-values for the repartition after the removal.
- Table S-32. Samples removed for the immunosuppressor equilibration, summary of the samples left and p-values for the repartition after the removal.

- Table S-33. Samples removed for the age equilibration, summary of the samples left and p-values for the repartition after the removal.
- Table S-34. Effect of the factors on the separation ability measured through the residual separation capability.
- Table S-35. Comparison of the separation ability (p-values) before and after the equilibration, for the candidates potentially affected.

S-18. Data Processing. A. Meta-markers. B. Testing of the potential biases between the training and test sets:

- Table S-36. Distribution of the samples and corresponding p-values between training and test sets for the various potential bias factors.

S-19. Separation between healthy controls and Crohn's disease samples. A. Performances. B. Selection order in the models:

- Table S-37. Discrimination performances achieved with the three different meta-markers using OPLS-DA, PLS-DA, LDA, sPLS-DA, SVM and RF models.
- Figure S-28. ROC curves for an increasing number of candidates. Multivariate statistical power for all the candidates at FDR 0.01.
- Figure S-29. HCA and PCA plots constructed on the most significant candidate biomarkers.
- Figure S-30. PCA and OPLS plots constructed on the three different.
- Table S-38. Ranking or selection order of the highlighted candidate biomarkers in neural network, Bayes classifier, PLS-DA, OPLS-DA, SVM and RF models.

S-20. Separation between the three Crohn's disease groups. A. Performances. B. Selection order in the models:

- Table S-39. Discrimination performances achieved with the three different meta-markers using PLS-DA, LDA, sPLS-DA and RF models.
- Figure S-31. 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 S-32. HCA and PCA plots constructed on all the candidate biomarkers.

- Table S-40. Ranking or selection order of the highlighted candidate biomarkers in NN, Bayes classifier, PLS-DA and RF models.

S-21. Annotation of the candidate biomarkers.

- Table S-41. Annotation of the candidate biomarkers.

S-22. Biological functions of the candidate biomarkers.

- Table S-42. Biological functions of the candidate biomarkers.

S-23. Literature review.

- Table S-43. Known and unknown metabolites in IBD through literature review.
- Table S-44. Literature review for the selected and annotated potential biomarkers.

**S-1. Samples.**

(A) Sample collection

The applied conditions, originally developed from proteomic studies but close to general biobanking SOPs, are quite strict concerning the collection, processing and conservation of samples. They meet the protocols reviewed in a white paper from the Metabolomics Society Initiative[S-1] as well as their guidelines concerning blood derived products. In brief, Venosafe (Terumo, NV, EU) 10 mL dry tubes were used for serum collection. Clotting time was set at 30 minutes at room temperature before a centrifugation step at 2,000 RCF for 10 minutes at 4 °C. Aliquoting of sera was performed in polypropylene cryotubes. The delay between sample intake and storage (-80 °C) was kept inferior to 2 hours. All patients recruited for the proof of concept study were controlled for IBD or any other known digestive cancer by an endoscopic examination of the bowel (rectum, colon and, when possible, last part of the ileum) performed by a trained endoscopist. When applicable, the diagnosis was confirmed by an anatomopathological examination of the biopsies or specimen resected.

(B) Clinical and analytical metadata for the Crohn's disease study

| | Healthy Controls | Low EA | High EA | Quiescent EA |
|---|---|---|---|---|
| Total Number | 33 | 14 | 12 | 9 |
| Male/Female (n/n) | 14/19 | 5/9 | 6/6 | 3/6 |
| Mean age at diagnosis (range) in years | 44 (26-70) | 39 (28-54) | 38 (22-71) | 51 (37-65) |
| Mean BMI (range) in kg/m$^2$ | 25 (16-37) | 26 (22-42) | 22 (15-26) | 27 (24-39) |
| Current smoker (n) | 6 | 6 | 5 | 2 |
| Current alcohol consummers (n) | 7 | 3 | 1 | 4 |
| Medication for gastro-oesophage reflux | 9 | 4 | 4 | 2 |
| L-Thyroxin | 6 | 3 | 1 | 3 |
| CD specific treatments anti-Anti-TNFα | 0 | 10 | 6 | 2 |
| Immunosuppressor | 0 | 0 | 4 | 1 |
| Mean drying batch | 4.5 | 4.3 | 5.3 | 5.7 |
| Mean extraction batch | 1.5 | 1.7 | 1.5 | 1.8 |
| Mean injection order (range) | 43 (3-94) | 59 (16-93) | 47 (8-92) | 53 (9-91) |

| | Low EA | High EA | Quiescent EA |
|---|---|---|---|
| Number | 14 | 12 | 9 |
| Median CDEIS (range) | 1.3 (0.5-2) | 9.1 (6-28) | 0 (0-0) |
| Median CRP (range / NA, in mg/l) | 1 (1-10 / 6) | 21.9 (1-67 / 7) | 2 (1-14 / 4) |
| Median Faecal calprotectin (range / NA, in µg/g) | 190 (38-1282 / 8) | 474 (63-1935 / 4) | 138 (25-538 / 6) |
| ***Disease location*** | | | |
| Ileal disease | 3 | 2 | 5 |
| Colonic disease | 2 | 4 | 0 |
| Ileocolonic disease | 6 | 3 | 0 |
| Pancolitis | 2 | 1 | 1 |
| ***Montreal*** | | | |
| A1 | 1 | 1 | 0 |
| A2 | 8 | 9 | 6 |
| A3 | 3 | 0 | 3 |
| B1 | 8 | 7 | 7 |
| B2 | 3 | 1 | 2 |
| B3 | 2 | 3 | 1 |
| L1 | 4 | 2 | 5 |
| L2 | 1 | 2 | 2 |
| L3 | 7 | 6 | 3 |
| L4 | 0 | 1 | 0 |
| p- | 6 | 7 | 7 |
| p+ | 6 | 5 | 2 |

Table S-1. Clinical metadata for the samples of the proof of concept study. Endoscopic activity (EA) indices include Montreal classification at diagnosis, Crohn's disease endoscopic index score (CDEIS), C-reactive protein and faecal calprotectin. NA indicates the number of samples for which the information was not available.

**S-2. Security precautions for the sample preparation.**

Working with serum samples represent a potential infection risk. In addition, some of the chemicals used are toxic, flammable, corrosive or irritant (methanol, hexane, pyridine, MSTFA and MeOX). Therefore, the staff was properly vaccinated[S-2] and all manipulations were conducted in a fume hood, with appropriate personal protection equipment including gloves and glasses[S-3], in a laboratory with restricted access.

**S-3. Box-Behnken design of experiment.**

Regarding the DoE, a Box-Behnken design was preferred to a central composite face-centered one because it considered the interactions between the variables as well as their quadratic variations and it required less experimental points. Moreover, the extreme regions were not crucial in the range tested[S-4]. MSTFA and MeOX temperatures of 20, 40 and 60 °C were tested along with MSTFA volumes of 10, 20 and 30 uL while the MeOX volume and duration of reaction remained constant at 10 µL and 30 min. The experimental plan consisted in 17 serum aliquot samples, including 5 center points. The *m/z* = 73, specific to TMS groups, was maximized for both the whole chromatogram (at the exclusion of the highly concentrated d-glucose and urea) and the most derivatized forms of amino acids that are representative of less reactive groups such as amines. Because the optimum was obtained for the lowest volume of MSTFA, a second DoE was conducted that tested the six combinations of MSTFA (5 and 10 µL) and MeOX volumes (5, 10 and 15 µL). Finally, four durations (0.5, 1, 2 and 4 hours) were evaluated at the optimized volumes and temperatures. Additional tests were made at low and high temperatures (4, 80 and 100 °C).

**S-4. List of the metabolites monitored in the internal QC samples. Intra and inter-batch variations.**

| QC Samples | | | | | Intra-Batch | | | Inter-Batch | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean Values | | | RSD (%) | | | RSD (%) | | |
| Automatic ID | Chemical class | $^1t_R$ (min) | $^2t_R$ (s) | Normal. V | $^1t_R$ | $^2t_R$ | Normal. V | $^1t_R$ | $^2t_R$ | Normal. V |
| / | / | 16.8 | 2.9 | 1.1E+6 | 0.2 | 1.1 | 5.4 | 0.2 | 0.6 | 18.2 |
| Leucine | Amino acid | 16.8 | 1.0 | 1.5E+6 | 0.2 | 1.8 | 5.1 | 0.2 | 2.0 | 9.5 |
| Glycine* | Amino acid | 24.3 | 1.9 | 1.7E+6 | 0.2 | 3.3 | 0.0 | 0.2 | 1.7 | 0.0 |
| Pentanoic acid. 3-methyl | Organic acid | 19.9 | 2.1 | 4.1E+4 | 0.2 | 3.2 | **42.2** | 0.1 | 2.1 | **60.6** |
| Succinic acid* | Organic acid | 25.1 | 2.1 | 4.3E+6 | 0.1 | 1.6 | 0.0 | 0.1 | 1.0 | 0.0 |
| / | / | 28.9 | 2.4 | 3.1E+5 | 0.1 | 2.0 | 19.0 | 0.1 | 0.8 | **54.2** |
| Benzoic acid | Organic acid | 33.2 | 3.0 | 5.9E+4 | 0.0 | 1.4 | 25.1 | 0.0 | 0.5 | 20.7 |
| Glutamic acid | Organic acid | 34.1 | 2.9 | 4.4E+5 | 0.1 | 1.7 | 8.3 | 0.1 | 0.7 | 13.3 |
| Cysteine | Amino acid | 35.1 | 1.6 | 2.7E+5 | 0.1 | 2.1 | **34.6** | 0.1 | 1.3 | **60.7** |
| / | / | 38.5 | 1.8 | 4.0E+4 | 0.1 | 2.1 | 23.8 | 0.1 | 1.6 | **45.3** |
| 4.4'-Dibromooctafluorobiphenyl* | Injection standard | 39.6 | 3.1 | 1.4E+7 | 0.1 | 1.1 | 0.0 | 0.1 | 2.6 | 0.0 |
| Ribitol | Sugar alcohol | 41.8 | 3.3 | 6.0E+5 | 0.1 | 1.2 | 3.0 | 0.1 | 1.3 | 10.6 |
| Tetradecanoic acid | Fatty acid | 45.1 | 2.1 | 1.4E+6 | 0.1 | 1.4 | 6.0 | 0.1 | 3.0 | 7.6 |
| Allofuranose | Sugar | 46.2 | 0.5 | 1.2E+5 | 0.1 | 1.0 | 10.1 | 0 | 2.4 | 22.2 |
| Tyrosine | Amino acid | 48.8 | 2.1 | 4.5E+6 | 0.1 | 1.6 | 11.1 | 0 | 1.7 | 14.2 |
| Gulose | Sugar | 52.2 | 0.7 | 3.8E+6 | 0.1 | 0.8 | 12.5 | 0 | 2.2 | 12.5 |
| Heptadecanoic acid | Fatty acid | 54.3 | 2.3 | 1.5E+5 | 0.1 | 2.3 | 9.2 | 0 | 1.1 | 13.9 |
| Linoleic acid | Fatty acid | 56.3 | 3.0 | 4.1E+6 | 0.1 | 0.8 | 3.7 | 0 | 1.0 | 5.9 |
| Tryptophan | Amino acid | 57.0 | 0.6 | 2.6E+6 | 0.1 | 0.9 | **33.5** | 0.1 | 1.7 | **32.2** |

Table S-2. List of the 19 representative metabolites monitored in the internal QC samples. Intra and inter-batch variations (RSD) are given for the normalized peak volume and the retention time in both dimensions. The values with a natural variability > 30% are in bold.

The monitored metabolites had various retention times in the two dimensions and various volumes. They also represented the main biological classes[S-5]: amino acids (5), organic acids (4), fatty acids (3), sugars (2) and sugar alcohols (1). For both intra and inter-batch observations, all $RSD_1$ were ≤ 0.2 %, which is equivalent to 2 times the $P_M$, and even < 0.1 % for 15 out of 19 metabolites followed (Table S-2). All $RSD_2$ were ≤ 5 %, even < 2 % for 14 metabolites out of the 19. This is in agreement with previously reported GC×GC performance[S-6,7].

## S-5. List of the metabolites assessed in the NIST SRM 1950[S-8].

Table 4. Certified Values for Amino Acids

| | Mass Fraction (mg/kg) | Molar Concentration (µmol/L) |
|---|---|---|
| Alanine | 26.2 ± 2.2 | 300 ± 26 |
| Glycine | 18.0 ± 1.2 | 245 ± 16 |
| Histidine | 11.04 ± 0.55 | 72.6 ± 3.6 |
| Isoleucine | 7.13 ± 0.42 | 55.5 ± 3.4 |
| Leucine | 12.90 ± 0.82 | 100.4 ± 6.3 |
| Lysine | 20.0 ± 1.9 | 140 ± 14 |
| Methionine | 3.26 ± 0.26 | 22.3 ± 1.8 |
| Proline | 19.9 ± 1.1 | 177 ± 9 |
| Serine | 9.87 ± 0.44 | 95.9 ± 4.3 |
| Threonine | 13.94 ± 0.70 | 119.5 ± 6.1 |
| Tyrosine | 10.17 ± 0.53 | 57.3 ± 3.0 |
| Valine | 20.9 ± 1.2 | 182.2 ± 10.4 |

Table 5. Reference Values for Amino Acids

| | Mass Fraction (mg/kg) | Molar Concentration (µmol/L) |
|---|---|---|
| Arginine | 13.89 ± 0.40 | 81.4 ± 2.3 |
| Cysteine | 5.26 ± 0.81 | 44.3 ± 6.9 |
| Cystine | 1.83 ± 0.08 | 7.8 ± 0.4 |
| Glutamic Acid | 9.7 ± 2.5 | 67 ± 18 |
| Ornithine | 6.7 ± 0.4 | 52.1 ± 2.8 |
| Phenylalanine | 8.2 ± 1.1 | 51 ± 7 |

Table 9. Certified Values for Selected Clinical Markers

| | Mass Concentration (mg/dL) | Molar Concentration (mmol/L) |
|---|---|---|
| Creatinine | 0.6789 ± 0.0108 | 0.0600 ± 0.0009 |
| Glucose | 82.16 ± 1.00 | 4.560 ± 0.056 |
| Urea | 23.45 ± 0.49 | 3.90 ± 0.08 |
| Uric Acid | 4.274 ± 0.089 | 0.254 ± 0.005 |

Table 2. Certified Values for Fatty Acids

| Lipid Name | Chemical Name (Common Name) | Mass Fraction (µg/g) | Molar Concentration (µmol/L) |
|---|---|---|---|
| C12:0 | Dodecanoic Acid (Lauric Acid) | 1.86 ± 0.11 | 9.47 ± 0.57 |
| C16:0 | Hexadecanoic Acid (Palmitic Acid) | 594 ± 19 | 2364 ± 77 |
| C16:1 n-7 | (Z)-9-Hexadecenoic Acid (Palmitoleic Acid) | 53.5 ± 6.4 | 215 ± 26 |
| C18:0 | Octadecanoic Acid (Stearic Acid) | 179 ± 12 | 644 ± 41 |
| C18:3 n-3 | (Z,Z,Z)-9,12,15-Octadecatrienoic Acid (α-Linolenic Acid) | 14.9 ± 1.0 | 54.6 ± 3.6 |
| C18:1 n-9 | (Z)-9-Octadecenoic Acid (Oleic Acid) | 447 ± 43 | 1614 ± 154 |
| C18:2 n-6 | (Z,Z)-9,12-Octadecadienoic Acid (Linoleic Acid) | 780 ± 39 | 2838 ± 143 |
| C22:0 | Docosanoic Acid (Behenic Acid) | 15.9 ± 1.5 | 47.8 ± 4.6 |

Table 3. Reference Values for Fatty Acids

| Lipid Name | Chemical Name (Common Name) | Mass Fraction (µg/g) | Molar Concentration (µmol/L) |
|---|---|---|---|
| C14:0 | Tetradecanoic Acid (Myristic Acid)[a,b] | 17.9 ± 3.8 | 80.1 ± 17.0 |
| C14:1 | (Z)-9-Tetradecenoic Acid (Myristoleic Acid)[c,d] | 1.57 ± 0.03 | 7.1 ± 0.1 |
| C15:0 | Pentadecanoic Acid[b,e] | 1.08 ± 0.01 | 4.56 ± 0.04 |
| C17:0 | Heptadecanoic Acid (Margaric Acid[d,e] | 4.7 ± 0.2 | 17.6 ± 0.7 |
| C18:3 n-6 | (Z,Z,Z)-6,9,12-Octadecatrienoic Acid (γ-Linolenic Acid)[a,b] | 10.9 ± 2.3 | 39.9 ± 8.5 |
| C18:1 n-7 | (Z)-11-Octadecenoic Acid (Vaccenic Acid)[c,d] | 37.7 ± 0.9 | 136 ± 3 |
| C20:0 | Eicosanoic Acid (Arachidic Acid)[c,d] | 5.5 ± 0.2 | 18.0 ± 0.5 |
| C20:1 | (Z)-11-Eicosenoic Acid (Gondolic Acid[c,d] | 3.5 ± 0.1 | 11.5 ± 0.5 |
| C20:2 | (Z,Z)-11,14-Eicosadienoic Acid[c,d] | 5.7 ± 0.2 | 18.8 ± 0.6 |
| C20:3 n-6 | (Z,Z,Z)-8,11,14-Eicosatrienoic Acid (Homo-γ-Linolenic Acid)[c,d] | 41.8 ± 1.1 | 139 ± 4 |
| C20:4 n-6 | (Z,Z,Z,Z)-5,8,11,14-Eicosatetraenoic Acid (Arachidonic Acid)[a,b] | 293 ± 54 | 984 ± 180 |
| C20:5 n-3 | (Z,Z,Z,Z,Z)-5,8,11,14,17-Eicosapentaenoic Acid (EPA)[c,d] | 11.4 ± 0.1 | 38.6 ± 0.5 |
| C22:1 | (Z)-13-Docosenoic Acid (Erucic Acid)[c,d] | 1.1 ± 0.4 | 3.4 ± 1.3 |
| C22:4 n-6 | (Z,Z,Z,Z)-7,10,13,16-Docosatetraenoic Acid[c,d] | 8.3 ± 0.2 | 25.5 ± 0.6 |
| C22:5 n-3 | (Z,Z,Z,Z,Z)-7,10,13,16,19-Docosapentaenoic Acid (DPA)[c,d] | 12.5 ± 0.2 | 38.5 ± 0.7 |
| C22:5 n-6 | (Z,Z,Z,Z,Z)-4,7,10,13,16-Docosapentaenoic Acid[c,d] | 6.3 ± 0.1 | 19.5 ± 0.4 |
| C22:6 n-3 | (Z,Z,Z,Z,Z,Z)-4,7,10,13,16,19-Docosahexaenoic Acid (DHA)[a,b] | 37.9 ± 6.8 | 118 ± 21 |
| C24:0 | Tetracosanoic Acid (Lignoceric Acid)[c,d] | 16.8 ± 0.9 | 46.6 ± 2.6 |
| C24:1 | (Z)-15-Tetracosenoic Acid (Nervonic Acid)[c,d] | 25.6 ± 1.2 | 71.3 ± 3.2 |

Figure S-1. Extracts from the NIST SRM 1950 (Metabolites in Frozen Human Plasma) certificate. The metabolites evaluated at the validation step are framed in blue.

**S-6. Method optimization. A. Sample preparation. B. Separation and detection.**

(A) Sample preparation

As expected, the optimization led to consequential improvements[S-9]. Methanol (3:1) ***extracted*** higher amounts of metabolites than the chloroform/methanol/water mixture (2:1:1:1) (Table S-5). Previously published results were mitigated about methanol[S-10,S-11,S-12]. Lyophilization, used in a some studies[S-13], was not tested, in order to limit the loss of the most volatile compounds[S-14] such as fatty acids.

As mentioned in the manuscript, the moderate optimal ***derivatization*** conditions compromised between reaction completion and side-effects that lead to degradation (Figure S-2). This is supported by the incomplete reactions at 4, 80 and 100 °C, with respectively 64, 86 and 71 % of the maximum TIC obtained at 40°C (Table S-3). And 37, 92 and 86 % of the normalized volume of amino compounds at the same temperature. Another observation is that the less reactive groups logically required harder conditions. When derivatization duration went from 30 min to 1 hour, the *m/z* = 73 was increased, in the total chromatogram and for the less reactive amino-groups. After 2 and 4 hours, however, the gain became marginal. Because the response near the optimal values of temperature and duration was stable, small changes should not affect much the process[S-11,S-15,S-16]. BSTFA was confirmed[S-17] to be less efficient than MSTFA, with a decreased global *m/z* 73 signal. On the contrary, MTBSTFA gave quite similar results and could therefore be appropriate in certain applications. Here, MSTFA was selected because, despite a reported lower stability of adducts[S-12] and a lower efficiency for the amino groups[S-18], the TMS groups are lighter and smaller and have lower steric hindrance that would limit the derivatization of carbohydrates[S-16, S-18]. In addition, MSTFA is the most reported in literature and mass spectral databases[S-19]. Despite our attempts, and its reported advantages[S-20], methyl chloroformate alkylation could not be performed efficiently, thus it could not be compared. To avoid impurities and preserve the syringe, only 15 µL of the supernatant (out of 25, MeOX + MSTFA) were taken before injection. Regarding the ***stability*** of the samples, the storage at -80 °C produced only minor variations, probably because it never exceeded 4 weeks.

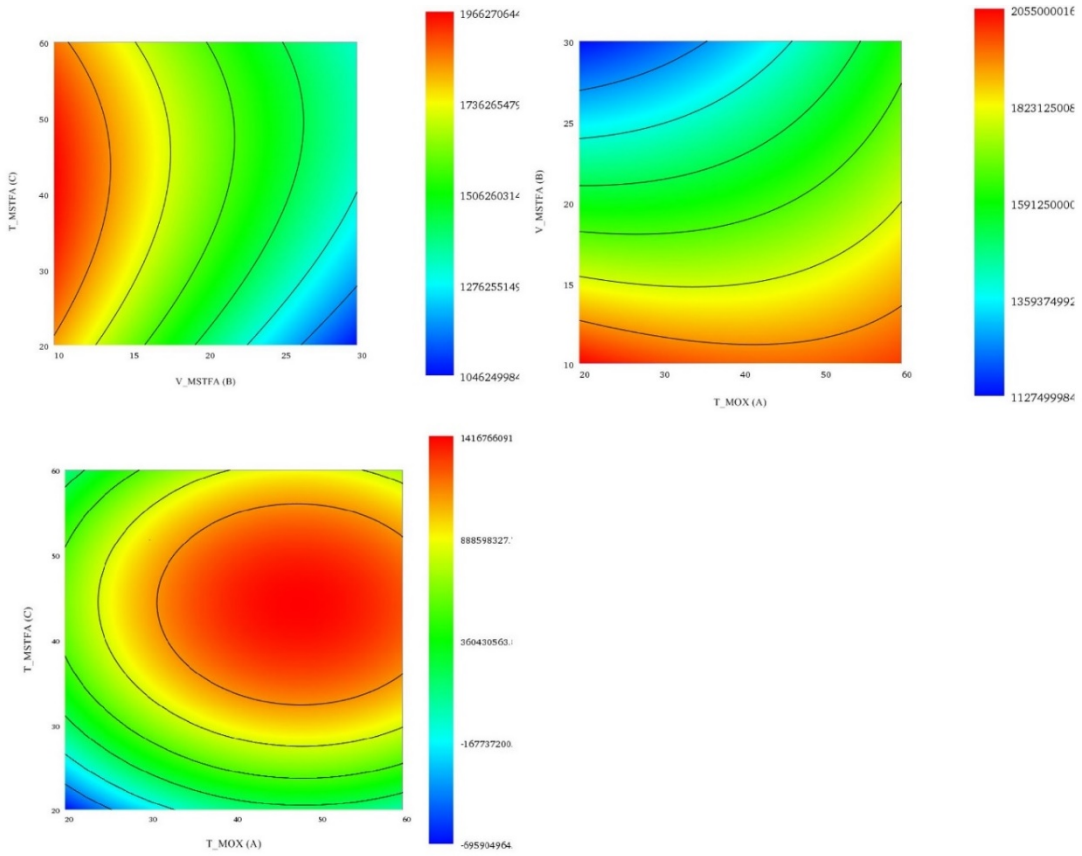



Figure S-2. DoE response surfaces for MSTFA and MeOX volumes and temperatures.

| Derivatization Agent | | | V. MTSFA / V. MTBSTFA |
|---|---|---|---|
| Putative ID | $^{1}t_R$ | $^{2}t_R$ | |
| Propanoic acid | 14.33 | 4.21 | 0.90 |
| DL-3-Phenyllactic acid | 14.97 | 0.83 | 0.53 |
| Butanoic acid | 17.24 | 0.71 | 0.56 |
| Aspartic acid | 19.70 | 2.78 | 1.96 |
| Ethanedioic acid | 25.71 | 2.14 | 2.35 |
| Aminomalonic acid | 32.18 | 1.59 | 1.09 |
| / | 36.56 | 2.62 | 0.37 |
| Oxalic acid | 37.49 | 1.23 | 0.39 |
| 2-Piperidinecarb. acid | 37.37 | 0.79 | 0.67 |
| D-Glucofuranose | 39.77 | 0.63 | 2.21 |
| 2-Hexenedioic acid | 40.47 | 1.03 | 1.23 |
| Undecanoic acid | 43.09 | 1.19 | 0.52 |
| 2-Butenoic acid | 47.82 | 1.07 | 4.71 |
| Dodecanoic acid | 48.40 | 1.55 | 2.40 |
| Cinnamic acid | 49.97 | 2.70 | 1.03 |
| Hexadecanoic acid | 51.78 | 1.98 | 1.07 |
| Uric acid | 54.47 | 3.06 | 0.28 |
| Alanine | 16.02 | 0.71 | 0.66 |
| Isoleucine | 24.60 | 0.87 | 0.83 |
| Proline | 24.60 | 1.39 | 1.30 |
| Glycine | 25.01 | 0.87 | 0.76 |
| Serine | 27.69 | 0.79 | 0.61 |
| threonine | 28.86 | 0.67 | 2.83 |
| Proline | 34.05 | 3.37 | 1.16 |
| Ethionine | 53.42 | 2.30 | 1.55 |

| Derivatization Agent | MSTFA | MTBSTFA | BSTFA |
|---|---|---|---|
| Total Volume | 100 | 102 | 88 |
| N-groups | 100 | 94 | 84 |

| Temperature | Total Norm. V | N-groups |
|---|---|---|
| 4°C | 64 | 37 |
| 40°C | 100 | 100 |
| 80°C | 86 | 92 |
| 100°C | 71 | 86 |

| Duration | 30 min | 60 min | 120 min | 240 min |
|---|---|---|---|---|
| Total Norm V | 100 | 110 | 112 | 108 |
| N-groups | 100 | 111 | 113 | 118 |

| Extraction Agent | MeOH 3:1 | $CCl_4$/MeOH (3:1:1) |
|---|---|---|
| Total V | 100 | 72 |
| N-groups | 100 | 75 |

Table S-3. Sample preparation optimization. MSTFA and MTBSTFA comparison (left). Derivatization reagents comparison (top right). Influence of the temperature and duration of methoxymation and silylation on the global derivatization (total normalized volume) as well as on less reactive amino groups (middle right). Extraction agents comparison (bottom right). To simplify the comparisons, the peak volumes were normalized to a reference value of 100.

(B) Separation and detection

To assess the ***column sets***, in addition to the direct number of unique compounds resolved, parameters linked to resolution, peak shape and peak size (such as symmetry, kurtosis or spread) were measured for 11 compounds representative for chemical classes, retention times and peak volumes (amino, organic and fatty acids, sugars). As a result, the non-polar/semi-polar (normal) set of columns gave the highest number of resolved metabolites (~500 peaks, Figure S-3). The alternative set worked equally well, with a different elution pattern that could be interesting in specific applications[S-21]. The reverse set (semi-polar/non-polar) gave poor results. Possibly because of a retention of the metabolites in the column due to increased interactions. The two columns used the same diameter and film thickness in order to have an optimal flow in both. The values chosen, 0.25 mm - 0.25 µm, compromised between sample

loadability and peak capacity in the second dimension, to fit with the objective of an untargeted method. For the same reason, a linear ***temperature program*** was chosen. It also allowed the use of linear retention indices for compound identification. This parameter revealed to be much more influential than the flow rate, at the tested values (Figures S-4 and S-5). The best resolution was obtained at 1 mL/min, in accordance to the Van Deemter equation, and 3 °C/min, as explained in the manuscript. This low ramp increased the run time to 72 minutes (Table S-4), the price to pay to take full advantage of the instrumentation in an exploratory mode.

