Harmonized quality assurance/quality control provisions to assess completeness and robustness of MS1 data preprocessing for LC-HRMS-based suspect screening and non-targeted analysis

Sarah Lennon[1], Jade Chaker[1], Elliott J. Price[2], Juliane Hollender[3], Carolin Huber[4], Tobias Schulze[5], Lutz Ahrens [6], Frederic Béen[7,8], Nicolas Creusot[9], Laurent Debrauwer[10], Gaud Dervilly[11], Catherine Gabriel[12,13], Thierry Guérin[14], Baninia Habchi[15], Emilien L Jamin[10], Jana Klánová[2], Tina Kosjek[16], Bruno Le Bizec[11], Jeroen Meijer[7], Hans Mol[17], Rosalie Nijssen[17], Herbert Oberacher[18], Nafsika Papaioannou[12,13], Julien Parinet[19], Dimosthenis Sarigiannis[12,13], Michael A. Stravs[3], Žiga Tkalec[2,16], Emma L. Schymanski[20], Marja Lamoree[7], Jean-Philippe Antignac[11], Arthur David[1]*

[1] Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-35000 Rennes, France.

[2] RECETOX, Faculty of Science, Masaryk University, Kotlarska 2, Brno, Czech Republic.

[3] Swiss Federal Institute of Aquatic Science and Technology, Eawag, 8600 Dübendorf, Switzerland & Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zürich, Switzerland.

[4] Helmholtz Center for Environmental Research – UFZ, Department of Exposure Science,, Permoser Straße 15, 04317 Leipzig, Germany.

[5] German Environment Agency (UBA), Colditzstr. 34, D-12099 Berlin, Germany.

[6] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), Box 7050, 75007, Uppsala, Sweden.

[7] Chemistry for Environment & Health, Amsterdam Institute for Life and Environment (A-LIFE), Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081HV Amsterdam, The Netherlands.

[8] KWR Water Research Institute, Groningenhaven 7, 3430BB Nieuwegein, The Netherlands.

[9] INRAE, French national research institute for agriculture, food & environment. UR1454 EABX, Bordeaux Metabolome, MetaboHub, Gazinet Cestas, France.

[10] Toxalim (Research Centre in Food Toxicology), INRAE UMR 1331, ENVT, INP-Purpan, Paul Sabatier University (UPS), Toulouse, France.

[11] Oniris, INRAE, LABERCA, Nantes, France.

[12] Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki

[13] HERACLES Research Center on the Exposome and Health, Center for Interdisciplinary Research and Innovation, Greece

[14] ANSES, Strategy and Programmes Department, F-94701 Maisons-Alfort, France.

[15] INRS, Département Toxicologie et Biométrologie Laboratoire Biométrologie 1, rue du Morvan - CS 60027 - 54519 Vandoeuvre Cedex.

[16] Jozef Stefan Institute, Department of Environmental Sciences, Ljubljana, Slovenia.

[17] Wageningen Food Safety Research (WFSR), Part of Wageningen University & Research, Wageningen, the Netherlands.

[18] Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Innsbruck, A-6020, Austria.

[19] ANSES, Laboratory for Food Safety, F-94701 Maisons-Alfort, France.

[20] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg.

*To whom correspondence should be addressed:

Tel: +33 299022885

email: arthur.david@ehesp.fr

Abstract

Non-targeted and suspect screening analysis using liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) holds great promise to comprehensively characterize complex chemical mixtures. Data preprocessing is a crucial part of the process, however, some limitations are observed: (i) peak-picking and feature extraction might be incomplete, especially for low abundant compounds, and (ii) limited reproducibility has been observed between laboratories and software for detected features and their relative quantification. We first conducted a critical review of existing solutions that could improve the reproducibility of preprocessing for LC-HRMS. Solutions include providing repositories and reporting guidelines, open and modular processing workflows, public benchmark datasets, tools to optimize the data preprocessing and to filter out false positive detections. We then propose harmonized quality assurance/quality control guidelines that would allow to assess the sensitivity of feature detection, reproducibility, integration accuracy, precision, accuracy, and consistency of data preprocessing for human biomonitoring, food and environmental communities.

Keywords

High-resolution mass spectrometry, exposomics, metabolomics, non-targeted analysis, suspect screening analysis, data preprocessing, contaminants of emerging concern, chemical exposome, harmonised QA/QC.

Abbreviations

NIEHS: National Institute of Environmental Health Sciences

SSA: suspect screening analysis

NTA: non-targeted analysis

GC-HRMS: gas chromatography coupled to high-resolution mass spectrometry

LC-HRMS: liquid chromatography coupled to high-resolution mass spectrometry

CEC: contaminants of emerging concern

QA: quality assurance

QC: quality control

DoE: design of experiment

CV: coefficient of variation

ROI: region of interest

CNN: convolutional neural network

m/z: mass - to -charge ratio

1. Introduction

Chemical pollution linked to novel entities is one of the nine planetary boundaries and is known to affect ecosystems and human health. According to a recent inventory, there are more than 350,000 chemicals registered for production and use, with 120,000 of them having substantial gaps in the chemical identity information[1]. Consequently, even though large and historical EU and US initiatives have been implemented to help map human and environmental exposures to chemicals, such as the European Human Biomonitoring Initiative-HBM4EU[2] or the National Institute of Environmental and Health Sciences initiatives -NIEHS, the number of substances (based on priority lists) for which human biomonitoring or toxicological data are reported in the literature remains limited. Moreover, for most of these chemicals, the knowledge of their environmental fate and exposure of animals and humans through food and environment are not well characterized. Toxicity data, especially below acute toxicity levels are also lacking, preventing an efficient risk assessment[3].

The potential association between chemical exposure and adverse effects on environmental and human/animal health is difficult to study because of the lack of knowledge on chemical exposures, which can be partly explained by the limitations of current monitoring methods. The conventional approach in monitoring methods is based on targeted quantitative measurements of selected contaminant/matrix combinations, using internal standard corrections and calibration curves. These methods are robust, accurate, precise, sensitive and reliable and will provide concentrations for the contaminants of interest. However they do not offer a comprehensive overview of the exposure, as they are limited to a subset of chemicals, often from the same chemical class[4]. Conversely, SSA (suspect screening analysis) and NTA (non-targeted analysis) using gas or liquid chromatography coupled to high resolution mass spectrometry (GC-HRMS or LC-HRMS) offer great promise to characterize the global exposure and identify chemicals of emerging concern (CECs) [5,6]. SSA/NTA studies are qualitative and aim at determining contaminant detection frequency in a population, and/or at quantifying these contaminants in a relative way to compare different populations and/or at following the detection and

relative quantification of particular compounds over time. This review will focus primarily on SSA/NTA studies using LC-HRMS.

SSA/NTA workflows typically include study design, sampling and sample preparation (extraction and concentration of the compounds of interest) followed by separation via LC, HRMS analysis, and finally data preprocessing followed by identification steps. The data preprocessing step aims at obtaining a list of detected signals (features) characterized by several pieces of information (e.g., at least by their *m/z*, intensity and/or area, and retention time). Depending on the samples investigated, thousands to tens thousands of features can be detected in a single analysis. They can then be aligned and grouped across batches and analyses. After the preprocessing step, in SSA, features of interest are annotated using a list of expected ("suspected") substances, while prioritization strategies (e.g., multivariate analysis) followed by identification steps are commonly used for NTA[7,8]. Although promising, the development and implementation of workflows for SSA/NTA are still affected by several analytical and informatics challenges. The large diversity in physicochemical properties hampers the use of only one analytical set-up to detect all the compounds of interest, whereas the wide dynamic range of concentrations in the sample prevents the detection of low abundant contaminants due to analytical interferences[5]. Furthermore, there are currently no universal solutions available to comprehensively preprocess the data generated with SSA/NTA. Finally, the annotation step is extremely time-consuming, and often remains incomplete[5]. This is in part due to the lack of standard compounds, which impacts the LC-HRMS libraries information available on molecules (MS/MS, retention time, logD) and consequently undermines the level of confidence in the annotation. Additionally, xenobiotics are usually detected at low level, and it can be difficult to acquire MS/MS data for those compounds, decreasing the body of proofs available for annotation.

Regarding the data preprocessing step, feature integration is dependent on the quality of feature detection, meaning unoptimized feature detection can lead to false positives (type I error, or noise being reported as a real feature) and/or more concerning false negatives (type II errors, or real peak

being missed) which can then compromise the exposure assessment[9]. Overall, the main limitations observed during the preprocessing step include the fact that: (i) peak-picking and features extraction might be incomplete, especially for low abundant compounds detection[10,11]. In that case, it is often difficult to distinguish actual signals from noise in complex samples with variable noisy backgrounds, especially if the chromatographic peak does not have a Gaussian shape[12]. Moreover, default data preprocessing parameter settings, often optimized for metabolomic application, can lead to significant false positive or false negative rates for exogenous chemicals present at trace levels [10,11]. (ii) reproducibility issues have been observed between laboratories and software for detected features and their relative quantification[12].

To harmonize the processes across laboratories and ensure that SSA/NTA can provide a list of confidently detected and integrated features, standardized data preprocessing quality assurance/quality control procedures (QA/QC) similar to the ones used for validation and monitoring of analytical methods for target screening are missing. QA aims to define all the activities and processes to ensure that all quality requirements will be fulfilled. QC describes the individual measures used to detect non-conformities regarding method performance[13]. We suggest that these QA/QC procedures could be applied in SSA and NTA to validate the efficiency, completeness and reproducibility of data preprocessing methods.

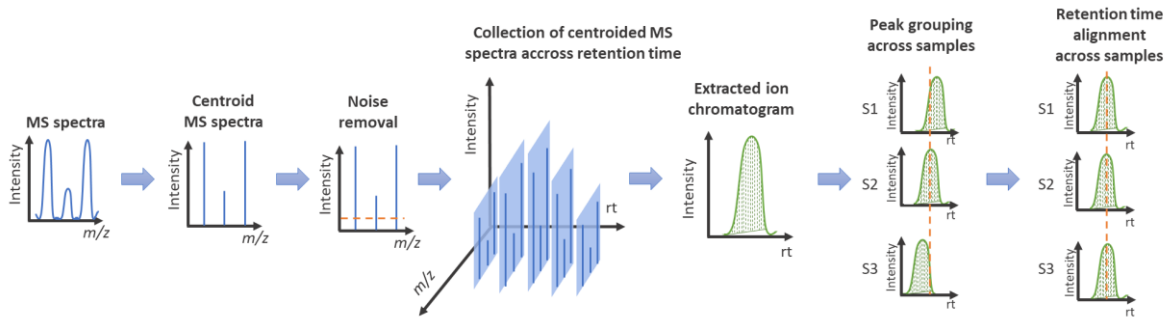
To address the current limitations related to data preprocessing, we first performed a literature review of existing solutions that aim to improve the reproducibility of data preprocessing and accurate detection of all true peaks in LC-HRMS data. Then, within the European Partnership for the Assessment of Risks from Chemicals (PARC), we propose harmonized QA/QC procedures for data preprocessing relevant for human biomonitoring (HBM), food and environment communities to ensure robust and reproducible detection of CECs. In this review, we focus on the data preprocessing step of SSA/NTS workflows using LC-HRMS, while other separation methods, e.g., gas chromatography was out of the scope of this study. Aspects linked to analytical reproducibility (sample preparation, correction across

batches) are already established[13–16] and will not be discussed in this paper. Furthermore, normalization, that aims to eliminate unwanted experimental and biological variation, might bring additional variabilities to the data. This step, further discussed in Misra et al.[17] and Cuevas-Delgado et al.[18], was not considered part of the data preprocessing.

