

Supporting Information

A large scale inter-laboratory DI-FT-ICR MS comparability study employing various systems.

Sara Forcisi^{1,2, ‡,*}, Franco Moritz^{1, ‡,*}, Christopher J. Thompson^{3, ‡,*}, Basem Kanawati¹, Jenny Uhl¹, Carlos Afonso⁴, Chantal D. Bader⁵, Aiko Barsch⁶, Berin A. Boughton^{7,‡}, Rosalie K. Chu⁸, Justine Ferey^{4,‡}, Francisco Fernandez-Lima^{9,10}, Céline Guéguen^{11,¶}, Dimitri Heintz¹², Mario Gomez-Hernandez^{9,10}, Kyoung-Soon Jang¹³, Nikolas Kessler⁶, Vaughn Mangal¹¹, Rolf Müller⁵, Ryo Nakabayashi¹⁴, Edith Nicol¹⁵, Simone Nicolardi¹⁶, Magnus Palmblad¹⁶, Ljiljana Paša-Tolić⁸, Jacob Porter^{9,10}, Isabelle Schmitz-Afonso⁴, Jong Bok Seo¹⁷, Eduardo Sommella¹⁸, Yuri E. M. van der Burgt¹⁶, Claire Villette¹², Matthias Witt⁶, Ashley Wittrig^{20,‡}, Jeremy J. Wolff³, Michael L. Easterling^{3,21}, Frank H. Laukien^{3,21}, Philippe Schmitt-Kopplin^{1,2,22,*}.

1. Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
2. German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany.
3. Bruker Daltonics Inc, Billerica, Massachusetts 01821, United States.
4. COBRA, UMR 6014 et FR 3038, INSA de Rouen, CNRS, IRCOF, Normandie Université, Université de Rouen, 76130 Mont Saint Aignan CEDEX, France.
5. Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarland University Campus, 66123 Saarbrücken, Germany and Department of Pharmacy, Saarland University, 66123 Saarbrücken, Germany
6. Bruker Daltonik GmbH, Fahrenheitstrasse 4, 28359 Bremen, Germany.
7. Metabolomics Australia, School of BioSciences, University of Melbourne, Melbourne, Victoria 3010, Australia.
8. Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, USA.
9. Department of Chemistry and Biochemistry, Florida International University, 11200 SW Eighth Street, AHC4-233, Miami, Florida 33199, United States.
10. Biomolecular Sciences Institute, Florida International University, 11200 Eighth Street, AHC4-211, Miami, Florida 33199, United States.
11. Chemistry Department, Trent University, 1600 West Bank Drive, Peterborough, ON, K9J 7B8, Canada.
12. Plant Imaging and Mass Spectrometry (PIMS), Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France.
13. Bio-Chemical Analysis Team, Korea Basic Science Institute, Cheongju, 28119, South Korea.
14. Metabolomics Research Group, RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan.
15. Laboratoire de Chimie Moléculaire (LCM), CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91128, Palaiseau, France.
16. Center for Proteomics and Metabolomics, Leiden University Medical Center Leiden, 2333 ZC Leiden, The Netherlands.
17. Seoul Center, Korea Basic Science Institute, 145, Anam-Ro, Seongbuk-Gu, 02841 Seoul, South Korea.
18. Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy.
19. Bruker Daltonik GmbH, Fahrenheitstrasse 4, 28359 Bremen, Germany.
20. ExxonMobil Research and Engineering Company, 1545 Route 22 East, Clinton, NJ 08869.
21. Visiting Scholar, Harvard University, Department of Chemistry & Chemical Biology, Cambridge, Massachusetts 02138, United States.
22. Analytical Food Chemistry, Technical University of Munich, 85354 Freising, Germany.

‡ S.F., F.M. and C.J. T. These authors contributed equally.

AUTHOR INFORMATION

Corresponding Author

* Email:sara.forcisi@tum.de

* Email:franco.fmoritz@helmholtz-muenchen.de

* Email:chris.thompson@brightspec.com

* Email:schmitt-kopplin@helmholtz-muenchen.de

Contents

SI-1: Details on data processing and statistics	3
SI-2: Criteria for exclusion of labs	4
SI-3: Conditions for univariate comparability	6
SI-4: Cumulative PCA-plots for co-presence analysis (CPA)	7
SI-5: Normalizations and Scaling	9
SI-6: Zoomed intra-CVs for Z-experiments	11
SI-7: Multivariate Reproducibility.	11
SI-8: PCA cross-validation map of experiment Y	12
SI-9: High Background in Laboratory D	13
SI-10: Analysis SOP. Excerpt relevant to the Manuscript	14
SI-11: REFERENCES	19

SI-1: Details on data processing and statistics

Scaling. MS-features were scaled per laboratory using the Euclidean norm prior to principal component analyses (PCA) and linear regression models. Data were unit variance scaled prior PCA when appropriate.

Imputation. NaN-entries in data that were subject to scaling were imputed feature following $X_{\text{imputed}} = 0.85 \cdot \min(x) + (0.9 \cdot \min(x) - 0.85 \cdot \min(x)) \cdot \epsilon$, with ϵ being random noise sampled from the uniform distribution.

Visualizations. All visualizations of results (Boxplots, PCAs, combined jitter-line plots with confidence intervals) were computed using the corresponding ggplot2 functionalities in R programming language (R version 3.6.0).

Principle component analyses. PCAs were computed using the corresponding matlab pca-function with or without mean centering and UV-scaling where specified.

Linear regression models. Linear regression models were computed using the matlab fitlm-function, which provides a data structure containing several quality measures such as R2, coefficients, sum of squared residuals and others.

Comparison of signal magnitudes. *Univariate.* Boxplots of raw signal intensities per laboratory and experiment were generated. *Multivariate.* The sequence of the 5th to 95th percentiles was computed per MS-spectrum (raw feature intensities; 91 percentile values per spectrum). Percentile lists were merged into one matrix and subjected to un-centered PCA.

SI-2: Criteria for exclusion of labs

1. Insufficient sensitivity

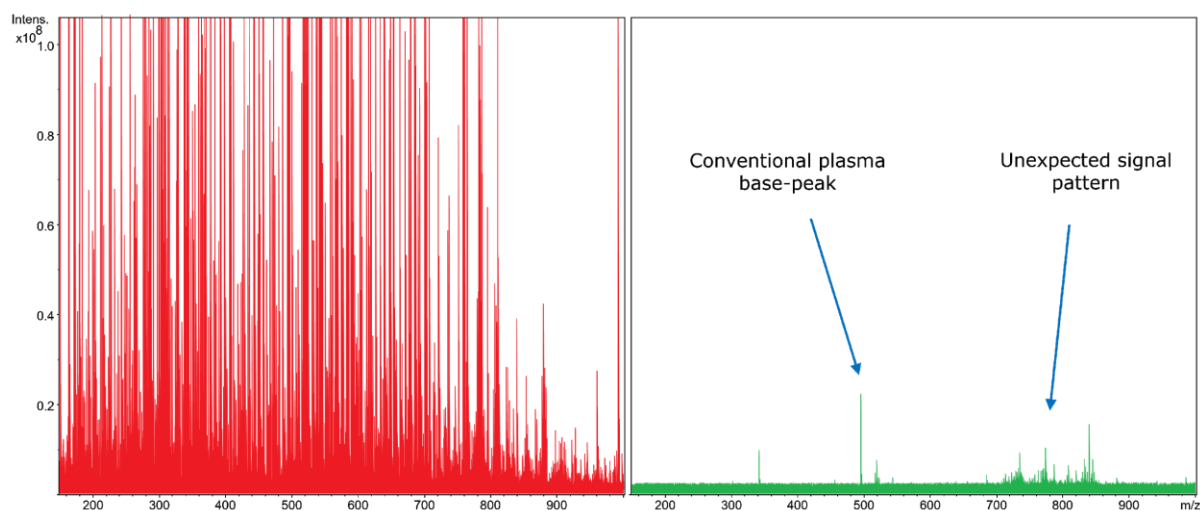


Figure S1. Example of 'low sensitivity' as a reason for exclusion of analyses from the inter-laboratory study. Typical signal magnitudes are displayed on the left panel, while signal magnitudes on the right panel are exemplary for insufficient sensitivity (magnitudes of counts were equalized). The right spectrum barely shows the base-peak typical for human plasma, while possessing an unexpected signal pattern.