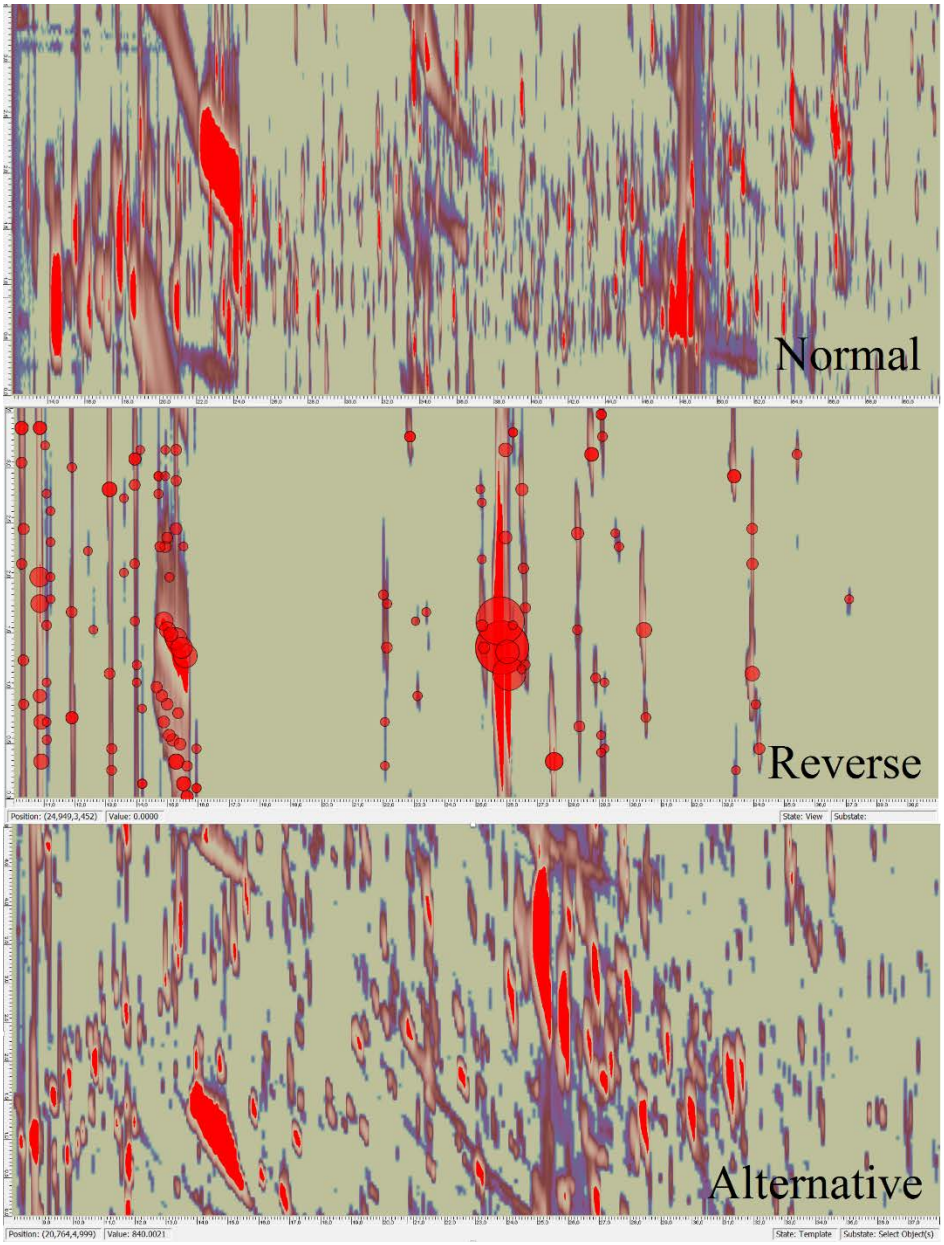



Figure S-3. Examples of chromatograms produced by the three sets of columns.

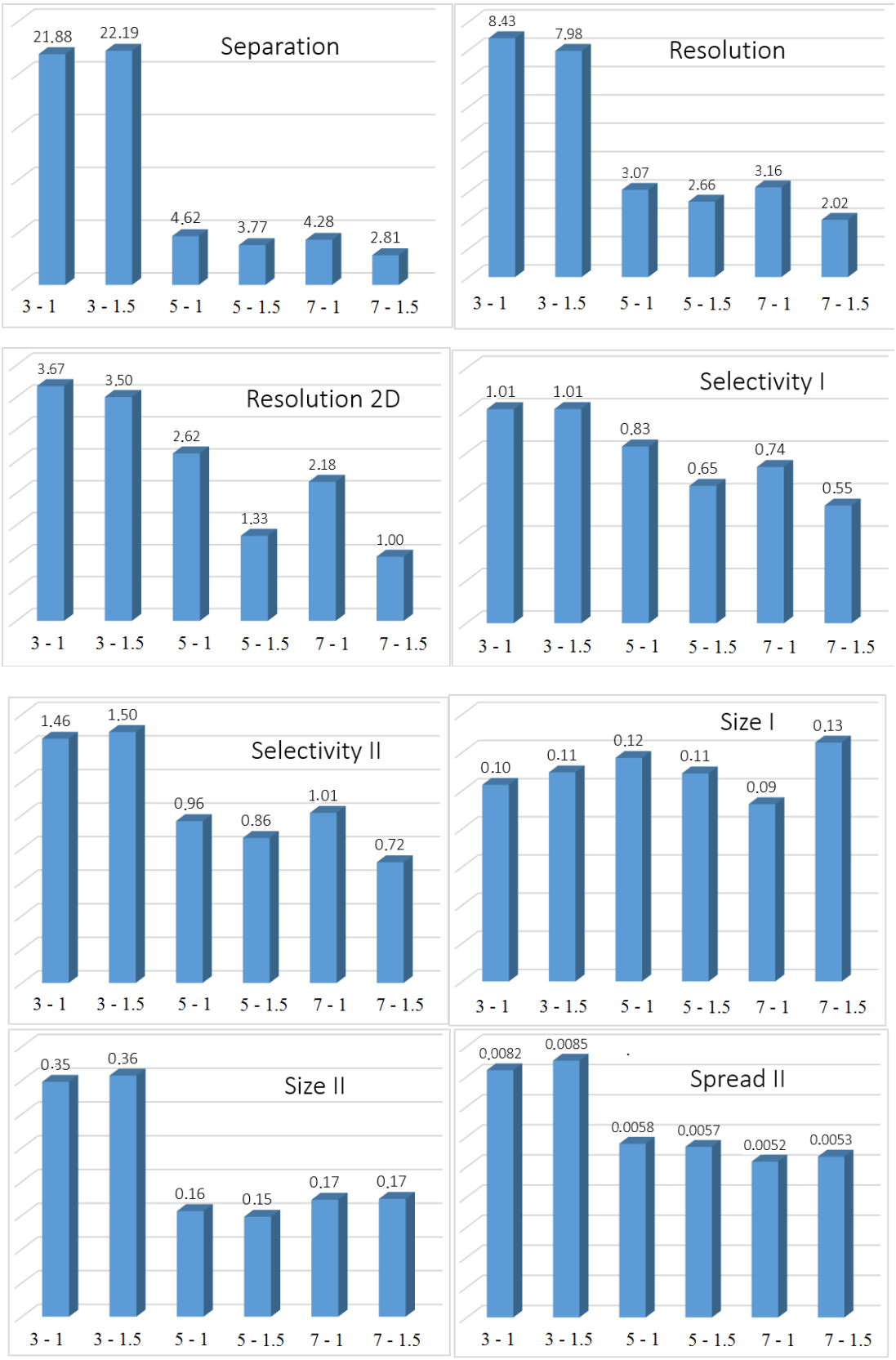



Figure S-4. Separation parameters for various flow rates and temperature ramps. 3 °C/min, either with 1 or 1.5 mL/min, provided the best separation, with higher resolution, selectivity and separation despite a higher spread of the peaks.

.

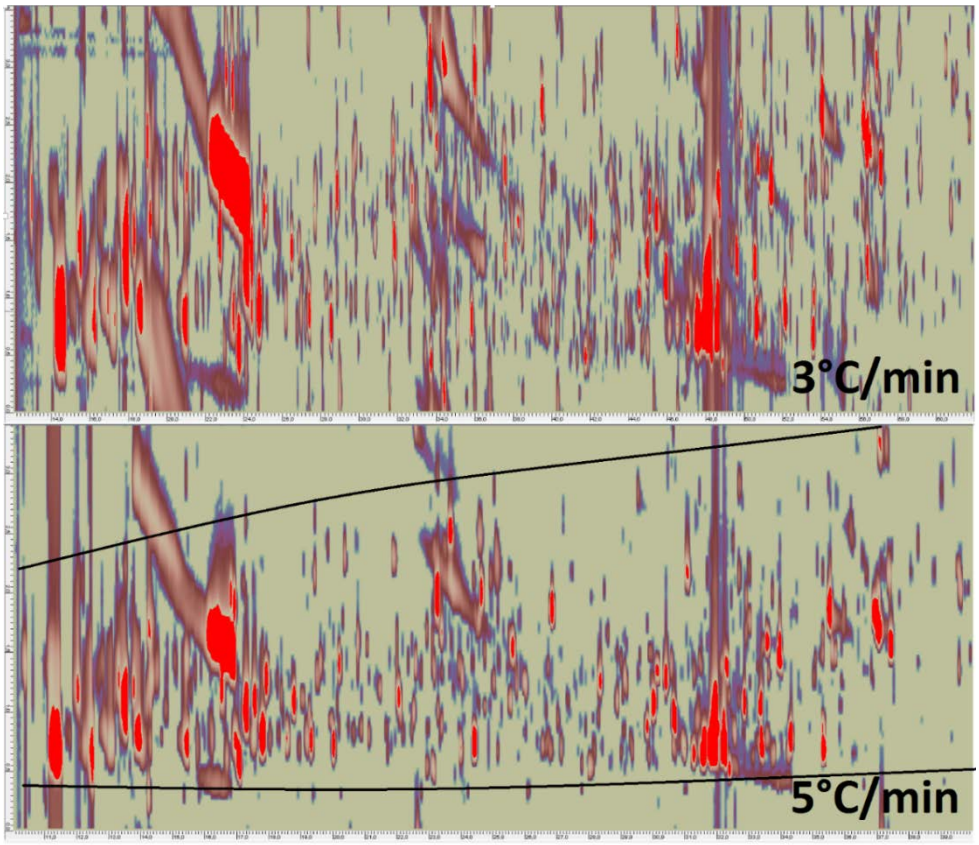



Figure S-5. Chromatograms obtained at 3 and 5 °C/min. The first provided a better use of the separation space, especially in the second dimension, increasing the overall spatial resolution.

| Hold time (min) | 1 | 0 | 5 |
|---|---|---|---|
| Temp (°C) | 50 | 240 | 300 |

| Ramp 1 : 50 → 240 °C |
|---|
| Ramp 2 : 240 → 300 °C |

| Ramp 2 Temp (°C/min) |
|---|
| 15 °C/min |

| Ramp 1 Temp (°C/min) | Ramp 1 time (min) | Total Time (min) |
|---|---|---|
| 3 | 63 | **73** |
| 5 | 38 | **48** |
| 7 | 27 | **37** |
| 10 | 19 | **29** |

Table S-4. Run times for the four temperature ramps evaluated, not including the cooling time of the oven and the columns.

The best ***initial temperature*** was 50 °C (Figure S-6), in agreement to the good practice rule that suggests to start a run 20 °C below the boiling point of the solvent[S-22], here hexane (BP : 69 °C). Indeed, 90 and 70 °C both led to badly focused first eluting compounds. 40 °C did not improve further the resolution.

Once the initial and ramp temperatures were known, the solvent delay was set at 8 minutes to preserve the detector.

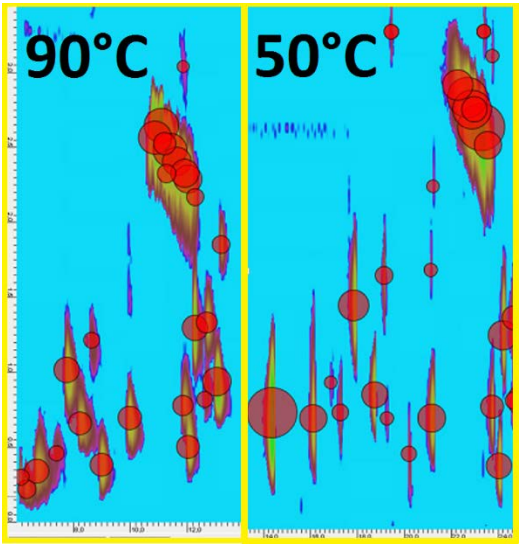



Figure S-6. Effect of the initial temperature on the early eluting compounds, that are broadened at 90°C.

The configuration of the ***modulation system*** depends on the samples to analyze. Here, a modulation period ($P_M$) of 3.5 seconds was found to maintain the separation of the peaks in the first dimension with minimal empty space and wraparound in the second dimension. Meanwhile, a 700 ms hot jet at 170 °C allowed an efficient reinjection, which is especially important for volatile compounds eluting at low temperatures, and a reduced breakthrough. 500 ms was found too short (Figure S-7), and a temperature of 100 °C too cold. There was no clear improvement at 900 ms and 200 °C.

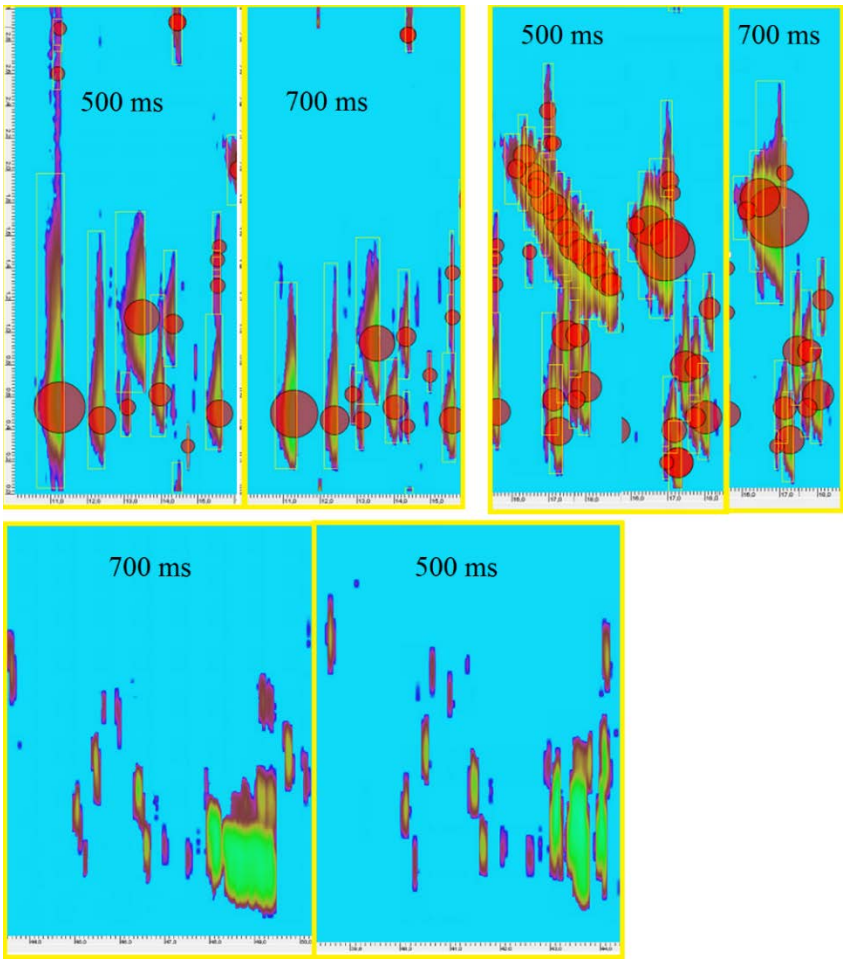



Figure S-7. Effect of the hot jet duration ($P_M$ 3.5 s) on the resolution in the overloaded zones of urea (above) and glucose (below).

The most effective ***injection*** temperature was 250 °C (Table S-5). 200 °C was too low to volatilize correctly the heaviest metabolites, especially in a biological matrix[S-23]. 280 °C was too destructive for the TMS adducts.

| Represent. Compounds | | Injection standard | |
|---|---|---|---|
| Injection Temp. | Norm. Volume | Injection Temp. | Norm. Volume |
| 200 | 92 | 200 | 97 |
| 250 | 100 | 250 | 100 |
| 280 | 88 | 280 | 101 |

Table S-5. Influence of the injection temperature on the normalized peak volume, here normalized to 100 at 250°C.

Split mode injection has a reported beneficial effect on peak shape and quantitation[S-24]. However, negative effects on peak ratios and the stability of amino-groups adducts have also been observed[S-25]. Here, a split as low as 10 was found too dilutive for untargeted analysis, preventing the detection of trace metabolites, while a split less than 10 was too unstable to be reproducible. The opposite alternative,

i.e. injecting 2 µL instead of 1 µL, increased the signals of trace metabolites, but induced an overloading of highly concentrated compounds such as urea and glucose. This resulted in an overall decrease in the number and the chromatographic quality of the metabolites detected in wide areas around them (Figure S-8). Not mentioning the possible backflash that would make the chromatogram unrepresentative of the initial sample. Backflash that likely explained the lower signals detected with the 'Gooseneck' liner which, by design, gives a lower volume to the gas phase to expand (Figure S-9).

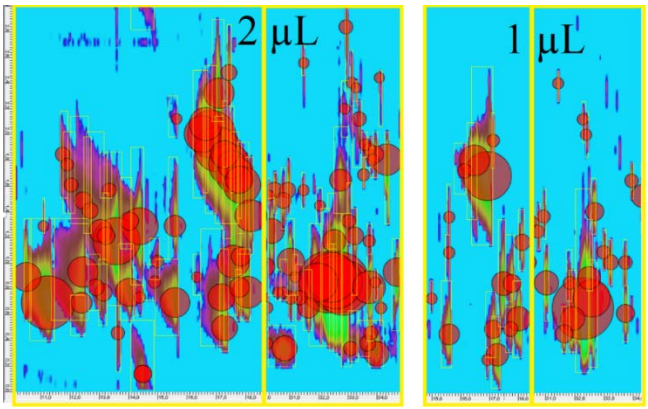



Figure S-8. Impact of the injection volume on the chromatographic resolution in an area at risk of overloading (urea).

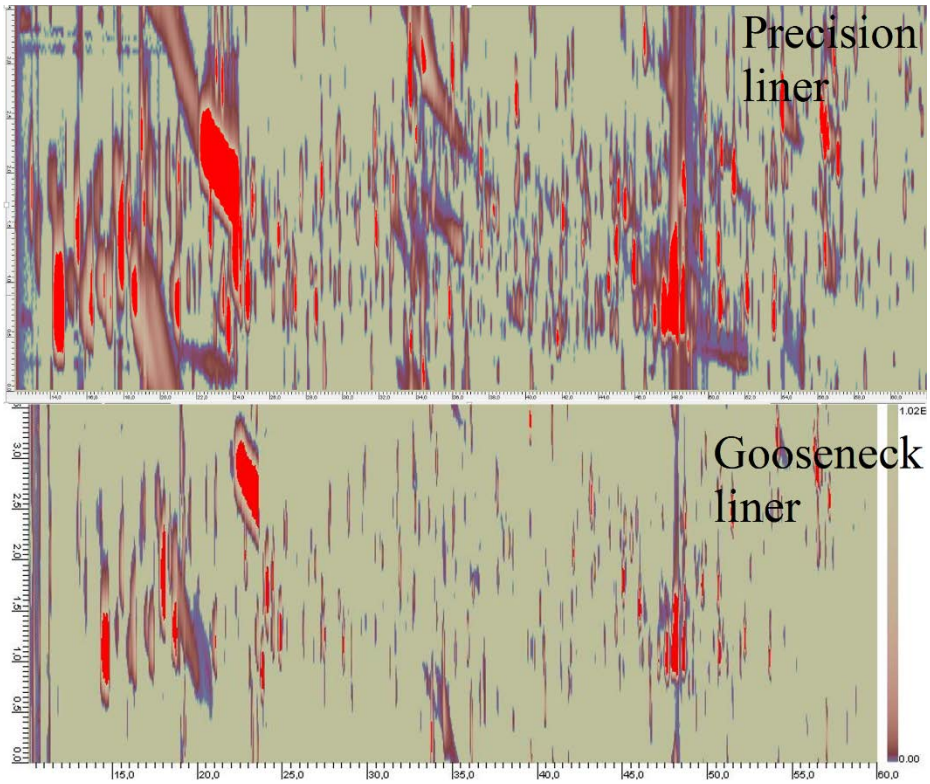



Figure S-9. Effect of the liner design on the global chromatogram. Higher number of metabolites and signal intensity were detected with the 'Precision' liner.

**EXPERIMENTAL**

Sample Prep

1. Extraction: methanol (3:1)
2. Derivatization (30 µL serum):
   MeOX 15 µL - 40 °C - 1h
   MSTFA 10 µL - 40 °C - 1h

Separation and detection

Rxi-5 - 30 m / Rxi-17 - 2 m
Both 0.25 mm, 0.25 µm

| | |
|---|---|
| 1 mL/min flow | 50 °C initial T |
| 3 °C/min ramp T | 3.5 seconds $P_M$ |
| 700 ms hot jet | 250 °C injection T |
| Precision Sky liner | 1 µL splitless |

Figure S-10. Summary of the optimized analytical conditions.

## S-7. Accuracy and precision assessment in NIST SRM 1950.

| | Mean (umol/L) | Precision - RSD (%) | | | Accuracy - Z-score | |
|---|---|---|---|---|---|---|
| | NIST | NIST | Intra-Batch | Inter-Batch | Intra-Batch | Inter-Batch |
| Alanine | 300 | 9 | 7.4 | 11.3 | 4.4 | 0.5 |
| Valine | 182.2 | 6 | 2.5 | 8.1 | 1.9 | 6.3 |
| Leucine | 100.4 | 6 | 4.2 | 8.7 | 1.0 | 2.5 |
| Isoleucine | 55.5 | 6 | 11.6 | 7.1 | 34.8 | 11.2 |
| Proline | 177 | 5 | 5.6 | 13.9 | 7.9 | 9.5 |
| Serine | 95.9 | 4 | 4.5 | 6.3 | 4.3 | 6.6 |
| Threonine | 119.5 | 5 | 9.8 | 10.4 | 0.5 | 24.4 |
| Phenylalanine | 51 | 14 | 4.2 | 10.7 | 11.3 | 13.1 |
| Methionine | 22.3 | 8 | 14.8 | 20.2 | 7.7 | 6.4 |
| Glycine | 245 | 7 | 4.1 | 4.2 | **0.5** | **0.9** |
| Tyrosine | 57.3 | 5 | 4.7 | 11.2 | 25.0 | 26.3 |
| Lysine | 140 | 10 | 11.4 | 8.4 | 8.0 | 9.2 |
| Cystéine | 44.3 | 16 | 14.4 | 26.4 | 5.6 | 5.7 |
| Ornithine | 133.5 | 4 | 5.3 | 8.7 | 5.8 | 5.1 |
| Urea | 3900 | 2 | 7.2 | 4.9 | 12.1 | 7.7 |
| Creatinine | 60 | 2 | 6.0 | 5.3 | 13.0 | 4.6 |
| Histidine | 72.6 | 5 | 9.0 | 7.7 | 17.8 | 16.2 |
| Glutamic acid | 67 | 27 | 3.0 | 9.2 | 1.0 | 1.2 |
| Uric acid | 254 | 2 | 2.8 | 16.2 | 27.9 | 26.2 |
| Dodecanoic acid | 9.5 | 6 | 3.7 | 3.6 | 13.6 | 12.0 |
| Myristic acid | 80.1 | 21 | 3.4 | 4.3 | 4.1 | 3.9 |
| Pentadecanoic acid | 4.6 | 1 | 7.5 | 14.7 | 84.0 | 72.3 |
| Palmitic Acid | 2364 | 3 | 7.3 | 7.7 | 29.2 | 28.9 |
| Heptadecanoic acid | 17.6 | 4 | 9.4 | 7.8 | 23.9 | 23.3 |
| Linoelaidic acid | 2838 | 5 | 5.1 | 4.5 | 19.3 | 19.7 |
| Oleic Acid. (Z)- | 1614 | 10 | 3.4 | 9.9 | 9.6 | 9.5 |
| Stearic acid | 644 | 6 | 3.5 | 7.1 | 14.4 | 14.0 |
| Linoleic Acid | 54.6 | 7 | 5.7 | 12.7 | 14.6 | 14.4 |

Table S-6. Accuracy and precision assessment in NIST SRM 1950 through intra and inter-batch measurements.

## S-8. Recovery assessment in NIST SRM 1950 and internal QC samples.

| **Recovery (%)** | Sample Preparation | | 1. Extraction | 2. Drying & Derivatization |
|---|---|---|---|---|
| | Glycine | Succinic acid | | |
| **QC** Inter-batch | 73 | 57 | 63 | -16 |
| | 74 | 60 | 69 | -20 |
| | 73 | 60 | 66 | -16 |
| | 75 | 59 | 66 | -18 |
| Mean | 74 | 59 | 66 | -17 |
| RSD (%) | 1.1 | 2.3 | 4.2 | 10.7 |
| **NIST** Intra-batch | 65 | 53 | 76 | -16 |
| | 61 | 50 | 74 | -16 |
| | 63 | 50 | 73 | -17 |
| | 61 | 50 | 67 | -22 |
| | 61 | 47 | 68 | -19 |
| Mean | 62 | 50 | 72 | -18 |
| RSD (%) | 2.9 | 3.8 | 5.5 | 15.3 |
| **NIST** Inter-batch | 65 | 46 | 59 | -12 |
| | 61 | 41 | 57 | -16 |
| | 57 | 40 | 55 | -15 |
| Mean | 61 | 43 | 57 | -14 |
| RSD (%) | 6.7 | 7.6 | 3.2 | 13.2 |
| Total Mean | 66 | 51 | 66 | -17 |
| Total RSD (%) | 9.7 | 13.3 | 10.1 | 15.6 |

Table S-7. Recovery assessment in NIST SRM 1950 and internal QC in intra and inter-batch measurements.

## S-9. Regression methods. Determination of the best fit.

| Succinic acid | | | | | | |
|---|---|---|---|---|---|---|
| | | Concentration (ng/µL) | Predicted Signal | Observed Signal | % Recovery | Relative Error |
| Slope | 9035 | 0 | 8.8E+2 | 0.0E+0 | / | / |
| Intersect | 882 | 0.05 | 1.3E+3 | 3.5E+1 | 3794 | 3694 |
| **Unweighted** | | 0.25 | 3.1E+3 | 1.5E+3 | 210 | 110 |
| | | 1.67 | 1.6E+4 | 2.1E+4 | 78 | 22 |
| $R^2$ | 0.9957 | 10 | 9.1E+4 | 9.1E+4 | **101** | 1 |
| | | | | | | 3827 |
| Slope | 9148 | 0 | 4.0E+0 | 0.0E+0 | / | / |
| Intersect | 4 | 0.05 | 4.6E+2 | 3.5E+1 | 1312 | 1212 |
| **√S** | | 0.25 | 2.3E+3 | 1.5E+3 | 153 | 53 |
| | | 1.67 | 1.5E+4 | 2.1E+4 | 75 | 25 |
| | | 10 | 9.1E+4 | 9.1E+4 | **101** | 1 |
| | | | | | | 1292 |
| Slope | 9136 | 0 | -8.7E+0 | 0.0E+0 | / | / |
| Intersect | -9 | 0.05 | 4.5E+2 | 3.5E+1 | 1275 | 1175 |
| **S** | | 0.25 | 2.3E+3 | 1.5E+3 | 152 | 52 |
| | | 1.67 | 1.5E+4 | 2.1E+4 | 74 | 26 |
| | | 10 | 9.1E+4 | 9.1E+4 | **101** | 1 |
| | | | | | | 1253 |
| Slope | 6649 | 0 | -1.4E-1 | 0.0E+0 | / | / |
| Intersect | 0 | 0.05 | 3.3E+2 | 3.5E+1 | 945 | 845 |
| **$S^2$** | | 0.25 | 1.7E+3 | 1.5E+3 | **111** | 11 |
| | | 1.67 | 1.1E+4 | 2.1E+4 | 54 | 46 |
| | | 10 | 6.6E+4 | 9.1E+4 | 73 | 27 |
| | | | | | | 929 |
| Slope | 1145 | 0 | -2.2E-4 | 0.0E+0 | / | / |
| Intersect | 0 | 0.05 | 5.7E+1 | 3.5E+1 | 163 | 63 |
| **$S^3$** | | 0.25 | 2.9E+2 | 1.5E+3 | 19 | 81 |
| | | 1.67 | 1.9E+3 | 2.1E+4 | 9 | 91 |
| | | 10 | 1.1E+4 | 9.1E+4 | 13 | 87 |
| | | | | | | 322 |
| Slope | 746 | 0 | -4.9E-7 | 0.0E+0 | / | / |
| Intersect | 0 | 0.05 | 3.7E+1 | 3.5E+1 | **106** | 6 |
| **$S^4$** | | 0.25 | 1.9E+2 | 1.5E+3 | 12 | 88 |
| | | 1.67 | 1.2E+3 | 2.1E+4 | 6 | 94 |
| | | 10 | 7.5E+3 | 9.1E+4 | 8 | 92 |
| | | | | | | **279** |
| Slope | 709 | 0 | -1.5E-9 | 0.0E+0 | / | / |
| Intersect | 0 | 0.05 | 3.5E+1 | 3.5E+1 | **101** | 1 |
| **$S^5$** | | 0.25 | 1.8E+2 | 1.5E+3 | 12 | 88 |
| | | 1.67 | 1.2E+3 | 2.1E+4 | 6 | 94 |
| | | 10 | 7.1E+3 | 9.1E+4 | 8 | 92 |
| | | | | | | 275 |

Table S-8. Determination of the most efficient regression for succinic acid. The framed cells indicate the sum of relative errors. The bold values are between 80 and 120% of recovery (ideal case). The lowest sums, in grey, were obtained with the $S^5$ and $S^4$ weights (standard deviation at the fifth and fourth power, respectively). The latter was chosen because it gave, for a very similar error, a simpler model. As showed in the percentage of recovery, this regression fitted well only at low concentrations, the ones used for the sensitivity assessment.