2. Raw data pre-processing steps and challenges

Multiple software and tools have been developed for preprocessing LC-HRMS data. The most used open-source data preprocessing software include XCMS[19], MS-DIAL[20], MZmine[21], OpenMS[22], …, while vendor software includes Compound Discoverer (ThermoFisher), MarkerView (ABSciex), MassHunter Profinder (Agilent), Metaboscape (Bruker) and Progenesis QI (Waters). A comprehensive overview of available data preprocessing software can be found in reviews from Misra et al.[23], Renner et al.[12], Spicer et al.[24], Stanstrup et al.[25] and Hollender et al.[6]. Although the detailed algorithms are different, the peak-picking or feature extraction strategy is generally based on the same principles: raw data are first centroided and noise is removed with a simple constant threshold, an adjustable region of interest (ROI)[26], or a more variable and complex intensity threshold. Then, extracted ion chromatograms (EIC) are generated and a peak-picking algorithm is used to identify features[27]. Features are grouped across the measurement sequence and retention time alignment is performed (**Figure 1**). At this stage, gap filling can be performed to recover peaks that were not integrated in all analyses to minimize the number of missing values. Gap filling is discussed further in Müller et al. [28] and Armitage et al[29].

198



199

200 *Figure 1 : Data preprocessing steps for one feature of interest. Raw mass spectrometry profile pattern*
201 *is first centroided and noise is removed. For the same feature, a collection of centroided MS spectra*
202 *across retention time is obtained. Extracted ion chromatograms are generated. Chromatograms are*
203 *grouped across the measurement sequence and retention time alignment is performed.*

204 Parameters for centroiding (smoothing algorithms), peak-picking (*m/z* error, estimated
205 chromatographic peak width, signal thresholds), retention time correction (alignment gap penalties)
206 and grouping algorithms (*m/z*, retention time deviation and minimum number of detections) are
207 critical. Multiple studies, particularly in the field of metabolomics, have shown that using different
208 parameters for data preprocessing can lead to three major issues[27]: (i) lack of reproducibility and
209 substantial differences in the list of all detected and integrated features[27], (ii) suboptimal detection of
210 low abundant features, even those with a Gaussian chromatographic peak[10,30] and (iii) reporting of
211 some features as multiple artifactual features (peak splitting) or merging of two features into one
212 because of poor peak shape linked to low abundance or chemical properties[31] with algorithms
213 struggling to locate the local intensity minima[32]. Examples illustrating common peak-picking issues are
214 shown in **Figure 2**.

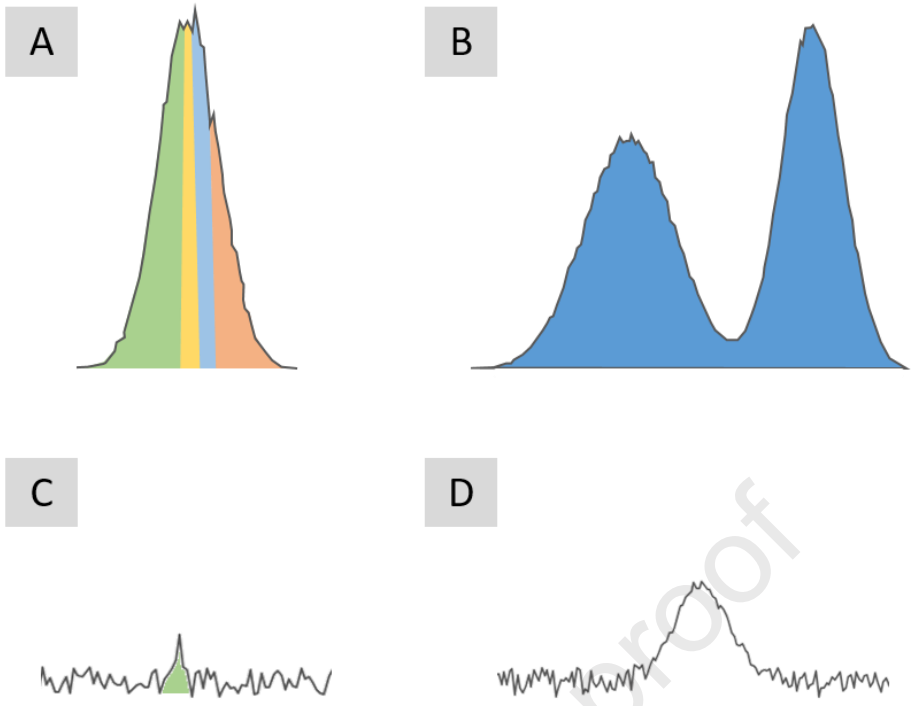


Figure 2 : Example of common peak-picking errors: (A) Artifactual splitting of a peak into multiple features, (B) merging of two peaks into one feature, (C) integration of noise, and (D) missing peak. The first two issues are generally related to selecting an inadequately low (A) or high (B) peak width value during the preprocessing step, whereas the last two issues are generally attributable to selecting an inappropriately low (C) or high (D) noise threshold.

Since data may be acquired in either centroid or profile mode, centroiding is generally only necessary in the data preprocessing workflow for the latter case. Additionally, centroiding may be performed after data acquisition on-the-fly by unpublished vendor algorithms with no accessible parameters. To the best of our knowledge, very few studies evaluating the impact of centroiding on data preprocessing have been reported[33].

Multiple studies have highlighted significant differences in feature detection, with as low as 10% overlapping features[27,34] and up to three times more detected features depending on the preprocessing software used[32,35–37]. A recent study from Guo et al. demonstrates variability between five different preprocessing software regarding the true positive rate (number of true positive features detected related to the total number of true positive features)[38]. It is important to acknowledge the

231 complexity of comparing detected features across different software, given that each preprocessing
232 software employs distinct algorithms that may not be implemented in the same sequence. Step by
233 step comparisons are consequently difficult to interpret. Hohrenk et al.[27] demonstrated that this
234 phenomenon is not only necessarily related to low abundant features, as they also observe only ~10%
235 overlap between the MZmine, XCMS Online and enviMass preprocessing of wastewater treated
236 samples for the top 100 most abundant features. Integration is also affected, and Coble et al. have
237 noted an absolute bias of up to 22% compared to manual integration with the vendor software[39].

238 Variability was also observed in the detection of spiked or standard compounds, with recall rate of
239 suspect ranging from 64 to 88%[27]. Li et al. have noted in their measurements and data evaluation of a
240 mixture of 1,100 compounds that the recall rate ranged from 85 to 95%, but the relative error in
241 integration ranged from 64 to 99%[36]. El Abiead et al. have shown that a minimal change in the XCMS
242 centWave maximum peak width parameters led to an increase in the proportion of missed spiked
243 compounds from 6 to 93%[11]. This phenomenon was also noticed by Chaker et al., who observed that
244 the lack of optimization of data preprocessing software such as XCMS can lead to a false negative rate
245 of up to 80% for chemicals spiked at low levels in blood[10].

246 Differences have also been observed in statistically significant potential biomarkers. For instance,
247 Baran reprocessed five untargeted metabolomics datasets from public repositories, and although the
248 study was not aimed to be exhaustive, the author could detect 50 biologically relevant omissions in
249 each dataset[40]. Chen et al. compared three preprocessing software and showed that altogether 14
250 markers were reported as statistically different, but only two were detected by all software[37]. Another
251 study independently performed on the same cancer proteomics dataset reported 17 biomarkers,
252 where only two were shared between the software approaches[41]. Li et al. and Horenk et al. also
253 mentioned the difficulty in matching detected features across samples and/or different processing
254 software due to failure in *m/z* and retention alignment[27,42].

To summarize, it appears difficult to be comprehensive in terms of feature detection. Even with carefully optimized parameters, some compounds that provide reliable signals (including the isotope profiles) will be missed (i.e. false negatives) by the algorithms[10]. For a number of software tools, extensively decreasing the thresholds in an attempt to increase the number of low abundant ions detected will increase the number of reported false positive features[43]. It could lead to excessively long preprocessing times (days to weeks), especiallyfor large scale application (> 1000 samples), where users will be technically limited by their computers (amount of RAM, hard disk space, numbers of CPU cores), their cloud based solutions (disk quotas) or the programming of the software, i.e., possibility of task parallelism [44]. Thus, it is necessary to (i) ensure that data preprocessing is well adapted to the scientific question and (ii) minimize discrepancies between data processing tools (i.e. via robust intercomparability using similar datasets or the same dataset processed with different tools or with different parameters within the same tool). Moreover, beyond data preprocessing using computational algorithms, differences are also observed in features classification performed by mass spectrometry experts (true peak or false peak issued from background contamination or electronic noise)[45]. It is therefore important to define QA/QC criteria that ensure that the data preprocessing step will provide the most accurate and reproducible results possible.

3. Initiatives for reproducible data preprocessing

Considering all the different possible analytical set-up and data preprocessing tools, it seems extremely difficult to propose a harmonized procedure and parametrization for data preprocessing[8]. However, to minimize computational irreproducibility between data processing pipelines and maximize the detection of real peaks, multiple initiatives are proposed: guidelines for the reporting of the data preprocessing parameters, online repositories to provide access to the data, reproducible computational workflows and provision of benchmarking datasets.