2. Activated Collision Energy.

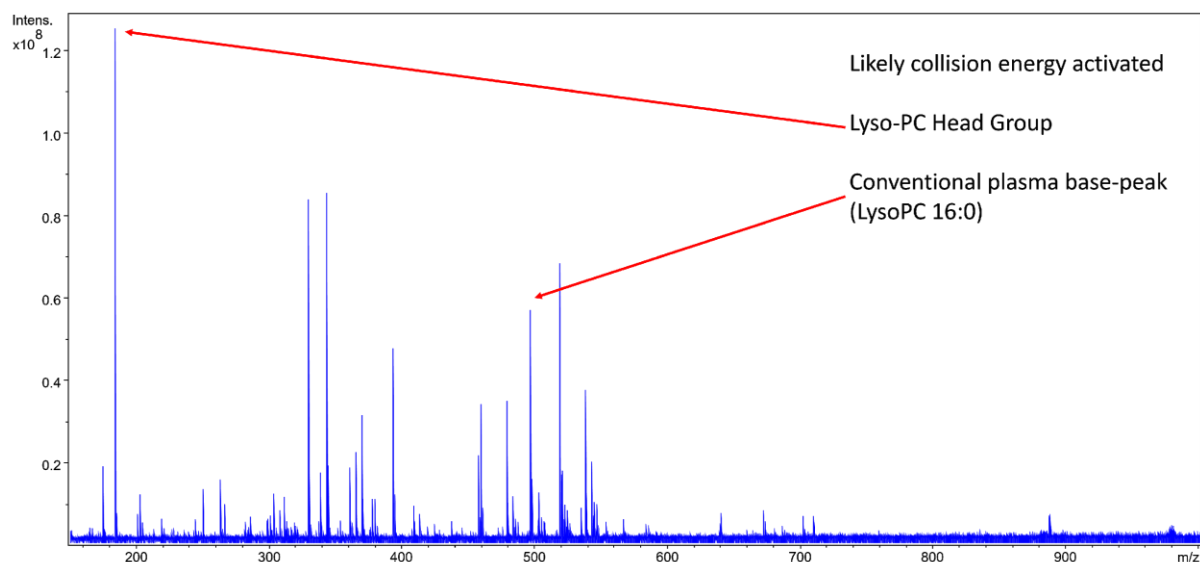


Figure S2. Likely activated collision energy. The base peak of human plasma spectra corresponds to Lyso-PC(16:0) normally. This peak produces a fragment ion at m/z 184, generated by the Lyso-PC head group. This fragment is the base peak in this excluded spectrum (indicating activated collision energy).

3. Strong signal variability due to unstable ionization.

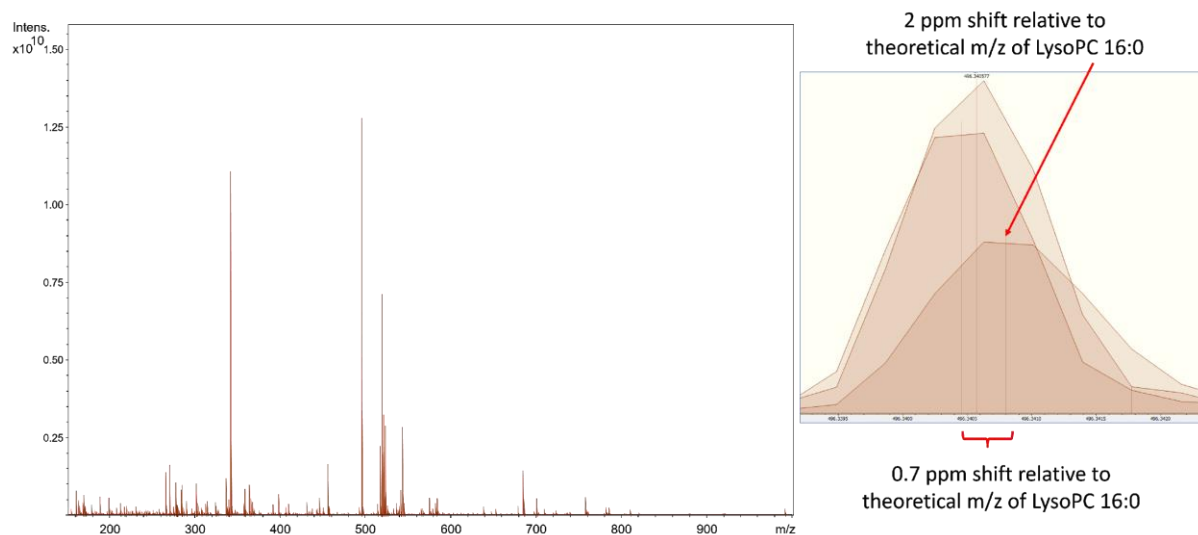


Figure S3. Specific ion sources produced spectra at varying signal magnitudes, leading to corresponding m/z shifts. m/z depends on the size of ion packages as well as FT-ICR MS; particularly when the older infinity cells are used for detection. Large intensity shifts can prevent successful calibration.

4. Peak splitting and peak deformation.

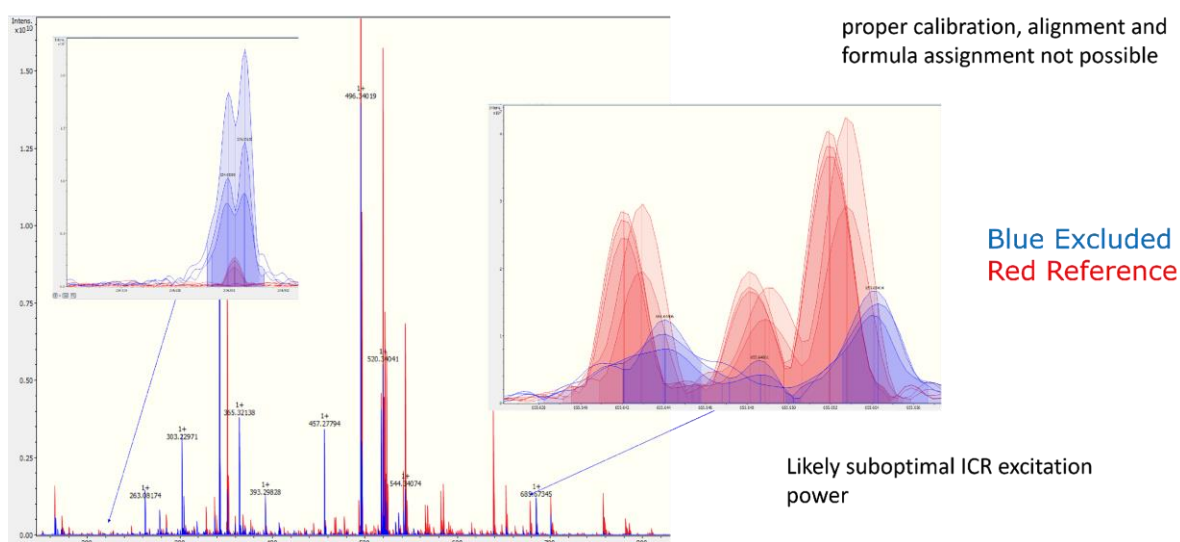


Figure S4. Sub-optimal ICR excitation powers and/or over-saturation of the ICR cell lead to deformed, and even split m/z peaks. Such data cannot be calibrated, let alone be aligned to other spectra.