| Fumaric acid | | | | | | |
|---|---|---|---|---|---|---|
| | | Concentration (ng/µL) | Predicted Signal | Observed Signal | % Recovery | Relative Error |
| Slope | 43158 | 0 | 7.0E+3 | 1.0E+1 | 67822 | 67722 |
| Intersect | 6964 | 0.05 | 9.1E+3 | 1.6E+3 | 577 | 477 |
| **Unweighted** | | 0.25 | 1.8E+4 | 1.2E+4 | 147 | 47 |
| | 0.9948 | 1.67 | 7.9E+4 | 1.0E+5 | 77 | 23 |
| $R^2$ | | 10 | 4.4E+5 | 4.3E+5 | **101** | 1 |
| | | | | | | 68271 |
| Slope | 43379 | 0 | 1.0E+3 | 1.0E+1 | 10203 | 10103 |
| Intersect | 1048 | 0.05 | 3.2E+3 | 1.6E+3 | 204 | 104 |
| **√S** | | 0.25 | 1.2E+4 | 1.2E+4 | **98** | 2 |
| | | 1.67 | 7.3E+4 | 1.0E+5 | 71 | 29 |
| | | 10 | 4.3E+5 | 4.3E+5 | **100** | 0 |
| | | | | | | 10237 |
| Slope | 44162 | 0 | 3.2E+1 | 1.0E+1 | 315 | 215 |
| Intersect | 32 | 0.05 | 2.2E+3 | 1.6E+3 | 142 | 42 |
| **S** | | 0.25 | 1.1E+4 | 1.2E+4 | **92** | 8 |
| | | 1.67 | 7.4E+4 | 1.0E+5 | 72 | 28 |
| | | 10 | 4.4E+5 | 4.3E+5 | **102** | 2 |
| | | | | | | 295 |
| Slope | 44021 | 0 | 8.8E+0 | 1.0E+1 | **86** | 14 |
| Intersect | 9 | 0.05 | 2.2E+3 | 1.6E+3 | 140 | 40 |
| **$S^2$** | | 0.25 | 1.1E+4 | 1.2E+4 | **91** | 9 |
| | | 1.67 | 7.4E+4 | 1.0E+5 | 71 | 29 |
| | | 10 | 4.4E+5 | 4.3E+5 | **101** | 1 |
| | | | | | | **93** |
| Slope | 37180 | 0 | 1.0E+1 | 1.0E+1 | **100** | 0 |
| Intersect | 10 | 0.05 | 1.9E+3 | 1.6E+3 | **118** | 18 |
| **$S^{2.5}$** | | 0.25 | 9.3E+3 | 1.2E+4 | 77 | 23 |
| | | 1.67 | 6.2E+4 | 1.0E+5 | 60 | 40 |
| | | 10 | 3.7E+5 | 4.3E+5 | **86** | 14 |
| | | | | | | 96 |
| Slope | 32201 | 0 | 1.0E+1 | 1.0E+1 | **100** | 0 |
| Intersect | 10 | 0.05 | 1.6E+3 | 1.6E+3 | **103** | 3 |
| **$S^3$** | | 0.25 | 8.1E+3 | 1.2E+4 | 67 | 33 |
| | | 1.67 | 5.4E+4 | 1.0E+5 | 52 | 48 |
| | | 10 | 3.2E+5 | 4.3E+5 | 74 | 26 |
| | | | | | | 110 |
| Slope | 30993 | 0 | 1.0E+1 | 1.0E+1 | **100** | 0 |
| Intersect | 10 | 0.05 | 1.6E+3 | 1.6E+3 | **99** | 1 |
| **$S^4$** | | 0.25 | 7.8E+3 | 1.2E+4 | 64 | 36 |
| | | 1.67 | 5.2E+4 | 1.0E+5 | 50 | 50 |
| | | 10 | 3.1E+5 | 4.3E+5 | 71 | 29 |
| | | | | | | 116 |

Table S-9. Determination of the most efficient regression for fumaric acid. The framed cells indicate the sum of relative errors. The bold values are between 80 and 120% of recovery (ideal case). The lowest sum, in grey, was obtained with the $S^2$ weight (square standard deviation). The obtained regression fitted well at all concentrations.

| Glycine | | | | | | |
|---|---|---|---|---|---|---|
| | | Concentration (ng/µL) | Predicted Signal | Observed Signal | % Recovery | Relative Error |
| Slope | 16841 | 0 | 5.0E+4 | 4.3E+4 | **116** | 16 |
| Intersect | 49606 | 0.05 | 5.0E+4 | 4.9E+4 | **104** | 4 |
| **Unweighted** | | 0.25 | 5.4E+4 | 5.0E+4 | **107** | 7 |
| | 0.9866 | 1.67 | 7.8E+4 | 9.2E+4 | **84** | 16 |
| $R^2$ | | 10 | 2.2E+5 | 2.2E+5 | **101** | 1 |
| | | | | | | **44** |
| Slope | 295307 | 0 | -3.1E+5 | 4.3E+4 | -722 | 822 |
| Intersect | -307991 | 0.05 | -2.9E+5 | 4.9E+4 | -602 | 702 |
| **√S** | | 0.25 | -2.3E+5 | 5.0E+4 | -465 | 565 |
| | | 1.67 | 1.9E+5 | 9.2E+4 | 200 | 100 |
| | | 10 | 2.6E+6 | 2.2E+5 | 1226 | 1126 |
| | | | | | | 3316 |
| Slope | 313520 | 0 | -1.9E+5 | 4.3E+4 | -445 | 545 |
| Intersect | -189820 | 0.05 | -1.7E+5 | 4.9E+4 | -357 | 457 |
| **S** | | 0.25 | -1.1E+5 | 5.0E+4 | -222 | 322 |
| | | 1.67 | 3.3E+5 | 9.2E+4 | 361 | 261 |
| | | 10 | 2.9E+6 | 2.2E+5 | 1366 | 1266 |
| | | | | | | 2851 |
| Slope | 161166 | 0 | -2.5E+4 | 4.3E+4 | -58 | 158 |
| Intersect | -24919 | 0.05 | -1.7E+4 | 4.9E+4 | -35 | 135 |
| **$S^2$** | | 0.25 | 1.5E+4 | 5.0E+4 | 31 | 69 |
| | | 1.67 | 2.4E+5 | 9.2E+4 | 264 | 164 |
| | | 10 | 1.6E+6 | 2.2E+5 | 736 | 636 |
| | | | | | | 1162 |

Table S-10. Determination of the most efficient regression for glycine. The framed cells indicate the sum of relative errors. The bold values are between 80 and 120% of recovery (ideal case). The unweighted regression gave the lowest sum and fitted very well at all the concentrations tested.

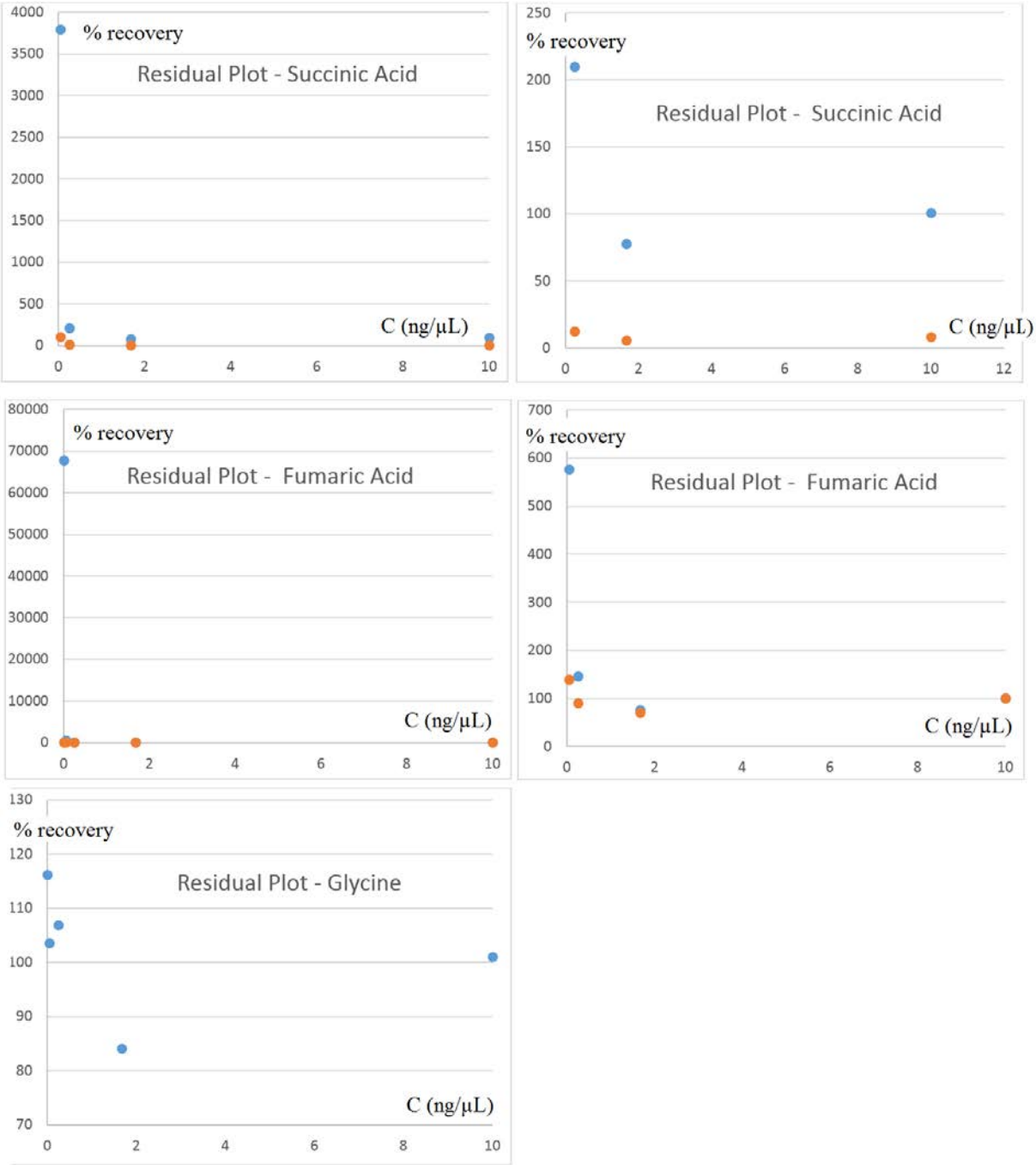



Figure S-11. Residual plots for the three IS using unweighted (blue points) and weighted (orange points) regressions. The complete plots are on the left. For succinic and fumaric acids, the plots on the right zoom-in in order to neglect the very high recovery percentage of the blank in the unweighted regression. The efficiency gain produced by weighting was especially important at low concentrations, where the limits of detection and quantification (LOD and LOQ) were assessed.

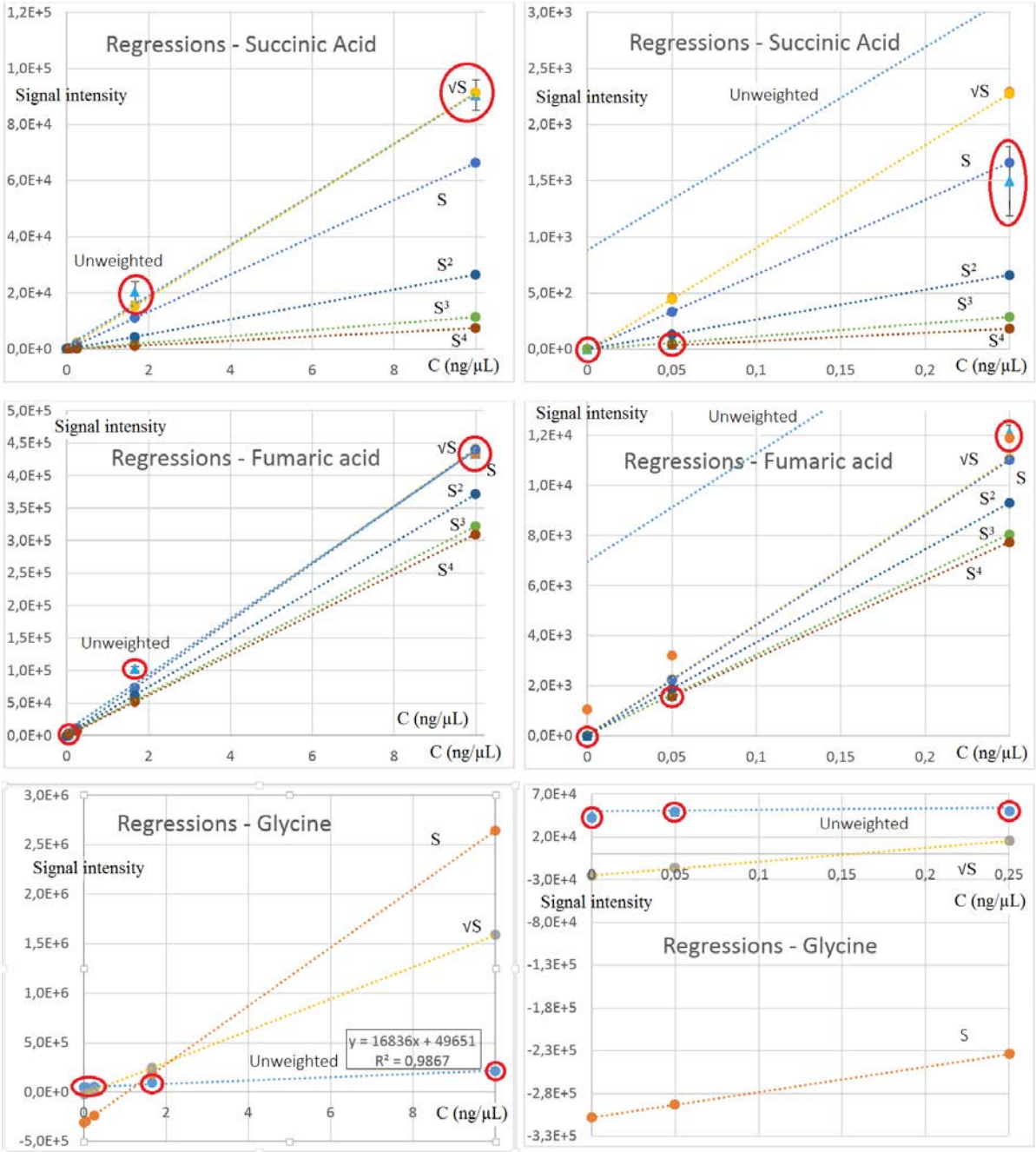



Figure S-12. Unweighted and weighted linear regression for succinic acid, fumaric acid and glycine with zoom-in to the lowest concentrations (right). The experimental values are encircled in red.

**S-10. Sensitivity assessment. A. Methods. B. Calculation of LOD and LOQ.**

(A) Methods

The selective *m/z* considered were: 247 for fumaric acid-2,3-$d_2$, 251 for succinic acid-2,2,3,3-$d_4$, 104 (2 TMS), 250 and 278 (3 TMS) for glycine-2,2-$d_2$.

Sensitivity was assessed in three ways. In the first, the analytical signals for LOD and LOQ, named yLOD and yLOQ, were defined as[S-26,S-27,S-28]:

$$yLOD = M_{blank} + 3.S_{blank}$$

$$yLOQ = M_{blank} + 10.S_{blank}$$

Where M is the mean signal and S the standard deviation. The regressions obtained earlier are of the (linear) form y = ax + b, where y is the analytical signal at the analyte concentration x. Therefore:

$$LOD/LOQ = \frac{yLOD\ /\ yLOQ\ - b}{a}$$

In the second, the signal from the blank was compared to the signal at the lowest concentration injected[S-29].

$$LOD = yLOD.\frac{Lowest\ Conc}{Mean\ Signal_{Lowest\ conc}}$$

$$LOQ = yLOQ.\frac{Lowest\ Conc}{Mean\ Signal_{Lowest\ conc}}$$

In the third, the LOQ was defined as the minimal amount really injected with a RSD below 30 %[S-30] and a signal over 5 times the blank signal[S-31]. S/N ratios were not tested because of the artificial loss of signal due to the early centroidization of the chromatograms that accounted for a factor around 2 despite optimization.

(B) Calculation of LOD and LOQ

| | LOD (pg/uL) **1.** | LOQ (pg/uL) **2.** | LOD (pg/uL) **3.** | LOQ (pg/uL) **4.** | 30% RSD & > 5*Blank Signal | Sensitivity (e-/pg.ul) |
|---|---|---|---|---|---|---|
| Fumaric Acid | 1.6 | 5.2 | 2.2 | 7.3 | 50 | 44.0 |
| Succinic Acid | 4 | 13 | 4.5 | 14 | 250 | 7.5 |
| Glycine | 125 | 1375 | 53.1 | 75 | 10000 | 16.8 |

| | On-column | | | |
|---|---|---|---|---|
| | LOD (pg/uL) **1.** | LOQ (pg/uL) **2.** | LOD (pg/uL) **3.** | LOQ (pg/uL) **4.** |
| Fumaric Acid | 1.9 | 6.3 | 2.7 | 8.8 |
| Succinic Acid | 5 | 16 | 5.4 | 17 |
| Glycine | 150 | 1650 | 63.7 | 90 |

**1.** LOD = 3 SD blank / slope regression
**2.** LOQ = 10 SD blank / slope regression
**3.** LOD = Min.Conc*yLOD / y Min.Conc
**4.** LOD = Min.Conc *yLOQ / y Min.Conc

Table S-11. LOD, LOQ and sensitivity assessment (in pg/uL injected) using three different methods: regression applied to the blank (1 and 2), blank and minimal concentration tested (3 and 4) and minimal concentration tested having RSD < 30% (5).

Because of the high signal of its blank, glycine had a lower sensitivity. It also exhibited higher variations from one method of evaluation to the other. Therefore, only the method that compared, and in a way corrected, the blank to the signal at the lowest concentration injected worked well (Table S-11). To achieve that, the blank had to be reliably measured, which was the case here with a RSD of 7%. This procedure seemed the most consistent overall, presumably because it uses signals from real injections while the regression curves modelize an approximated behavior. Nevertheless, these results also emphasize the need for m/z that are very specific of the IS, therefore the need for highly deuterated standards, to reliably quantify compounds at low concentrations. Because it is based on real amounts injected, here the third method suffered from the low number of concentrations tested near the LOQ and thus was not very informative.

Overall, the discrepancies observed between the three IS as well as the deviations from an ideal behaviour, for both linearity and sensitivity, show that the capacity to quantify, even when limited to semi-quantitation, cannot be systematically achieved. Hence, it is important, when establishing a method, to evaluate these parameters for compounds representative of the various metabolites present in the samples in terms of chemical classes, signal intensities and retention times.

The sensitivity value is based on the working principle of the microchannel plate (MCP) detector used in the TOF spectrometer. It makes the assumption that each unity of the signal (given by the slopes of the regressions) corresponds to an electron produced through the electron amplifier of the detector. Thus, theoretically 7, 17 and 44 electrons were produced at the detector per pg of compound (respectively succinic acid, fumaric acid and glycine) per µL of sample.

**S-11. QC system. A. Acceptance/rejection criteria. B. Effect of the QC correction**

(A) Acceptance / rejection criteria

Any injected sample was rejected if the raw peak volume of any standard –internal or injection- was outside the QC mean value ± 30 %.

A whole batch of 5 study samples was rejected if the following QC sample was rejected or if at least 2 study samples were rejected.

A QC sample was rejected if:

- The raw peak volume of any standard –internal or injection- was outside the QC mean value ± 30%.
- Any retention time of any representative metabolite (out of 19) was outside the action limits, or at least 3 of the same kind (first or second dimension) were outside the warning limits.
- The corrected peak volume was outside the action limits for at least 2 representative metabolites (out of 16, IS excluded), was outside the warning limits for at least 4 of them, or was outside the action limits for at least 1 of them and outside the warning limits for at least 2 of them.

No trends criteria were used in this study but could be interesting in others.

Any study or QC sample rejected was directly reinjected. A second rejection, in the absence of any other clue, was considered a derivatization failure. In this circumstance, a biological replicate was prepared and injected. Another rejection was considered a system failure and stopped the sequence of injection. Once the issue had been found and fixed, and the system reconditioned, the sequence started again. In the case of a study sample, a biological replicate was prepared and injected. In the case of a QC sample, the whole batch was rejected and biological replicates were prepared and injected.

(B) Effect of the QC correction


Figure S-13. Effect of the LOESS (partial) correction on the QC samples for two metabolites of the final dataset.

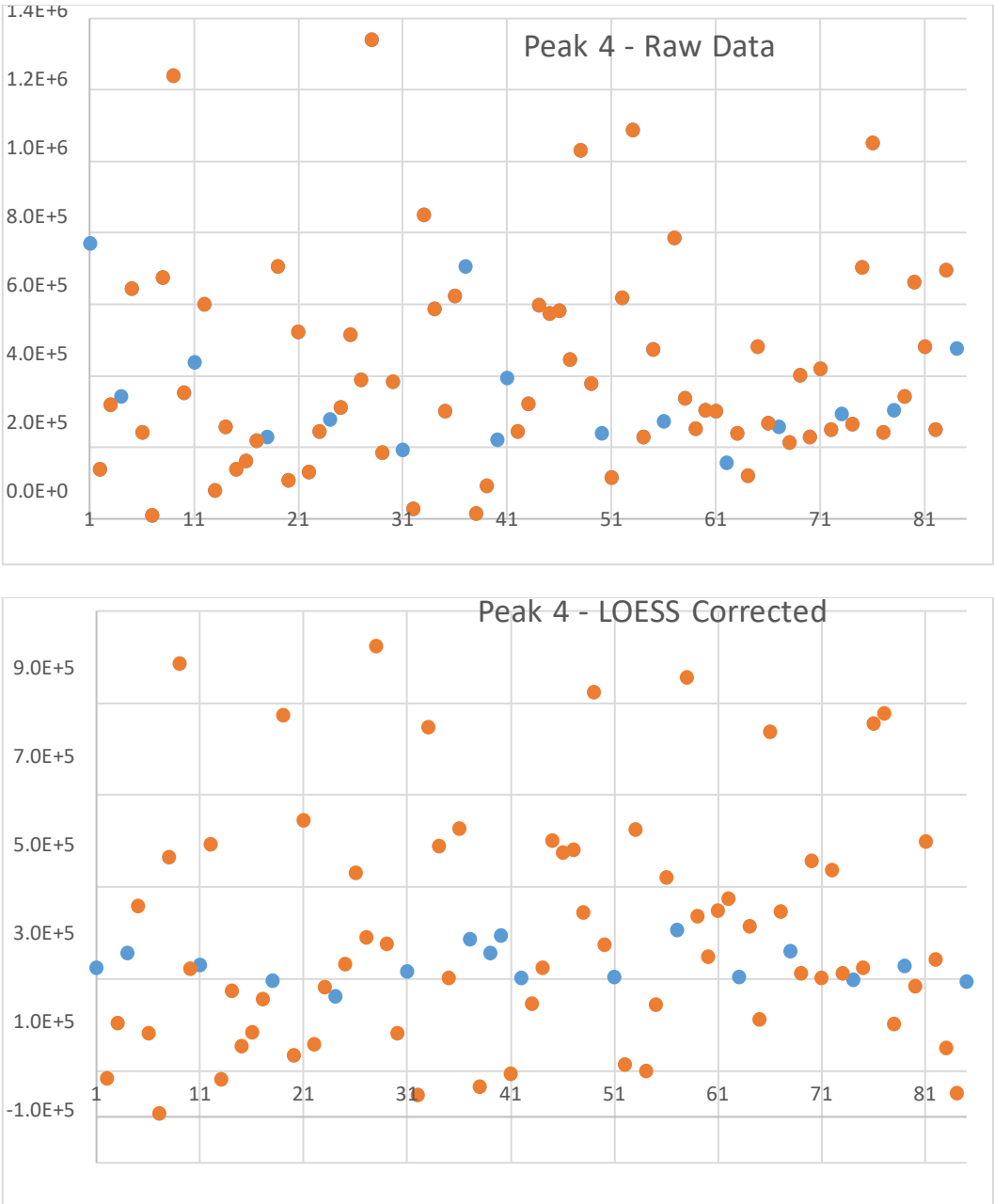



Figure S-14. QC (blue points) and study samples (orange points) signals before (above) and after (below) the LOESS (partial) correction.

The LOESS procedure corrected, albeit not completely, the analytical variations measured in the QC samples (Figure S-4, blue points in Figure S-5), while maintaining the biological variation (orange points in S-5). Indeed, the RSD in the QC samples went respectively from 48 and 25% to 12 and 11%. However, whatever the effectiveness of the procedure, the analytical variations cannot be totally corrected a posteriori. Trying to do so is risky because it would most probably lead to overfitting and to the reduction of the biological variation of interest. Since it is also more effective and less risky to correct for small variations, a prior optimization of the analytical method is very advisable.

## S-12. Data Scaling.

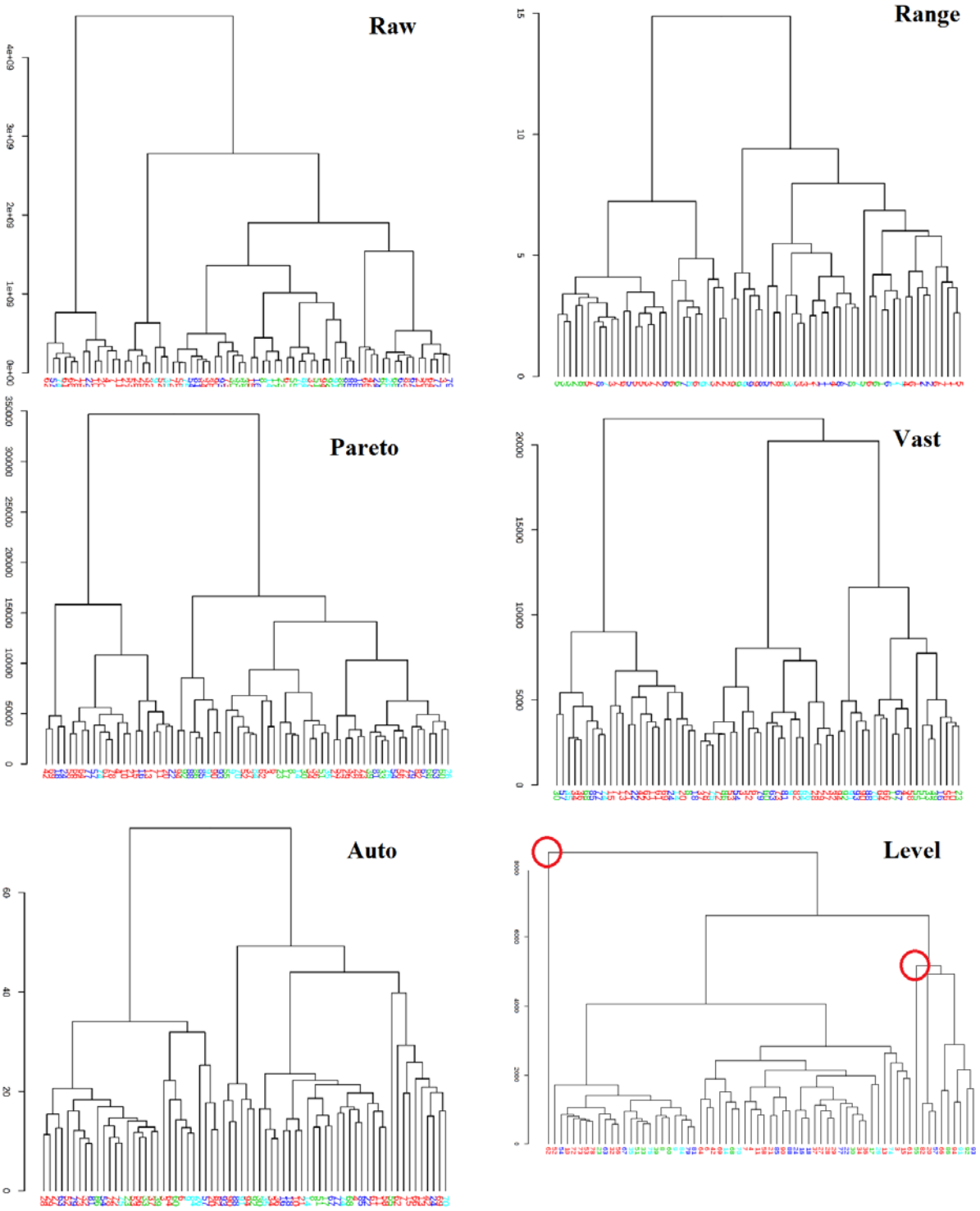



Figure S-15. HCA plots for the various scaling methods. Healthy controls are in red, high endoscopic activity samples in green, low endoscopic activity samples in royal blue and quiescent samples in light blue. The red circles emphasize the potential outliers and their dissimilarity.

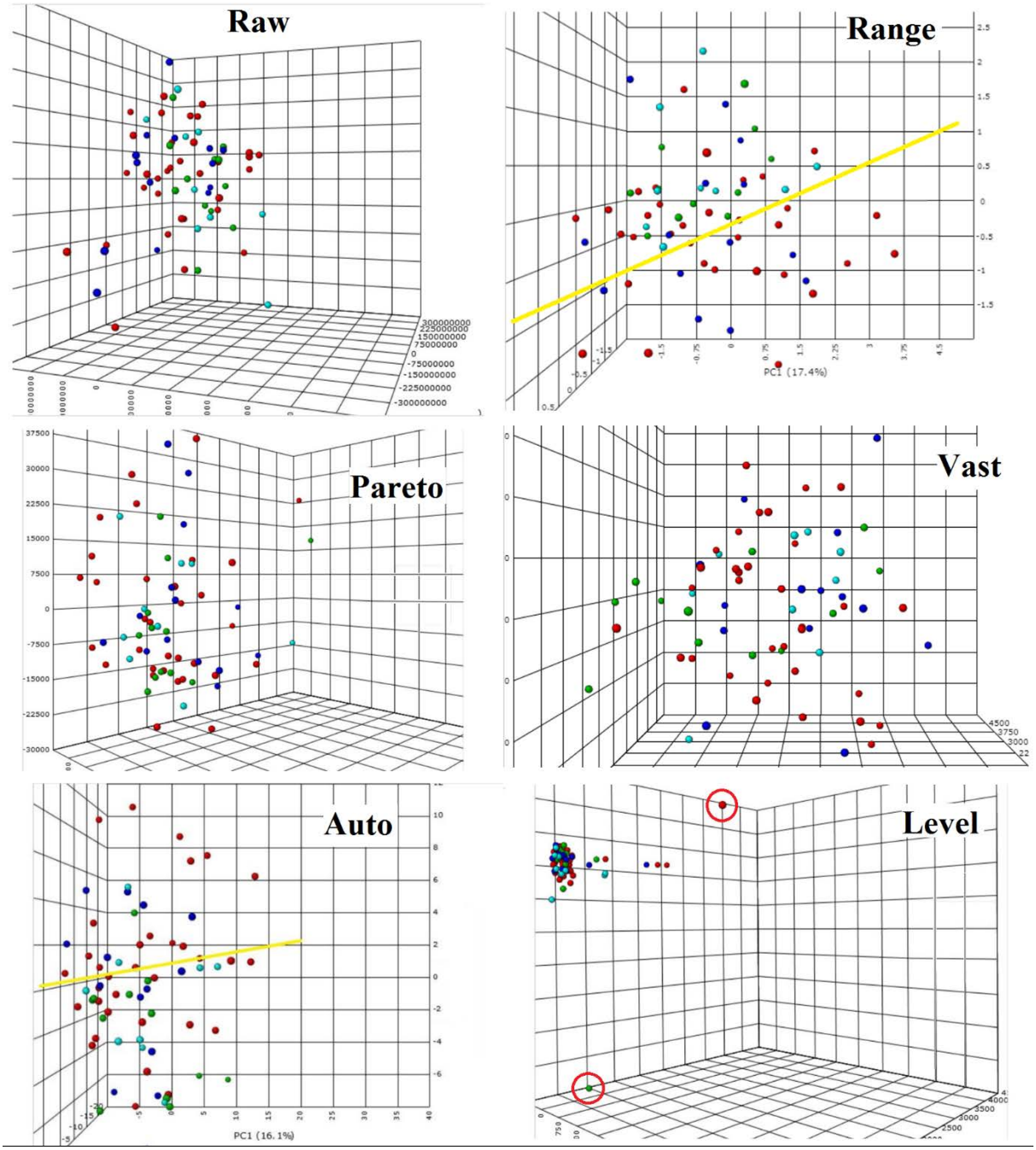



Figure S-16. PCA 3D plots for the various scaling methods. Healthy controls are in red, high endoscopic activity samples in green, low endoscopic activity samples in royal blue and quiescent samples in light blue. The red circles emphasize the potential outliers.The yellow lines emphasize an early separation.