3.1. Guidelines for data preprocessing reporting

279 The first guidelines for NTA data reporting were published in 2007 by the metabolomics community[46].
280 However, Considine et al. reviewed 17 studies published between 2008 and 2014 and concluded that
281 the guidelines were not followed, as the description of the data processing parameters was too vague,
282 making it impossible or very difficult to replicate the data preprocessing workflow[47]. In 2019, in
283 collaboration with the mQACC consortium, the MERIT project detailed best practice guidelines,
284 method performance standards, and minimal reporting standards for the acquisition, processing and
285 statistical analysis to encourage usage of metabolomics analysis in the regulatory toxicology context[48].
286 The 2023 OECD guidelines (number 390) were published with the aim to provide a clear and consistent
287 framework for reporting each element of an omics study intended for use in regulatory toxicology,
288 from study design through to data analysis. However, the OECD guidelines only define the workflow
289 parameters/steps that need to be described. There is no mention of QA/QC for data preprocessing. In
290 2022, the mQACC consortium[13] published a paper with the aim to encourage the reporting of QA/QC
291 procedures (i.e., description of the criteria used to define acceptable performances and data used to
292 demonstrate, that the results are indeed acceptable). A framework is provided for consistent reporting
293 of QA/QC sample information and quality metrics. These guidelines were designed for metabolomics
294 studies and are not detailed enough for the regulatory context. There are no defined metrics, and the
295 provided template is organized following the type of QA/QC rather than checked metrics. In parallel,
296 the NTA Study Reporting Tool was developed by the Benchmarking and Publications for Non-Targeted-
297 Analysis (BP4NTA) working group[49]. This tool aims to help reviewers to evaluate work submitted for
298 publication by providing a score to assess the quality of NTA study reporting. More recently, the
299 Norman study groups has also published guidance for reporting of SSA/NTA data preprocessing
300 parameters[6].

301 These five documents aim at ensuring that all critical elements of a study are reported. In particular
302 for data preprocessing, the software and its source, and the peak-picking parameters (*m/z* tolerances,
303 intensity thresholds, signal-to-noise ratio, noise filtering settings) are required. MERIT,

OECD,mQACC,BP4NTA and the Norman study group provide guidelines in terms of QA/QC reporting. Nonetheless, they do not explicitly cover QA/QC metrics for data preprocessing.

3.2. Data repositories

Like guidelines, data repositories aim to ensure data reproducibility and re-use. Repositories such as the Metabolomics Workbench[50], MetaboLights[51] and GNPS integrated within MassIVE[52] aim to standardize data submission and disseminate public MS data, ensuring data reproducibility and re-use. However, in contrast to the proteomics field[53], where metadata for more than 30,000 datasets are accessible on ProteomeXchange[54], only 3998 datasets were available on MetabolomeXchange[55]. In 2019, the NORMAN Association established the partially public Digital Sample Freezing Platform[56] to provide the first repository tailored for environmental mass spectral data. It currently contains 60 public datasets[57]. The discrepancies in the availability of public datasets in the different domains might be explained by the challenging and time-consuming process associated with publication of a small molecule dataset. In addition, the divergent commitments of the communities on standardization and reproducibility of research and open science are a strong push factor for the development, operation and use of common repositories.

3.3. Processing Workflow

Open source processing workflows, allowing the data processing from preprocessing to statistical analysis and data annotation, have been developed to increase reproducibility and reduce the influence of manual intervention on the final results[58,59]. Modular workflows, where new tools can be implemented as modules, facilitate usage by the analyst, increases reproducibility and favors data sharing[41].

Platforms gathering all the tools necessary for data processing have been implemented for Metabolomics. Examples include Workflow4Metabolomics[60], MetaboAnalyst[61], MZmine[21], the metaRbolomics Toolbox[25] and RforMassSpectrometry[62]. For environmental studies, patRoon[63] was released for comprehensive NTA data processing of environmental samples. Although having these

different approaches is a great way forward, not all software tools are compatible with the same platform and choices have to be made. Software interoperability should be improved (e.g., modularization), where possible, to widen user access to different approaches.

3.4. Benchmark datasets

Benchmark datasets are useful to evaluate the efficiency of data preprocessing and quality of peak-picking[64,65]. Benchmark datasets can also be used to compare algorithms and better understand the key parameters[66,67]. Few benchmark datasets have been published to date for exposomic, food and environmental sciences. One example is the dataset published by Schulze et al. comprising 4 water samples analyzed by 21 laboratories on a wide range of instruments and with different analysis conditions (column, gradients, acquisition mode…)[68]. Another dataset is a collection of 255,000 extracted ion chromatograms, manually classified as being a peak or not, to improve, for example, peak picking or gap filling algorithms[45]. Existing open data repositories can also be a source of benchmark datasets. For instance, the data preprocessing evaluation tool mzRAPP was assessed using datasets downloaded from MetaboLights[11]. Although very useful to develop, improve and evaluate data preprocessing algorithms, benchmark datasets are not necessarily representative of the nature of the specific study data, so data preprocessing parameters cannot be optimized.

4. Existing tools to minimize true and false negative peak-picking results

To detect a maximum of true features without introducing too much noise, two types of strategies have been investigated to date: optimization of the data preprocessing parameters and filtering of the data after preprocessing. Examples of tools that can be used to finely tune the data preprocessing parameters and minimize true and false negatives are listed in **Table 1**. In parallel, preprocessing software using alternative methods for peak-picking have been explored.

352 *Table 1 - List of tools that can be used to finely tune the LC-HRMS data preprocessing parameters and to minimize true and false negatives peak picking. N/A*
353 *= not applicable, DoE = design of experiment, QC = quality control, ROI = region of interest, %CV = coefficient of variation, CNN = convolutional neural*
354 *network.*

Name	Authors	Method	Criteria	Comments	Data preprocessing software compatibility
Tools to optimize data preprocessing					
N/A	Eliasson et al.[69]	Iterative DoE based on dilution of pooled QC samples	Reliability index metrics: evaluate repeatability of peaks using correlation between diluted compounds and integrated peak area		Any
N/A	Zheng et al.[70]	Iterative DoE based on dilution of pooled QC samples	Reliability index metrics: evaluate repeatability of peaks using correlation between diluted compounds and integrated peak area	Use a Plackett Burman design for screening and central composite design for optimization	Any
N/A	Kiefer et al.[71] Chaker et al.[10] Dom et al.[72] Hu et al.[73]	Iterative DoE based on spiked compounds	Settings are optimized until a defined percentage of target spiked compounds are detected.	Low abundant isotopes of internal standards can be used to cover low abundant peaks	Any
N/A	Manier et al.[74]	Iterative DoE	Coefficient of variation (%CV) on replicates measurements of samples		Any
XMSanalyzer	Uppal et al.[75]	Merge features from best sets of data preprocessing parameters	Coefficient of variation (%CV) on replicates measures of samples. Features merge of multiple data preprocessing results	For redundant features, best results (highest %CV) are kept	apLCMS[26] and XCMS[19]
FFRGD	Ju et al.[76]	Merge features from best sets of data preprocessing parameters	Fuses features and removes redundancy based on graph density	A graph is defined to cover the features generated from different software, in which nodes and edges represent the features and their similarity relationships	XCMS[19], Sieve (ThermoFisher), MZmine[21]
N/A	Brodsky et al.[77]	DoE	Z-transformed Pearson correlation coefficient between intensity profiles of sample replicates		Any
IPO	Libiseller et al.[78]	Iterative DoE	Peak-picking score based on reliability of a peak. Retention time correction score depending on deviation to the mean of all peaks after correction. Grouping score based on classification of peaks as reliable or not	Reliable peak belongs to an isotopologue (13C)	centWave XCMS[19]

MetaboAnalystR	Pang et al.[61]	Iterative DoE on ROI (region of interest) of raw data enriched for real peaks	Quality score based on the 3 scores of IPO taking into account peaks with low-abundant isotopes, Gaussian shape of the peaks and coefficient of variation between areas of the same compounds	Reliable peak belongs to an isotopologue (13C)	MetaboAnalystR[61]
SLAW	Delabriere et al.[79]	Iterative DoE based on ROI (region of interest) of raw data containing the most abundant features	S_{iso} = similar to IPO peak-picking score S_{integ} = based on detection in other sample and %CV S_{align} = retention time deviation across samples	Use surface models to select the best parameters	OpenMS[22], MZmine[21], XCMS[19]
mzRAPP	El Abiead et al.[11]	DoE	Completeness and accuracy of integration evaluated from a benchmark dataset of compounds for which all peaks have been manually integrated	High-quality of benchmark dataset ensured by comparing manually integrated isotopologue ratios to theoretical ones	Any
Autotuner	McLean et al.[80]	Direct determination of best parameters in a single step, using raw data	Parameters are derived from shape of chromatograms	Take a sample of peaks from data using slicing windows	centWave XCMS[19] and Mzmine[21]
Paramounter	Guo et al.[81]	Direct determination of best parameters in ROIs	Define universal parameters based on raw data (mass tolerance, peak heights, peak width, instrument shift)		Any, but automated conversion of parameters only for XCMS[19], MS-Dial[20], Mzmine[21]
EVA	Guo et al.[82]	CNN	Recognition of false positive metabolic features with poor EIC peak shape	Training on 25 000 manually recognized EIC peaks and output true or false values.	XCMS[19], MS-Dial[20], OpenMS[22], MZmine[21]
False positive peak filtering					
N/A	Want et al.[83]	%CV across QCs	Filter out features with %CV <30%		Any
N/A	Schiffman et al.[84]	Adaptative filtering	Filters based on blank samples, % of missing values, ICC (inter class correlation coefficient)	Determine the filtering thresholds and evaluate the effectiveness of the filtering based on the training set (900 features evaluated as high or low quality)	Any
rFPF	Ju et al.[85]	Entropy index and %CV across QCs	1. Peaks must be reproducible in 80% of the samples. 2. An entropy index is used to recognize real peaks. %CV on the rest should be <30%		Sieve (Thermofisher) and XCMS[19]

MS-CleanR	Fraisier-Vannier et al.[86]	Adaptative filtering	Filtering based on blank samples, unusual and relative Mass Defect, relative standard deviation among sample class	Filters are user tunable	MS-DIAL[20]
CPC	Pirtilla et al.[87]	Comprehensive Peak Characterization after extraction in raw data from XCMS tables	Determine peak area, signal to noise ratio, FWHM, width at base and 5,10%	User based filters settings on the peak parameters	XCMS[19]
NeatMS	Gloaguen et al.[88]	Deep learning-based peak filter tool (CNN)	Classify peaks in 3 quality peak classes: high, acceptable, poor quality/noise	Requires a training set which can be defined by the user	Any
N/A	Kantz et al.[89]	Deep neural network	Classify peaks as true or false signals	Training sets contain 1 304 manually classified LC peaks	MZmine[21]
N/A	Kantz et al.[89]	Multiple logistic regression model	Classify peaks as true or false signals using 6 peak shape attributes associated in 59 peak group factors	Distinguish true from false signals	MZmine[21]
MetaClean	Chetnik et al.[90]	Combination of Machine learning (AdaBoost algorithm) and 22 peak quality metrics	Classify peaks as pass or fail	Performed after initial filtering based on %CV (<30%).	XCMS[19]
EVA	Guo et al.[82]	CNN	Classify peaks as true or false.	Model was trained on 25,000 manually recognized EIC peaks	XCMS[19], MS-Dial[20], OpenMS[22], MZmine[21]

4.1. Tools to optimize data preprocessing

Although the algorithms vary, the most important parameters for peak-picking are *m/z* errors on different mass spectra of the same feature, chromatographic peak width (for instance, full width at half maximum (FWHM), minimum/maximum peak width) and signal thresholds[32]. For retention time, alignment gap penalties applied to the alignment score have to be defined. The gap penalty allows evaluating the deviation from the diagonal of the similarity matrices. Finally, maximum *m/z* and retention time deviations, and the minimum number of samples in which a peak should be detected, must be established[81]. Altogether, about 10 to 15 parameters must be defined, making the data processing cumbersome for less experienced users.