5. Polymeric Contamination.

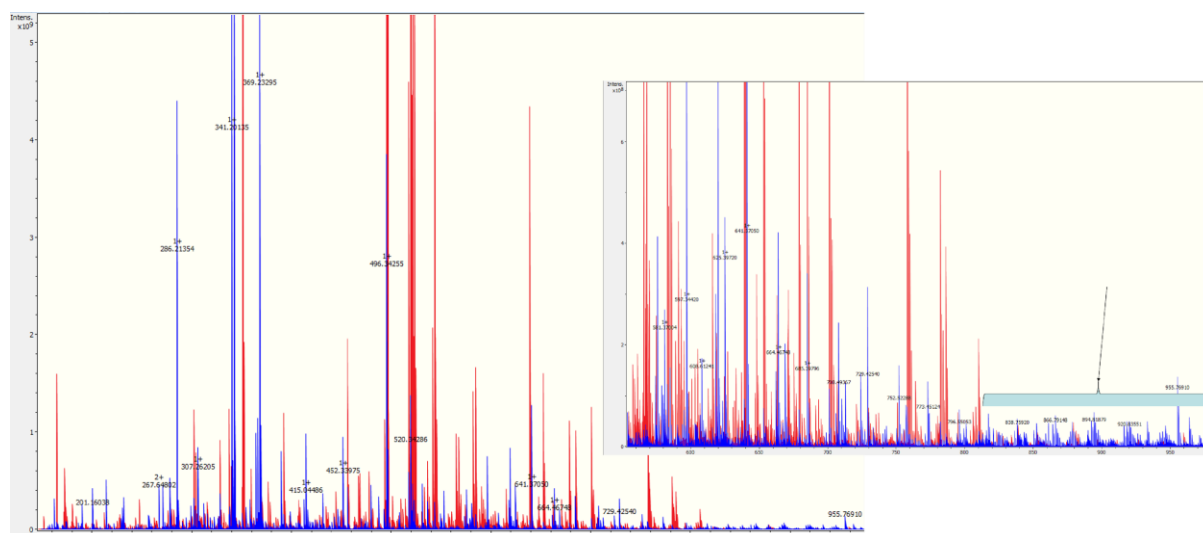


Figure S5. Polymeric contamination bears the potential of producing numerous artefact m/z peaks and may serve to suppress the analytical matrix. The can further distort m/z-error distributions. (Blue: Excluded, Red: Reference)

SI-3: Conditions for univariate comparability

The prime condition for univariate comparability is that a variable X in lab One ‘lives’ on the same distribution as the same variable X in lab Two. That is that $\mu(X_{\text{One}}) = \mu(X_{\text{Two}})$ and $\sigma(X_{\text{One}}) = \sigma(X_{\text{Two}})$ approximately holds for the means and standard deviations of X in lab One and lab Two, respectively. This condition is realized by means of calibration against an authentic standard S in targeted analyses. Given a calibration function $X = aS + b$, any value of intensity X is mapped to known concentrations of S by $S = (X - b)/a$. Comparing this expression to UV-scaling: ${}^hX = (X - \mu(X))/\sigma(X)$ – with index h meaning ‘hat’ – shows, that applying the inverse calibration function on an MS-feature is analogous to (UV-)scaling the MS-feature. Applying the calibration slope is equivalent to normalizing by standard deviation and subtracting the y-intercept is analogous to mean centering in that offsets on the y-axis are compensated for. Both respective actions align the variables’ spreads (a, σ) and positions (b, μ). It follows that quantitation is equivalent to scaling with the difference being that quantitation scales into the space of molar concentration, while scaling does not map into physical dimensions. If analytical accuracy is subject to scaling, then analytical precision is subject to normalization. The second condition for univariate comparability is the analytical precision by which an analytical method produces signals given the same sample. In both quantitative targeted analysis and untargeted analysis it is necessary to normalize each sample’s spectrum to cancel external effectors on precision¹⁻³. Such effectors may be differing dilution factors for (bio) fluids or different sample weights for biopsies, both of which are generally assumed to exert a linear shift or scaling that is the same for all MS-peaks in a spectrum. A common approach is to divide individual signal intensities either by the sum of signal intensities in an MS spectrum (L1 norm or Manhattan norm), by the square root of the sum of squared signal intensities (L2 norm or Euclidean norm), any other scaling method applied to samples (mostly without mean centering) or probabilistic quotient normalization (PQN)^{4,5}, to name the most prominent methods. These normalizations assume that an entire MS spectrum is shifted, compressed or stretched by the joint action of external factors linearly. This is, for example, the assumption by which PQN was conceptualized. The assumption was that the major factor influencing Nuclear Magnetic Resonance (NMR)-peak areas other than biology was sample dilution, in their work. Similarly, it is common to correct spectral intensities for the dry-weight of tissue samples or dissolved

organic carbon concentration in fluids. Normalization has the same task both in targeted and untargeted analysis, i.e. to correct for linear effects external factors exert on the analytical process. Methods that can cope with non-linear effects model individual signal trajectories in the context of batch-correction and only apply to features detected in quality controls (QCs)⁴. Indeed, providing univariate reproducibility is the task of minimizing the difference both, of spread and position between spectra and between features. Reproducibility itself is quantified by means of coefficients of variation (CV; relative standard deviation) on triplicate measurements within a laboratory (intra-CV) or on corresponding signal characteristics across laboratories (inter-CV). The minimum CV achievable in untargeted measurements is significantly higher than in targeted, quantitative analyses, since untargeted approaches aim to find a compromise between recovery of maximum information (minimum selectivity) and optimum analytical quality measures. The literature on untargeted ring trials discards features at intra-CV > 20%, while equivalent measures in targeted analytics are smaller by a factor of 10. The above conditions – the scaling and normalization conditions – can be broken down into two more fundamental conditions: 1) samples to be compared have to produce the same peaks; 2) samples to be compared should be generated on instruments with similar m/z-dependent ion transmission. Failure in point 1) cannot be corrected mathematically, while failure in point two disables linear spectral normalization. Here, two strategies are followed to select the laboratories for comparison: 1) Analysis of presence-absence patterns on binary data and 2) recovery of lab-specific biases by means of smoothed quotient correction.