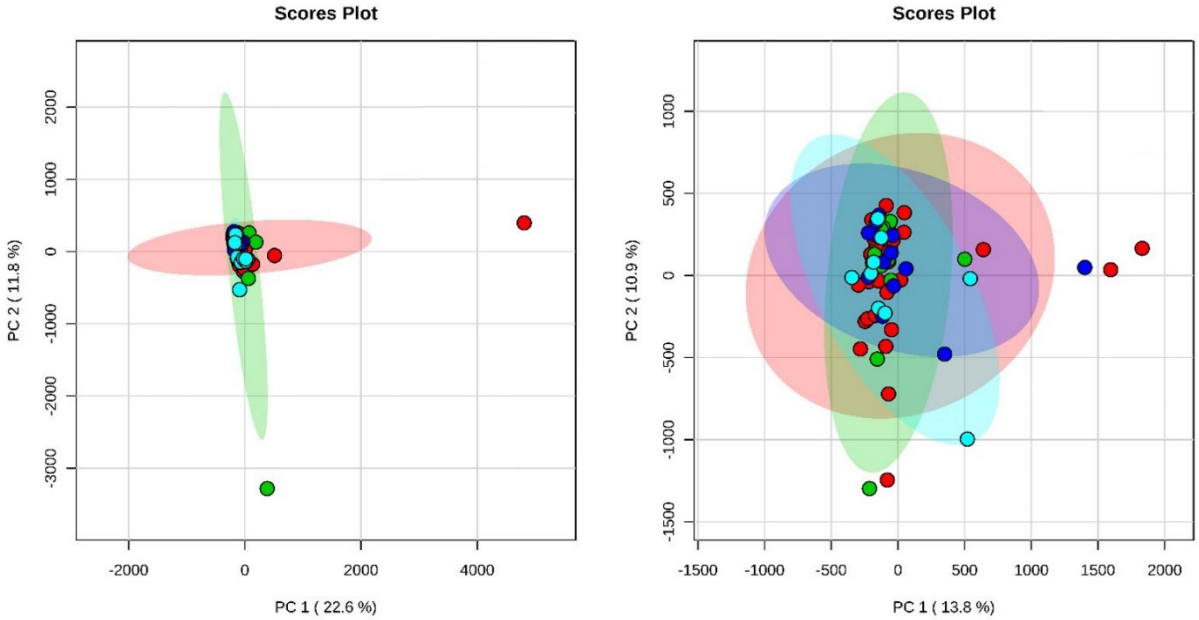



Figure S-17. PCA 2D plots for level scaling with all samples (left) and without the two alleged outliers (right), where new ones appear. Healthy controls are in red, high endoscopic activity samples in green, low endoscopic activity samples in royal blue and quiescent samples in light blue.

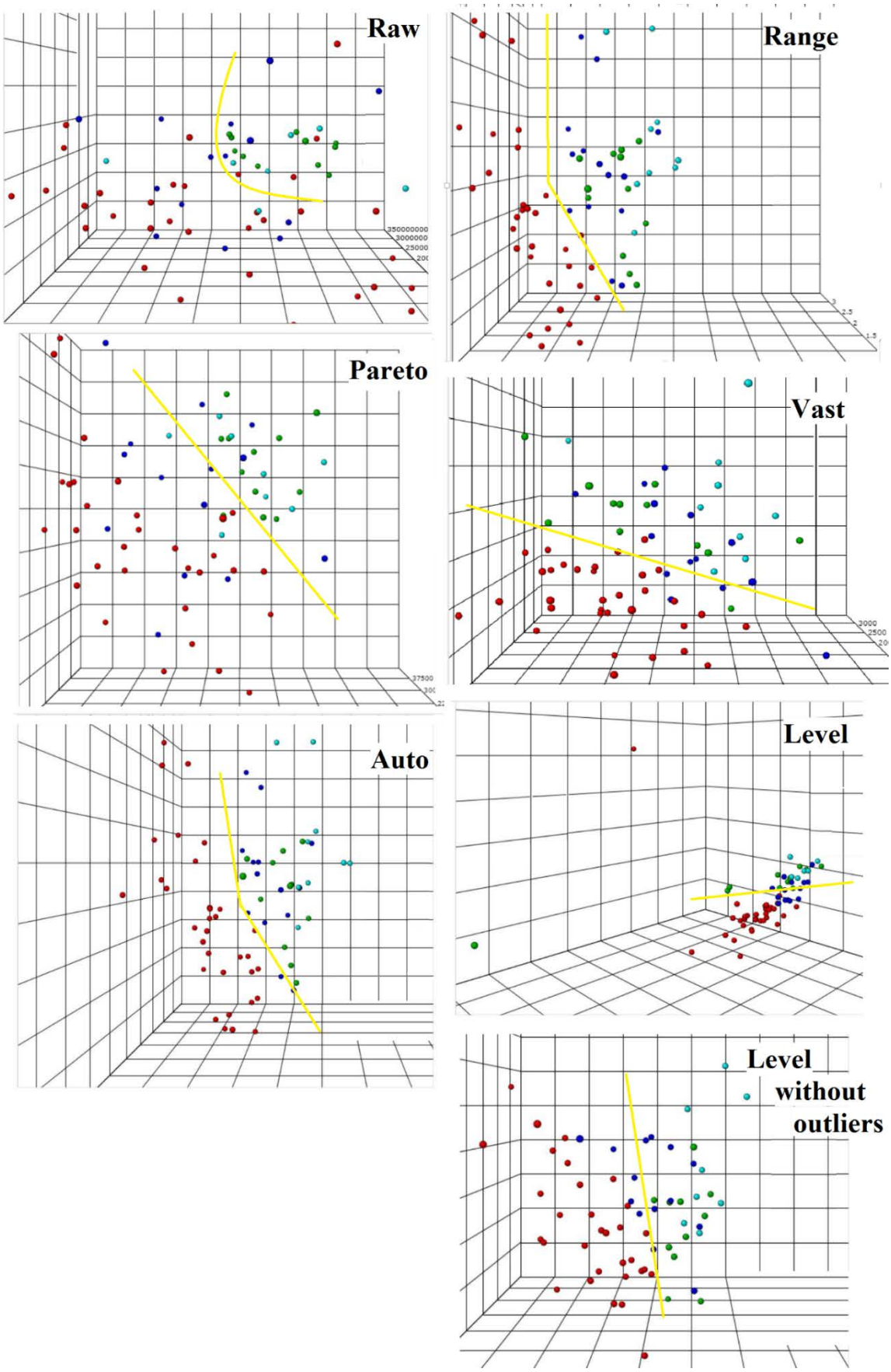



Figure S-18. PLS-DA 3D plots for the various scaling methods. Healthy controls are in red, high endoscopic activity samples in green, low endoscopic activity samples in royal blue and quiescent samples in light blue. The yellow lines emphasize an early separation.

| **Raw** | OOB 0.456 | **Auto** | OOB 0.426 | **Pareto** | OOB 0.426 | **Range** | OOB 0.412 | **Vast** | OOB 0.397 |
|---|---|---|---|---|---|---|---|---|---|
| Peak | MDA | Peak | MDA | Peak | MDA | Peak | MDA | Peak | MDA |
| 56 | 0.013 | 144 | 0.017 | 144 | 0.017 | 144 | 0.015 | 144 | 0.017 |
| 144 | 0.012 | 56 | 0.015 | 56 | 0.013 | 56 | 0.012 | 56 | 0.016 |
| 87 | 0.008 | 90 | 0.010 | 90 | 0.010 | 90 | 0.010 | 90 | 0.010 |
| 114 | 0.008 | 87 | 0.006 | 11 | 0.008 | 11 | 0.007 | 81 | 0.007 |
| 90 | 0.008 | 11 | 0.006 | 154 | 0.005 | 154 | 0.006 | 114 | 0.006 |
| 11 | 0.007 | 114 | 0.006 | 114 | 0.005 | 87 | 0.005 | 11 | 0.006 |
| 154 | 0.006 | 154 | 0.005 | 63 | 0.005 | 63 | 0.004 | 87 | 0.006 |
| 122 | 0.005 | 63 | 0.005 | 87 | 0.004 | 114 | 0.004 | 63 | 0.005 |
| 63 | 0.004 | 81 | 0.004 | 54 | 0.004 | 161 | 0.004 | 109 | 0.005 |
| 29 | 0.004 | 122 | 0.004 | 122 | 0.004 | 14 | 0.003 | 154 | 0.005 |
| 14 | 0.004 | 14 | 0.004 | 14 | 0.003 | 69 | 0.003 | 122 | 0.004 |
| 81 | 0.003 | 54 | 0.003 | 161 | 0.003 | 81 | 0.003 | 161 | 0.003 |
| 109 | 0.003 | 136 | 0.003 | 182 | 0.003 | 105 | 0.003 | 29 | 0.002 |
| 136 | 0.003 | 68 | 0.003 | 81 | 0.002 | 29 | 0.003 | 133 | 0.002 |
| 75 | 0.002 | 105 | 0.003 | 4 | 0.002 | 122 | 0.002 | 136 | 0.002 |

Table S-12. Effect of scaling on random forests' (RF) ability to highlight the most significant metabolites. All methods gave similar results, either for the significant metabolites or their associated mean decrease accuracy values (MDA). All methods also improved the significance of the potential biomarkers (higher MDA and lower out-of-bag error, OOB) in comparison to the raw data.

| | PLS-DA $Q^2$ | PLS-DA $R^2$ |
|---|---|---|
| Raw | 0.01 | 0.35 |
| Pareto | 0.02 | 0.47 |
| Auto | **0.28** | 0.77 |
| Range | **0.31** | 0.77 |
| Vast | 0.17 | 0.73 |
| Level | Outliers | |

Table S-13. PLS-DA $Q^2$ and $R^2$ for the various scaling methods. Auto and range scalings provided the best performances, calculated on all variables.

| Variation explained | PCA (%) | PLS-DA (%) |
|---|---|---|
| Raw | 70 | 61 |
| Pareto | 42 | 34 |
| Auto | **24** | **18** |
| Range | **26** | **20** |
| Vast | 37 | **21** |
| Level | Outliers | |

Table S-14. Percentages of variation explained by the first two axis in PCA and PLS-DA plots, that are lower for auto and range scalings. Again, the models were constructed on all variables.

**S-13. Statistics for biomarker research. A. Statistical treatment details. B. Significance.**

(A) Statistical treatment details

ANOVA consisted of Welch ANOVA and Kruskal-Wallis test. Random forests (RF) produced 500 trees and applied both[S-32] Gini Index[S-33] and Gain Ratio in order to improve the splitting and the variable selection[S-34]. PLS-DA performed cross-validation to define the optimal number of latent variables. PLS-DA and OPLS-DA performed permutation testing (2000 permutations). Naïve Bayes did not assume the homoscedasticity of the variables. SVM used a polynomial kernel that was adapted (often linear) to maximize the separation. Neural network (NN) used no hidden layer because this configuration revealed to be more efficient to limit overfitting as well as to select the significant variables[S-35]. Multivariate ROC curves used SVM, RF, PLS-DA and logistic regression algorithms, with Monte-Carlo cross-validation. Since the global statistical power is directly linked to the p-value[S-36, S-37], it was interpreted in a similar way. Here, it was only partially informative because the calculation implemented in MetaboAnalyst, the only free resource we found, suffered from the non-consideration of the covariance between the biomarkers[S-38].

(B) Significance

To be able to simultaneously take into account different statistical tests, thresholds corresponding to moderate and strong significance were defined for each of them, based on general statistical knowledge and specific sources. A variable meeting one or both thresholds was attributed respectively one or two points. The points were summed and the metabolites were ranked and selected accordingly. The thresholds were: 0.05 and 0.01 for any test giving a p-value, such as Pearson r, Spearman ρ, Kendall τ, Welch or Mann-Whitney/Kruskal-Wallis ANOVA and Fisher ratios (FR). 2 and 3 for PLS-DA VIP. 0.0020 and 0.0025 for RF[S-39]. 0.7 and 0.8 for AUC[S-40]. 1.5 and 2 for fold change (FC). 70 and 80% for univariate two-tailed statistical power. To be selected and to be selected among the first ten variables for feature selection methods and OPLS-DA. To be among the most significant 10% or 5% based on

weight or distance for Bayes classifier, NN and SVM. For the three CD subgroups separation, the biological proximity of the samples made more difficult their separation and therefore lowered the statistical values. The PLS-DA VIP thresholds were adapted to 1.4 and 1.7. Bonferroni correction for multiple testing[S-40] was mostly informative. Indeed, it was interesting to see which metabolites met this conservative threshold for significance but we did not want to restrict the selection process. Especially since the ability of the metabolites was further assessed at the following performance assessment step.

**S-14. Data control and selection process for the Crohn's disease samples and the healthy controls.**

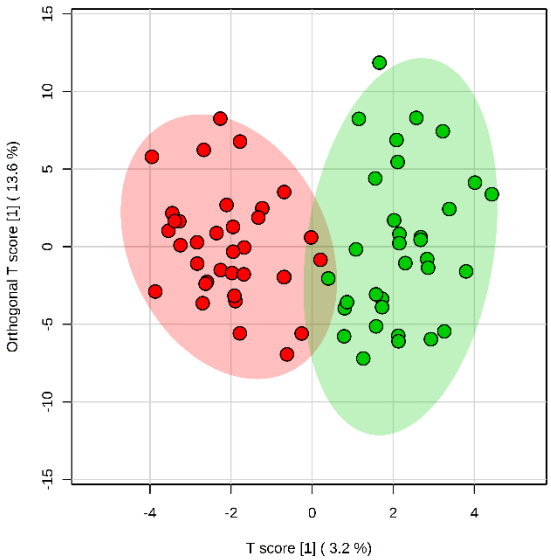



Figure S-19. OPLS-DA plots for the separation between CD (red) and HC (green) samples, constructed on all the 183 metabolites taken in the final dataset.

The good discrimination achieved in Figure S-19 shows the existence, in the data set, of metabolic information capable to separate the Crohn's disease (CD) samples from the healthy controls (HC). Thus the interest to try to extract it through the selection process.

The main statistical values for the candidate biomarkers are given in the article.

| 15 tests | Score |
|---|---|
| Peak | / 30 |
| 144 | 24 |
| 114 | 23 |
| 29 | 22 |
| 69 | 21 |
| 154 | 20 |
| 56 | 19 |
| 87 | 19 |
| 63 | 18 |
| 81 | 15 |
| 11 | 14 |
| 105 | 14 |
| 54 | 12 |
| 132 | 12 |
| 136 | 11 |
| 109 | 10 |
| 13 | 9 |
| 55 | 9 |
| 122 | 8 |
| 133 | 8 |
| 168 | 8 |
| 4 | 7 |
| 90 | 7 |
| 126 | 7 |
| 97 | 6 |
| 124 | 6 |
| 86 | 5 |

| 8 tests | Score |
|---|---|
| Peak | / 16 |
| 144 | 15 |
| 114 | 14 |
| 56 | 13 |
| 29 | 12 |
| 63 | 12 |
| 69 | 12 |
| 154 | 12 |
| 87 | 11 |
| 81 | 10 |
| 105 | 9 |
| 11 | 7 |
| 54 | 6 |
| 55 | 6 |
| 90 | 5 |
| 132 | 5 |
| 133 | 5 |
| 136 | 5 |
| 168 | 5 |
| 4 | 4 |
| 97 | 4 |
| 109 | 4 |
| 122 | 4 |
| 126 | 4 |
| 161 | 4 |
| 13 | 3 |
| 124 | 3 |

| 15 tests | Score |
|---|---|
| Peak | / 30 |
| 4 | 7 |
| 11 | 14 |
| 13 | 9 |
| 29 | 22 |
| 54 | 12 |
| 55 | 9 |
| 56 | 19 |
| 63 | 18 |
| 69 | 21 |
| 81 | 15 |
| 86 | 5 |
| 87 | 19 |
| 90 | 7 |
| 97 | 6 |
| 105 | 14 |
| 109 | 10 |
| 114 | 23 |
| 122 | 8 |
| 124 | 6 |
| 126 | 7 |
| 132 | 12 |
| 133 | 8 |
| 136 | 11 |
| 144 | 24 |
| 154 | 20 |
| 168 | 8 |

| 8 tests | Score |
|---|---|
| Peak | / 16 |
| 4 | 4 |
| 11 | 7 |
| 13 | 3 |
| 29 | 12 |
| 54 | 6 |
| 55 | 6 |
| 56 | 13 |
| 63 | 12 |
| 69 | 12 |
| 81 | 10 |
| 87 | 11 |
| 90 | 5 |
| 97 | 4 |
| 105 | 9 |
| 109 | 4 |
| 114 | 14 |
| 122 | 4 |
| 124 | 3 |
| 126 | 4 |
| 132 | 5 |
| 133 | 5 |
| 136 | 5 |
| 144 | 15 |
| 154 | 12 |
| 161 | 4 |
| 168 | 5 |

Table S-15. Global significance scores obtained through the aggregation of the individual statistical significance values (see section S-12). Two cases are compared. In the first, all 15 tests were performed and considered (first and third frames). In the second, only the 8 most effective tests, the ones recommended in the main article, were considered (second and fourth frames): FC, ROC, Pearson/Spearman, Kendall, Welch/Mann-Whitney ANOVA, Stepdisc, PLS-DA, RF. The first two frames are ordered according to the global score while the third and fourth ones are ordered according to the peak number ID to facilitate the comparison. The peak number is the one in the final data set, among the 183 final variables.

Table S-15 shows that using only the 8 most effective tests, instead of all 15 tests, did not change the candidate biomarkers that were selected. The peaks with the highest scores were the same in both cases (peaks 86 and 161 had different rankings but also a low significance). For this reason, these 8 tests are recommended to simplify the selection process. Regarding the Bonferroni criterion, many candidates met it for some of the statistical tests (Table S-16). The most significant even met it for all the 4 tests considered.

| Bonferroni | |
|---|---|
| CB | / 4 tests |
| 1 | 4 |
| 2 | 4 |
| 3 | 4 |
| 4 | 3 |
| 5 | 4 |
| 6 | 0 |
| 7 | 1 |
| 8 | 3 |
| 9 | 1 |
| 10 | 1 |
| 11 | 2 |
| 12 | 0 |
| 13 | 0 |
| 14 | 0 |
| 15 | 0 |
| 17 | 0 |
| 18 | 0 |
| 19 | 2 |
| 20 | 3 |
| 22 | 4 |

Table S-16. Candidates meeting the significance threshold of 0.05 after Bonferroni correction for multiple testing (0.05/183) for four statistical tests: Pearson/Spearman and Kendall correlations, Welch/Kruskal-Wallis ANOVA and FR.

| | PLS-DA | OPLS-DA | RF |
|---|---|---|---|
| Peak Number | 137 | 46 | 161 |
| | 16 | 12 | 14 |
| | 46 | 16 | 130 |
| | 12 | 182 | 12 |
| | 76 | 162 | 75 |
| | 182 | 137 | 160 |
| | 3 | 143 | 41 |

Table S-17. Additional biomarker research through PLS-DA, OPLS-DA and RF. The grey cells indicate the metabolites highlighted multiple times. The peak number is the one in the final data set, among the 183 final variables.

| Peak Number | Ranking |
|---|---|
| 46 | 23 |
| 12 | 27 |
| 137 | 28 |
| 16 | 29 |
| 182 | 86 |

Table S-18. Ranking of the potential additional candidate biomarkers in the initial selection process.

Regarding the search for additional candidates, some metabolites were highlighted by at least two methods (Table S-17). Nevertheless, their boxplots were not ‘better’ than the ones of the selected candidate biomarkers (Figures S-21) and they were not selected. This supported the comprehensiveness and efficiency of the selection process. Just as did the comparison of the OPLS-DA plots before and

after the selection where a degradation of the separation indicated that a large part of the relevant information had been successfully extracted (Figure S-20).

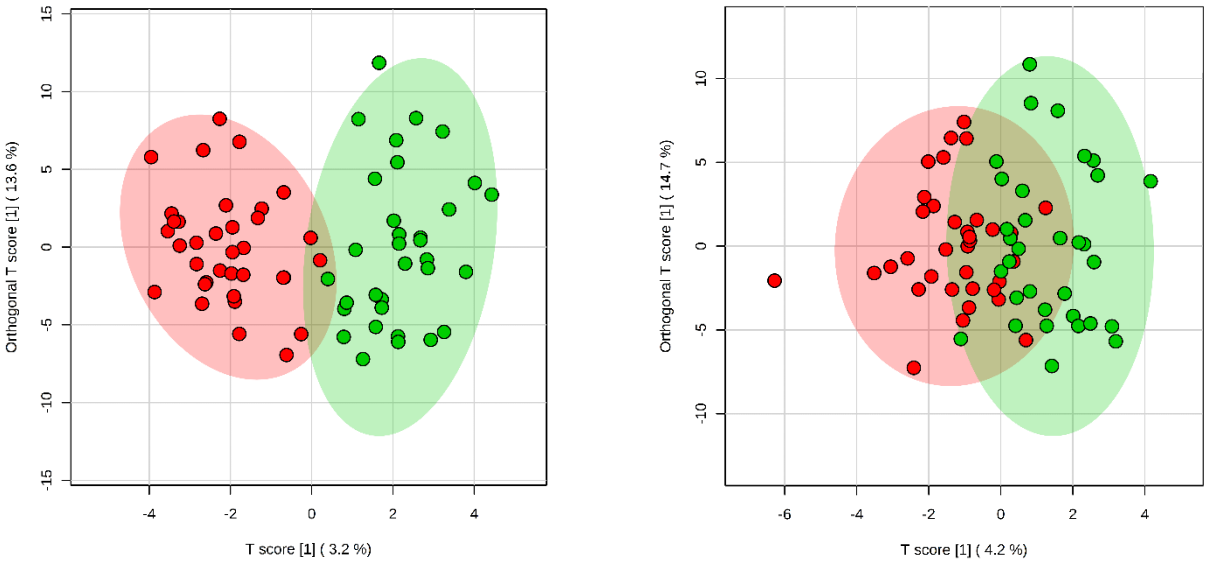
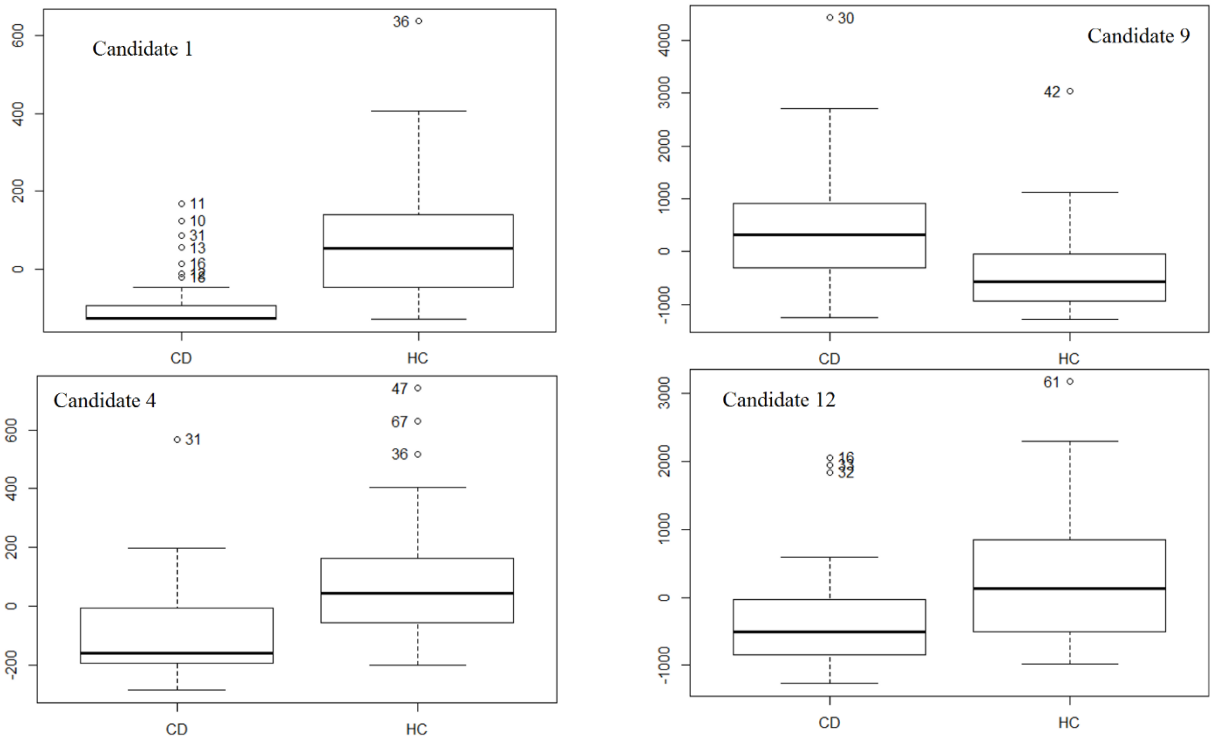



Figure S-20. OPLS-DA plots constructed on all 183 metabolites of the final dataset (left) and on the 161 metabolites left after the selection of the candidate biomarkers (right).



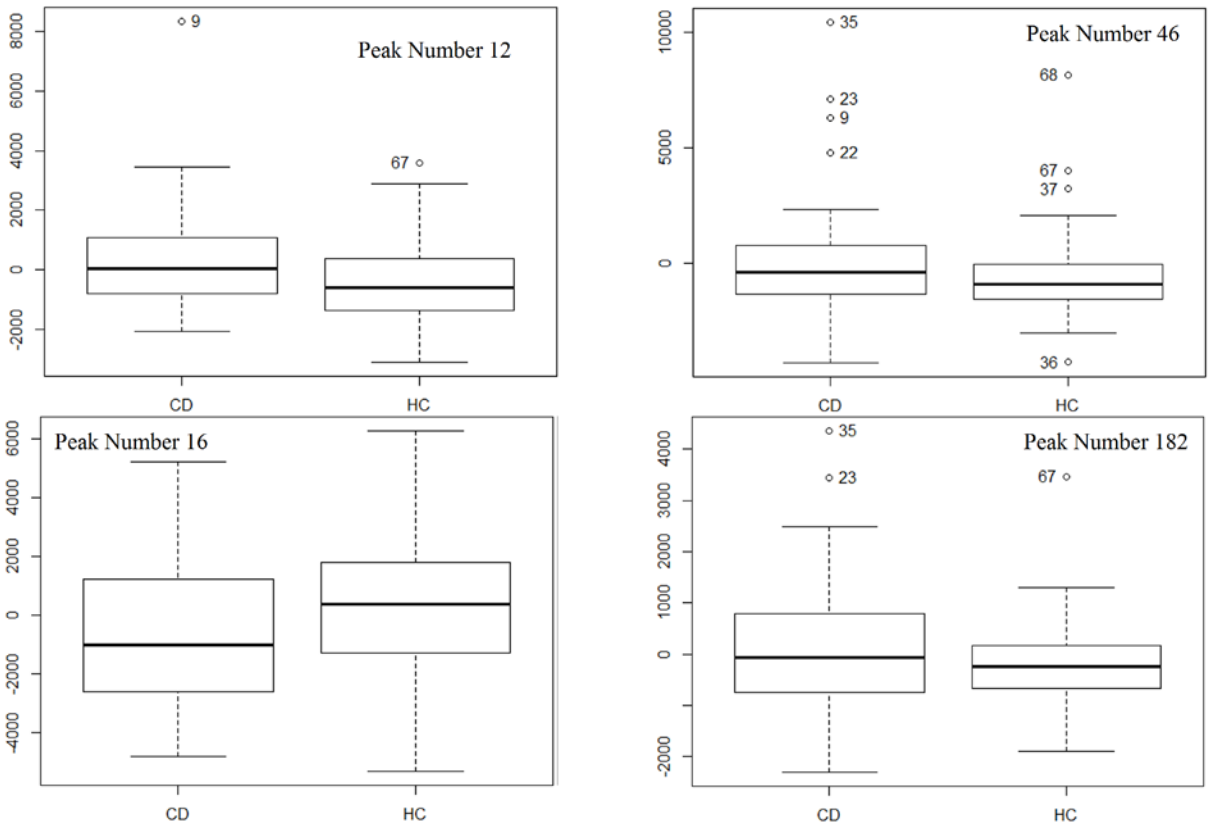



Figure S-21. Examples of boxplots for some of the candidate biomarkers (4 boxplots above) and for the potential additional ones (4 boxplots below).