Numerous tools are available to help with the selection of the best parameters and to easily optimize the data preprocessing step. These have primarily been developed for high throughput metabolomics applications, where reliable detection of the most abundant high-quality peaks is favored. Most of these tools apply a Design of Experiment (DoE) approach, where one or multiple outputs reflecting the quality of peak-picking are measured and parameters are adjusted depending on the results. Eliasson et al.[69] first introduced the concept for metabolomics data preprocessing using diluted pooled urine samples. They proposed to measure the correlation between diluted compounds and integrated peak area, assuming peak linearity. This method was improved by Zhang and al., who developed a Plackett Burman design for fast parameters screening and a central composite design for optimization. This reduces the time needed to determine the best parameter values[70]. Others suggested monitoring the coefficient of variation (%CV) on ten replicates to reflect data variability, assuming that an improved peak integration and lower missing rate correlates with a lower %CV[74,75]. The optimization of settings until a defined percentage model of target spiked compounds are detected is quite common in the environmental field[10,71–73].

XMSanalyzer and FFRGD go further by merging the results of different software. For redundant features, the best results are kept[75,76]. Brodsky et al. determine the average Pearson correlation

coefficient between intensity profiles of sample replicates and apply a Z-transformation to obtain a normal distribution. The algorithm is run multiple times, and the best combination is chosen based on the Zcorr score[77]. IPO uses stable ^{13}C isotopic peaks to calculate a peak-picking score by the ratio of reliable peaks to the total number of peaks minus the number of low abundant peaks. An iterative DoE process is performed until the optimal processing parameters allowing the best peak-picking score are determined[78]. IPO has been shown to work well for abundant features with good LC-MS performance[91]. However, it might provide unrealistic parameter settings for low abundant peak detection or for data with lower LC-MS quality[10,81]. MetaboAnalystR uses a strategy similar to IPO with few modifications: instead of using the full dataset, regions of interest enriched for real peaks are selected. The score includes parameters to consider the Gaussian shape of the peak, as well as the retention time correction score and grouping score[61]. SLAW also selects regions with the most abundant features and uses a score with two terms. The first term is similar to the IPO score, while the second term considers the reproducible integration across QCs files[79]. Finally, with mzRAPP, users have to manually integrate a benchmark dataset of known compounds and manual isotopologue area ratios are compared to experimental ones to ensure high quality of the benchmark dataset. Recovery and accuracy of integration using isotopologues are calculated after preprocessing and are used to evaluate the performance of the data preprocessing procedure[11]. These methods optimize the parameters in an undirected way and are data-driven, rather than relying on parameters derived from analytical chemistry domain experience[81].

Other optimization algorithms directly determine the best parameters using the raw data. Autotuner, for instance, derives parameters by sampling a set of peaks (slicing windows) and by assessing the shapes of the extracted ion chromatogram[80]. Paramounter[81] also defines universal parameters based on raw data (mass tolerance, peak heights, peak width and instrument shift). These universal parameters can then be converted to be used in XCMS, MS-DIAL and MZmine. However, even though based on direct determination of the best parameters, AutoTuner has been shown to be biased towards high quality abundant features[81]. Moreover, if detected, the integration of low abundant

features is not as reproducible as shown by Chaker et al. (fewer than 20% of serum spiked compounds have a CV<20%)[10].

Finally, machine learning algorithms are emerging. For example, EVA uses CNN (convolutional neural network) for peak quality evaluation. The model was trained on 25,000 manually annotated peaks (false and true). This allows the software to recognize false positive metabolic features with poor EIC peak shape[82]. The software is compatible with four different software (XCMS, MS-DIAL, OpenMS and MZmine).

All these optimization algorithms are interesting approaches to choose the best parameters for data preprocessing. However, they also need to be considered with care as some of these optimization strategies have been shown to discard low abundant and rare peaks, which are critical when performing environment, food safety and human biomonitoring analysis[10,11,92].

4.2. False positive peak filtering

After data preprocessing, features can be filtered to remove the maximum number of false positive peaks and only keep the real features. Common strategies to evaluate the quality of a peak and decide for filtering are based on repeatability metrics, blank subtraction, peak metrics, mass defect and machine learning.

Repeatability metrics include %CV[83] on spiked and/or on all detected compounds, interclass correlation coefficient (ICC)[84], entropy index which allows to evaluate noise[85] and percentage of missing values calculated on repeated injections of the same sample, like pooled QCs. For instance, Schiffman et al. manually evaluated 900 features as high or low quality, tested multiple filters and compared the results in terms of high- and low-quality features filtered out[84]. They concluded that a data-adaptive filtering outperforms methods based on non-specific thresholds.

Blank subtraction, included for instance in the tool MS-CleanR[86], will evaluate background ions and feature height ratio in samples vs QC.

Peak metrics are used for instance by the tool CPC, which calculates peak characteristics (peak area, signal-to-noise ratios, FWHM, width at base, 5% and 10%) and filters out features with no characteristic peak signatures in the second derivative[87]. MS-CleanR[86] also incorporates mass defects (unusual and relative mass defect calculation).

Machine learning aims to classify detected peaks as true or false based on a training set of manually classified peaks (binary classification). Image recognition algorithms, including deep learning[88], deep neural network[89] and CNN[82] have been used. Other strategies based on boosting have been suggested. For instance, MetaClean combines machine learning using the AdaBoost algorithm and 22 peak quality metrics[90]. A simpler multiple logistic model, including six peak shape attributes associated with 59 peak group factors, has been shown to provide reasonable results, although it did not perform as well as an image-based deep neural network on the same sample set[89].

4.3. New data preprocessing strategies

New types of algorithms are currently emerging to provide an alternative to the peak-picking approaches described above. For instance, Li et al. developed the algorithm Asari which aligns samples before peak detection using a composite mass track (LC-MS data points with the same consensus *m/z* value spanning the full retention time across all analysis). In commonly used software such as XCMS and MZmine for instance, peaks are aligned after the peak detection, which will cause a small variation of reported *m/z* values in each sample and the algorithms will have to ensure that correct peaks are grouped. By aligning before peak detection, a decrease in computational time and improvement in reproducibility was demonstrated, as there was no need to align elution peaks between samples and mass resolution was the only parameter requiring tuning[42].

The software HERMES foregoes classical peak detection by considering a vast array of possible molecular formulas and adducts, detecting information-rich signals independently of chromatographic peak shape[93]. IDSL.IPA uses the isotopologues $^{12}C/^{13}C$ in a similar way to the optimization tool IPO to define and isolate peaks of organic compounds[94].

Other approaches get rid of the centroiding step and directly work on raw data acquired in profile mode. Examples include machine learning algorithms using pattern recognition such as artificial neural networks (ANN) and deep neural networks to recognize features[95,96] or CNN to define peak integration and product separation region (peakOnly, PeakBot)[97,98]. These approaches, however, depend on the quality of the training set. The SAFD algorithm also works directly on profile raw data. A three-dimensional Gaussian distribution is fitted onto the profile data. This allows to consider all the measured points within one feature at the expense of computational time and difficulties in integrating irregular peak shapes[99]. Another approach uses a Bayesian probabilistic peak detection algorithm that weighs the data according to the probability of being affected by a chromatographic peak or noise[100]. Additionally, retention time alignment is also investigated to allow to correct for non-monotonic shifts. Examples include DeepRTAlign[101], that combines a pseudo warping function and a deep learning-based model and Alignstein[102], that uses a feature matching method.

5. Suggestion of harmonized QA/QC procedures for data preprocessing

5.1. Overview of current QA/QC approaches

QA/QC would complement all the previously described actions and certify that the data preprocessing of SSA/NTA meets some defined quality criteria. This will ensure the best possible detection of all true features and minimize false positives.

QA/QC has successfully been implemented for all analytical and instrumental drifts aspects for SSA and NTA[103]. Multiple papers discussed implementing and adopting common QA/QC practices. Still, to the best of our knowledge, no set of provisions has actually been proposed and defined to assess specifically the performance of data preprocessing[16,58,104]. Knolhoff et al. have experimented with QC practices to test the whole workflow, from sample analysis to data processing, using QC pooled samples spiked at low, medium and high level[105]. Satisfactory results were obtained with identification rates of 70% and a precision ranging from 30 to 50% for all spiked compounds in all QCs.

5.2. Harmonized QA/QC procedures

Here, building on the Knolhoff et al. initiative, we propose a set of QA/QC criteria that could be used to evaluate the quality of data processing of SSA/NTA analysis and more particularly: (i) sensitivity of feature detection, (ii) reproducibility, (iii) integration accuracy, (iv) mass and retention time accuracy (after realignment and calibration), and (v) consistency. All the parameters, criteria and provisions are described in **Table 2**.

At this stage, it is important to mention that the quality of the analytical design (inclusion of blank and quality control samples along the sequence, randomization of the samples in batches), and process (performance, stability, repeatability) needs to be thoroughly checked as it will impact the data quality in general and thus affect the data preprocessing. This is the only way to distinguish issues related to either analytical or data preprocessing errors.

QA/QC for data preprocessing should be evaluated on representative samples, e.g., pooled QC samples spiked with a set of known compounds relevant to the study at two concentration levels (high and low) injected multiple times, one after the other and across multiple batches. These types of QCs and blanks are usually included in large-scale non-targeted studies of human specimens[103], environmental[6] and food samples to monitor analytical performance and consistency of the instrument and thus will not require additional analysis. At this stage, standardized reference materials could also be used to support data preprocessing intercomparison between various studies from different laboratories.

Ideally, the data preprocessing should not include any gap filling or imputation (it will improve the detection frequency), grouping of the degenerate features, i.e., adducts, fragment ions (it will impact the integration results) or normalization of the data (it will affect the integration results of the compounds).