SI-4: Cumulative PCA-plots for co-presence analysis (CPA)

Experiments X

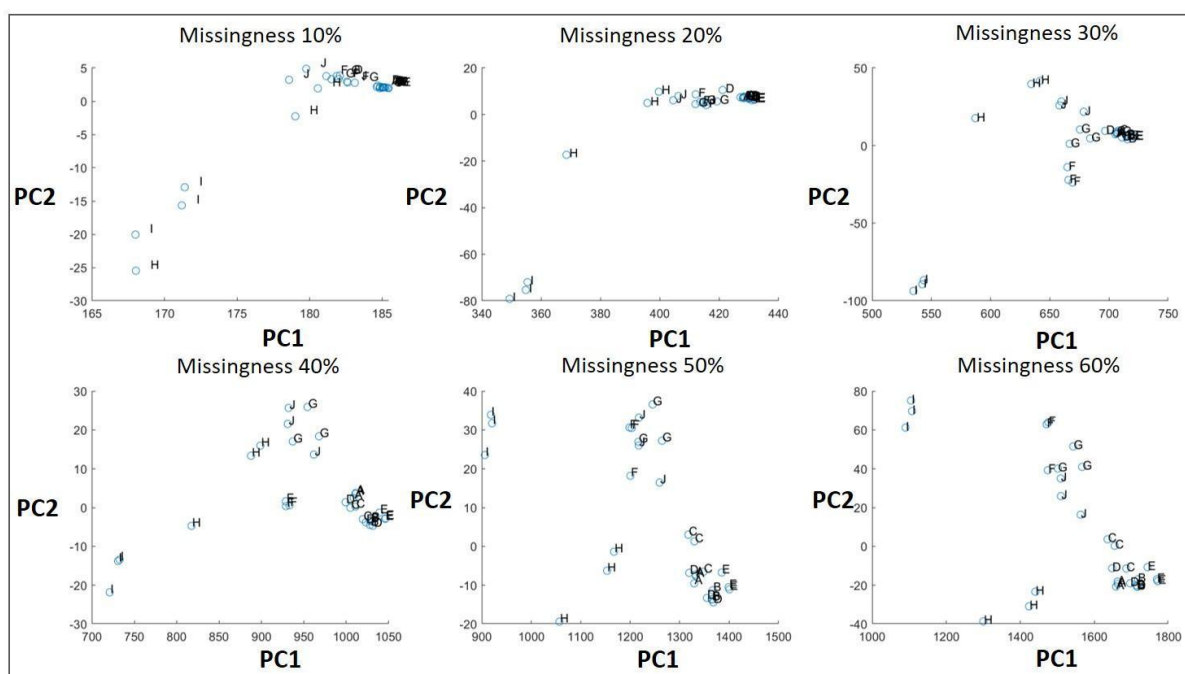


Figure S6. PCA analyses of binary data in X-experiments. Six data-subsets were generated given the indicated thresholds of missingness. Five PCA-models were computed per missingness subset; each for every dilution level. PC1 and PC2 of any dilution were cumulated respectively. This process generated an average mapping of

presence-absence distances between the laboratories. It can be seen, that laboratories {A, B, C, D, E} form a dense co-presence cluster that is stable until 40% of missingness, with density dropping at missingness 50% and 60%.

Experiments Y

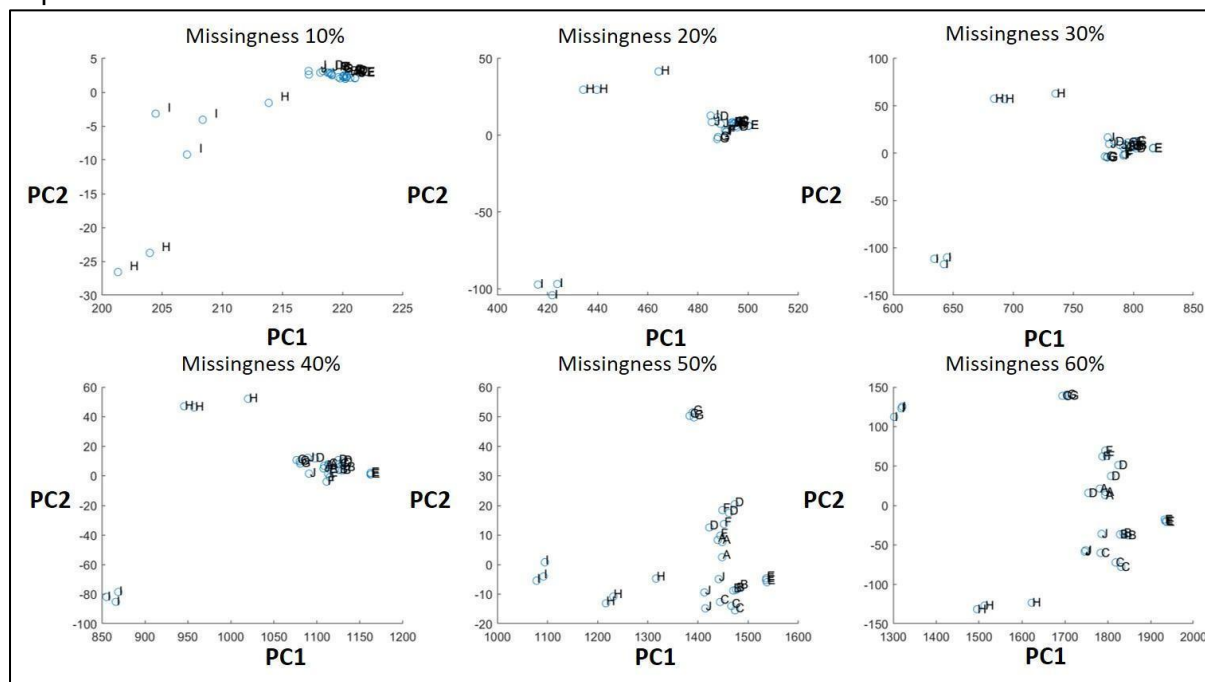


Figure S7. PCA analyses of binary data in Y-experiments. Six data-subsets were generated given the indicated thresholds of missingness. Five PCA-models were computed per missingness subset; each for every pesticide spiking level. PC1 and PC2 of any pesticide concentration were cumulated respectively. This process generated an average mapping of presence-absence distances between the laboratories. It can be seen, that laboratories {A,B,C,D,E,F,J} form a dense co-presence cluster that is stable until 40% of missingness, with density dropping strongly at missingness 50% and 60%. This behavior can be inspected on Fig. 2 in the main text.

Experiments Z

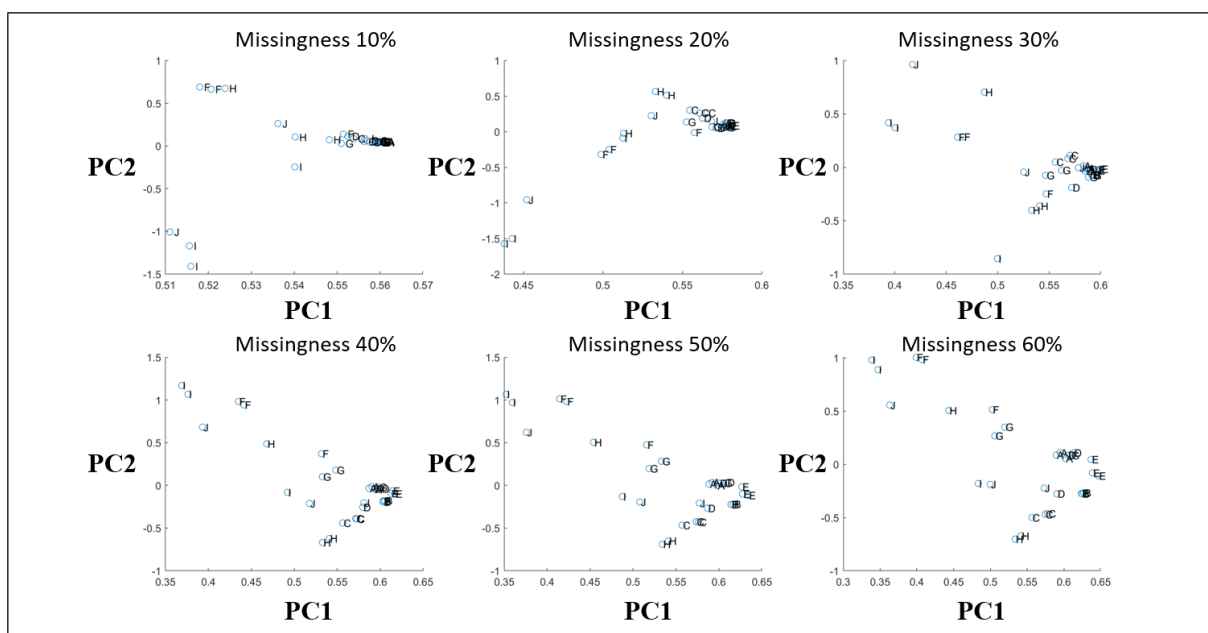


Figure S8. PCA analyses of binary data in Z-experiments. Six data-subsets were generated given the indicated thresholds of missingness. Five PCA-models were computed per missingness subset; each for every pesticide spiking level. PC1 and PC2 of any pesticide concentration were cumulated respectively. This process generated an average mapping of presence-absence distances between the laboratories. It can be seen, that laboratories