**S-15. Data control and selection process for the three subgroups of Crohn's disease samples.**

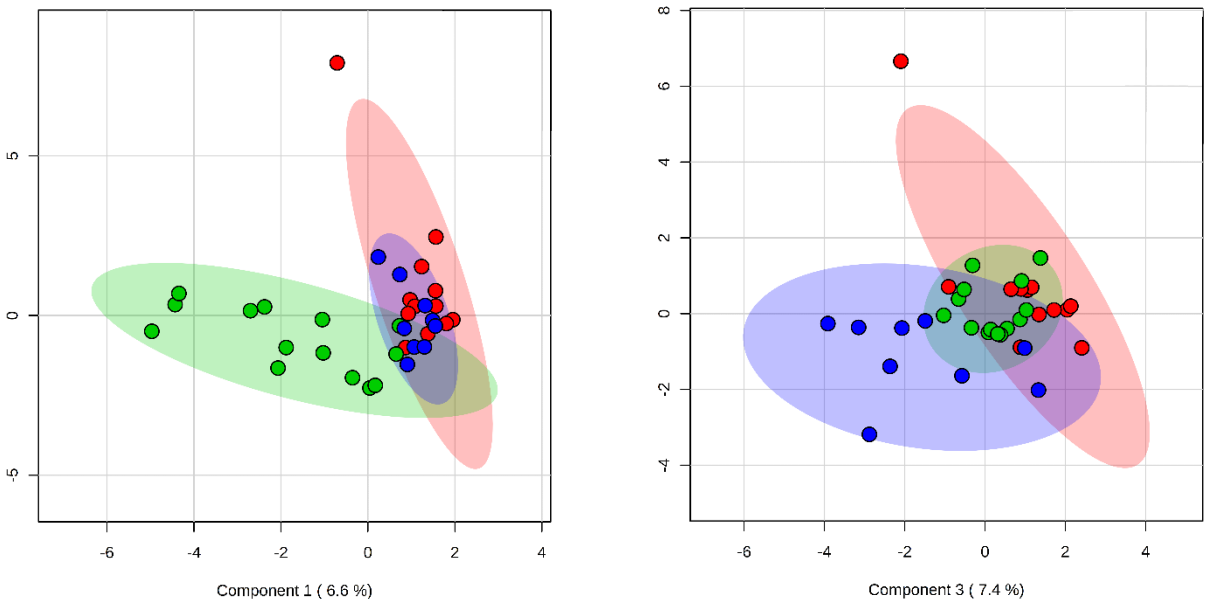



Figure S-22. sPLS-DA plots (latent variables 1 and 2 on the left, 3 and 4 on the right) for the separation between CD subgroups samples. High endoscopic activity samples are in red, low endoscopic activity samples in green and quiescent samples in blue.

Figure S-22 shows the existence of metabolic information capable to separate the CD subgroups. However, the difficulty to completely do so, in comparison to the direct separation with the healthy controls, is obvious. This is supported by the fact that, in the selection process (Table S-19), only the candidate highlighted as the most significant (candidate 11) met the Bonferroni criterion, for only one of the two univariate tests considered. The high endoscopic activity sample outside the 95% confidence ellipse in sPLS-DA will be investigated in a future biological interpretation of the results.

| Candidate | Normality | Stepdisc | Fisher p-value | Welch / K-W | PLS-DA VIP | RF MDA | Sense of variation |
|---|---|---|---|---|---|---|---|
| 11 | 0.00 | | **0.00** | **0.00** | **1.9** | **0.0039** | High > Quiescent > Low |
| 23 | 0.00 | | **0.01** | **0.01** | **1.7** | **0.0028** | High > Quiescent > Low |
| 1 | 0.00 | | **0.02** | **0.04** | **1.5** | **0.0099** | Low > High > Quiescent |
| 15 | 0.46 | | **0.04** | **0.05** | 1.3 | **0.0022** | High > Quiescent > Low |
| 18 | 0.20 | | **0.01** | **0.01** | **1.8** | **0.0078** | Quiescent > Low > High |
| 24 | 0.01 | | **0.00** | **0.04** | **2.0** | 0.0012 | Quiescent > Low > High |
| 25 | 0.00 | | **0.01** | **0.04** | **1.5** | **0.0097** | Low > High > Quiescent |
| 8 | 0.00 | | **0.01** | **0.03** | **1.5** | **0.0071** | Low > High > Quiescent |
| 5 | 0.02 | | 0.06 | **0.04** | **1.6** | 0.0008 | High > Quiescent > Low |
| 14 | 0.00 | | 0.16 | 0.10 | 1.2 | 0.0013 | Low > Quiescent > High |
| 7 | 0.15 | | **0.03** | **0.04** | 1.3 | **0.0029** | Low > High > Quiescent |
| 26 | 0.00 | | 0.34 | 1.00 | 0.9 | 0.0014 | High > Quiescent > Low |
| 9 | 0.00 | | **0.02** | **0.02** | 1.3 | **0.0034** | High > Quiescent > Low |
| 27 | 0.00 | | 0.12 | | 1.1 | 0.0009 | Quiescent > High > Low |
| 28 | 0.00 | | **0.01** | **0.01** | **1.6** | 0.0000 | Quiescent > High > Low |
| 6 | 0.00 | | **0.03** | **0.03** | 1.3 | **0.0033** | Quiescent > High > Low |
| 29 | 0.00 | **5** | 0.67 | 0.59 | 0.9 | 0.0000 | Quiescent > High > Low |
| 30 | 0.01 | **2** | 0.08 | 0.23 | **1.4** | 0.0003 | Quiescent > High > Low |
| 31 | 0.00 | | 0.05 | 0.07 | 1.4 | 0.0008 | Quiescent > High > Low |
| 2 | 0.00 | | **0.02** | **0.05** | 1.4 | 0.0006 | Low > High > Quiescent |
| 32 | 0.00 | | **0.02** | 0.06 | **1.4** | **0.0039** | Low > High > Quiescent |
| 33 | 0.23 | **7** | **0.05** | 0.10 | **1.4** | 0.0003 | Quiescent > High > Low |
| 34 | 0.57 | | 0.05 | **0.01** | 1.3 | 0.0011 | Low > High > Quiescent |
| 35 | 0.01 | | **0.05** | 0.18 | **1.4** | 0.0008 | Quiescent > High > Low |
| 36 | 0.01 | | 0.05 | **0.03** | 1.4 | 0.0005 | Quiescent > High > Low |
| 10 | 0.26 | **12** | 0.05 | 0.08 | 1.3 | **0.0038** | High > Quiescent > Low |
| 37 | 0.00 | **13** | 0.06 | 1.00 | **1.5** | 0.0003 | Quiescent > Low > High |
| 38 | 0.00 | | **0.03** | 0.12 | **1.7** | 0.0010 | Quiescent > High > Low |
| 39 | 0.00 | | **0.03** | 1.00 | **1.5** | 0.0004 | High > Low > Quiescent |

Table S-19. Main statistical values for the candidate biomarkers selected to separate the three CD subgroups. The candidates are ordered according to their significance in the selection process. Univariate methods: Welch and Kruskal-Wallis tests, FR. Multivariate methods: Stepdisc feature selection, PLS-DA, RF (mean decrease accuracy). The values meeting the statistical thresholds for significance are in bold.

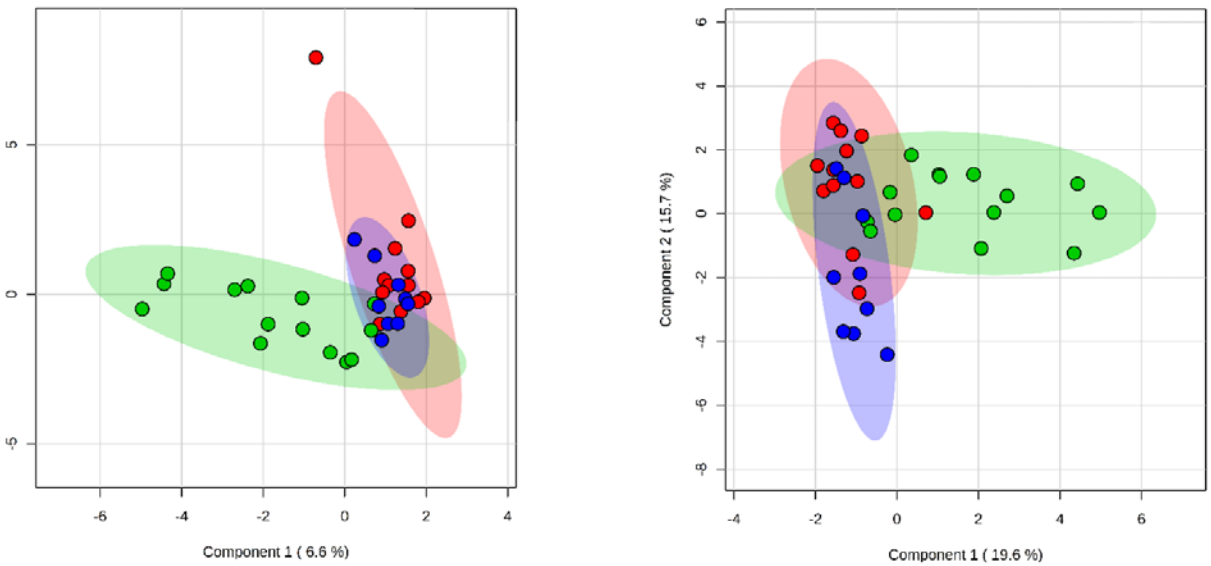



Figure S-23. sPLS-DA plots constructed on the 183 metabolites of the final dataset (left) and on the 154 metabolites left after selection of the candidate biomarkers (right).

Figure S-23, through the clear degradation of the separation when the candidates were removed, indicates that the information relevant for the biological question studied was successfully extracted from the data set by the selection process.

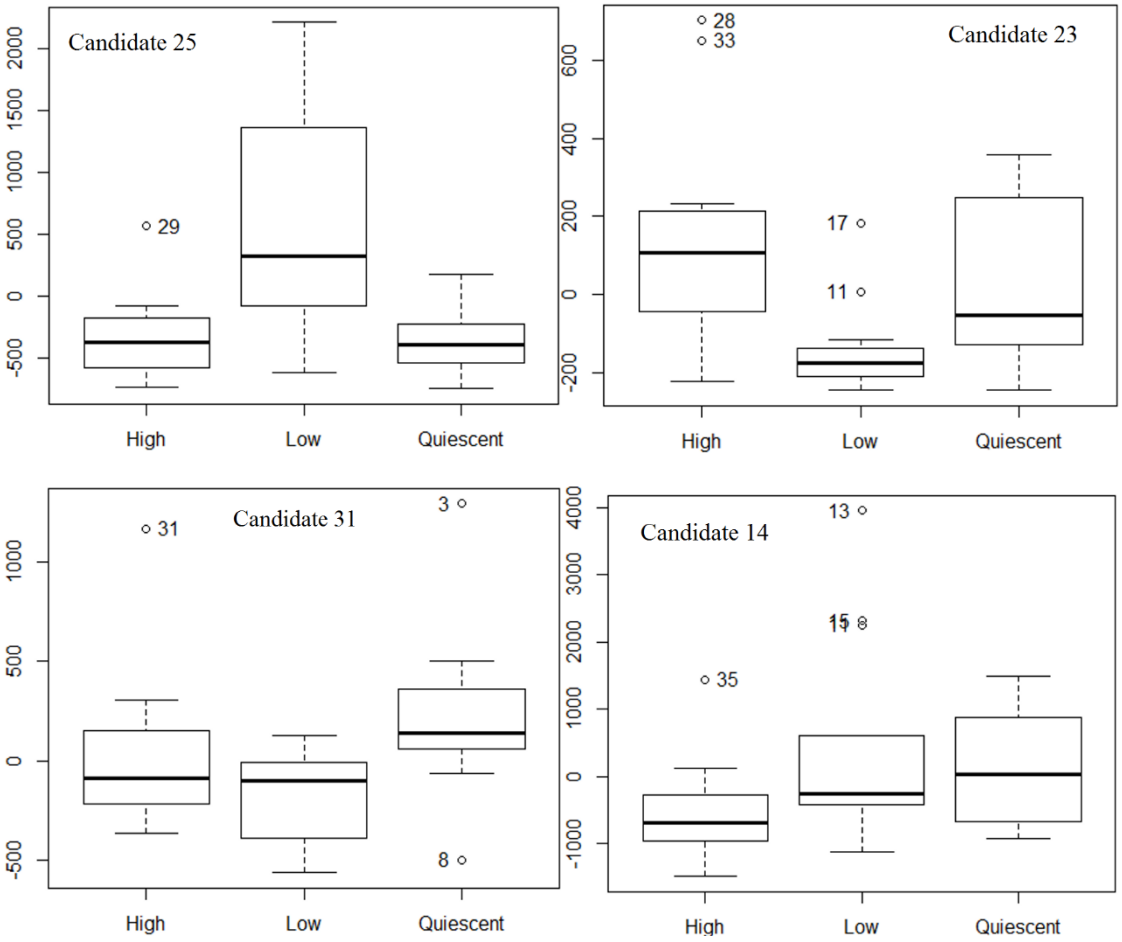



Figure S-24. Examples of boxplots for some of the candidate biomarkers.

**S-16. Bias control for the CD candidates. A. Repartition of the factors between the two biological classes. B. Relationships between the candidate biomarkers and the bias factors. C. Residual separation ability.**

The medications considered were the ones with enough representation in the groups or the ones linked to CD: Elthyrone© and L-Thyroxine© (thyroid regulation), Pantomed© (gastroesophageal reflux and ulcer), Remicade© and Humira© (anti-TNFα) and Imuran© (immunosuppression).

(A) Repartition of the factors between the two biological classes

| % | Gender (F) | Tobacco | Alcohol | Hemolysis | Gastro. reflux | L-Thyroxin | Anti-TNFα | Immunosupp. |
|---|---|---|---|---|---|---|---|---|
| HC | 58 | 18 | 21 | 15 | 27 | 18 | 0 | 0 |
| CD | 60 | 37 | 23 | 14 | 29 | 20 | 51 | 14 |

| Mean | Inj Order | Age | BMI |
|---|---|---|---|
| HC | 43 ± 27 | 44 ± 12 | 25 ± 5 |
| CD | 53 ± 28 | 42 ± 12 | 25 ± 5 |

| Drying | | |
|---|---|---|
| Number | HC | CD |
| Batch 1 | 4 | 4 |
| Batch 2 | 3 | 5 |
| Batch 3 | 6 | 2 |
| Batch 4 | 5 | 3 |
| Batch 5 | 4 | 4 |
| Batch 6 | 3 | 5 |
| Batch 7 | 3 | 5 |
| Batch 8 | 2 | 6 |
| Batch 9 | 3 | 1 |

| Drying | | |
|---|---|---|
| % | HC | CD |
| Batch 1 | 12 | 11 |
| Batch 2 | 9 | 14 |
| Batch 3 | 18 | 6 |
| Batch 4 | 15 | 9 |
| Batch 5 | 12 | 11 |
| Batch 6 | 9 | 14 |
| Batch 7 | 9 | 14 |
| Batch 8 | 6 | 17 |
| Batch 9 | 9 | 3 |

| Extraction | | |
|---|---|---|
| Number | HC | CD |
| Batch 1 | 16 | 16 |
| Batch 2 | 17 | 15 |
| Batch 3 | 0 | 4 |

| Extraction | | |
|---|---|---|
| % | HC | CD |
| Batch 1 | 48 | 46 |
| Batch 2 | 52 | 43 |
| Batch 3 | 0 | 11 |

Table S-20. Repartition of the bias factors between the CD samples and the healthy controls, given by the numbers of samples, the percentages, means and medians.

| | | | | | |
|---|---|---|---|---|---|
| | $Chi^2$ | 0.54 | | $Chi^2$ | 0.33 |
| Drying | Good.-Krusk. Tau | 0.55 | Hemolysis | Good.-Krusk. Tau | 0.33 |
| | Theil U | 0.52 | | Theil U | 0.27 |
| | $Chi^2$ | 0.13 | | $Chi^2$ | 0.91 |
| Extraction | Good.-Krusk. Tau | 0.13 | Gastro. reflux | Good.-Krusk. Tau | 0.91 |
| | Theil U | 0.06 | | Theil U | 0.91 |
| | $Chi^2$ | 0.84 | | $Chi^2$ | 0.85 |
| Gender | Good.-Krusk. Tau | 0.84 | Thyroid | Good.-Krusk. Tau | 0.85 |
| | Theil U | 0.84 | | Theil U | 0.85 |
| | $Chi^2$ | 0.08 | | $Chi^2$ | 0.00 |
| **Tobacco** | Good.-Krusk. Tau | 0.08 | **Anti-TNFα** | Good.-Krusk. Tau | 0.00 |
| | Theil U | 0.08 | | Theil U | 0.00 |
| | $Chi^2$ | 0.87 | | $Chi^2$ | 0.02 |
| Alcohol | Good.-Krusk. Tau | 0.87 | **Immunosupp.** | Good.-Krusk. Tau | 0.03 |
| | Theil U | 0.87 | | Theil U | 0.01 |

| | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|---|
| Inject. Order | 0.00 | 0.11 | 0.05 | 0.11 | 0.10 |
| Age | 0.03 | 0.44 | 0.30 | 0.44 | 0.40 |
| BMI | 0.00 | 0.92 | 0.80 | 0.92 | 0.83 |

Table S-21. p-values for the potential imbalances. The grey cells indicate the significant values. The factors with a statistical imbalance are in bold.

Anti-TNFα and immunosuppression medications had p-values < 0.05 and were potential biases for the selected candidates (Tables S-20 and S-21).

(B) Relationships between the candidate biomarkers and the bias factors

| **Anti-TNFα** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|---|
| **1** | 0.00 | 0.05 | 0.01 | 0.02 | 0.05 |
| 2 | 0.00 | 0.64 | 0.40 | 0.65 | 0.64 |
| **3** | 0.00 | 0.08 | 0.00 | 0.06 | 0.08 |
| **4** | 0.00 | 0.03 | 0.02 | 0.01 | 0.03 |
| 5 | 0.00 | 0.19 | 0.01 | 0.24 | 0.19 |
| 6 | 0.00 | 0.44 | 0.05 | 0.43 | 0.44 |
| **7** | 0.02 | 0.06 | 0.02 | 0.08 | 0.06 |
| 8 | 0.00 | 0.79 | 0.61 | 0.79 | 0.79 |
| **9** | 0.00 | 0.06 | 0.03 | 0.10 | 0.06 |
| 10 | 0.00 | 0.17 | 0.03 | 0.15 | 0.17 |
| 11 | 0.00 | 0.82 | 0.75 | 0.83 | 0.82 |
| **12** | 0.00 | 0.08 | 0.02 | 0.08 | 0.08 |
| 13 | 0.00 | 0.29 | 0.08 | 0.23 | 0.29 |
| 14 | 0.00 | 0.32 | 0.16 | 0.25 | 0.32 |
| 15 | 0.06 | 0.66 | 0.59 | 0.66 | 0.58 |
| **16** | 0.00 | 0.01 | 0.01 | 0.03 | 0.01 |
| 17 | 0.82 | 0.38 | 0.28 | 0.38 | 0.36 |
| 18 | 0.03 | 0.46 | 0.19 | 0.49 | 0.46 |
| 19 | 0.00 | 0.32 | 0.07 | 0.15 | 0.32 |
| **20** | 0.00 | 0.10 | 0.00 | 0.02 | 0.10 |
| 21 | 0.00 | 0.52 | 0.42 | 0.54 | 0.52 |
| **22** | 0.00 | 0.03 | 0.00 | 0.05 | 0.03 |

| **Immuno.** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|---|
| **1** | 0.00 | 0.07 | 0.01 | 0.00 | 0.07 |
| 2 | 0.00 | 0.18 | 0.11 | 0.13 | 0.18 |
| **3** | 0.00 | 0.17 | 0.00 | 0.00 | 0.17 |
| **4** | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 |
| **5** | 0.00 | 0.04 | 0.01 | 0.00 | 0.04 |
| 6 | 0.00 | 0.04 | 0.25 | 0.31 | 0.04 |
| **7** | 0.02 | 0.05 | 0.02 | 0.08 | 0.05 |
| **8** | 0.00 | 0.07 | 0.00 | 0.00 | 0.07 |
| **9** | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 |
| 10 | 0.00 | 0.30 | 0.40 | 0.51 | 0.30 |
| **11** | 0.00 | 0.06 | 0.00 | 0.00 | 0.06 |
| 12 | 0.00 | 0.92 | 0.97 | 0.94 | 0.92 |
| 13 | 0.00 | 0.26 | 0.10 | 0.02 | 0.26 |
| **14** | 0.00 | 0.23 | 0.04 | 0.06 | 0.23 |
| **15** | 0.06 | 0.01 | 0.00 | 0.01 | 0.01 |
| 16 | 0.00 | 0.49 | 0.16 | 0.45 | 0.49 |
| 17 | 0.82 | 0.19 | 0.11 | 0.18 | 0.26 |
| 18 | 0.03 | 0.14 | 0.13 | 0.23 | 0.14 |
| **19** | 0.00 | 0.52 | 0.02 | 0.04 | 0.52 |
| 20 | 0.00 | 0.38 | 0.08 | 0.08 | 0.38 |
| 21 | 0.00 | 0.86 | 0.95 | 0.84 | 0.86 |
| 22 | 0.00 | 0.24 | 0.61 | 0.72 | 0.24 |

Table S-22. p-values for the relationships between the candidate biomarkers and both anti-TNFα and immunosuppression medications. The grey cells indicate the significant values. The candidates significantly linked

to the potential bias factors are in bold. P/Sp stands for Pearson or Spearman correlation, according to the normality of the variable.

Candidates 1, 3, 4, 7, 9, 12, 16, 20, 22 and 1, 3-9, 11, 14, 15, 19, respectively, were found to be linked to, i.e. were potentially influenced by, the factors, with p-values < 0.05 or 0.01, % FC < 90 or 80, Δ AUC < -0.05 or -0.1 (Table S-22). The next section aimed to evaluate how the bias factors could affect the separation ability of these candidates.

(C) Residual separation ability

Sample Removal

To equilibrate the classes regarding the potentially active bias factors, different combinations were tested. As a result, despite the fact that it left aside a substantial proportion of the samples, all samples with anti-TNFα or immunosuppression medication were removed. This had the advantage to completely equalize the distributions (resulting p-values = 0). Details are given in Tables S-23 and S-24.

| Clinical ID | Class | **Anti-TNFα** |
|---|---|---|
| 228-VH | CD | Yes |
| 354-FD | CD | Yes |
| 356-OL | CD | Yes |
| 366-AT | CD | Yes |
| 368-FP | CD | Yes |
| 390-JT | CD | Yes |
| 404-AC | CD | Yes |
| 407-BM | CD | Yes |
| 425-SM | CD | Yes |
| 426-CF | CD | Yes |
| 429-ID | CD | Yes |
| 434-SLR | CD | Yes |
| 453-LR | CD | Yes |
| 458-SDB | CD | Yes |
| 465-PM | CD | Yes |
| 466-VLC | CD | Yes |
| 488-GA | CD | Yes |
| 508-DJ | CD | Yes |

| **Number** | Before | After |
|---|---|---|
| HC | 33 | 33 |
| CD | 35 | 17 |

Table S-23. Samples removed for the anti-TNFα equilibration and summary of the samples left.

| Clinical ID | Class | **Immunosup.** |
|---|---|---|
| 133-MD | CD | Yes |
| 351-JT | CD | Yes |
| 228-VH | CD | Yes |
| 465-PM | CD | Yes |
| 247-SK | CD | Yes |

| **Number** | Before | After |
|---|---|---|
| HC | 33 | 33 |
| CD | 35 | 30 |

Table S-24. Samples removed for the immunosuppressor equilibration and summary of the samples left.

Results

| Order (/ 22) | **Anti-TNFα** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis | Partial P / Sp | Partial Kendall | FC Ratio | ROC Difference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 203 | 0.06 |
| 3 | 3 | 0.00 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 68 | -0.04 |
| 4 | 4 | 0.00 | 0.02 | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 97 | -0.01 |
| 7 | 7 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 115 | 0.08 |
| 9 | 9 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 100 | 0.05 |
| 11 | 12 | 0.00 | 0.03 | 0.00 | 0.02 | 0.02 | 0.06 | 0.04 | 103 | 0.03 |
| 15 | **16** | 0.00 | 0.47 | 0.31 | 0.46 | 0.40 | 0.35 | 0.31 | 85 | -0.07 |
| 20 | **20** | 0.00 | 0.90 | 0.00 | 0.93 | 0.00 | 0.94 | 0.00 | 79 | -0.03 |
| 22 | **22** | 0.00 | 0.12 | 0.30 | 0.21 | 0.39 | 0.11 | 0.29 | 94 | -0.08 |

| Order (/ 22) | **Immuno.** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis | Partial P / Sp | Partial Kendall | FC Ratio | ROC Difference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 87 | -0.02 |
| 3 | 3 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 89 | -0.02 |
| 4 | 4 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 95 | -0.03 |
| 4 | 5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 97 | -0.02 |
| 6 | 6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 93 | 0.00 |
| 7 | 7 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 96 | -0.02 |
| 8 | 8 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 87 | -0.03 |
| 9 | 9 | 0.00 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 | 90 | -0.04 |
| 11 | 11 | 0.00 | 0.05 | 0.00 | 0.05 | 0.05 | 0.04 | 0.00 | 97 | -0.03 |
| 14 | **14** | 0.00 | 0.11 | 0.01 | 0.09 | 0.09 | 0.09 | 0.02 | 97 | -0.02 |
| 15 | 15 | 0.13 | 0.03 | 0.01 | 0.03 | 0.03 | 0.02 | 0.04 | 95 | -0.03 |
| 19 | **19** | 0.00 | 0.12 | 0.00 | 0.12 | 0.12 | 0.11 | 0.00 | 97 | -0.02 |

Table S-25. Effect of the factors on the separation ability measured through the residual separation capability (p-values for Pearson/Spearman and Kendall correlations, Welch/Kruskal-Wallis test and Fisher ratios) and on the capability variation (fold change percentage and ROC curve delta). The order is the ranking of the candidates in the set of candidate biomarkers. The grey cells indicate the significant p-values (< 0.05) while the candidates in bold are the ones with a low residual separation capability (p-values > 0.1).

| | Order | Candidate | | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis | FC | FC Ratio | Uni ROC | ROC Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Anti-TNFα** | 15 | **16** | Before | 0.00 | 0.03 | 0.01 | 0.03 | 0.03 | 1.29 | 85 | 0.65 | -0.07 |
| | | | After | 0.00 | 0.47 | 0.31 | 0.46 | 0.40 | 1.10 | | 0.58 | |
| | 20 | **20** | Before | 0.00 | 0.30 | 0.00 | 0.29 | 0.00 | 1.31 | 79 | 0.80 | -0.03 |
| | | | After | 0.00 | 0.90 | 0.00 | 0.93 | 0.00 | 1.03 | | 0.77 | |
| | 22 | **22** | Before | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 | 1.30 | 94 | 0.66 | -0.08 |
| | | | After | 0.00 | 0.12 | 0.30 | 0.21 | 0.39 | 1.22 | | 0.58 | |
| | | | | | | | | 0.01 | | | | |
| **Immuno.** | 14 | 14 | Before | 0.00 | 0.05 | 0.00 | 0.05 | 0.03 | 1.31 | 97 | 0.68 | -0.02 |
| | | | After | 0.00 | 0.11 | 0.01 | 0.10 | 0.00 | 1.27 | | 0.66 | |
| | 19 | 19 | Before | 0.00 | 0.08 | 0.00 | 0.03 | 0.00 | 1.88 | 97 | 0.75 | -0.02 |
| | | | After | 0.00 | 0.12 | 0.00 | 0.11 | 0.03 | 1.83 | | 0.73 | |

Table S-26. Comparison of the separation ability (p-values, FC, AUC) before and after the equilibration, for the candidates potentially affected. The grey cells indicate the significant variations. The order is the ranking of the candidates in the set of candidate biomarkers. The candidates with substantial deterioration due to the bias factors are in grey.

The residual separation capability evaluation worked in two steps, with first the direct assessment of the residual ability (Table S-25) and second, for the candidates with p-values > 0.1, an investigation of the variation resulting from the bias effect by comparing the ability with and without the consideration of the possible bias (Table S-26). Here, candidates 16, 20, 22 as well as 14 and 19, respectively, had p-values > 0.1 (Table S-25). Candidate 3, linked to anti-TNFα bias, had a high decrease in fold change but

its other statistical values were quite good. Table S-26 shows that the candidates 16, 20 and 22 were deteriorated in all five tests performed and had a low residual ability. These were therefore removed from the set of candidate biomarkers.

**S-17. Bias Control for the three CD subgroups candidates. A. Repartition of the factors between the biological classes. B. Relationships between the candidate biomarkers and the bias factors. C. Residual separation ability.**

The medications considered were the ones with enough representation in the groups or the ones linked to CD : Elthyrone© and L-Thyroxine© (thyroid regulation), Pantomed© (gastroesophageal reflux and ulcer), Remicade© and Humira© (anti-TNFα) and Imuran© (immunosuppression).