511 *Table 2 – Proposed harmonized QA/QC criteria to evaluate performances of data preprocessing for qualitative and quantitative SSA/NTA analysis. For each*
512 *parameter, criteria, provision, base for thresholds/tolerances, actions to be taken if failed criteria and useful tools are described.*

Parameters	Type of SSA/NTA study	Criteria	Provision	Base for thresholds/tolerance	Actions if failed criteria	Possible Tools
Sensitivity of feature detection	Qualitative Quantitative	False negative detection rate (Spiked compounds)	Compare the number of detected spiked compounds between manual accurate processing and automatized preprocessing using a suspect screening strategy	Proportion of compounds detected in low level spiked QCs, Proportion of compounds detected in high level spiked QCs	Optimize peak-picking parameters	Skyline[106], mzRAPP[11], Scannotation[107]
Reproducibility	Qualitative Quantitative	False negative detection rate (Spiked compounds)	Compare the false negative rate detection across repeated samples	Proportion of compounds detected in low level spiked QCs across samples. Proportion of compounds detected in high level spiked QCs across samples.	Optimize peak-picking parameters	Skyline[106], mzRAPP[11], Scannotation[107]
	Quantitative	Reproducibility of integration across all repeated samples analysis (All features)	Calculate the coefficient of variation on integrated areas for all compounds after data preprocessing	Coefficient of variation values (%CV)	Optimize peak-picking parameters	MetaboanalystR[61]
Integration accuracy	Quantitative	Proximity to curated integration (Spiked compounds)	Compare curated integration of isotope ratios to automatized preprocessing integration	Correlation between curated and automatized preprocessing integration	Optimize peak-picking parameters	mzRAPP[11]
	Quantitative	Relative quantification accuracy (Spiked compounds)	Calculate all the area ratios high vs. low level spiked compounds (Area at level 2 – Area in the procedural blank)/(Area at level 1 – Area in the procedural blank) and apply univariate statistics and plot a volcano plot	Spiked compounds should be highlighted as differential (p-value<0.01 and log2FC>2)	Check the full data preprocessing workflow	MetaboanalystR[61]
Precision/ accuracy	Qualitative Quantitative	Recalibration and time alignment quality (Spiked compounds)	Calculate the standard deviation in mass and retention time	Deviation in *m/z*<5 ppm or less depending on instrument and concentration of the spiked analytes. Relative deviation on retention time within reasonable limits	Check recalibration, grouping and realignment parameters	Scannotation[107]

Consistency	Qualitative Quantitative	Identification with the annotation workflow using 1) a suspect list containing only the standard compounds 2) the complete suspect list (Spiked compounds)	Run the suspect screening workflow with (1) a suspect list containing only the standard compounds and (2) the suspect list that will be used to answer the scientific questions and compare the rate of annotated spiked compounds vs. detected spiked compounds after data preprocessing	Proportion of compounds annotated in low level spiked QCs Proportion of compounds annotated in high level spiked QCs	Check the full data preprocessing workflow	Scannotation[107], patRoon[63], MS-Dial[20]

513

514 5.2.1. Sensitivity of feature detection

515 The sensitivity of feature detection parameter aims to evaluate the rate of false negative and false
516 positive detected compounds. These parameters can be evaluated by monitoring the percentage of
517 recovered spiked compounds compared to manually curated data. Beforehand, it is necessary to check
518 for the absence or at least a much lower detection (e.g., ratio 1:10) of the spiked compounds in
519 procedural blanks to avoid affecting detection frequency. A suspect screening strategy can then be
520 used to compare the number of detected spiked compounds between manual accurate and
521 automatized preprocessing.

522 5.2.2. Reproducibility

523 The reproducibility parameter evaluates the variability linked to data preprocessing of repeated
524 analysis of the same sample (i.e.; repeated injections of the same QC preparation) within a defined
525 time period (one or multiple batches). It aims to compare i) the false negative rate of detection of
526 spiked compounds across repeated samples and ii) the integration of all features across all repeated
527 analysis of the same sample. To evaluate this last point, following the metabolomics guidelines[15], we
528 suggest to keep only the compounds with a detection rate higher than 70% in all quality control
529 samples.

530 5.2.3. Integration accuracy

531 The integration accuracy aims to evaluate (i) the proximity to manual integration results on the set of
532 spiked compounds and (ii) the reproducibility and accuracy of integration on all features across all QC
533 runs. Integration accuracy can be evaluated, as suggested by El Abiead et al.[11], on spiked compounds
534 with the isotopic ratio for low abundant isotopologue (LAIT) and most abundant isotopologue (MAIT)
535 using the third isotopologue for halogenated compounds and the second for all the other compounds.
536 Manually curated integration can then be compared to automatized data preprocessing integration.
537 In parallel, the relative quantification accuracy will be evaluated by comparing the spiked compounds
538 areas at least at two concentration levels. The ratios ((Area at level 2 – Area in the procedural blank)/
539 (Area at level 1 – Area in the procedural blank)) are unlikely to be accurate, but they should be

highlighted as differential by univariate statistical analysis. Representation as volcano plots (results of the statistical test, e.g., p-value vs. logarithm in base 2 of fold change) could be used for easy visualization.

5.2.4. Precision / accuracy

Precision and accuracy of mass and retention time on spiked compounds must be checked to ensure proper data recalibration and time alignment. Mass and retention time deviations will heavily depend on the analytical configuration used for instance, the type of HRMS (QTOF vs. Orbitrap), the abundance of compounds, column stationary phase or flow rate (nano, micro, standard). For regulatory purposes, we advocate strict guidelines concerning mass deviation and define a strict limit of less than 5 ppm. Modern mass spectrometers generally significantly undercut this limit. To determine the retention time deviation limits, reference data should be collected on standard compounds over a minimum span of 10 days or column run time of 200 samples[103]. Retention time drifts should also be continuously monitored, using a set of internal standards spanning the whole elution window and/or routine measurement of a set of unlabeled compounds and/or reference matrices also spanning the whole elution window.

5.2.5. Consistency

The consistency parameter will evaluate (i) the ability to identify the compounds with the subsequent annotation workflow using a suspect list containing only the standards and ii) the ability to identify the compounds with the subsequent annotation workflow using the most relevant suspect list to answer the scientific question. Thus, the proportion of correct identifications among spiked compounds after running the annotation workflow will be compared to the detected spiked compounds after data preprocessing.

5.3. Tools for QA/QC automatic evaluation

Algorithms have been developed to investigate data quality and could be used to support and help monitoring the various parameters defined earlier. In addition to vendor software, Skyline[106], for instance, is an open-source software allowing targeted extraction of compounds that could be used

for rapid manual integration of spiked compounds. mzRAPP[11] has been designed to support routine assessment of the detection and integration of non-target features. It calculates metrics such as benchmark recovery and isotopic ratio accuracy based on the most abundant isotopologue (MAIT) and the lowest abundant isotopologue (LAIT). It might be sometimes difficult to see a consistent isotopic pattern for low level contaminants. Scannotation[107] compares experimental isotopic ratios to theoretical ones. In addition, Scannotation provides a confidence index based on multiple parameters (retention times, mass accuracy and isotopic ratios) and a detection frequency of each feature in the dataset and could be used to evaluate consistency. MetaboanalystR[61] offers multiple statistical tools and can be used for instance to calculate coefficients of variation on peak integration. MetaboanalystR also provides univariate analysis that could be used to evaluate the semi-quantification accuracy. Finally, multiple annotation software could be used to ensure that correctly preprocessed spiked compounds are also identified. Examples of tools include patRoon[63] and MS-Dial[20].

6. Conclusion

Non-targeted LC-HRMS environmental, food and human biomonitoring data preprocessing suffers from type I errors (false positive detection), type II errors (false negative detection) and poor reproducibility, depending on the preprocessing software, preprocessing parameters and user experience. Currently, there is no ideal tool capable of preprocessing the data in a non-linear way and allowing the peak-picking of a diverse array of chromatographic peaks. Solutions have been proposed to mitigate these issues: (i) repositories, (ii) guidelines for reporting data preprocessing, (ii) implementation of semi-automated preprocessing workflows, (iii) provision of benchmark datasets, and (iv) development of tools to minimize true and false negative peak-picking (optimization of data preprocessing parameters and filtering of false positive features). To add to these ongoing initiatives, we propose a set of harmonized QA/QC procedures to ensure optimal detection of all true features and minimize false positives. This QA/QC set checks for sensitivity of feature detection, reproducibility, integration accuracy, precision/accuracy and consistency. We recommend these criteria to be carefully

591 checked before further investigating the results. We did not provide any thresholds in this review, as
592 the decision of what is acceptable depends on the study design and objectives, the instrument and the
593 preprocessing tool, as well as the compromise the user is ready to accept between preprocessing time
594 and detection of compounds of interest. Further collaborative studies will be needed to determine
595 thresholds and tolerances for these QA/QCs.
596 In any case, the results of QA/QC should be reported in the SSA/NTA data preprocessing workflow, as
597 a table for instance, to ensure transparency and ease of data reusability of any published study.
598 Interpretable criteria will also help to communicate confidence of data in the regulatory context. We
599 envision that these QA/QC set will evolve with time to incorporate the last technology advancements,
600 for instance ion mobility measurements and derived collision cross section (CCS) that are started to be
601 evaluated for application in the regulatory context, and for which reporting guides are already
602 available[108,109]. We hope that these QA/QC approaches will help to develop a new generation of tools
603 and benchmark datasets aiming to assess efficiently the quality of SSA and NTA data preprocessing.
604 Providing high quality preprocessed datasets with robust feature annotation is a mandatory step to
605 provide proper training datasets for the next-generation machine learning tools that will help to
606 automate the processing of complex HRMS datasets in the near future.
607

608 **Acknowledgment**

609 This work was supported by the project Partnership for the Assessment of Risks from Chemicals (PARC)
610 funded by the European Union research and innovation program Horizon Europe [grant numbers
611 101057014]. SL, JC and AD acknowledge the research infrastructure France Exposome. EJP, JK and ŽT
612 acknowledge the research infrastructure RECETOX RI (LM2023069), H2020 CETOCOEN Excellence
613 857560 and OP RDE CZ.02.1.01/0.0/0.0/17_043/0009632).

614 **Author contributions:**

615 Conceptualization; SL, JC, AD

616 Investigation; SL, JC, AD, EJP, JH, CH, ELS, ML, JPA

Methodology; SL, JC, AD, EJP, JH, CH, ELS, ML, JPA

Validation; SL, JC, AD

Visualization; SL, JC

Roles/Writing - original draft; SL

Writing - review & editing; SL, JC, AD, EJP, JH, CH, TS, LA, FB, NC, LD, GD, CG, TG, BH, EJ, JK, TK, BLB, HM, RN, HO, NP, JP, DS, MS, ŽT, ELS, ML, JPA

References

(1) Wang, Z.; Walker, G. W.; Muir, D. C. G.; Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ. Sci. Technol.* **2020**, *54* (5), 2575–2584. https://doi.org/10.1021/acs.est.9b06379.

(2) Ganzleben, C.; Antignac, J.-P.; Barouki, R.; Castaño, A.; Fiddicke, U.; Klánová, J.; Lebret, E.; Olea, N.; Sarigiannis, D.; Schoeters, G. R.; Sepai, O.; Tolonen, H.; Kolossa-Gehring, M. Human Biomonitoring as a Tool to Support Chemicals Regulation in the European Union. *International Journal of Hygiene and Environmental Health* **2017**, *220* (2, Part A), 94–97. https://doi.org/10.1016/j.ijheh.2017.01.007.