{A,B,C,D,E,F,J} form a dense co-presence cluster that is stable until 40% of missingness, with density dropping strongly at missingness 50% and 60%. This behavior can be inspected on Fig. 2 in the main text

SI-5: Normalizations and Scaling

Here, neither classical method for feature normalization¹ served to reduce inter-CVs to acceptable levels without obstructing the density-distributions of intra-CVs (results not shown). At the same time, feature-scaling (UV, Euclidean) reduced inter-CVs markedly, yet not to an acceptable degree if all laboratories were considered.

The fact that scaling reduced inter-CVs (leaving intra-CVs untouched) while normalizations did not, implied that spectral biases manifested not in terms of univariate measures of position (e.g. mean) and spread (e.g. standard deviation) over spectral intensity distributions. While polynomial methods for spectral normalization could have been an option, hypothesizing m/z -dependent characteristics due to instrumental tuning to be interfering with spectral normalization, presented a hypothesis that could be tested more tangibly. The m/z -dependency of spectral characteristics were investigated by generating a reference plasma spectrum $refX_{ij}$ (with indices i and j for features and labs, respectively) over the non-zero intensities of all laboratories. Quotients $Q_{ij} = refX_{ij}/X_{ij}$ were computed (as performed in probabilistic quotient normalization – PQN) and found to show laboratory-specific patterns as a function of m/z (**Figures S9 and S10**).

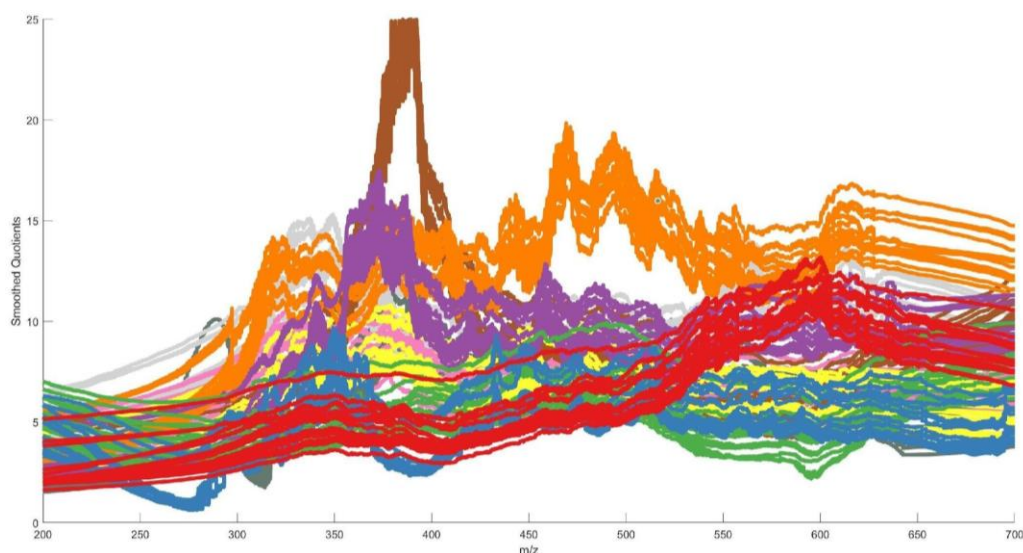


Figure S9. Smoothed quotients over m/z range (abscissa: m/z ; ordinate: smoothed quotients). Laboratories show specific characteristics in terms of deviations from the median spectrum computed from all laboratories. Laboratory colors codes as in Figure 1a.

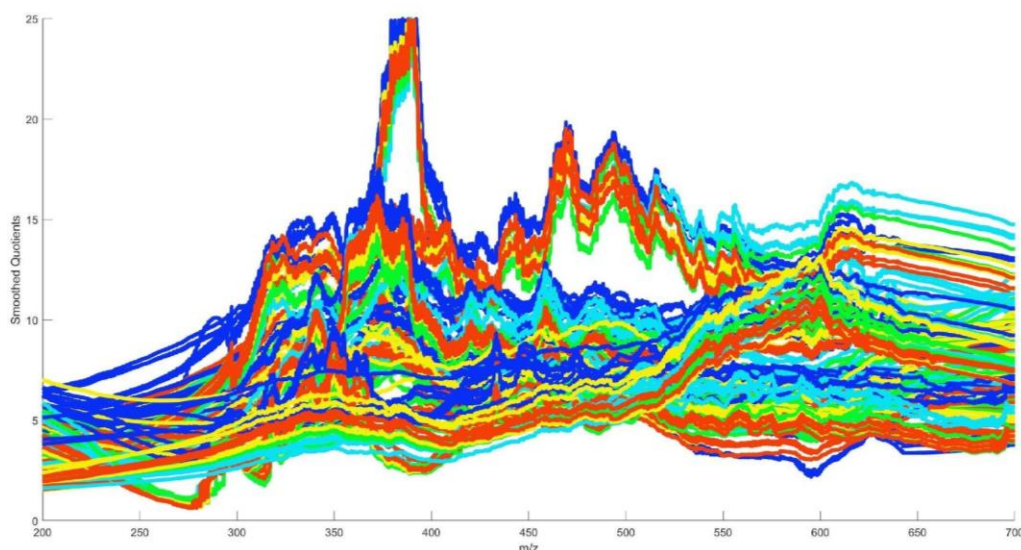


Figure S10. Smoothed quotients over m/z range (abscissa: m/z ; ordinate: smoothed quotients). Laboratories profiles are colored following the different dilution levels 1, 2, 3, 4 and 5.

Subsequently, quotients were smoothed per spectrum using the Matlab expression $SQ_{ij} = \text{smoothdata}(Q_{ij}, 'sgolay', \text{Window}, 'omitnan')$ with window-size per spectrum encompassing 50% of corresponding feature count. A PCA on the smoothed quotients (SQs) of laboratories revealed systematic deviations from the global median spectrum in a laboratory specific fashion (**Figure S11**).