(A) Repartition of the factors between the biological classes

| % | Gender (F) | Tobacco | Alcohol | Hemolysis | Gastro. reflux | L-Thyroxin | Anti-TNFα | Immunosupp. |
|---|---|---|---|---|---|---|---|---|
| High | 50 | 42 | 8 | 8 | 33 | 8 | 50 | 33 |
| Low | 64 | 43 | 21 | 7 | 29 | 21 | 71 | 0 |
| Quiescent | 67 | 22 | 44 | 33 | 22 | 33 | 22 | 11 |

| **Mean** | Inj Order | Age | BMI |
|---|---|---|---|
| High | 47 ± 26 | 38 ± 16 | 22 ± 3 |
| Low | 59 ± 28 | 39 ± 9 | 26 ± 5 |
| Quiescent | 53 ± 32 | 51 ± 8 | 27 ± 5 |

| **Drying** % | High | Low | Quiescent |
|---|---|---|---|
| Batch 1 | 8 | 14 | 11 |
| Batch 2 | 25 | 14 | 0 |
| Batch 3 | 0 | 7 | 11 |
| Batch 4 | 8 | 14 | 0 |
| Batch 5 | 0 | 21 | 11 |
| Batch 6 | 17 | 7 | 22 |
| Batch 7 | 8 | 14 | 22 |
| Batch 8 | 25 | 7 | 22 |
| Batch 9 | 8 | 0 | 0 |

| **Extraction** % | High | Low | Quiescent |
|---|---|---|---|
| A | 58 | 36 | 44 |
| B | 33 | 57 | 33 |
| C | 8 | 7 | 22 |

| **Location** % | High | Low | Quiescent |
|---|---|---|---|
| Ileitis | 17 | 21 | 56 |
| Colitis | 25 | 14 | 0 |
| Ileocolitis | 33 | 43 | 0 |
| Pancolitis | 8 | 14 | 11 |
| Other | 17 | 7 | 33 |

Table S-27. Repartition of the bias factors considered between the three CD groups (high, low and quiescent endoscopic activities) given by the numbers of samples, the percentages, the means and the medians.

| | | | | | |
|---|---|---|---|---|---|
| | $Chi^2$ | 0.71 | | $Chi^2$ | 0.17 |
| Drying | Good.-Krusk. Tau | 0.72 | Hemolysis | Good.-Krusk. Tau | 0.22 |
| | Theil U | 0.42 | | Theil U | 0.21 |
| | $Chi^2$ | 0.54 | | $Chi^2$ | 0.86 |
| Extraction | Good.-Krusk. Tau | 0.55 | Gastro. reflux | Good.-Krusk. Tau | 0.87 |
| | Theil U | 0.58 | | Theil U | 0.85 |
| | $Chi^2$ | 0.68 | | $Chi^2$ | 0.36 |
| Gender | Good.-Krusk. Tau | 0.68 | Thyroid | Good.-Krusk. Tau | 0.40 |
| | Theil U | 0.68 | | Theil U | 0.34 |
| | $Chi^2$ | 0.56 | | $Chi^2$ | 0.01 |
| Tobacco | Good.-Krusk. Tau | 0.61 | **Anti-TNFα** | Good.-Krusk. Tau | 0.01 |
| | Theil U | 0.54 | | Theil U | 0.01 |
| | $Chi^2$ | 0.15 | | $Chi^2$ | 0.05 |
| Alcohol | Good.-Krusk. Tau | 0.19 | **Immunosupp.** | Good.-Krusk. Tau | 0.04 |
| | Theil U | 0.15 | | Theil U | 0.03 |
| | | | Disease Location | $Chi^2$ | 0.12 |
| | | | | Good.-Krusk. Tau | 0.25 |
| | | | | Theil U | 0.06 |

| | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|
| Inject; Order | 0.01 | 0.56 | 0.58 |
| **Age** | 0.17 | 0.01 | 0.01 |
| **BMI** | 0.00 | 0.01 | 0.01 |

Table S-28. p-values for the potential imbalances. The grey cells indicate the significant values. The factors with statistical imbalance are in bold.

Anti-TNFα and immunosuppression medications as well as age and BMI had p-values < 0.05 (Tables S-27 and S-28).

Unsupervised plots were also drawn to visualize the impact of categorical bias factors, either on all metabolites in the final dataset or on the selected candidates. The one for disease location is given at Figure S-16. Here, only a small effect between ileitis and ileocolitis can be seen. The problem of such plots, however, is that the metabolites are considered together. This dilutes the effect of each metabolite individually.

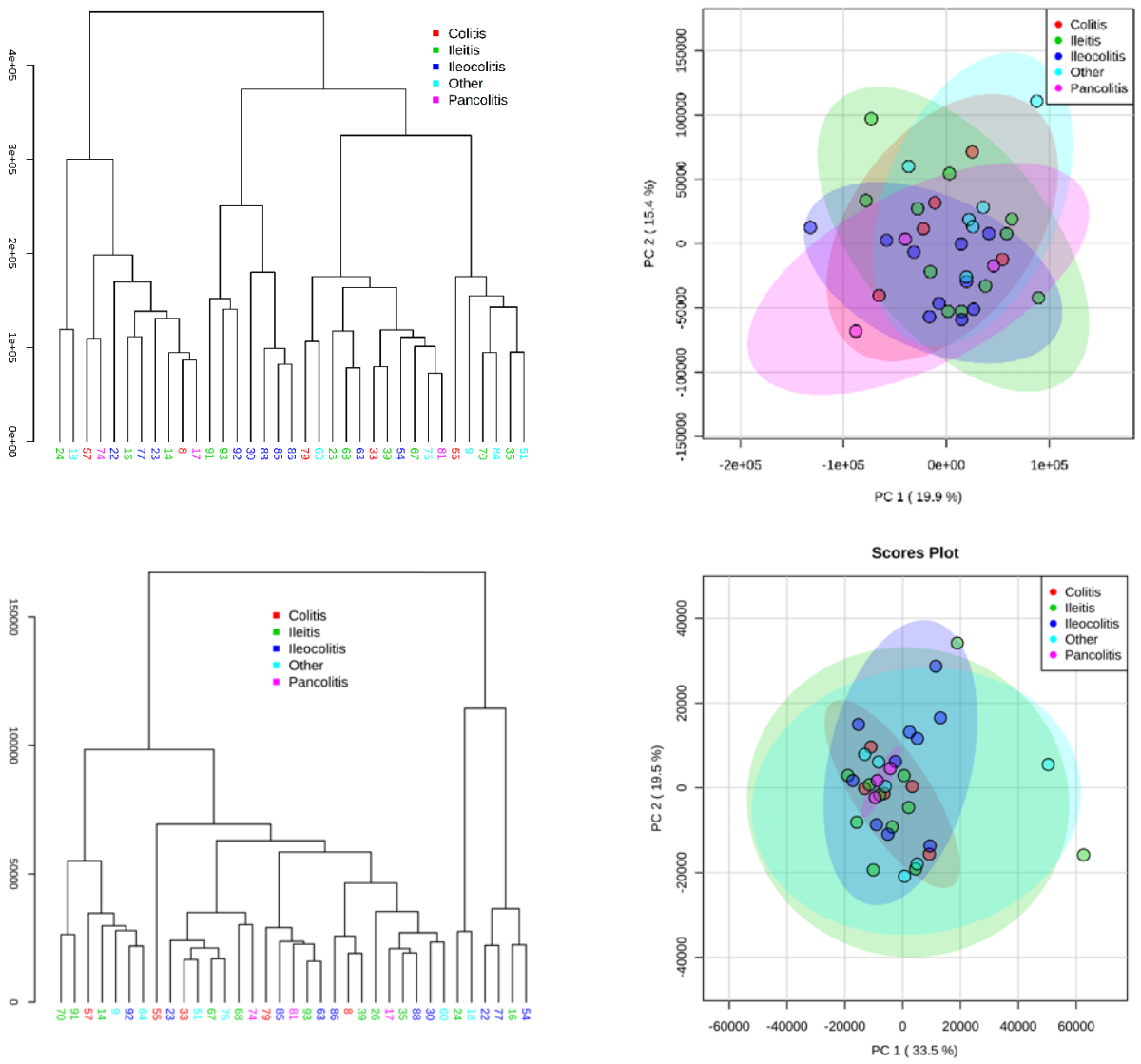



Figure S-25. HCA and PCA plots constructed on all 183 metabolites (above) and the selected candidates (n=29, below) to see the influence of disease location.

(B) Relationships between the candidate biomarkers and the bias factors

| **Anti-TNFα** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|---|
| **11** | 0.00 | 0.01 | 0.01 | 0.01 | 0.04 |
| 23 | 0.00 | 0.39 | 0.47 | 0.41 | 0.55 |
| 1 | 0.00 | 0.13 | 0.06 | 0.10 | 0.15 |
| 15 | 0.46 | 0.11 | 0.05 | 0.09 | 0.11 |
| 18 | 0.20 | 0.92 | 0.90 | 0.95 | 0.92 |
| 24 | 0.01 | 0.07 | 0.06 | 0.09 | 0.11 |
| **25** | 0.00 | 0.05 | 0.02 | 0.03 | 0.06 |
| **8** | 0.00 | 0.03 | 0.00 | 0.01 | 0.00 |
| 5 | 0.00 | 0.34 | 0.78 | 0.29 | 0.82 |
| 14 | 0.00 | 0.68 | 0.22 | 0.68 | 0.30 |
| 7 | 0.15 | 0.47 | 0.38 | 0.35 | 0.46 |
| 26 | 0.00 | 0.41 | 0.81 | 0.37 | 0.84 |
| 9 | 0.00 | 0.95 | 0.94 | 0.95 | 0.95 |
| 27 | 0.00 | 0.21 | 0.23 | 0.26 | 0.32 |
| 28 | 0.00 | 0.10 | 0.48 | 0.16 | 0.56 |
| 6 | 0.00 | 0.23 | 0.38 | 0.26 | 0.46 |
| 29 | 0.00 | 0.15 | 0.32 | 0.20 | 0.40 |
| 30 | 0.01 | 0.39 | 0.84 | 0.43 | 0.87 |
| 31 | 0.00 | 0.06 | 0.14 | 0.09 | 0.22 |
| **2** | 0.00 | 0.04 | 0.04 | 0.02 | 0.08 |
| **32** | 0.00 | 0.05 | 0.01 | 0.03 | 0.04 |
| 33 | 0.23 | 0.20 | 0.12 | 0.14 | 0.19 |
| 34 | 0.57 | 0.51 | 0.42 | 0.76 | 0.50 |
| 35 | 0.01 | 0.39 | 0.20 | 0.39 | 0.29 |
| 36 | 0.01 | 0.12 | 0.05 | 0.14 | 0.10 |
| 10 | 0.26 | 0.49 | 0.40 | 0.37 | 0.48 |
| 37 | 0.00 | 0.17 | 0.13 | 0.19 | 0.21 |
| 38 | 0.00 | 0.03 | 0.13 | 0.06 | 0.21 |
| 39 | 0.00 | 0.24 | 0.11 | 0.20 | 0.18 |

| **Immuno.** | Norm. | P / Sp | Kendall | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|---|
| 11 | 0.00 | 0.20 | 0.07 | 0.02 | 0.13 |
| 23 | 0.00 | 0.02 | 0.40 | 0.28 | 0.48 |
| 1 | 0.00 | 0.22 | 0.37 | 0.00 | 0.48 |
| **15** | 0.46 | 0.02 | 0.00 | 0.01 | 0.02 |
| 18 | 0.20 | 0.38 | 0.28 | 0.43 | 0.37 |
| 24 | 0.01 | 0.59 | 0.21 | 0.70 | 0.30 |
| 25 | 0.00 | 0.28 | 0.06 | 0.01 | 0.12 |
| **8** | 0.00 | 0.27 | 0.01 | 0.01 | 0.04 |
| 5 | 0.00 | 0.25 | 0.21 | 0.06 | 0.30 |
| 14 | 0.00 | 0.36 | 0.16 | 0.23 | 0.24 |
| 7 | 0.15 | 0.31 | 0.21 | 0.28 | 0.30 |
| 26 | 0.00 | 0.37 | 0.43 | 0.05 | 0.51 |
| **9** | 0.00 | 0.01 | 0.01 | 0.14 | 0.03 |
| 27 | 0.00 | 0.91 | 0.69 | 0.91 | 0.74 |
| 28 | 0.00 | 0.91 | 0.95 | 0.89 | 0.96 |
| 6 | 0.00 | 0.31 | 0.57 | 0.50 | 0.64 |
| 29 | 0.00 | 0.42 | 0.53 | 0.16 | 0.60 |
| 30 | 0.01 | 0.78 | 0.57 | 0.87 | 0.64 |
| 31 | 0.00 | 0.66 | 0.71 | 0.43 | 0.76 |
| 2 | 0.00 | 0.68 | 0.78 | 0.62 | 0.81 |
| 32 | 0.00 | 0.27 | 0.14 | 0.02 | 0.22 |
| 33 | 0.23 | 0.78 | 0.73 | 0.91 | 0.78 |
| **34** | 0.57 | 0.05 | 0.02 | 0.11 | 0.05 |
| 35 | 0.01 | 0.47 | 0.28 | 0.54 | 0.37 |
| 36 | 0.01 | 0.58 | 0.43 | 0.61 | 0.51 |
| 10 | 0.26 | 0.93 | 0.91 | 0.93 | 0.92 |
| 37 | 0.00 | 0.43 | 0.43 | 0.24 | 0.51 |
| 38 | 0.00 | 0.60 | 0.53 | 0.55 | 0.60 |
| 39 | 0.00 | 0.23 | 0.09 | 0.01 | 0.16 |

| **Age** | Norm. | P / Sp | Kendall |
|---|---|---|---|
| 11 | 0.00 | 0.89 | 0.81 |
| 23 | 0.00 | 0.51 | 0.62 |
| 1 | 0.00 | 0.19 | 0.50 |
| 15 | 0.46 | 0.48 | 0.50 |
| **18** | 0.20 | 0.01 | 0.01 |
| **24** | 0.01 | 0.00 | 0.00 |
| 25 | 0.00 | 0.89 | 0.58 |
| 8 | 0.00 | 0.76 | 0.76 |
| 5 | 0.00 | 0.34 | 0.09 |
| 14 | 0.00 | 0.16 | 0.11 |
| 7 | 0.15 | 0.57 | 0.53 |
| 26 | 0.00 | 0.99 | 0.62 |
| 9 | 0.00 | 0.49 | 0.43 |
| 27 | 0.00 | 0.31 | 0.56 |
| **28** | 0.00 | 0.00 | 0.02 |
| 6 | 0.00 | 0.61 | 0.62 |
| **29** | 0.00 | 0.01 | 0.05 |
| 30 | 0.01 | 0.02 | 0.15 |
| **31** | 0.00 | 0.01 | 0.01 |
| 2 | 0.00 | 0.93 | 0.68 |
| 32 | 0.00 | 0.35 | 0.87 |
| 33 | 0.23 | 0.09 | 0.12 |
| 34 | 0.57 | 0.59 | 0.48 |
| 35 | 0.01 | 0.02 | 0.11 |
| **36** | 0.01 | 0.05 | 0.04 |
| 10 | 0.26 | 0.70 | 0.72 |
| 37 | 0.00 | 0.51 | 0.68 |
| **38** | 0.00 | 0.03 | 0.05 |
| 39 | 0.00 | 0.50 | 0.70 |

| **BMI** | Norm. | P / Sp | Kendall |
|---|---|---|---|
| 11 | 0.00 | 0.20 | 0.01 |
| 23 | 0.00 | 0.31 | 0.41 |
| 1 | 0.00 | 0.47 | 0.14 |
| 15 | 0.45 | 0.90 | 0.77 |
| **18** | 0.31 | 0.00 | 0.00 |
| 24 | 0.01 | 0.35 | 0.24 |
| 25 | 0.00 | 0.57 | 0.12 |
| 8 | 0.00 | 0.47 | 0.80 |
| 5 | 0.00 | 0.80 | 0.29 |
| 14 | 0.00 | 0.14 | 0.63 |
| 7 | 0.23 | 0.67 | 0.55 |
| 26 | 0.00 | 0.27 | 0.52 |
| 9 | 0.00 | 0.56 | 0.39 |
| 27 | 0.00 | 0.03 | 0.89 |
| 28 | 0.00 | 0.80 | 0.17 |
| 6 | 0.00 | 0.71 | 0.86 |
| 29 | 0.00 | 0.31 | 0.19 |
| 30 | 0.02 | 0.96 | 0.93 |
| 31 | 0.00 | 0.09 | 0.38 |
| 2 | 0.00 | 0.97 | 0.63 |
| 32 | 0.00 | 0.79 | 0.19 |
| 33 | 0.22 | 0.82 | 0.91 |
| **34** | 0.32 | 0.05 | 0.05 |
| 35 | 0.03 | 0.93 | 0.61 |
| 36 | 0.01 | 0.45 | 0.47 |
| 10 | 0.17 | 0.22 | 0.24 |
| 37 | 0.00 | 0.11 | 0.43 |
| 38 | 0.00 | 0.31 | 0.89 |
| **39** | 0.00 | 0.01 | 0.00 |

Table S-29. p-values for the relationships between the candidate biomarkers and the four potential bias factors. The grey cells indicate the significant values. The candidates significantly linked to the bias factor are in bold.

Candidates 11, 25, 8, 2, 32; 15, 8, 9, 34; 18, 24, 28, 29, 31, 36, 38 and 18, 34, 39, respectively, were found to be linked to the factors (p-values < 0.05; Table S-28). The potential impact was further assessed through the residual ability to separate the biological groups and, for continuous bias factors, through scatterplots (Figures S-26 and S-27). These figures confirmed the links between age or BMI and the candidates. The relationship for candidate 28 seemed due to atypical values (outliers), but it was kept for the remainder of the process since it was significant in Kendall correlation, a non-parametric measure.

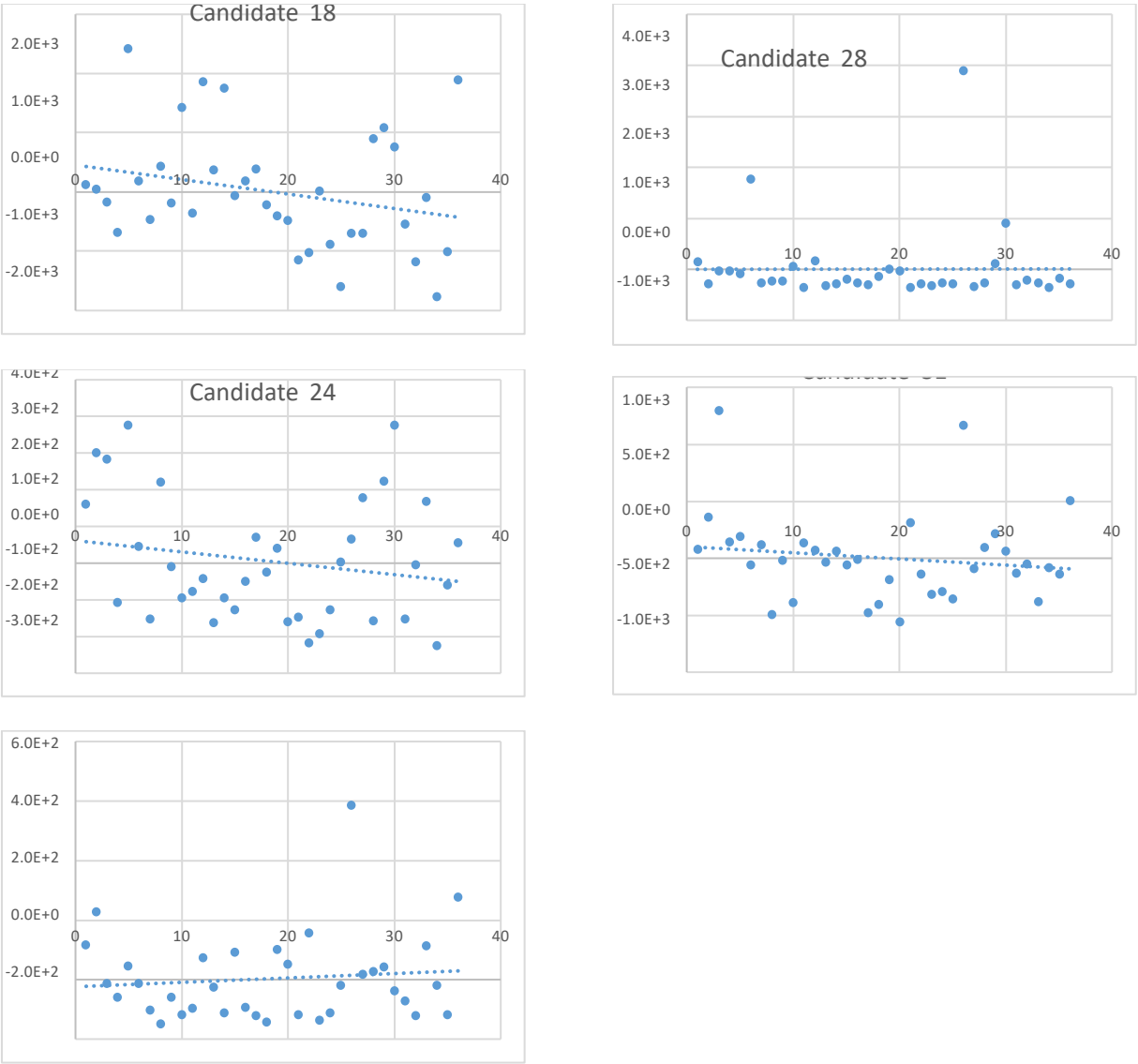



Figure S-26. Scatterplots of the candidates' detected signal against age with the corresponding linear regression.

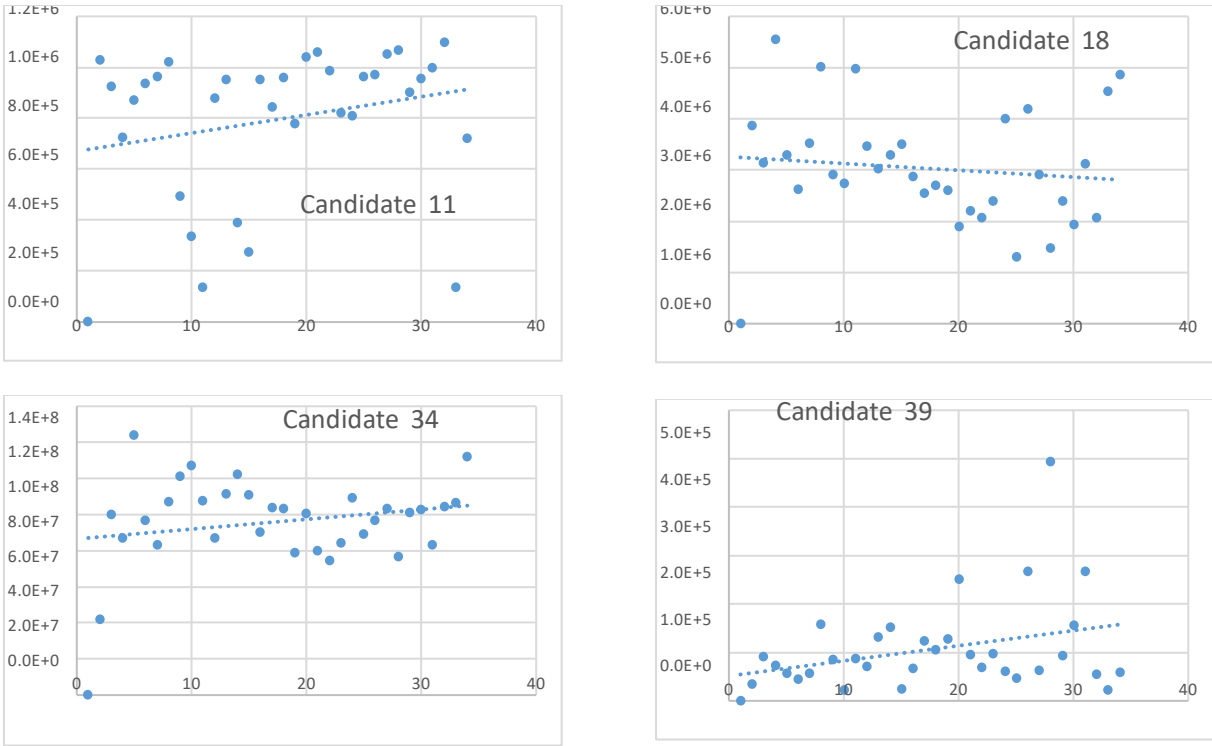



Figure S-27. Scatterplots of the candidates' detected signal against BMI with the corresponding linear regression.

(C) Residual separation ability

<u>Sample Removal</u>

To equilibrate the classes regarding the potentially active bias factors, different combinations were tested. Details of the final results are given at Table S-30, 31, 32 and 33.

| | Group | BMI |
|---|---|---|
| 390-JT | Quiescent | / |
| 434-SLR | "Low" | / |
| 429-ID | "High" | 15.4 |
| 83-ML | "High" | 17.5 |
| 508-DJ | "High" | 19.6 |
| 502-SA | "High" | 20.7 |
| 466-VLC | "High" | 20.5 |
| 367-VW | Quiescent | 39.2 |
| 354-FD | "Low" | 30.1 |
| 357-WF | "Low" | 41.5 |

| Number | Before | After |
|---|---|---|
| Quiescent | 9 | 7 |
| "High" | 12 | 7 |
| "Low" | 14 | 11 |

| N | W / K-W | Fisher |
|---|---|---|
| 0.98 | 0.29 | 0.27 |

Table S-30. Samples removed for the BMI equilibration (left), summary of the samples left (top right) and p-values for the repartition after the removal (bottom right).

| Clinical ID | Class | **Anti-TNFα** |
|---|---|---|
| 404-AC | Low | Yes |
| 354-FD | Low | Yes |
| 453-LR | Low | Yes |
| 356-OL | Low | Yes |
| 426-CF | Low | Yes |
| 425-SM | Low | Yes |
| 133-MD | Quiescent | No |
| 155-PC | Quiescent | No |
| 28-CL | Quiescent | No |

| **Number** | Before | After |
|---|---|---|
| Quiescent | 9 | 6 |
| High | 12 | 12 |
| Low | 14 | 8 |

| | |
|---|---|
| $Chi^2$ | 0.38 |
| Good.-Krusk. Tau | 0.46 |
| Theil U | 0.37 |

| Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|
| 0.00 | 0.43 | 0.42 |

Table S-31. Samples removed for the anti-TNFα equilibration (left), summary of the samples left (top right) and p-values for the repartition after the removal (bottom right).

| Clinical ID | Class | Immunosupp. |
|---|---|---|
| 465-PM | High | Yes |
| 247-SK | High | Yes |

| **Number** | Before | After |
|---|---|---|
| Quiescent | 9 | 9 |
| High | 12 | 10 |
| Low | 14 | 14 |

| | |
|---|---|
| $Chi^2$ | 0.24 |
| Good.-Krusk. Tau | 0.21 |
| Theil U | 0.15 |

| Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|
| 0 | 0.88 | 0.25 |

Table S-32. Samples removed for the immunosuppressor equilibration (left), summary of the samples left (top right) and p-values for the repartition after the removal (bottom right).

| Clinical ID | Class | **Age** |
|---|---|---|
| 133-MD | Quiescent | 65 |
| 426-CF | Low | 28 |
| 366-AT | High | 22 |
| 390-JT | Low | 29 |
| 247-SK | High | 26 |
| 228-VH | High | 26 |
| 28-CL | Quiescent | 56 |

| **Number** | Before | After |
|---|---|---|
| Quiescent | 9 | 7 |
| High | 12 | 9 |
| Low | 14 | 12 |

| Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|
| 0.17 | 0.13 | 0.38 |

Table S-33. Samples removed for the age equilibration (left), summary of the samples left (top right) and p-values for the repartition after the removal (bottom right).

Results

| **Anti-TNFα** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|
| 11 | 0.00 | 0.10 | 0.01 |
| **25** | 0.00 | 0.30 | 0.01 |
| **8** | 0.00 | 0.15 | 0.11 |
| **2** | 0.00 | 0.48 | 0.65 |
| **32** | 0.02 | 0.28 | 0.25 |

| **Immuno.** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|
| 15 | 0.58 | 0.08 | 0.08 |
| 8 | 0.00 | 0.03 | 0.00 |
| 9 | 0.00 | 0.03 | 0.04 |
| 34 | 0.63 | 0.02 | 0.01 |

| **Age** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|
| **18** | 0.18 | 0.16 | 0.07 |
| **24** | 0.01 | 0.11 | 0.11 |
| **28** | 0.00 | 0.51 | 0.34 |
| **29** | 0.00 | 0.54 | 0.79 |
| **31** | 0.00 | 0.19 | 0.14 |
| 36 | 0.01 | 0.08 | 0.07 |
| **38** | 0.00 | 0.28 | 0.61 |

| **BMI** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|
| 11 | 0.00 | 0.03 | 0.02 |
| **18** | 0.45 | 0.14 | 0.05 |
| **34** | 0.56 | 0.12 | 0.11 |
| **39** | 0.00 | 0.40 | 0.41 |

Table S-34. Effect of the factors on the separation ability measured through the residual separation capability (p-values for Pearson/Spearman and Kendall correlations, Welch/Kruskal-Wallis test and Fisher ratios). The order is the ranking of the candidates in the set of candidate biomarkers. The grey cells indicate the significant p-values (< 0.05) while the candidates in bold are the ones with a low residual separation capability (p-values > 0.1).

| Order (/ 29) | **Anti-TNFα** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|
| 7 | 25 | 0.00 | 0.04 | 0.01 |
| | | 0.00 | 0.30 | 0.01 |
| 7 | 8 | 0.00 | 0.03 | 0.01 |
| | | 0.00 | 0.15 | 0.11 |
| 20 | **2** | 0.00 | 0.05 | 0.02 |
| | | 0.00 | 0.48 | 0.25 |
| 20 | **32** | 0.00 | 0.00 | 0.02 |
| | | 0.02 | 0.28 | 0.15 |

| Order (/ 29) | **Age** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|
| 4 | **18** | 0.20 | 0.01 | 0.01 |
| | | 0.18 | 0.07 | 0.14 |
| 4 | 24 | 0.01 | 0.04 | 0.00 |
| | | 0.01 | 0.11 | 0.02 |
| 10 | 28 | 0.00 | 1.00 | 0.55 |
| | | 0.00 | 0.51 | 0.61 |
| 17 | 29 | 0.00 | 0.59 | 0.67 |
| | | 0.00 | 0.54 | 0.44 |
| 20 | 31 | 0.00 | 0.07 | 0.05 |
| | | 0.00 | 0.19 | 0.16 |
| 25 | **38** | 0.00 | 0.12 | 0.03 |
| | | 0.00 | 0.28 | 0.17 |

| Order (/ 29) | **BMI** | Norm. | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|
| 4 | **18** | 0.20 | 0.01 | 0.01 |
| | | 0.45 | 0.05 | 0.05 |
| 20 | 34 | 0.57 | 0.01 | 0.05 |
| | | 0.56 | 0.11 | 0.26 |
| 25 | **39** | 0.00 | 0.13 | 0.03 |
| | | 0.00 | 0.40 | 0.26 |

Table S-35. Comparison of the separation ability (p-values) before and after the equilibration, for the candidates potentially affected. The grey cells indicate the significant values. The order is the ranking of the candidates in the set of candidate biomarkers. The candidates with substantial deterioration due to the bias factor are in bold.