(3) Brack, W.; Aissa, S. A.; Backhaus, T.; Dulio, V.; Escher, B. I.; Faust, M.; Hilscherova, K.; Hollender, J.; Hollert, H.; Müller, C.; Munthe, J.; Posthuma, L.; Seiler, T.-B.; Slobodnik, J.; Teodorovic, I.; Tindall, A. J.; de Aragão Umbuzeiro, G.; Zhang, X.; Altenburger, R. Effect-Based Methods Are Key. The European Collaborative Project SOLUTIONS Recommends Integrating Effect-Based Methods for Diagnosis and Monitoring of Water Quality. *Environmental Sciences Europe* **2019**, *31* (1), 10. https://doi.org/10.1186/s12302-019-0192-2.

(4) Luijten, M.; Vlaanderen, J.; Kortenkamp, A.; Antignac, J.-P.; Barouki, R.; Bil, W.; van den Brand, A.; den Braver-Sewradj, S.; van Klaveren, J.; Mengelers, M.; Ottenbros, I.; Rantakokko, P.; Kolossa-Gehring, M.; Lebret, E. Mixture Risk Assessment and Human Biomonitoring: Lessons Learnt from HBM4EU. *International Journal of Hygiene and Environmental Health* **2023**, *249*, 114135. https://doi.org/10.1016/j.ijheh.2023.114135.

(5) David, A.; Chaker, J.; Price, E. J.; Bessonneau, V.; Chetwynd, A. J.; Vitale, C. M.; Klánová, J.; Walker, D. I.; Antignac, J.-P.; Barouki, R.; Miller, G. W. Towards a Comprehensive Characterisation of the Human Internal Chemical Exposome: Challenges and Perspectives. *Environ Int* **2021**, *156*, 106630. https://doi.org/10.1016/j.envint.2021.106630.

(6) Hollender, J.; Schymanski, E. L.; Ahrens, L.; Alygizakis, N.; Béen, F.; Bijlsma, L.; Brunner, A. M.; Celma, A.; Fildier, A.; Fu, Q.; Gago-Ferrero, P.; Gil-Solsona, R.; Haglund, P.; Hansen, M.; Kaserzon, S.; Kruve, A.; Lamoree, M.; Margoum, C.; Meijer, J.; Merel, S.; Rauert, C.; Rostkowski, P.; Samanipour, S.; Schulze, B.; Schulze, T.; Singh, R. R.; Slobodnik, J.; Steininger-Mairinger, T.; Thomaidis, N. S.; Togola, A.; Vorkamp, K.; Vulliet, E.; Zhu, L.; Krauss, M. NORMAN Guidance on Suspect and Non-Target Screening in Environmental Monitoring. *Environmental Sciences Europe* **2023**, *35* (1), 75. https://doi.org/10.1186/s12302-023-00779-4.

(7) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

(8) Pourchet, M.; Debrauwer, L.; Klanova, J.; Price, E. J.; Covaci, A.; Caballero-Casero, N.; Oberacher, H.; Lamoree, M.; Damont, A.; Fenaille, F.; Vlaanderen, J.; Meijer, J.; Krauss, M.; Sarigiannis, D.; Barouki, R.; Le Bizec, B.; Antignac, J.-P. Suspect and Non-Targeted Screening of Chemicals of Emerging Concern for Human Biomonitoring, Environmental Health Studies and Support to Risk Assessment: From Promises to Challenges and Harmonisation Issues. *Environment International* **2020**, *139*, 105545. https://doi.org/10.1016/j.envint.2020.105545.

(9) Rampler, E.; Abiead, Y. E.; Schoeny, H.; Rusz, M.; Hildebrand, F.; Fitz, V.; Koellensperger, G. Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics-Standardization, Coverage, and Throughput. *Anal Chem* **2021**, *93* (1), 519–545. https://doi.org/10.1021/acs.analchem.0c04698.

(10) Chaker, J.; Gilles, E.; Léger, T.; Jégou, B.; David, A. From Metabolomics to HRMS-Based Exposomics: Adapting Peak Picking and Developing Scoring for MS1 Suspect Screening. *Anal Chem* **2021**, *93* (3), 1792–1800. https://doi.org/10.1021/acs.analchem.0c04660.

(11) El Abiead, Y.; Milford, M.; Schoeny, H.; Rusz, M.; Salek, R. M.; Koellensperger, G. Power of mzRAPP-Based Performance Assessments in MS1-Based Nontargeted Feature Detection. *Anal. Chem.* **2022**, *94* (24), 8588–8595. https://doi.org/10.1021/acs.analchem.1c05270.

(12) Renner, G.; Reuschenbach, M. Critical Review on Data Processing Algorithms in Non-Target Screening: Challenges and Opportunities to Improve Result Comparability. *Anal Bioanal Chem* **2023**, *415* (18), 4111–4123. https://doi.org/10.1007/s00216-023-04776-7.

(13) Kirwan, J. A.; Gika, H.; Beger, R. D.; Bearden, D.; Dunn, W. B.; Goodacre, R.; Theodoridis, G.; Witting, M.; Yu, L.-R.; Wilson, I. D.; the metabolomics Quality Assurance and Quality Control Consortium (mQACC). Quality Assurance and Quality Control Reporting in Untargeted Metabolic Phenotyping: mQACC Recommendations for Analytical Quality Management. *Metabolomics* **2022**, *18* (9), 70. https://doi.org/10.1007/s11306-022-01926-3.

(14) Oberacher, H.; Sasse, M.; Antignac, J.-P.; Guitton, Y.; Debrauwer, L.; Jamin, E. L.; Schulze, T.; Krauss, M.; Covaci, A.; Caballero-Casero, N.; Rousseau, K.; Damont, A.; Fenaille, F.; Lamoree, M.; Schymanski, E. L. A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries. *Environmental Sciences Europe* **2020**, *32* (1), 43. https://doi.org/10.1186/s12302-020-00314-9.

(15) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72. https://doi.org/10.1007/s11306-018-1367-3.

(16) Dudzik, D.; Barbas-Bernardos, C.; García, A.; Barbas, C. Quality Assurance Procedures for Mass Spectrometry Untargeted Metabolomics. a Review. *Journal of Pharmaceutical and Biomedical Analysis* **2018**, *147*, 149–173. https://doi.org/10.1016/j.jpba.2017.07.044.

693 (17) Misra, B. B. Data Normalization Strategies in Metabolomics: Current Challenges, Approaches,
694 and Tools. *Eur J Mass Spectrom (Chichester)* **2020**, *26* (3), 165–174.
695 https://doi.org/10.1177/1469066720918446.
696 (18) Cuevas-Delgado, P.; Dudzik, D.; Miguel, V.; Lamas, S.; Barbas, C. Data-Dependent
697 Normalization Strategies for Untargeted Metabolomics-a Case Study. *Anal Bioanal Chem* **2020**,
698 *412* (24), 6391–6405. https://doi.org/10.1007/s00216-020-02594-9.
699 (19) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass
700 Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and
701 Identification. *Anal. Chem.* **2006**, *78* (3), 779–787. https://doi.org/10.1021/ac051437y.
702 (20) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst,
703 J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive
704 Metabolome Analysis. *Nat Methods* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.
705 (21) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.;
706 Murray, K. J.; Hoffmann, N.; Lu, M.; Sarvepalli, A.; Zhang, Z.; Fleischauer, M.; Dührkop, K.; Wesner,
707 M.; Hoogstra, S. J.; Rudt, E.; Mokshyna, O.; Brungs, C.; Ponomarov, K.; Mutabdžija, L.; Damiani, T.;
708 Pudney, C. J.; Earll, M.; Helmer, P. O.; Fallon, T. R.; Schulze, T.; Rivas-Ubach, A.; Bilbao, A.; Richter,
709 H.; Nothias, L.-F.; Wang, M.; Orešič, M.; Weng, J.-K.; Böcker, S.; Jeibmann, A.; Hayen, H.; Karst, U.;
710 Dorrestein, P. C.; Petras, D.; Du, X.; Pluskal, T. Integrative Analysis of Multimodal Mass
711 Spectrometry Data in MZmine 3. *Nat Biotechnol* **2023**, *41* (4), 447–449.
712 https://doi.org/10.1038/s41587-023-01690-2.
713 (22) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.;
714 Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.;
715 Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.;
716 Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible
717 Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat Methods* **2016**, *13* (9),
718 741–748. https://doi.org/10.1038/nmeth.3959.

719 (23) Misra, B. B. New Software Tools, Databases, and Resources in Metabolomics: Updates from
720 2020. *Metabolomics* **2021**, *17* (5), 49. https://doi.org/10.1007/s11306-021-01796-1.
721 (24) Spicer, R.; Salek, R. M.; Moreno, P.; Cañueto, D.; Steinbeck, C. Navigating Freely-Available
722 Software Tools for Metabolomics Analysis. *Metabolomics* **2017**, *13* (9), 106.
723 https://doi.org/10.1007/s11306-017-1242-7.
724 (25) Stanstrup, J.; Broeckling, C. D.; Helmus, R.; Hoffmann, N.; Mathé, E.; Naake, T.; Nicolotti, L.;
725 Peters, K.; Rainer, J.; Salek, R. M.; Schulze, T.; Schymanski, E. L.; Stravs, M. A.; Thévenot, E. A.;
726 Treutler, H.; Weber, R. J. M.; Willighagen, E.; Witting, M.; Neumann, S. The metaRbolomics
727 Toolbox in Bioconductor and Beyond. *Metabolites* **2019**, *9* (10), 200.
728 https://doi.org/10.3390/metabo9100200.
729 (26) Yu, T.; Park, Y.; Johnson, J. M.; Jones, D. P. apLCMS--Adaptive Processing of High-Resolution
730 LC/MS Data. *Bioinformatics* **2009**, *25* (15), 1930–1936.
731 https://doi.org/10.1093/bioinformatics/btp291.
732 (27) Hohrenk, L. L.; Itzel, F.; Baetz, N.; Tuerk, J.; Vosough, M.; Schmidt, T. C. Comparison of
733 Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in
734 Nontarget Screening of Environmental Samples. *Anal. Chem.* **2020**, *92* (2), 1898–1907.
735 https://doi.org/10.1021/acs.analchem.9b04095.
736 (28) Müller, E.; Huber, C. E.; Brack, W.; Krauss, M.; Schulze, T. Symbolic Aggregate Approximation
737 Improves Gap Filling in High-Resolution Mass Spectrometry Data Processing. *Anal. Chem.* **2020**, *92*
738 (15), 10425–10432. https://doi.org/10.1021/acs.analchem.0c00899.
739 (29) Armitage, E. G.; Godzien, J.; Alonso-Herranz, V.; López-Gonzálvez, Á.; Barbas, C. Missing Value
740 Imputation Strategies for Metabolomics Data. *Electrophoresis* **2015**, *36* (24), 3050–3060.
741 https://doi.org/10.1002/elps.201500352.
742 (30) Clark, T. N.; Houriet, J.; Vidar, W. S.; Kellogg, J. J.; Todd, D. A.; Cech, N. B.; Linington, R. G.
743 Interlaboratory Comparison of Untargeted Mass Spectrometry Data Uncovers Underlying Causes
744 for Variability. *J. Nat. Prod.* **2021**, *84* (3), 824–835. https://doi.org/10.1021/acs.jnatprod.0c01376.