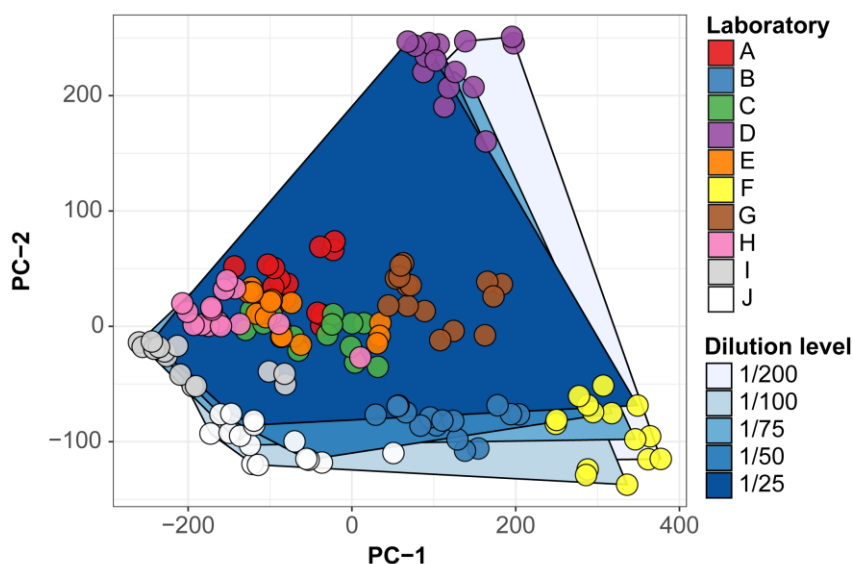


Figure S11. PCA plot on mean centered smoothed quotients (SQs). SQs show strong specificity for laboratory behavior while leaving responses towards pesticide concentration levels untouched (Polygons overlap almost completely). Note that SQs were computed on experiment Y, constituting the stronger effect as compared to experiment X.

Figure S11 shows that SQ's are more specific for laboratory characteristics than for the strong effect of pesticide spiking. Laboratories tend to cluster among themselves while the polygons that are specific for spiking levels overlap. Smoothed quotient correction (SQC) of the data was subsequently performed by multiplying each feature X_{ij} with its corresponding SQ_{ij} . The effects that SQC and Euclidean normalization over the removal of laboratories with aberrant detection patterns exert on both intra-CV and inter-CV are discussed in the subsequent section.

SI-6: Zoomed intra-CVs for Z-experiments



Figure S12. Zoomed-in complement to Fig. 3, showing the convolution of inter-CVs and intra-CVs as a function of lab removal. It can be seen, that inter-CVs seem to outperform intra-CVs at the point where merely the 4 most similar labs in terms of co-presence structure are present. This fact arises because inter-CVs were computed on the triplicate means shared by the laboratories, effectively meaning that the relative standard error of the mean is reported here. This is a valid procedure, as it would be a triplicate's mean value that would be reported by laboratories, so the mean is the value that would be taken into account for a meta-analysis. Intra-CVs represent the variability that is subject to the single measurements within a triplicate. The present result is in agreement with the central limit theorem, where the standard error of the means that estimates how far the means are away from the true value must be smaller than the variation allotted to single measurements that determine the means themselves.

SI-7: Multivariate Reproducibility.

Univariate reproducibility is the focus of optimization in inter-laboratory comparisons, because univariate reproducibility guarantees multivariate reproducibility (e.g. QCs clustering together in a PCA). Is univariate reproducibility an absolute requirement for multivariate reproducibility? What is multivariate reproducibility? From a univariate perspective, multivariate reproducibility would be achieved if the scores of all samples in a PCA clustered into experimental characteristics and QCs perfectly as all laboratories to be examined are part of the same dataset. If all data of all laboratories were present in the same dataset, PCA (whose primary task is the reconstruction of the dataset) would have to model all laboratory specificities including missing data too. This task would 'absorb' considerable proportions of (co)variance, leading to shifts that are caused by sample missingness and most generally, univariate differences. Yet, multivariate reproducibility can be phrased differently: 'Does the multivariate model on my data lead to the same outcome when it is applied to the data generated by another laboratory?'. This task is equivalent to building a multivariate classifier for biomarkers given a clinical study followed by cross-validation against data raised on an equivalent study conducted on a new set of specimens and/or by another laboratory. If such a classifier generalizes well enough to make the same statements on equivalent datasets, the studies captured the underlying effects in a multivariate reproducible manner.

SI-8: PCA cross-validation map of experiment Y

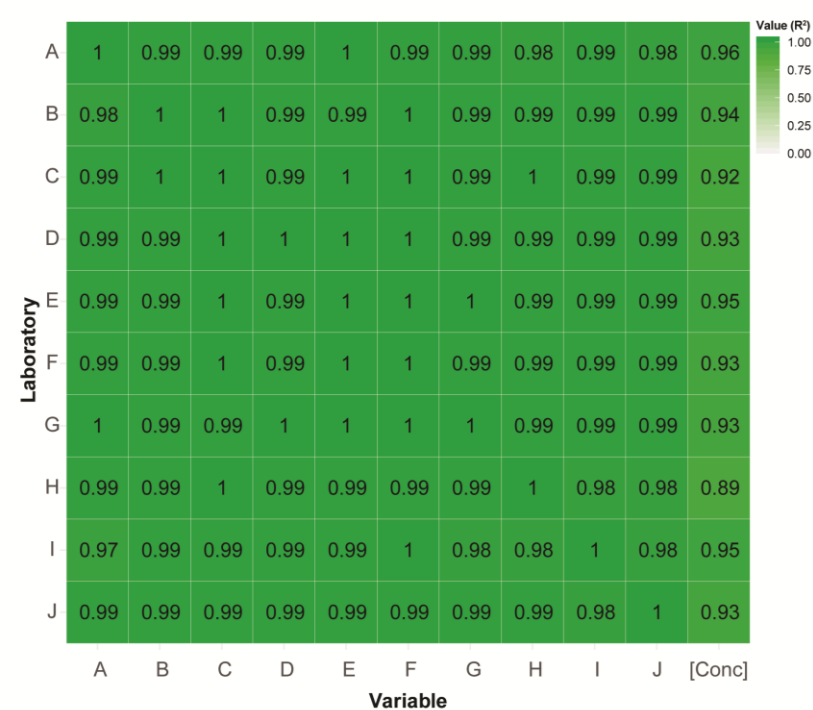


Figure S13. PCA cross-validation map for experiment Y.

SI-9: Scatter plots supporting the PCA cross-validation maps

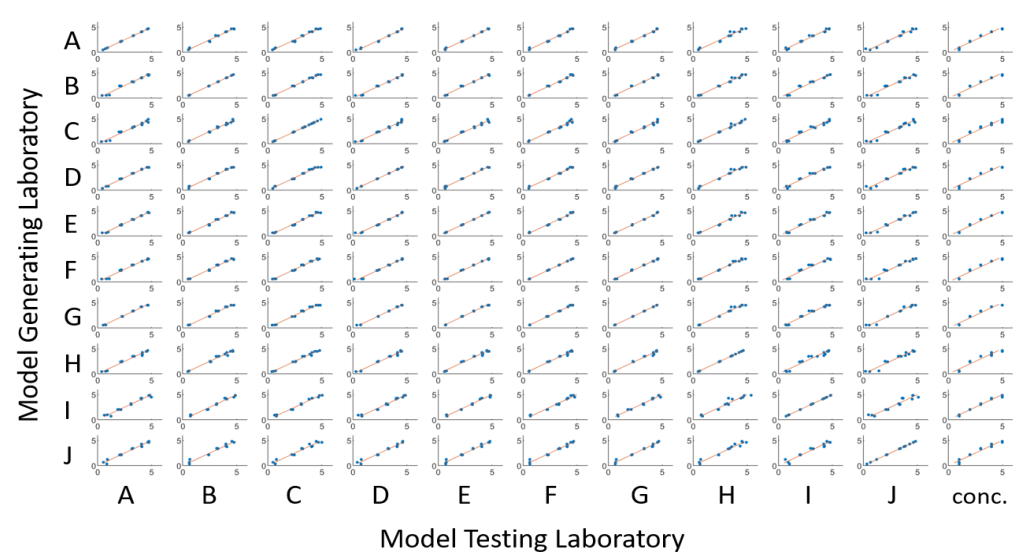


Figure S14. Scatter plot supporting Figure 4.