The residual separation capability evaluation worked in two steps. First, the direct assessment of the residual ability (Table S-34). Second, for the candidates with p-values > 0.1, an investigation of the variation resulting from the bias effect by comparing the ability with and without the consideration of the possible bias (Table S-35). Overall, candidates 2, 32, 38 and 39 were considered deteriorated and thus removed from the set of candidate biomarkers. The candidate 18 was impacted by two different factors and was also removed. Candidates 8 and 34, as far as we could evaluate, were significantly

impacted (with initial low p-values, $< 0.05$) but had remaining p-values around 0.1. It was decided to keep them as biological markers but to leave them aside when evaluating the separation performance.

**S-18. Data Processing. A. Meta-markers. B. Testing of potential bias between training and test sets.**

(A) Meta-markers

Three meta-markers were considered. The first (CB1) used the most significant candidates for each separation (n=6 for both). The second (CB2) added the very significant ones (n=12 and n=15, respectively). The third (CB3) used all the selected and not biased candidates (n= 19 and n=22).

(B) Testing of potential bias between training and test sets for all samples

| Number of samples | | |
|---|---|---|
| Class | Training | Test |
| Healthy Controls | 22 | 11 |
| Low | 10 | 4 |
| High | 8 | 4 |
| Quiescent | 6 | 3 |

| Extraction | Training | Test |
|---|---|---|
| Batch 1 | 21 | 11 |
| Batch 2 | 23 | 9 |
| Batch 3 | 2 | 2 |

| Drying | Training | Test |
|---|---|---|
| Batch 1 | 7 | 1 |
| Batch 2 | 6 | 2 |
| Batch 3 | 4 | 4 |
| Batch 4 | 5 | 3 |
| Batch 5 | 5 | 3 |
| Batch 6 | 6 | 2 |
| Batch 7 | 5 | 3 |
| Batch 8 | 7 | 1 |
| Batch 9 | 1 | 3 |

| | Number | | | | | |
|---|---|---|---|---|---|---|
| | Inj. Order | Gender | Age at sampling | Tobacco | Alcohol | BMI |
| Training | 48.4 | 0.39 | 41.3 | 0.26 | 0.24 | 24.70 |
| Test | 47.4 | 0.45 | 46.2 | 0.27 | 0.18 | 25.32 |

| | Gastro. reflux | L-Thyroxin | Anti-TNFα | Immuno. |
|---|---|---|---|---|
| Training | 0.30 | 0.13 | 0.26 | 0.11 |
| Test | 0.23 | 0.32 | 0.27 | 0.00 |

| | | p-values |
|---|---|---|
| Drying | $Chi^2$ | 0.43 |
| | Wilk's Lambda | 0.09 |
| | Good-Krusk. Tau | 0.44 |
| | Theil U | 0.41 |
| Extraction | $Chi^2$ | 0.64 |
| | Wilk's Lambda | 0.00 |
| | Good-Krusk. Tau | 0.64 |
| | Theil U | 0.65 |
| Gender | $Chi^2$ | 0.62 |
| | Wilk's Lambda | 0.00 |
| | Good-Krusk. Tau | 0.62 |
| | Theil U | 0.62 |

| | Normality | P / Sp | ANOVA | Kruskal-Wallis |
|---|---|---|---|---|
| Inj. Order | 0.00 | 0.89 | 0.89 | 0,89 |
| Age | 0.03 | 0.11 | 0.11 | 0,09 |
| Tobacco | 0.00 | 0.92 | 0.92 | 0,92 |
| Alcohol | 0.00 | 0.60 | 0.59 | 0,60 |
| Gastro. reflux | 0.00 | 0.51 | 0.50 | 0,51 |
| L-Thyroxin | 0.00 | 0.07 | 0.11 | 0,07 |
| Anti-TNFα | 0.00 | 0.91 | 0.91 | 0,91 |
| Immunosupp. | 0.00 | 0.11 | 1.00 | 0.11 |

Table S-36. Distribution of the samples (numbers of samples and percentages) and corresponding p-values between training and test sets for the various potential bias factors.

**S-19. Separation between healthy controls and Crohn's disease samples. A. Performances. B. Selection order in the models.**

(A) Performances

| | Predictive | | Orthogonal | | Permutation | | Probability | |
|---|---|---|---|---|---|---|---|---|
| **OPLS-DA** | $Q^2$ | $R^2Y$ | $Q^2$ | $R^2Y$ | $Q^2$ | $R^2Y$ | $Q^2$ | $R^2Y$ |
| CB 1 | 0.45 | 0.48 | 0.00 | 0.03 | 0.46 | 0.51 | < 0.0005 | < 0.0005 |
| CB 2 | 0.46 | 0.49 | 0.00 | 0.09 | 0.46 | 0.58 | < 0.0005 | < 0.0005 |
| CB 3 | 0.48 | 0.53 | -0.01 | 0.09 | 0.47 | 0.62 | < 0.0005 | < 0.0005 |

| **PLS-DA** | $Q^2$ | $R^2Y$ | Probability | Nber LV |
|---|---|---|---|---|
| CB 1 | / | / | / | / |
| CB 2 | 0.33 | 0.49 | < 0.0005 | 3 |
| CB 3 | 0.31 | 0.39 | < 0.0005 | 1 |

| **LDA** | Rao value | p-value |
|---|---|---|
| CB 1 | 10.7 | < 0.0001 |
| CB 2 | 7.3 | < 0.0001 |
| CB 3 | 4.5 | < 0.0001 |

| **Error rates** | sPLS-DA | RF | SVM |
|---|---|---|---|
| CB 1 | / | 0.15 | / |
| CB 2 | 0.16 | 0.19 | 0.21 |
| CB 3 | 0.18 | 0.15 | 0.23 |

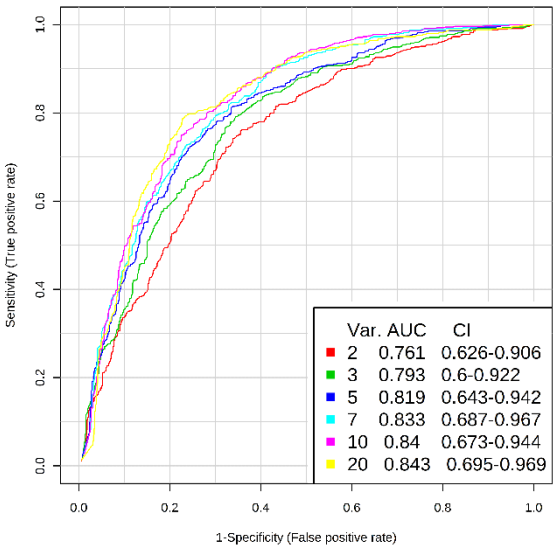
| **Error rates** | Re-subst. | Re-sampling | Test |
|---|---|---|---|
| CB 1 | 0.19 | 0.22 | 0.16 |
| CB 2 | 0.14 | 0.20 | 0.14 |
| CB 3 | 0.11 | 0.21 | 0.16 |

| **ROC** | Predict. | Holdout | Test | Permut. |
|---|---|---|---|---|
| CB 1 | 0.85 | 0.90 | 19 | 0.01 |
| CB 2 | 0.78 | 0.93 | 19 | 0.05 |
| CB 3 | 0.75 | 0.92 | 17 | 0.09 |

Table S-37. Discrimination performances achieved with the three different meta-markers using OPLS-DA, PLS-DA, LDA, sPLS-DA, SVM and RF models. Plus ROC curves and mean error rates for Bayes classifier, NN, SVM and PLS-DA. The grey cells indicate the best values obtained. Nber LV is the number of latent variables required to obtain the most efficient separation in PLS-DA.

The results summarized in Table S-37 showed that the CD samples were effectively separated from the healthy controls. Rao p-value was < 0.001 for all three potential meta-markers. Mean error rates were around 0.2 (range 0.16-0.29, depending on the meta-marker and the method) and 0.15 (range 0.09-0.23) in re-sampling and test validation, respectively. This was confirmed by sPLS-DA cross-validation (0.16-0.18) and RF out-of-bag error (0.17-0.19). As well as by ROC test validation (between 0.15 and 0.2). The predictive $Q^2$ was moderate in PLS-DA model (0.31-0.33) and good in OPLS-DA (0.45-0.48)[40]. Importantly[S-41], there was very little overfitting [S-42] since $Q^2$ was close to $R^2Y$ in both PLS-DA and OPLS-DA : 0.39/0.49 and 0.48/0.53 respectively. In OPLS-DA, we observed an explicative power much higher than noise, with the orthogonal $Q^2$ equal to 0. Moreover, the permutation $Q^2$ was close to the predictive $Q^2$ (0.46/0.47) and its p-value was low (< 0.0005), rejecting the hypothesis of a discrimination

due to bare chance. The diagnosis ability measured by ROC AUC was good [S-40, S-43] with a mean value of 0.80, a range of 0.75-0.85, a confidence interval (CI) mean range of 0.70 - 0.99, a CI global range of 0.57 - 1 and a probability p-value < 0.1. The global statistical power was over 0.8 at FDR 0.1 when all the candidates where employed (Figure S-28). Re-sampling and test validations also indicated the potential ability of the meta-markers to be generalized to new samples. The ratio between the information valuable to separate and the noise can be evaluated by comparing the supervised (that use only mostly the first) and the unsupervised plots (that use both). Here, as in most cases, the successive additions of candidates brought valuable information (better discrimination in supervised plots) but at the cost of noise addition (poorer separation in unsupervised plots; Figure S-30). The performances attained were particularly satisfying, if one considers that the CD samples consisted in three endoscopic activity subgroups, which most probably increased their overall group variance.



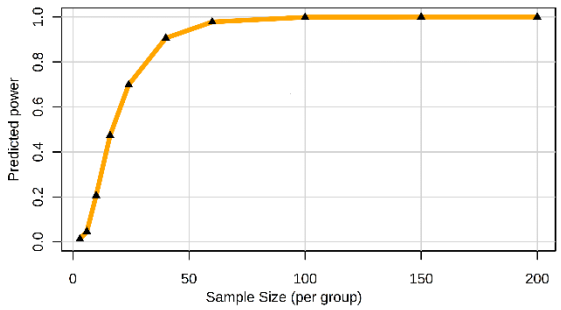




Figure S-28. ROC curves (100 cross-validation, SVM algorithm) for an increasing number of candidates (left). Adding candidates to the meta-marker increased the diagnosis capability marginally. Multivariate statistical power for all the candidates at FDR 0.01, > 0.8 for the minimal n (= 33).

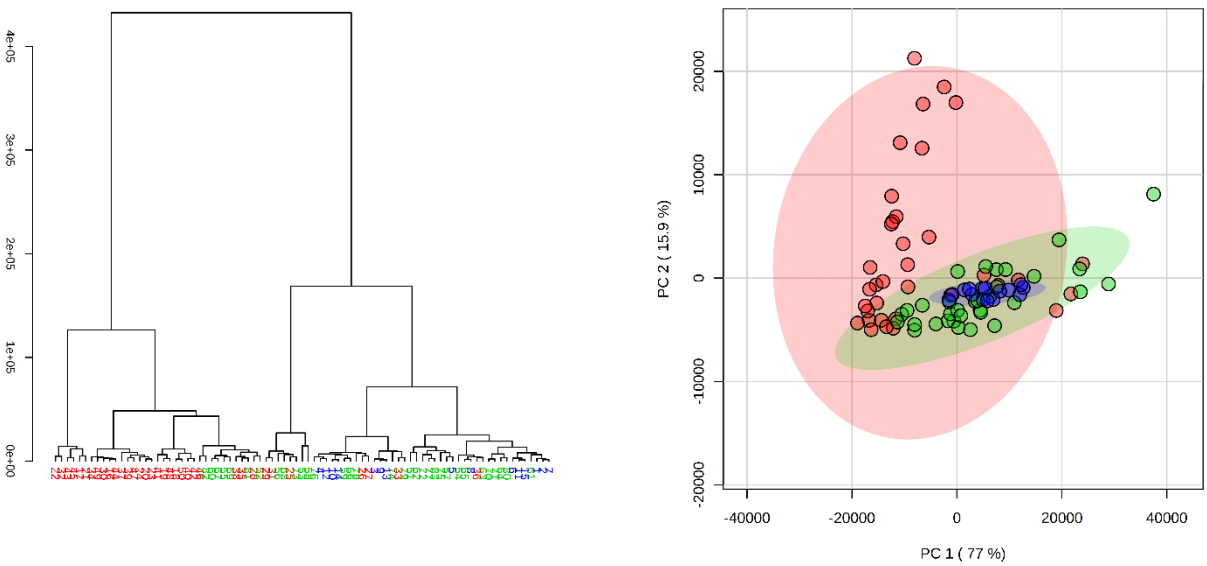



Figure S-29. HCA and PCA plots constructed on the most significant candidate biomarkers. The CD samples are in red, the HC in green and the QC samples in blue.

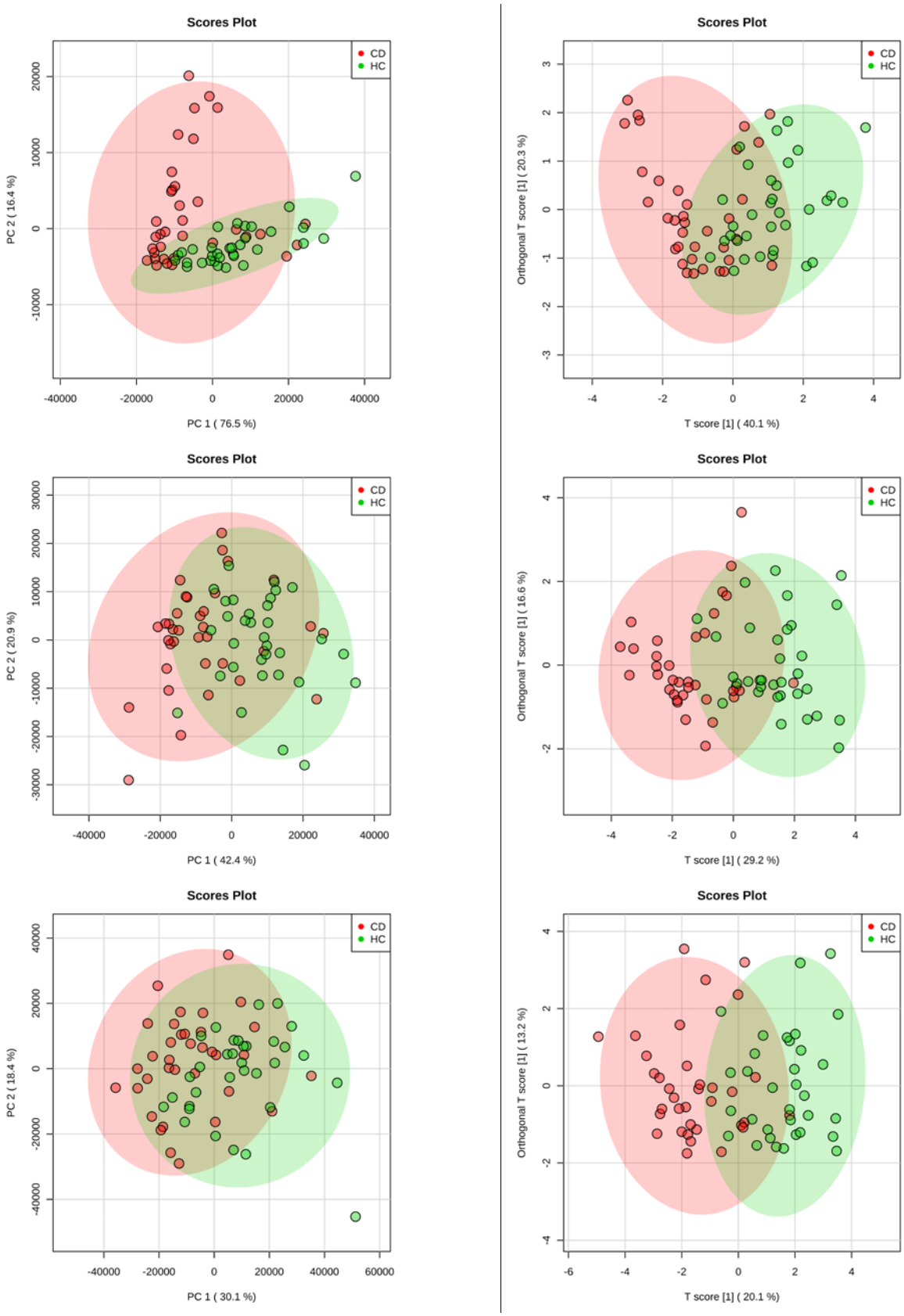



Figure S-30. PCA (left) and OPLS (right) plots constructed on the three different meta-markers (CB1, above, CB2, middle, and CB3, below).

(B) Selection order in the models

| CB | NN | Bayes | PLS-DA | SVM | OPLS-DA | RF | CB | Mean | Median | Best | Worst | LDA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 2 | 1 | 3 | 1 | 1 | 1 | 2 | 2 | 1 | 5 | 1 |
| 2 | 11 | 14 | 5 | 16 | 3 | 7 | 2 | 9 | 9 | 3 | 16 | |
| 3 | 2 | 1 | 2 | 2 | 7 | 4 | 3 | 3 | 2 | 1 | 7 | 3 |
| 4 | 6 | 5 | 6 | 4 | 10 | 14 | 4 | 8 | 6 | 4 | 14 | 4 |
| 5 | 18 | 7 | 3 | 7 | 2 | 3 | 5 | 7 | 5 | 2 | 18 | |
| 6 | 4 | 6 | 10 | 8 | 6 | 11 | 6 | 8 | 7 | 4 | 11 | 6 |
| 7 | 7 | 11 | 8 | 5 | 4 | 6 | 7 | 7 | 7 | 4 | 11 | |
| 8 | 17 | 8 | 4 | 9 | 5 | 2 | 8 | 8 | 7 | 2 | 17 | 8 |
| 9 | 10 | 15 | 14 | 15 | 8 | 9 | 9 | 12 | 12 | 8 | 15 | |
| 10 | 12 | 9 | 7 | 6 | 9 | 16 | 10 | 10 | 9 | 6 | 16 | 10 |
| 11 | 14 | 13 | 12 | 13 | 15 | 5 | 11 | 12 | 13 | 5 | 15 | |
| 12 | 3 | 16 | 9 | 10 | 11 | 8 | 12 | 10 | 10 | 3 | 16 | |
| 13 | 16 | 12 | 18 | 11 | 16 | 13 | 13 | 14 | 15 | 11 | 18 | |
| 14 | 15 | 18 | 16 | 18 | 19 | 12 | 14 | 16 | 17 | 12 | 19 | |
| 15 | 13 | 10 | 13 | 14 | 12 | 19 | 15 | 14 | 13 | 10 | 19 | |
| 17 | 19 | 19 | 15 | 12 | 17 | 18 | 17 | 17 | 18 | 12 | 19 | |
| 18 | 8 | 17 | 17 | 19 | 13 | 17 | 18 | 15 | 17 | 8 | 19 | |
| 19 | 9 | **3** | 19 | 17 | 18 | 10 | 19 | 13 | 14 | 3 | 19 | |
| 22 | **1** | **4** | 11 | **1** | 14 | 15 | 22 | 8 | 8 | 1 | 15 | |

Table S-38. Ranking or selection order of the candidate biomarkers in neural network (NN), Bayes classifier, PLS-DA, OPLS-DA, SVM and RF models. In the far right column, the metabolites found significant in the LDA model. The grey cells emphasize the highest ranks.

As expected, the candidates selected as the most significant had the highest ranks in the models (Table S-38). As already noticed in the selection process, Bayes classifier, neural network and SVM worked differently, particularly regarding the candidates 19 and 22 (in bold), an issue worth to mention but outside the scope of this study.

**S-20. Separation between the three Crohn's disease groups. A. Performances. B. Selection order in the models.**

(A) Performances

| PLS-DA | $Q^2$ | $R^2Y$ | Probability | Nber LV |
|---|---|---|---|---|
| CB 1 | / | / | / | / |
| CB 2 | -0.20 | 0.07 | 0.003 | 1 |
| CB 3 | -0.25 | 0.11 | < 0.0005 | 1 |

| LDA Rao | | |
|---|---|---|
| CB 1 | 3.7 | 0.000 |
| CB 2 | 1.9 | 0.032 |
| CB 3 | 1.7 | 0.069 |

| Error rates | sPLS-DA | RF |
|---|---|---|
| CB 1 | / | 0.29 |
| CB 2 | 0.31 | 0.37 |
| CB 3 | 0.29 | 0.37 |

| Error rates | Re-subst. | Re-sampling | Test |
|---|---|---|---|
| CB 1 | 0.24 | 0.36 | 0.41 |
| CB 2 | 0.11 | 0.45 | 0.57 |
| CB 3 | 0.07 | 0.44 | 0.57 |

Table S-39. Discrimination performances achieved with the three different meta-markers using PLS-DA, LDA, sPLS-DA and RF models, and mean error rates for Bayes classifier, NN, SVM and PLS-DA.

Error rates in re-substitution were acceptable, either for sPLS-DA, RF or the mean of SVM, Bayes classifier, NN, and PLS-DA models (0.20-0.30, but as low as 0.07, Table S-39). However, the re-sampling and test error rates increased to 0.36 and 0.41 respectively. And the PLS-DA models were not able to separate effectively the three CD subgroups (negative $Q^2$). Reasons for this have been mentioned in the manuscript (section 3.3), particularly the difficulty to separate three classes with one rule (section 5.5). To support this, the Rao p-value indicated the possibility to distinguish at least two classes, with a much higher effectiveness ($p < 0.0001$). And HCA and PLS plots clearly showed that the candidates were able to separate (Figure S-31).

Despite the low number of samples available (n=9 for quiescent CD), the variations of concentration of the candidates were sufficient to achieve a statistical power around 0.8 at FDR 0.1.

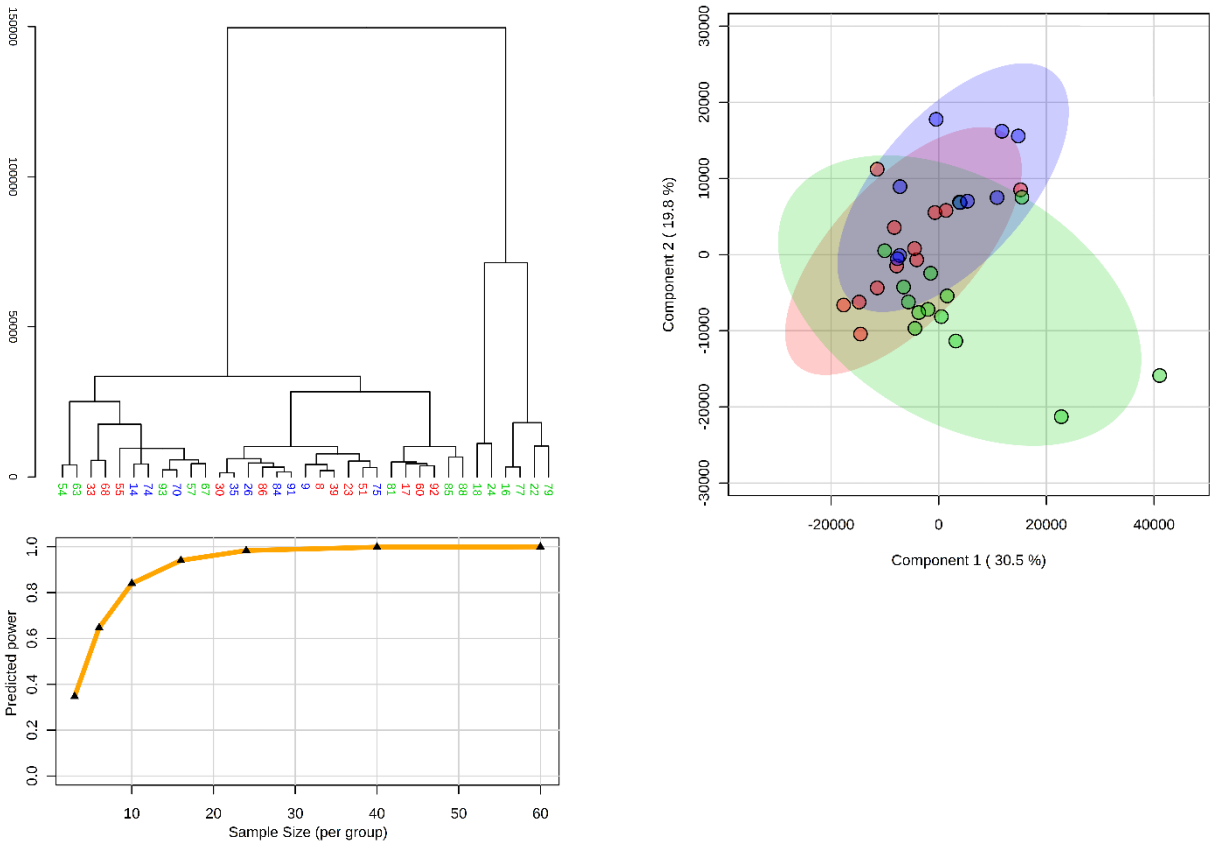



Figure S-31. HCA and PLS plots for the separation between the three CD groups (above), constructed on the most significant candidates (CB1) and all the selected and not biased candidates (CB3), respectively. The high endoscopic activity samples are in red, the low endoscopic activity ones in green, the quiescent ones in blue. Statistical power for the 15 most significant candidates (CB2) at FDR 0.1 (below).

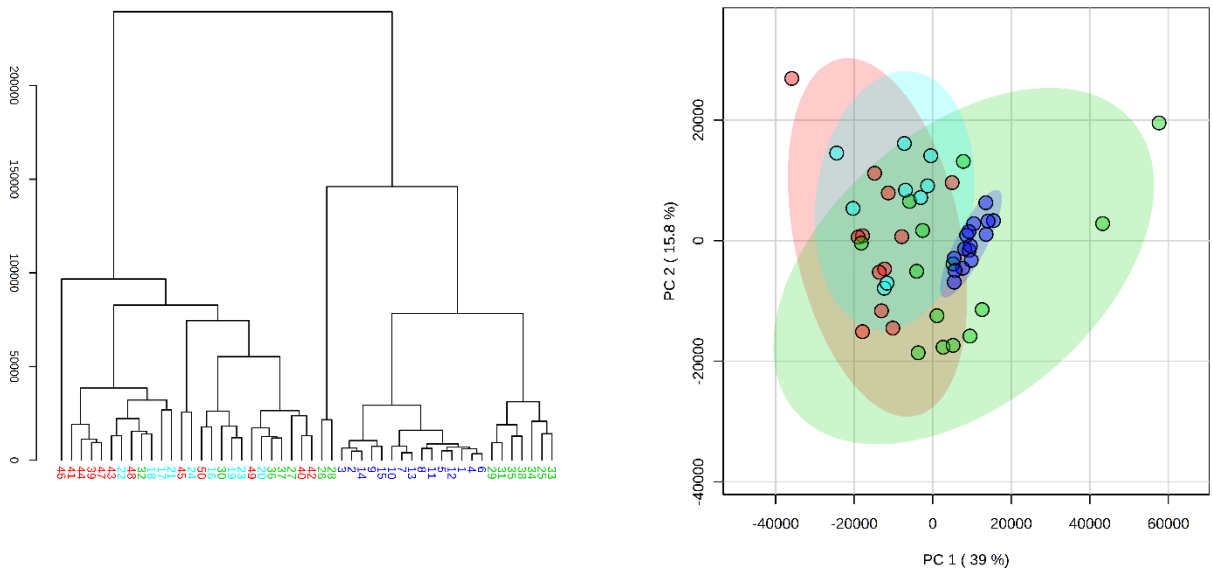



Figure S-32. HCA and PCA plots constructed on all the candidate biomarkers. The high endoscopic activity samples are in red, the low endoscopic activity ones are in green, the quiescent ones in light blue, the QC samples in royal blue.

(B) Selection order in the models

| CB | NN | Bayes | PLS-DA | RF | CB | Mean | Median | Best | Worst | LDA |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 2 | 6 | 1 | 2 | 11 | 3 | 2 | 1 | 6 | 1 |
| 23 | 8 | 7 | 3 | 6 | 23 | 6 | 7 | 3 | 8 | |
| 1 | 15 | 1 | 7 | 1 | 1 | 6 | 4 | 1 | 15 | |
| 15 | 4 | 9 | 9 | 4 | 15 | 7 | 7 | 4 | 9 | |
| 25 | 16 | 4 | 6 | 3 | 25 | 7 | 5 | 3 | 16 | |
| 5 | 17 | 10 | 8 | 16 | 5 | 13 | 13 | 8 | 17 | |
| 14 | 18 | 20 | 15 | 14 | 14 | 17 | 17 | 14 | 20 | |
| 7 | 5 | 14 | 4 | 9 | 7 | 8 | 7 | 4 | 14 | |
| 26 | 3 | 17 | 17 | 13 | 26 | 13 | 15 | 3 | 17 | |
| 9 | 9 | 19 | 16 | 7 | 9 | 13 | 13 | 7 | 19 | |
| 27 | 19 | 2 | 18 | 8 | 27 | 12 | 13 | 2 | 19 | |
| 28 | 10 | 11 | 19 | 12 | 28 | 13 | 12 | 10 | 19 | |
| 6 | 11 | 15 | 14 | 20 | 6 | 15 | 15 | 11 | 20 | |
| 29 | 12 | 13 | 20 | 15 | 29 | 15 | 14 | 12 | 20 | |
| 30 | 13 | 8 | 13 | 11 | 30 | 11 | 12 | 8 | 13 | |
| 33 | 20 | 18 | 10 | 10 | 33 | 15 | 14 | 10 | 20 | |
| 35 | 14 | 5 | 11 | 18 | 35 | 12 | 13 | 5 | 18 | |
| 36 | 6 | 12 | 5 | 17 | 36 | 10 | 9 | 5 | 17 | |
| 10 | 7 | 16 | 12 | 19 | 10 | 14 | 14 | 7 | 19 | |
| 37 | **1** | **3** | **2** | **5** | 37 | **3** | **3** | **1** | **5** | |

Table S-40. Ranking or selection order of the candidate biomarkers in NN, Bayes classifier, PLS-DA and RF models. In the far right column, the significant metabolites in the LDA model. The grey cells emphasize the highest ranks.