(31) Baker, E. S.; Patti, G. J. Perspectives on Data Analysis in Metabolomics: Points of Agreement and Disagreement from the 2018 ASMS Fall Workshop. *J Am Soc Mass Spectrom* **2019**, *30* (10), 2031–2036. https://doi.org/10.1007/s13361-019-02295-3.

(32) Smith, R.; Tostengard, A. R. Quantitative Evaluation of Ion Chromatogram Extraction Algorithms. *J. Proteome Res.* **2020**, *19* (5), 1953–1964. https://doi.org/10.1021/acs.jproteome.9b00768.

(33) Reuschenbach, M.; Hohrenk-Danzouma, L. L.; Schmidt, T. C.; Renner, G. Development of a Scoring Parameter to Characterize Data Quality of Centroids in High-Resolution Mass Spectra. *Anal Bioanal Chem* **2022**, *414* (22), 6635–6645. https://doi.org/10.1007/s00216-022-04224-y.

(34) Rafiei, A.; Sleno, L. Comparison of Peak-Picking Workflows for Untargeted Liquid Chromatography/High-Resolution Mass Spectrometry Metabolomics Data Analysis. *Rapid Communications in Mass Spectrometry* **2015**, *29* (1), 119–127. https://doi.org/10.1002/rcm.7094.

(35) Liao, J.; Zhang, Y.; Zhang, W.; Zeng, Y.; Zhao, J.; Zhang, J.; Yao, T.; Li, H.; Shen, X.; Wu, G.; Zhang, W. Different Software Processing Affects the Peak Picking and Metabolic Pathway Recognition of Metabolomics Data. *Journal of Chromatography A* **2023**, *1687*, 463700. https://doi.org/10.1016/j.chroma.2022.463700.

(36) Li, Z.; Lu, Y.; Guo, Y.; Cao, H.; Wang, Q.; Shui, W. Comprehensive Evaluation of Untargeted Metabolomics Data Processing Software in Feature Detection, Quantification and Discriminating Marker Selection. *Analytica Chimica Acta* **2018**, *1029*, 50–57. https://doi.org/10.1016/j.aca.2018.05.001.

(37) Chen, Y.; Xu, J.; Zhang, R.; Shen, G.; Song, Y.; Sun, J.; He, J.; Zhan, Q.; Abliz, Z. Assessment of Data Pre-Processing Methods for LC-MS/MS-Based Metabolomics of Uterine Cervix Cancer. *Analyst* **2013**, *138* (9), 2669–2677. https://doi.org/10.1039/C3AN36818A.

(38) Guo, J.; Huan, T. Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2023**, *95* (14), 5894–5902. https://doi.org/10.1021/acs.analchem.2c04887.

(39) Coble, J. B.; Fraga, C. G. Comparative Evaluation of Preprocessing Freeware on Chromatography/Mass Spectrometry Data for Signature Discovery. *Journal of Chromatography A* **2014**, *1358*, 155–164. https://doi.org/10.1016/j.chroma.2014.06.100.

(40) Baran, R. Untargeted Metabolomics Suffers from Incomplete Raw Data Processing. *Metabolomics* **2017**, *13* (9), 107. https://doi.org/10.1007/s11306-017-1246-3.

(41) Smith, R.; Ventura, D.; Prince, J. T. Controlling for Confounding Variables in MS-Omics Protocol: Why Modularity Matters. *Briefings in Bioinformatics* **2014**, *15* (5), 768–770. https://doi.org/10.1093/bib/bbt049.

(42) Li, S.; Siddiqa, A.; Thapa, M.; Chi, Y.; Zheng, S. Trackable and Scalable LC-MS Metabolomics Data Processing Using Asari. *Nat Commun* **2023**, *14* (1), 4113. https://doi.org/10.1038/s41467-023-39889-1.

(43) Houriet, J.; Vidar, W. S.; Manwill, P. K.; Todd, D. A.; Cech, N. B. How Low Can You Go? Selecting Intensity Thresholds for Untargeted Metabolomics Data Preprocessing. *Anal. Chem.* **2022**, *94* (51), 17964–17971. https://doi.org/10.1021/acs.analchem.2c04088.

(44) Hajjar, G.; Barros Santos, M. C.; Bertrand-Michel, J.; Canlet, C.; Castelli, F.; Creusot, N.; Dechaumet, S.; Diémé, B.; Giacomoni, F.; Giraudeau, P.; Guitton, Y.; Thévenot, E.; Tremblay-Franco, M.; Junot, C.; Jourdan, F.; Fenaille, F.; Comte, B.; Pétriacq, P.; Pujos-Guillot, E. Scaling-up Metabolomics: Current State and Perspectives. *TrAC Trends in Analytical Chemistry* **2023**, *167*, 117225. https://doi.org/10.1016/j.trac.2023.117225.

(45) Müller, E.; Huber, C.; Beckers, L.-M.; Brack, W.; Krauss, M.; Schulze, T. A Data Set of 255,000 Randomly Selected and Manually Classified Extracted Ion Chromatograms for Evaluation of Peak Detection Methods. *Metabolites* **2020**, *10* (4), 162. https://doi.org/10.3390/metabo10040162.

(46) Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F. Proposed Minimum Reporting Standards for Data

796 Analysis in Metabolomics. *Metabolomics* **2007**, *3* (3), 231–241. https://doi.org/10.1007/s11306-
797 007-0081-3.
798 (47) Considine, E. C.; Thomas, G.; Boulesteix, A. L.; Khashan, A. S.; Kenny, L. C. Critical Review of
799 Reporting of the Data Analysis Step in Metabolomics. *Metabolomics* **2017**, *14* (1), 7.
800 https://doi.org/10.1007/s11306-017-1299-3.
801 (48) Viant, M. R.; Ebbels, T. M. D.; Beger, R. D.; Ekman, D. R.; Epps, D. J. T.; Kamp, H.; Leonards, P.
802 E. G.; Loizou, G. D.; MacRae, J. I.; van Ravenzwaay, B.; Rocca-Serra, P.; Salek, R. M.; Walk, T.;
803 Weber, R. J. M. Use Cases, Best Practice and Reporting Standards for Metabolomics in Regulatory
804 Toxicology. *Nat Commun* **2019**, *10*, 3041. https://doi.org/10.1038/s41467-019-10900-y.
805 (49) Peter, K. T.; Phillips, A. L.; Knolhoff, A. M.; Gardinali, P. R.; Manzano, C. A.; Miller, K. E.;
806 Pristner, M.; Sabourin, L.; Sumarah, M. W.; Warth, B.; Sobus, J. R. Nontargeted Analysis Study
807 Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. *Anal Chem*
808 **2021**, *93* (41), 13870–13879. https://doi.org/10.1021/acs.analchem.1c02621.
809 (50) Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi,
810 R.; Nair, K. S.; Sumner, S.; Subramaniam, S. Metabolomics Workbench: An International
811 Repository for Metabolomics Data and Metadata, Metabolite Standards, Protocols, Tutorials and
812 Training, and Analysis Tools. *Nucleic Acids Res* **2016**, *44* (Database issue), D463–D470.
813 https://doi.org/10.1093/nar/gkv1042.
814 (51) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C.
815 MetaboLights: A Resource Evolving in Response to the Needs of Its Scientific Community. *Nucleic*
816 *Acids Research* **2020**, *48* (D1), D440–D444. https://doi.org/10.1093/nar/gkz1019.
817 (52) Leao, T. F.; Clark, C. M.; Bauermeister, A.; Elijah, E. O.; Gentry, E.; Husband, M.; Faria de
818 Oliveira, M.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Quick-Start for Untargeted Metabolomics
819 Analysis in GNPS. *Nat Metab* **2021**, *3* (7), 880–882. https://doi.org/10.1038/s42255-021-00429-0.
820 (53) Deutsch, E. W.; Bandeira, N.; Perez-Riverol, Y.; Sharma, V.; Carver, J. J.; Mendoza, L.; Kundu,
821 D. J.; Wang, S.; Bandla, C.; Kamatchinathan, S.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.;

822 Kawano, S.; Okuda, S.; Watanabe, Y.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaíno,
823 J. A. The ProteomeXchange Consortium at 10 Years: 2023 Update. *Nucleic Acids Res* **2023**, *51* (D1),
824 D1539–D1548. https://doi.org/10.1093/nar/gkac1040.
825 (54) *ProteomeCentral Datasets*. https://proteomecentral.proteomexchange.org/ (accessed 2023-
826 11-26).
827 (55) *MetabolomeXchange*. http://www.metabolomexchange.org/site/ (accessed 2023-11-26).
828 (56) Alygizakis, N. A.; Oswald, P.; Thomaidis, N. S.; Schymanski, E. L.; Aalizadeh, R.; Schulze, T.;
829 Oswaldova, M.; Slobodnik, J. NORMAN Digital Sample Freezing Platform: A European Virtual
830 Platform to Exchange Liquid Chromatography High Resolution-Mass Spectrometry Data and
831 Screen Suspects in "Digitally Frozen" Environmental Samples. *TrAC Trends in Analytical Chemistry*
832 **2019**, *115*, 129–137. https://doi.org/10.1016/j.trac.2019.04.008.
833 (57) *Home - Digital Sample Freezing Platform*. https://dsfp.norman-data.eu/ (accessed 2023-11-
834 26).
835 (58) Schulze, B.; Jeon, Y.; Kaserzon, S.; Heffernan, A. L.; Dewapriya, P.; O'Brien, J.; Gomez Ramos,
836 M. J.; Ghorbani Gorji, S.; Mueller, J. F.; Thomas, K. V.; Samanipour, S. An Assessment of Quality
837 Assurance/Quality Control Efforts in High Resolution Mass Spectrometry Non-Target Workflows
838 for Analysis of Environmental Samples. *TrAC Trends in Analytical Chemistry* **2020**, *133*, 116063.
839 https://doi.org/10.1016/j.trac.2020.116063.
840 (59) Hites, R. A.; Jobst, K. J. Is Nontargeted Screening Reproducible? *Environ. Sci. Technol.* **2018**,
841 *52* (21), 11975–11976. https://doi.org/10.1021/acs.est.8b05671.
842 (60) Giacomoni, F.; Le Corguillé, G.; Monsoor, M.; Landi, M.; Pericard, P.; Pétéra, M.; Duperier, C.;
843 Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; Goulitquer, S.; Thévenot, E. A.; Caron, C.
844 Workflow4Metabolomics: A Collaborative Research Infrastructure for Computational
845 Metabolomics. *Bioinformatics* **2015**, *31* (9), 1493–1495.
846 https://doi.org/10.1093/bioinformatics/btu813.