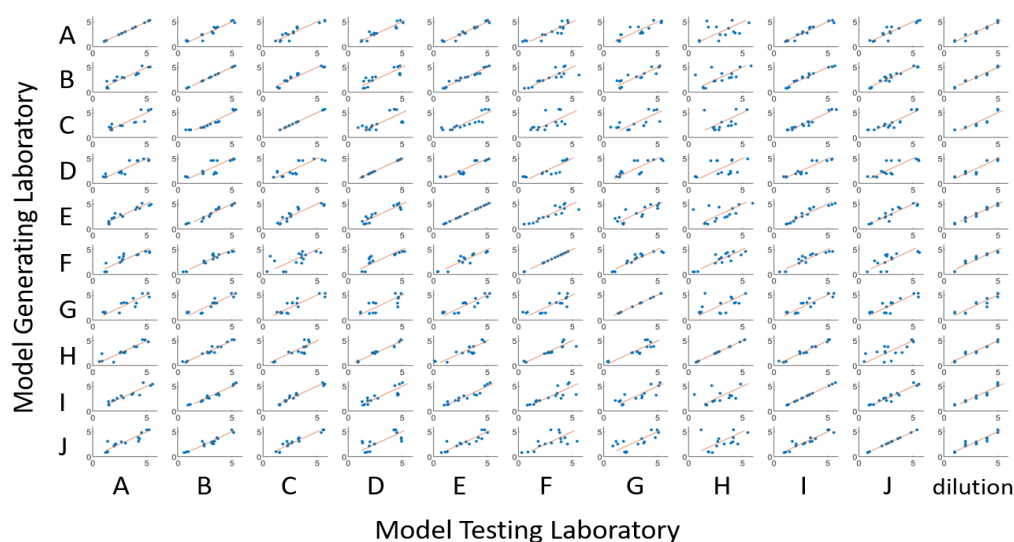


Figure S15. Scatter plot supporting Figure S13.

SI-9: High Background in Laboratory D

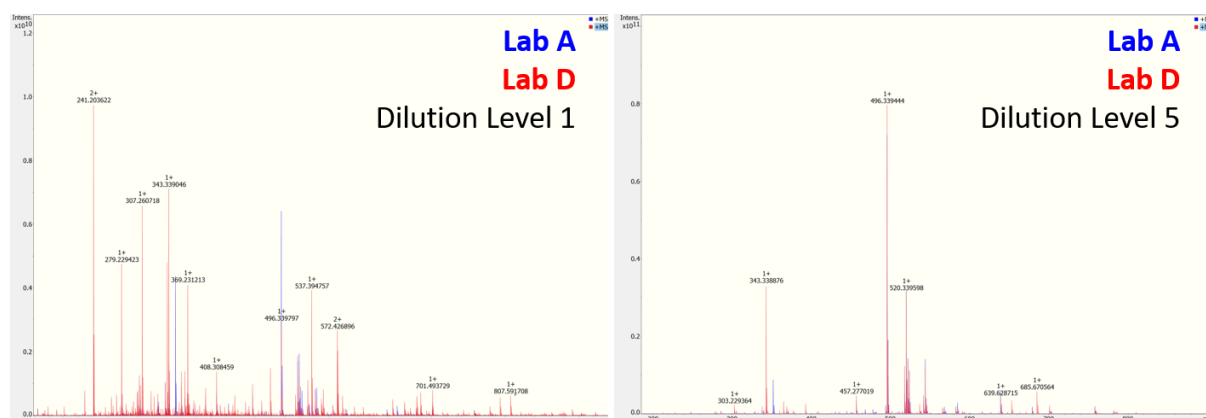


Figure S16. High background in laboratory D. Left panel: Laboratory D shows substantial background/contamination signals, strongly overpowering signals of blood plasma. The typical Lyso-PC(16:0) base-peak has strongly diminished magnitude as compared to laboratory A. Right panel: The background signals in laboratory D are suppressed and the expected plasma pattern, including the base-peak, is restored. Signal magnitude of the base-peak in laboratory D is approximately 10% higher than in laboratory A.

SI-10: Analysis SOP. Excerpt relevant to the Manuscript

1) ICR status check

Before starting the cleaning procedure, check the status of the ICR cell in order to define whether a bake out cycle is necessary. The following criteria are applied at BGC.

- a) Assure that the cell pressure is in the optimal range (between $1 \cdot 10^{-9}$ mbar and $1 \cdot 10^{-10}$ mbar).
- b) Inject a solution of 5 ppm Arginine (Sample B is described in section 5a) from the RingTrial Kit (Table 3) and check the resolution at 523 m/z, the resolution should be above: 204,000 (7T), 350,000 (12T), 437,500 (15T).
- c) If both points are not met, please consider baking out the ICR cell.
- d) Set peak-picking parameter to 20,000 peaks in the FTMS Control

2) Instrumental parameters

The experimenter should check and optimize all the instrument parameters for achieving the highest sensitivity in the mass range m/z **150-1000**. Please note that the low mass will slightly vary with magnetic field. Enter 150 and it will be set accordingly (e.g. 147.32 for 12T). The following table list the parameters that are required for the experimental method.

Parameter (Required)	Value	Parameter (Required)	Value
Data Size	4 MW	Peak Picking Method	Apex
Mass Range	150-1,000 m/z	Maximum Peaks	20,000
Accumulated Scans	100	Abs. Min. Intens.	1.0e5
Accumulation Time	0.2 s	Rel. Min. Intens.	0
Save FID	On	S/N	2
Save Profile Spectrum	On	Data Reduction	Off

As every instrument is slightly different, the ion optic tuning can vary. The following table lists recommended values for key parameters. These values are not required, and you are asked to use the parameters that provide the highest sensitivity within the stated mass range.

Parameter (Suggested)	Value	Parameter (Suggested)	Value
Funnel RF Amp	100 Vpp	Collision Cell RF Freq	2 MHz
Q1 Mass	100 m/z	CC RF Amplitude	1500 Vpp
Collision Voltage	-2.0 V	TOF	0.7 ms
DC Extract Bias	0.6 V	Transfer Frequency	6 MHz
		Xfer RF Amplitude	350 Vpp

Suggested ESI source conditions:

Parameter (Suggested)	Value	Parameter (Suggested)	Value
Flow Rate	2 μ L/min	Nebulizer	1.0 bar
Capillary	4500 V	Dry Gas	4.0 L/min
End Plate Offset	-500	Dry Temp	220 C

3) Wash the ESI-system (PEEK-tube, ESI-needle, shield and cone)

Each participant should be aware that the mass spectrometers that we all use are extremely sensitive towards contaminations. It is necessary to clean the ESI-line before a new sample is injected to maximize the inter-laboratory comparability. The following cleaning sequence for direct injection via syringe is applied at BGC:

a) Wash the ESI cone and ESI shield

Transfer the ESI shield and the ESI cone into different glass beakers. Wash the ESI cone and ESI shield in a glass beaker with a solution of 10 % formic acid in MilliQ water. Sonicate it for 15 minutes in a sonication bath. Discard the liquid and replace it with pure MilliQ water in order to remove all the traces of formic acid. Sonicate for 15 minutes. Replace H₂O with MeOH and sonicate it for 15 minutes before drying the shield and the cone with pure air or N₂. Clean the ESI door with MeOH if you find some residuals.

b) Wash the syringe and ESI needle

First, rinse your syringe in the following sequence:

MilliQ water (3 x 250 μ L) → MeOH (3 x 250 μ L)

Connect the syringe to the ESI sprayer and rinse the unit (source door open) in the following sequence:

MilliQ water (3 x 250 μ L) → MeOH (3 x 250 μ L)

If you perceive any blockage by flushing the different solvents, please consider to remove the ESI needle, reverse it and flush it with the same sequence of solvents in order to remove the blockage.

c) The Ring Trial study applies two washing sequences:

Weak washing:

MilliQ water (3 x 250µL) → MeOH (3 x 250µL)

Strong washing:

MilliQ water (3 x 250µL) → MeOH (3 x 250µL) → IsoPrOH (3 x 250µL) → MeOH (3 x 250µL)

During each cleaning injection, please keep the source door open.