Again, the candidates selected as the most significant had the highest ranks in the models, a convergence that tended to support the selection process (Table S-40). Candidate 37, however, seemed important to discriminate despite the fact that it was at first selected in a low position. This discrepancy could again be due to the complexity of separating three groups at once.

## S-21. Annotation of the candidate biomarkers.

| Candidate | ID | Specific Peaks | Match | Proba. | Δ RT | Δ (ppm) | Criteria Met |
|---|---|---|---|---|---|---|---|
| 1 | / | 205, 103 | | | | | 0 |
| 2 | Erythronic acid / Threonic Acid | | **902** | 49 | 57 | **0.1** | **2** |
| 3 | / | 259, 198, 167, 139, 114 | | | | | 0 |
| 4 | / | 216, 188, 172 | | | | | 0 |
| 5 | / | 449, 408, 393, 333, 318, 305 | | | | | 0 |
| 6 | / | 205, 147, 117 | | | | | 0 |
| 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** |
| 11 | / | 203, 156, 112, 89 | | | | | 0 |
| 12 | / | 364, 277, 189, 172 | | | | | 0 |
| 13 | Methionine | | **678** | **86** | **0** | **0.1** | **3** |
| 14 | 2-Hydroxybutyric acid | | **846** | 27 | **1** | **0.1** | **3** |
| 15 | / | 249, 233, 207, 175, 133 | | | | | 0 |
| 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** |
| 22 | / | 56 | | | | | 0 |
| 23 | Erythrose / Threose | | 572 | 12 | **24** | **0.4** | **2** |
| 24 | / | 325, 174 | | | | | 0 |
| 25 | / | 203, 188, 172 | | | | | 0 |
| 26 | Capric acid | | **795** | **93** | **0** | **0.3** | **3** |
| 27 | / | 292 | | | | | 0 |
| 28 | / | 384, 369, 327, 266, 237 | | | | | 0 |
| 29 | / | 217, 191, 147 | | | | | 0 |
| 30 | / | 327, 145, 129 | | | | | 0 |
| 31 | Erythritol / Threitol | | **823** | 23 | **16** | / | **2** |
| 33 | Myristic acid | | **879** | **96** | **14** | **0.1** | **3** |
| 34 | Glucose | | **899** | 15 | **7** | **0.6** | **3** |
| 35 | / | 255 | | | | | 0 |
| 36 | Lauric acid | | **806** | **90** | **8** | **0.1** | **3** |
| 37 | Terephthalic acid** | | 644 | 0 | 10 | 0.2 | 2 |

Table S-41. Annotation of the candidate biomarkers. The candidates highlighted in the CD vs HC separation are above. The candidates highlighted in the 3 CD subgroups separation are below. The values that meet the significance thresholds are in bold.

## S-22. Biological functions of the candidate biomarkers.

| Compound | 2-Hydroxybutyric acid | 1,5-anhydroglucitol | Citric acid | Galactose | Erythrose / Threose | Decanoic acid |
|---|---|---|---|---|---|---|
| HMDB | 0000008 | 0002712 | 0000094 | 0000143 | 0002649 | 0000511 |
| CHEBI | 1148 | 16070 | 30769 | 28061 | 23956 | 30813 |
| Pubchem | 11266 | 64960 | 311 | 6036 | 439574 | 2969 |
| KEGG | C05984 | C07326 | C00158 | C00124 | C01796 | C01571 |
| CAS | 565-70-8 | 154-58-5 | 77-92-9 | 3646-73-9 | 1758-51-6 | 334-48-5 |
| Source | Endogenous | Endogenous | Endogenous, food | Endogenous, food | Food | Endogenous, food |
| Tissues and organs | Prostate | Prostate | All Tissues<br>Prostate | Brain<br>Liver<br>Prostate<br>Blood<br>Breast Milk<br>Cellular Cytoplasm<br>Feces<br>Saliva | Cartilage | Liver<br>Prostate<br>Stratum Corneum<br>Thyroid Gland |
| Biofluid | Blood<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine | Blood<br>CSF<br>Saliva<br>Sweat<br>Urine | Blood<br>Breast Milk<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine | Urine<br>Extracellular<br>Lysosome | Blood<br>Feces | Blood<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine |
| Cell | Cytoplasm<br>Extracellular | Cytoplasm (predicted) | Cytoplasm<br>Extracellular<br>Mitochondria | na | Cytoplasm (predicted) | Cytoplasm<br>Extracellular<br>Membrane (predicted) |
| Biological role | Metabolite | na | na | na | na | Membrane stabilizer<br>Energy source<br>Energy storage<br>Nutrient |
| Biochemical process | na | na | na | na | na | Lipid transport<br>Lipid metabolism<br>Fatty acid metabolism |
| Cellular process | na | na | na | na | na | Cell signaling |
| Chemical reaction | na | na | na | na | na | Beta Oxidation of LC Fatty Acids |
| Environmental role | na | na | na | na | na | Lipid peroxidation |
| Industrial application | na | na | Pharmaceutical<br>Antimicrobial agent<br>Food and nutrition<br>Personal care products<br>Chelating agent | na | na | Food and nutrition<br>Surfactant<br>Emulsifier<br>Pharmaceutical |
| **Pathways** | | | | | | |
| HMDB | Yes | na | Yes | Yes | na | Yes |
| KEGG | Yes | na | Yes | Yes | na | Yes |
| **Biomarker** | | | | | | |
| CD | na | na | Yes | na | na | na |
| IBD | na | na | Yes | na | na | na |
| UC | na | na | Yes | na | na | na |

| Compound | Erythronic acid | Threonic acid | Aspartic acid | Threonolactone* | Glutamic acid | Xylose | Methionine |
|---|---|---|---|---|---|---|---|
| HMDB | 0000613 | 0000943 | 0000191 | 0000940 | 0000148 | 0000098 | 0000696 |
| CHEBI | 37655 | 15908 | 17053 | 71176 | 16015 | 53455 | 16643 |
| Pubchem | 2781043 | 5460407 | 5960 | 2724794 | 33032 | 135191 | 6137 |
| KEGG | / | C01620 | C00049 | / | C00025 | C00181 | C00073 |
| CAS | 13752-84-6 | 7306-96-9 | 56-84-8 | 21730-93-8 | 56-86-0 | 58-86-6 | 63-68-3 |
| Source | Endogenous | na | Endogenous, food | Endogenous | Endogenous, food | Endogenous, food | Endogenous, food |
| Tissues and organs | na | na | All tissues<br>Prostate | na | Adrenal Medulla<br>Epidermis<br>Fibroblasts<br>Intestine<br>Kidney<br>Muscle<br>Myelin<br>Nerve Cells<br>Neuron<br>Pancreas<br>Placenta<br>Platelet<br>Prostate<br>Skeletal Muscle<br>Spleen<br>Stratum Corneum | Erythrocyte<br>Small Intestine | Fibroblasts<br>Kidney<br>Liver<br>Muscle<br>Pancreas<br>Prostate<br>Spleen |
| Biofluid | Blood<br>CSF<br>Feces<br>Saliva<br>Urine | Blood<br>Feces<br>Saliva<br>Sweat<br>Urine | Blood<br>Breast milk<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine | na | Blood<br>Cellular Cytoplasm<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine | Blood<br>Feces<br>Saliva<br>Urine | CSF<br>Feces<br>Saliva<br>Sweat<br>Urine |
| Cell | Cytoplasm (predicted) | Cytoplasm (predicted) | Cytoplasm<br>Mitochondria<br>Extracellular | na | Extracellular<br>Mitochondria<br>Lysosome<br>Endoplasmic reticulum | Cytoplasm<br>Extracellular<br>Lysosome | Extracellular |
| Biological role | na | na | Essential amino acid | na | Metabolite<br>Trace element<br>Molecular messenger<br>Neurotransmitter | na | Drug metabolite<br>Essential amino acid<br>Waste product<br>Trace element<br>Atherogen (indirect)<br>Metabotoxin (indirect) |
| Biochem. proc. | na | na | na | na | na | na | na |
| Cellular process | na | na | na | na | na | na | na |
| Chemical reaction | na | na | na | na | na | na | na |
| Environmental role | na | na | na | na | na | na | na |
| Industrial application | na | Pharmaceutical | Pharmaceutical | na | Nutritional supplement<br>Pharmaceutical | Food and nutrition<br>Personal care products<br>Pharmaceutical | Nutritional supplement<br>Pharmaceutical |
| **Pathways** | | | | | | | |
| HMDB | na | na | Yes | na | Yes | na | Yes |
| KEGG | na | Yes | Yes | na | Yes | Yes | Yes |
| **Biomarker** | | | | | | | |
| CD | na | na | Yes | na | Yes | Yes | Yes |
| IBD | na | na | Yes | na | Yes | na | Yes |
| UC | na | na | Yes | na | Yes | na | Yes |

| Compound | Erythritol | Threitol | Myristic acid | Glucose | Dodecanoic acid | Terephthalic acid* |
|---|---|---|---|---|---|---|
| HMDB | 0002994 | 0004136 | 0000806 | 0000122 | 0000638 | 0002428 |
| CHEBI | 17113 | 48300 | 28875 | 4167 | 30805 | 15702 |
| Pubchem | 222285 | 169019 | 11005 | 5793 | 3893 | 7489 |
| KEGG | C00503 | C16884 | C06424 | C00031 | C02679 | C06337 |
| CAS | 149-32-6 | 2418-52-2 | 544-63-8 | 50-99-7 | 143-07-7 | 100-21-0 |
| Source | Endogenous, food | Endogenous, food | Endogenous, food | Endogenous, food | Endogenous, food | Endogenous |
| Tissues and organs | Epidermis<br>Prostate | na | Adipose Tissue<br>Epidermis<br>Prostate<br>Spleen | Adipose Tissue<br>Adrenal Cortex<br>Adrenal Gland<br>Adrenal Medulla<br>Beta Cell<br>Bladder<br>Brain<br>Brain Plaques<br>Epidermis<br>Eye Lens<br>Fetus<br>Fibroblasts<br>Gonads<br>Gut<br>Intestine<br>Kidney<br>Liver<br>Lung<br>Mouth<br>Muscle<br>Myelin<br>Nerve Cells<br>Neuron<br>Pancreas<br>Placenta<br>Prostate | Stratum Corneum | Fibroblasts<br>Platelet |
| Biofluid | Blood<br>CSF<br>Feces<br>Saliva<br>Urine | Blood<br>CSF<br>Feces<br>Urine | Blood<br>CSF<br>Feces<br>Saliva<br>Urine | Blood<br>Breast Milk<br>CSF<br>Feces<br>Saliva<br>Sweat<br>Urine | Blood<br>Breast Milk<br>Feces<br>Saliva<br>Sweat<br>Urine | Blood<br>Saliva<br>Urine |
| Cell | Cytoplasm (predicted) | na | Cytoplasm<br>Extracellular<br>Membrane (predicted) | Extracellular<br>Lysosome<br>Endoplasmic reticulum<br>Golgi apparatus | Extracellular<br>Membrane (predicted) | na |
| Biological role | na | na | Membrane stabilizer<br>Energy source<br>Energy storage<br>Nutrient | Metabolite | Membrane stabilizer<br>Energy source<br>Energy storage<br>Nutrient | na |
| Biochemical process | na | na | Lipid transport<br>Lipid metabolism<br>Fatty acid metabolism | na | Lipid transport<br>Lipid metabolism<br>Fatty acid metabolism | na |
| Cellular process | na | na | Cell signaling | na | Cell signaling | na |
| Chemical reaction | na | na | Fatty Acid Biosynthesis | na | na | na |
| Environmental role | na | na | Lipid peroxidation | na | Lipid peroxidation | na |
| Industrial application | na | na | Food and nutrition<br>Surfactant | Food and nutrition | Food and nutrition<br>Personal care products | na<br>PET plastic (manual) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | Emulsifier<br>Pharmaceutical | | Surfactant<br>Emulsifier<br>Antibacterial<br>Pharmaceutical | |
| **Pathways** | | | | | | |
| HMDB | na | na | Yes | na | Yes | na |
| KEGG | Yes | na | Yes | Yes | Yes | All microbial |
| **Biomarker** | | | | | | |
| CD | na | na | na | na | na | na |
| IBD | na | na | na | na | na | na |
| UC | na | na | na | Yes | na | na |

Table 42. Biological functions of the candidate biomarkers. * Threonolactone has been excluded as a product of degradation of L-ascorbic acid. Terephthalic acid is most probably an analytical artefact. Another evidence for it is that all pathways it is involved in are microbial ones.

## S-23. Literature review.

Despite the variety of matrices, phenotypes and populations reported, the variations of concentration reported are often in agreement with our observations.

| Known Metabolites in IBD | | |
|---|---|---|
| 2-Hydroxybutyric-acid | Galactose | Methionine |
| Aspartic Acid | Glucose | Xylose |
| Citric Acid | Glutamic acid | Erythronic / Threonic Acid |
| Erythritol /Threitol | | |

| Metabolites unknown in IBD | | |
|---|---|---|
| 1,5-Anhydroglucitol | Erythrose / Threose | Myristic acid |
| Capric acid | Terephthalic acid* | Threonolactone* |
| Lauric acid | | |

Table S-43. Metabolites known and unknown to play a role in IBD. The first ones tend to confirm the results obtained. The others could lead, after validation, to new knowledge that would be helpful in diagnosis and to understand better the CD phenotypes and their functioning. Threonolactone and terephthalic acid are probable artefacts (*).

**2-hydroxybutyrate**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Minamoto | GC-TOFMS | Serum, Feces | IBD Dogs | ↑ |
| Schicho 2012 | NMR | Serum | UC | ↑ |
| Schicho 2012 | NMR | Plasma | UC | ↑ |
| Schicho 2012 | NMR | Plasma | CD | ↑ |
| Dawiskiba | NMR | Urine | CD & UC | ↓ |
| Scoville | UPLC-MS/MS | Serum | CD & UC | ↓ |

**Aspartate**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Ooi | GC-MS | Tissues | CD & UC | ↑ |
| Scoville | UPLC-MS/MS | Serum | CD & UC | ↓ |
| Kolho | UPLC-MS/MS | Feces | CD & UC | Significant |
| Ponnusamy | GC-MS | Feces | IBS | Significant |

**Citrate**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Minamoto | GC-TOFMS | Serum, Feces | IBD Dogs | ↑ |
| Martin | NMR | Plasma | IL-10 mice (vs time) | ↑ |
| Dawiskiba | NMR | Serum | CD & UC | ↓ |
| Dawiskiba | NMR | Urine | CD & UC | ↓ |
| Schicho 2010 | NMR | Serum | DSS-induced (mice) | ↓ |
| Schicho 2012 | NMR | Serum | UC | ↓ |
| Schicho 2012 | NMR | Serum | CD | ↓ |

| | | | | |
|---|---|---|---|---|
| Schicho 2012 | NMR | Urine | UC | ↓ |
| Schicho 2012 | NMR | Urine | CD | ↓ |
| Scoville | UPLC-MS/MS | Serum | CD & UC | ↓ |
| Stephens | NMR | Urine | IBD | Significant |
| Williams 2009 | NMR | Urine | CD & UC | Significant |
| Murdoch | NMR | Urine | IL-10 mice | Significant |

**Galactose**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Schicho 2012 | NMR | Urine | CD | ↑ |
| Schicho 2012 | NMR | Urine | UC | ↓ |

**Glucose**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Williams 2012 | NMR | Serum | CD & UC | ↑ |
| Schicho 2012 | NMR | Serum | UC | ↑ |
| Zhang | NMR | Serum | Early stage colitis | ↑ |
| Balasubramanian | 1H-MRS | CMT | (In)active CD & UC | ↑ |
| Sharma | NMR | CMT | IBD | ↑ |
| Le Gall | NMR | Feces | CD & UC | ↑ |
| Schicho 2010 | NMR | Serum | DSS-induced (mice) | ↓ |
| Schicho 2012 | NMR | Plasma | CD | ↓ |
| Martin | NMR | Plasma | IL-10 mice (vs time) | ↓ |
| Dawiskiba | GC-MS | Urine | IL-10 mice | Significant |
| Scoville | UPLC-MS/MS | Serum | CD & UC | Not Sign. |
| Lai | LC-QTOFMS | Serum | In(active) CD | Not Sign. |

**Glutamate**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Hisamatsu | AA Analyze | Plasma | CD & UC | ↑ |
| Dawiskiba | NMR | Feces | CD & UC | ↑ |
| Bjerrum | NMR | Feces | (In)active CD & UC | ↓ |
| Ooi | GC-MS | Tissues | CD & UC | ↓ |
| Scoville | UPLC-MS/MS | Serum | CD & UC | Not Sign. |
| Le Gall | NMR | Feces | CD & UC | Not Sign. |

**Methionine**

| Reference | Instrumentation | Matrix | Samples | Variation |
|---|---|---|---|---|
| Ooi | GC-MS | Tissues | CD & UC | ↑ |
| Schicho 2012 | NMR | Serum | UC | ↑ |
| Martin | NMR | Plasma | IL-10 mice (vs time) | ↑ |
| Schicho 2010 | NMR | Urine | DSS-induced (mice) | ↓ |
| Schicho 2010 | NMR | Serum | DSS-induced (mice) | ↓ |
| Scoville | UPLC-MS/MS | Serum | CD & UC | Not Sign. |
| Lai | LC-QTOFMS | Serum | In(active) CD | Not Sign. |

| | | | | |
|---|---|---|---|---|
| Hisamatsu | AA Analyze | Plasma | CD & UC | Not Sign. |

| **Xylose** | | | | |
|---|---|---|---|---|
| **Reference** | **Instrumentation** | **Matrix** | **Samples** | **Variation** |
| Minamoto | GC-TOFMS | Serum, Feces | IBD Dogs | ↑ |
| Schicho 2012 | NMR | Urine | CD | ↑ |

| **Erythronic acid / Threonic Acid** | | | | |
|---|---|---|---|---|
| **Reference** | **Instrumentation** | **Matrix** | **Samples** | **Variation** |
| Minamoto | GC-TOFMS | Serum, Feces | IBD Dogs | ↑ |

| **Erythritol / Threitol** | | | | |
|---|---|---|---|---|
| **Reference** | **Instrumentation** | **Matrix** | **Samples** | **Variation** |
| Minamoto | GC-TOFMS | Serum, Feces | IBD Dogs | ↑ |

Table S-44. Literature review for the selected and annotated potential biomarkers. Comparison of experimental and reported variations. Threonolactone and terephthalic acid are probable artefacts (*). CMT stands for colonic mucosal tissue. The metabolites in grey are the ones with opposite reported variations.

**REFERENCES**


(1) Kirwan, J. A.; Brennan, L.; Broadhurst, D.; Fiehn, O.; Cascante, M.; Dunn, W. B.; Schmidt, M. A.; Velagapudi, V. Preanalytical Processing and Biobanking Procedures of Biological Samples for Metabolomics Research: A White Paper, Community Perspective (for "Precision Medicine and Pharmacometabolomics Task Group"—The Metabolomics Society Initiative). *Clin. Chem.* **2018**, *64* (8), 1158 LP – 1182. https://doi.org/10.1373/clinchem.2018.287045.

(2) Service public fédéral. Emploi travail et concertation sociale. Code du bien-être au travail. Livre VII. - Agents biologiques. Titre 1er. - Dispositions générales. http://www.emploi.belgique.be/DownloadAsset.aspx?id=46059 (accessed Jul 18, 2019).

(3) 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* (7), 1060–1083. https://doi.org/10.1038/nprot.2011.335.

(4) Ferreira, S. L. C.; Bruns, R. E.; da Silva, E. G. P.; dos Santos, W. N. L.; Quintella, C. M.; David, J. M.; de Andrade, J. B.; Breitkreitz, M. C.; Jardim, I. C. S. F.; Neto, B. B. Statistical Designs and Response Surface Techniques for the Optimization of Chromatographic Systems. *J. Chromatogr. A* **2007**, *1158* (1–2), 2–14. https://doi.org/10.1016/j.chroma.2007.03.051.

(5) 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* (2). https://doi.org/10.1371/journal.pone.0016957.

(6) Castillo, S.; Mattila, I.; Miettinen, J.; Orešič, M.; Hyötyläinen, T. Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2011**, *83* (8), 3058–3067. https://doi.org/10.1021/ac103308x.

(7) 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* (14), 5731–5739. https://doi.org/10.1021/ac900528b.

(8) 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).

(9) Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G. Stable Isotope Coded Derivatizing Reagents as Internal Standards in Metabolite Profiling. *J. Chromatogr. A* **2013**, *1296*, 196–203. https://doi.org/https://doi.org/10.1016/j.chroma.2013.03.072.

(10) 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* (7), 1–10. https://doi.org/10.1371/journal.pone.0040459.

(11) 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* (24), 8086–8094.

(12) Huang, X.; Regnier, F. E. Differential Metabolomics Using Stable Isotope Labeling and Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2008**, *80* (1),

107–114. https://doi.org/10.1021/ac071263f.

(13) Dunn, W. B.; Summers, A.; Brown, M.; Goodacre, R.; Lambie, M.; Johnson, T.; Wilkie, M.; Davies, S.; Topley, N.; Brenchley, P. Proof-of-Principle Study to Detect Metabolic Changes in Peritoneal Dialysis Effluent in Patients Who Develop Encapsulating Peritoneal Sclerosis. *Nephrol. Dial. Transplant.* **2012**, *27* (6), 2502–2510. https://doi.org/10.1093/ndt/gfr662.

(14) de Castro, M. D.; Izquierdo, A. Lyophilization: A Useful Approach to the Automation of Analytical Processes? *J. Automat. Chem.* **1990**, *12* (6), 267–279. https://doi.org/10.1155/S1463924690000347.

(15) Lien, S. K.; Kvitvang, H. F. N.; Bruheim, P. Utilization of a Deuterated Derivatization Agent to Synthesize Internal Standards for Gas Chromatography–Tandem Mass Spectrometry Quantification of Silylated Metabolites. *J. Chromatogr. A* **2012**, *1247*, 118–124. https://doi.org/https://doi.org/10.1016/j.chroma.2012.05.053.

(16) 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* (4), 1473–1482. https://doi.org/10.1016/j.talanta.2008.09.043.

(17) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry. *Anal. Chem.* **2000**, *72* (15), 3573–3580. https://doi.org/10.1021/ac991142i.

(18) Fiehn, O. Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr. Protoc. Mol. Biol.* **2016**, *114* (1), 30.4.1-30.4.32. https://doi.org/10.1002/0471142727.mb3004s114.

(19) 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* (1), 44–66. https://doi.org/10.1007/s11306-012-0434-4.

(20) 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* (1), 3–20. https://doi.org/10.3390/metabo1010003.

(21) Kowalski, J.; Misselwitz, M.; Cochran, J. Investigating GCxGC Separations Using Selective Column Chemistry and Compound Derivatization Pairings for Common Metabolomics Chemical Compounds. In *4th Multidimensional Chromatography Workshop*; Toronto, Canada, 2013; p 36.

(22) LCGC Editors. GC Temperature Programming—10 Things You Absolutely Need to Know. http://www.chromatographyonline.com/gc-temperature-programming-10-things-you-absolutely-need-know-0 (accessed Jul 18, 2019).

(23) 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. https://doi.org/10.1007/s11306-007-0057-3.

(24) Cochran, J. Split Injection GC: The Benefits of “Shoot-and-Dilute” GC. http://www.chromatographyonline.com/split-injection-gc-benefits-shoot-and-dilute-gc (accessed Jul 18, 2019).

(25) 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).

(26) Gustavo González, A.; Ángeles Herrador, M. A Practical Guide to Analytical Method Validation, Including Measurement Uncertainty and Accuracy Profiles. *TrAC Trends Anal. Chem.* **2007**, *26* (3), 227–238. https://doi.org/https://doi.org/10.1016/j.trac.2007.01.009.

(27) Ducauze, C.; Baillet-Guffroy, A.; Bui, T. X. Choix et Validation d'une Méthode d'Analyse http://www2.agroparistech.fr/IMG/pdf/Choix_et_Validation_Methode-2.pdf (accessed Jul 18, 2019).

(28) 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* (8), 535–552. https://doi.org/10.1016/j.trac.2004.04.001.

(29) 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).

(30) 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. https://doi.org/https://doi.org/10.1016/j.chroma.2014.08.015.

(31) 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).

(32) Raileanu, L. E.; Stoffel, K. Gini Index and Information Gain. *Ann. Math. Artif. Intell.* **2004**, *41* (2100), 77–93. https://doi.org/10.1023/B:AMAI.0000018580.96245.c6.

(33) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. A Tutorial Review: Metabolomics and Partial Least Squares-Discriminant Analysis - a Marriage of Convenience or a Shotgun Wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. https://doi.org/10.1016/j.aca.2015.02.012.

(34) Robnik-Šikonja, M. Improving Random Forests. In *European Conference on Machine Learning*; Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D., Eds.; Springer, 2004; pp 359–370.

(35) Tu, J. V. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *J. Clin. Epidemiol.* **1996**, *49* (11), 1225–1231. https://doi.org/https://doi.org/10.1016/S0895-4356(96)00002-9.

(36) Hoenig, J. M.; Heisey, D. M. The Abuse of Power. *Am. Stat.* **2001**, *55* (1), 19–24. https://doi.org/10.1198/000313001300339897.

(37) Thomas, L. Retrospective Power Analysis. *Conserv. Biol.* **1997**, *11* (1), 276–280. https://doi.org/10.1046/j.1523-1739.1997.96102.x.

(38) Blaise, B. J.; Correia, G.; Tin, A.; Young, J. H.; Vergnaud, A.-C.; Lewis, M.; Pearce, J. T. M.; Elliott, P.; Nicholson, J. K.; Holmes, E.; et al. Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Anal. Chem.* **2016**, *88* (10), 5179–5188. https://doi.org/10.1021/acs.analchem.6b00188.

(39) 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.

(40) Broadhurst, D. I.; Kell, D. B. Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments. *Metabolomics* **2006**, *2* (4), 171–196. https://doi.org/10.1007/s11306-006-0037-z.

(41) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; Velzen, E. J. J.; Duijnhoven, J. P. M.; Dorsten, F. A. Assessment of PLSDA Cross Validation. *Metabolomics* **2008**, *4* (1), 81–89. https://doi.org/10.1007/s11306-007-0099-6.

(42) Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2013**, *1* (1), 92–107. https://doi.org/10.2174/2213235X11301010092.

(43) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational Biomarker Discovery in Clinical Metabolomics: An Introductory Tutorial. *Metabolomics* **2013**, *9* (2), 280–299. https://doi.org/10.1007/s11306-012-0482-9.

(44) Schicho, R.; Shaykhutdinov, 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* (6), 3344–3357. https://doi.org/10.1021/pr300139q.

(45) Wu, H.; Xue, R.; Tang, Z.; Deng, C.; Liu, T.; Zeng, H.; Sun, Y.; Shen, X. Metabolomic Investigation of Gastric Cancer Tissue Using Gas Chromatography/Mass Spectrometry. *Anal. Bioanal. Chem.* **2010**, *396* (4), 1385–1395. https://doi.org/10.1007/s00216-009-3317-4.

(46) 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* (1), 33–47. https://doi.org/10.1080/19490976.2014.997612.

(47) 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* (14), 5524–5531. https://doi.org/10.1021/ac8005236.

(48) 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* **2014**, *11* (1), 122–133. https://doi.org/10.1007/s11306-014-0677-3.

(49) 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* (2), e42–e48. https://doi.org/10.1016/j.crohns.2012.04.019.

(50) Dawiskiba, T.; Deja, S.; Mulak, A.; Zabek, A.; Jawień, E.; Pawełka, D.; Banasik, M.; Mastalerz-Migas, A.; Balcerzak, W.; Kaliszewski, K.; et al. Serum and Urine Metabolomic Fngerprinting in Diagnostics of Inflammatory Bowel Diseases. *World J. Gastroenterol.* **2014**, *20* (1), 163–174. https://doi.org/10.3748/wjg.v20.i1.163.

(51) 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* (6), 1435–1444. https://doi.org/10.1038/ajg.2009.175.

(52) 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* (3), 321–334. https://doi.org/10.1093/ecco-jcc/jjw158.

(53) 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 1 H NMR Spectroscopy. *J. Proteome Res.* **2010**, *9* (12), 6265–6273. https://doi.org/10.1021/pr100547y.

(54) Lin, H. M.; Helsby, N. A.; Rowan, D. D.; Ferguson, L. R. Using Metabolomic Analysis to Understand Inflammatory Bowel Diseases. *Inflamm. Bowel Dis.* **2011**, *17* (4), 1021–1029. https://doi.org/10.1002/ibd.21426.

(55) Hisamatsu, T.; Okamoto, S.; Hashimoto, M.; Muramatsu, T.; Andou, A.; Uo, M.; Kitazume, M. T.; Matsuoka, K.; Yajima, T.; Inoue, N.; et al. Novel, Objective, Multivariate Biomarkers Composed of Plasma Amino Acid Profiles for the Diagnosis and Assessment of Inflammatory Bowel Disease. *PLoS One* **2012**, *7* (1). https://doi.org/10.1371/journal.pone.0031131.

(56) 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. *Inflamm. Res.* **2011**, *60* (9), 831–840. https://doi.org/10.1007/s00011-011-0340-7.

(57) 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* (9), 4208–4218. https://doi.org/10.1021/pr2003598.

(58) 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* (2), 546–551. https://doi.org/10.1021/pr060470d.