4) System calibration

Infuse a solution of 5 ppm Arginine (Sample B described in section 5a) at a flow rate of 120 µl/h. Calibrate the system. The linear calibration error should be minor 100 ppb for at least 3 Arginine cluster ions.

Table 1: List of the theoretical arginine cluster ions to use as reference calibrant. The optimal error (ppm) for a 12T solarix system, in both polarities, is described as well.

Ion (positive polarity)	Theoretical m/z (positive polarity)	Ion (negative polarity)	Theoretical m/z (negative polarity)	Desired m/z error (ppb)
[M+H] ⁺	175.118952	[M-H] ⁻	173.104399	<100
[2M+H] ⁺	349.230628	[2M-H] ⁻	347.216075	<100
[3M+H] ⁺	523.342304	[3M-H] ⁻	521.327751	<100
[4M+H] ⁺	697.45398	[4M-H] ⁻	695.439428	<100
[5M+H] ⁺	871.565656	[5M-H] ⁻	869.551104	<100

After Arginine injection, rinse the ESI-system with the following sequence:

MilliQ water (3*250µL) → MeOH/MilliQ water (50:50) (3*250µL) → MeOH (3*250µL)

5) Experiment

a) Material

Recommended chemicals to use during the experiment

Note the security precautions mentioned in the safety sheets of the various solvents

-Water for chromatography (LC-MS) LiChrosolv® (Merck Chemicals, prod no. 115333)

-Methanol hypergrade for LC-MS LiChrosolv® (Merck Chemicals; prod no. 106035)

-Formic acid, for mass spectrometry (Honeywell Fluka™; prod no. 94318)

-Isopropanol hypergrade für LC-MS LiChrosolv® (Merck chemicals; prod no. 102781)

List of the samples and sample ID

- Sample A: MeOH,
- Sample B: 5 ppm Arginine solution in 70%MeOH/30% H₂O

- Sample C-1: Sigma plasma at 1/200 dilution, sample C-2: plasma at 1/100 dilution, sample C-3: plasma at 1/75 dilution, sample C-4: plasma at 1/50 dilution, sample C-5: plasma at 1/25 dilution.
- Sample D-1: 1/50 diluted plasma spiked with 1 ppb pesticides mix, sample D-2: 1/50 diluted plasma spiked with 5 ppb pesticides mix, sample D-3: 1/50 diluted plasma spiked with 10 ppb pesticides mix, sample D-4: 1/50 diluted plasma spiked with 15 ppb pesticides mix, sample D-5: 1/50 diluted plasma spiked with 20 ppb pesticides mix.
- Sample F: NIST SRM 1950 plasma at 1/50 dilution

b) Analysis sequence

1. **After each injection (within each block)**: clean the syringe, peek tube and ESI needle using weak washing (see section 2.c).

2. **After each analysis block (inter-blocks)** (C1-C5, D1-D5 and F when the sequence from table 2 is completed): rinse the ESI-system with strong washing (see section 2.c).

3. If the analyses in table 2 cannot be completed in one day, please clean the syringe, peek tube and ESI needle.

At the end of the day: apply strong washing (see section 3.c).

At the beginning of the next day: apply weak washing, calibrate with sample B and apply weak washing.

The acquisitions should follow the order described in **Table 2**

Table 2: Analysis sequence of the study

Participant ID	Sample ID	Replicate number	Polarity	Scans number (NS)
x	A	1	+	100
x	A	1	-	100
x	B (calibration)	1	+	20
x	A	1	+	100
x	A	2	+	100
x	A	3	+	100
x	C-1	1	+	100
x	C-1	2	+	100
x	C-1	3	+	100
x	C-2	1	+	100
x	C-2	2	+	100
x	C-2	3	+	100
x	C-3	1	+	100
x	C-3	2	+	100
x	C-3	3	+	100
x	C-4	1	+	100
x	C-4	2	+	100
x	C-4	3	+	100
x	C-5	1	+	100
x	C-5	2	+	100

x	C-5	3	+	100
x	A	1	+	100
x	D-1	1	+	100
x	D-1	2	+	100
x	D-1	3	+	100
x	D-2	1	+	100
x	D-2	2	+	100
x	D-2	3	+	100
x	D-3	1	+	100
x	D-3	2	+	100
x	D-3	3	+	100
x	D-4	1	+	100
x	D-4	2	+	100
x	D-4	3	+	100
x	D-5	1	+	100
x	D-5	2	+	100
x	D-5	3	+	100
x	A	1	+	100
x	F	1	+	50
x	F	2	+	50
x	F	3	+	50
x	F	1	+	100
x	F	2	+	100
x	F	3	+	100
x	F	1	+	300
x	F	2	+	300
x	F	3	+	300
x	A	1	+	100

Table 3: Ring Trial KIT

Sample ID	Unit Number	Volume (per unit)	Total Volume
A	2	1 ml	2 ml
B	1	1 ml	1 ml
C-1	1	0.5 ml	0.5 ml
C-2	1	0.5 ml	0.5 ml
C-3	1	0.5 ml	0.5 ml
C-4	2	1 ml	2 ml
C-5	1	0.5 ml	0.5 ml
D-1	1	0.5 ml	0.5 ml
D-2	1	0.5 ml	0.5 ml
D-3	1	0.5 ml	0.5 ml
D-4	1	0.5 ml	0.5 ml
D-5	1	0.5 ml	0.5 ml
F	1	0.5 ml	0.5 ml

SI-11: REFERENCES

- (1) Chen, J.; Zhang, P.; Lv, M.; Guo, H.; Huang, Y.; Zhang, Z.; Xu, F. Influences of Normalization Method on Biomarker Discovery in Gas Chromatography-Mass Spectrometry-Based Untargeted Metabolomics: What Should Be Considered? *Anal. Chem.* **2017**, 89 (10), 5342–5348.
- (2) Yang, Q.; Hong, J.; Li, Y.; Xue, W.; Li, S.; Yang, H.; Zhu, F. A Novel Bioinformatics Approach to Identify the Consistently Well-Performing Normalization Strategy for Current Metabolomic Studies. *Briefings Bioinf.* **2019**, 21 (6), 2142-2152.
- (3) Fan, S.; Kind, T.; Cajka, T.; Hazen, S. L.; Tang, W. H. W.; Kaddurah-Daouk, R.; Irvin, M. R.; Arnett, D. K.; Barupal, D. K.; Fiehn, O. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal. Chem.* **2019**, 91 (5), 3590–3596.
- (4) Kirwan, J. A.; Weber, R. J. M.; Broadhurst, D. I.; Viant, M. R. Direct Infusion Mass Spectrometry Metabolomics Dataset: A Benchmark for Data Processing and Quality Control. *Sci. Data* **2014**, 1, 140012.
- (5) Chekmeneva, E.; Dos Santos Correia, G.; Gómez-Romero, M.; Stamler, J.; Chan, Q.; Elliott, P.; Nicholson, J. K.; Holmes, E. Ultra-Performance Liquid Chromatography-High-Resolution Mass Spectrometry and Direct Infusion-High-Resolution Mass Spectrometry for Combined Exploratory and Targeted Metabolic Profiling of Human Urine. *J. Proteome Res.* **2018**, 17 (10), 3492–3502.