(61) Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. Using MetaboAnalyst 5.0 for LC-HRMS Spectra Processing, Multi-Omics Integration and Covariate Adjustment of Global Metabolomics Data. *Nat Protoc* **2022**, *17* (8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w.

(62) An Open Software Development-Based Ecosystem of R Packages for Metabolomics Data Analysis. https://doi.org/10.5281/zenodo.7936787.

(63) Helmus, R.; ter Laak, T. L.; van Wezel, A. P.; de Voogt, P.; Schymanski, E. L. patRoon: Open Source Software Platform for Environmental Mass Spectrometry Based Non-Target Screening. *Journal of Cheminformatics* **2021**, *13* (1), 1. https://doi.org/10.1186/s13321-020-00477-w.

(64) Navarro, P.; Kuharev, J.; Gillet, L. C.; Bernhardt, O. M.; MacLean, B.; Röst, H. L.; Tate, S. A.; Tsou, C.-C.; Reiter, L.; Distler, U.; Rosenberger, G.; Perez-Riverol, Y.; Nesvizhskii, A. I.; Aebersold, R.; Tenzer, S. A Multicenter Study Benchmarks Software Tools for Label-Free Proteome Quantification. *Nat Biotechnol* **2016**, *34* (11), 1130–1136. https://doi.org/10.1038/nbt.3685.

(65) Ramus, C.; Hovasse, A.; Marcellin, M.; Hesse, A.-M.; Mouton-Barbosa, E.; Bouyssié, D.; Vaca, S.; Carapito, C.; Chaoui, K.; Bruley, C.; Garin, J.; Cianférani, S.; Ferro, M.; Van Dorssaeler, A.; Burlet-Schiltz, O.; Schaeffer, C.; Couté, Y.; Gonzalez de Peredo, A. Benchmarking Quantitative Label-Free LC–MS Data Processing Workflows Using a Complex Spiked Proteomic Standard Dataset. *Journal of Proteomics* **2016**, *132*, 51–62. https://doi.org/10.1016/j.jprot.2015.11.011.

(66) Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* **2017**, *89* (17), 8689–8695. https://doi.org/10.1021/acs.analchem.7b01069.

(67) Henning, J.; Tostengard, A.; Smith, R. A Peptide-Level Fully Annotated Data Set for Quantitative Evaluation of Precursor-Aware Mass Spectrometry Data Processing Algorithms. *J. Proteome Res.* **2019**, *18* (1), 392–398. https://doi.org/10.1021/acs.jproteome.8b00659.

872 (68) Schulze, B.; van Herwerden, D.; Allan, I.; Bijlsma, L.; Etxebarria, N.; Hansen, M.; Merel, S.;
873 Vrana, B.; Aalizadeh, R.; Bajema, B.; Dubocq, F.; Coppola, G.; Fildier, A.; Fialová, P.; Frøkjær, E.;
874 Grabic, R.; Gago-Ferrero, P.; Gravert, T.; Hollender, J.; Huynh, N.; Jacobs, G.; Jonkers, T.; Kaserzon,
875 S.; Lamoree, M.; Le Roux, J.; Mairinger, T.; Margoum, C.; Mascolo, G.; Mebold, E.; Menger, F.;
876 Miège, C.; Meijer, J.; Moilleron, R.; Murgolo, S.; Peruzzo, M.; Pijnappels, M.; Reid, M.; Roscioli, C.;
877 Soulier, C.; Valsecchi, S.; Thomaidis, N.; Vulliet, E.; Young, R.; Samanipour, S. Inter-Laboratory
878 Mass Spectrometry Dataset Based on Passive Sampling of Drinking Water for Non-Target Analysis.
879 *Sci Data* **2021**, *8* (1), 223. https://doi.org/10.1038/s41597-021-01002-w.
880 (69) Eliasson, M.; Rännar, S.; Madsen, R.; Donten, M. A.; Marsden-Edwards, E.; Moritz, T.;
881 Shockcor, J. P.; Johansson, E.; Trygg, J. Strategy for Optimizing LC-MS Data Processing in
882 Metabolomics: A Design of Experiments Approach. *Anal. Chem.* **2012**, *84* (15), 6869–6876.
883 https://doi.org/10.1021/ac301482k.
884 (70) Zheng, H.; Clausen, M. R.; Dalsgaard, T. K.; Mortensen, G.; Bertram, H. C. Time-Saving Design
885 of Experiment Protocol for Optimization of LC-MS Data Processing in Metabolomic Approaches.
886 *Anal. Chem.* **2013**, *85* (15), 7109–7116. https://doi.org/10.1021/ac4020325.
887 (71) Kiefer, K.; Du, L.; Singer, H.; Hollender, J. Identification of LC-HRMS Nontarget Signals in
888 Groundwater after Source Related Prioritization. *Water Research* **2021**, *196*, 116994.
889 https://doi.org/10.1016/j.watres.2021.116994.
890 (72) Dom, I.; Biré, R.; Hort, V.; Lavison-Bompard, G.; Nicolas, M.; Guérin, T. Extended Targeted and
891 Non-Targeted Strategies for the Analysis of Marine Toxins in Mussels and Oysters by (LC-HRMS).
892 *Toxins* **2018**, *10* (9), 375. https://doi.org/10.3390/toxins10090375.
893 (73) Hu, M.; Krauss, M.; Brack, W.; Schulze, T. Optimization of LC-Orbitrap-HRMS Acquisition and
894 MZmine 2 Data Processing for Nontarget Screening of Environmental Samples Using Design of
895 Experiments. *Anal Bioanal Chem* **2016**, *408* (28), 7905–7915. https://doi.org/10.1007/s00216-016-
896 9919-8.

(74) Manier, S. K.; Keller, A.; Meyer, M. R. Automated Optimization of XCMS Parameters for Improved Peak Picking of Liquid Chromatography–Mass Spectrometry Data Using the Coefficient of Variation and Parameter Sweeping for Untargeted Metabolomics. *Drug Testing and Analysis* **2019**, *11* (6), 752–761. https://doi.org/10.1002/dta.2552.

(75) Uppal, K.; Soltow, Q. A.; Strobel, F. H.; Pittard, W. S.; Gernert, K. M.; Yu, T.; Jones, D. P. xMSanalyzer: Automated Pipeline for Improved Feature Detection and Downstream Analysis of Large-Scale, Non-Targeted Metabolomics Data. *BMC Bioinformatics* **2013**, *14* (1), 15. https://doi.org/10.1186/1471-2105-14-15.

(76) Ju, R.; Liu, X.; Zheng, F.; Zhao, X.; Lu, X.; Lin, X.; Zeng, Z.; Xu, G. A Graph Density-Based Strategy for Features Fusion from Different Peak Extract Software to Achieve More Metabolites in Metabolic Profiling from High-Resolution Mass Spectrometry. *Analytica Chimica Acta* **2020**, *1139*, 8–14. https://doi.org/10.1016/j.aca.2020.09.029.

(77) Brodsky, L.; Moussaieff, A.; Shahaf, N.; Aharoni, A.; Rogachev, I. Evaluation of Peak Picking Quality in LC−MS Metabolomics Data. *Anal. Chem.* **2010**, *82* (22), 9177–9187. https://doi.org/10.1021/ac101216e.

(78) Libiseller, G.; Dvorzak, M.; Kleb, U.; Gander, E.; Eisenberg, T.; Madeo, F.; Neumann, S.; Trausinger, G.; Sinner, F.; Pieber, T.; Magnes, C. IPO: A Tool for Automated Optimization of XCMS Parameters. *BMC Bioinformatics* **2015**, *16*, 118. https://doi.org/10.1186/s12859-015-0562-8.

(79) Delabriere, A.; Warmer, P.; Brennsteiner, V.; Zamboni, N. SLAW: A Scalable and Self-Optimizing Processing Workflow for Untargeted LC-MS. *Anal. Chem.* **2021**, *93* (45), 15024–15032. https://doi.org/10.1021/acs.analchem.1c02687.

(80) McLean, C.; Kujawinski, E. B. AutoTuner: High Fidelity and Robust Parameter Selection for Metabolomics Data Processing. *Anal. Chem.* **2020**, *92* (8), 5724–5732. https://doi.org/10.1021/acs.analchem.9b04804.

(81) Guo, J.; Shen, S.; Huan, T. Paramounter: Direct Measurement of Universal Parameters To Process Metabolomics Data in a "White Box." *Anal. Chem.* **2022**, *94* (10), 4260–4268. https://doi.org/10.1021/acs.analchem.1c04758.

(82) Guo, J.; Shen, S.; Xing, S.; Chen, Y.; Chen, F.; Porter, E. M.; Yu, H.; Huan, T. EVA: Evaluation of Metabolic Feature Fidelity Using a Deep Learning Model Trained With Over 25000 Extracted Ion Chromatograms. *Anal. Chem.* **2021**, *93* (36), 12181–12186. https://doi.org/10.1021/acs.analchem.1c01309.

(83) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Global Metabolic Profiling Procedures for Urine Using UPLC–MS. *Nat Protoc* **2010**, *5* (6), 1005–1018. https://doi.org/10.1038/nprot.2010.50.

(84) Schiffman, C.; Petrick, L.; Perttula, K.; Yano, Y.; Carlsson, H.; Whitehead, T.; Metayer, C.; Hayes, J.; Rappaport, S.; Dudoit, S. Filtering Procedures for Untargeted LC-MS Metabolomics Data. *BMC Bioinformatics* **2019**, *20* (1), 334. https://doi.org/10.1186/s12859-019-2871-9.

(85) Ju, R.; Liu, X.; Zheng, F.; Zhao, X.; Lu, X.; Zeng, Z.; Lin, X.; Xu, G. Removal of False Positive Features to Generate Authentic Peak Table for High-Resolution Mass Spectrometry-Based Metabolomics Study. *Analytica Chimica Acta* **2019**, *1067*, 79–87. https://doi.org/10.1016/j.aca.2019.04.011.

(86) Fraisier-Vannier, O.; Chervin, J.; Cabanac, G.; Puech, V.; Fournier, S.; Durand, V.; Amiel, A.; André, O.; Benamar, O. A.; Dumas, B.; Tsugawa, H.; Marti, G. MS-CleanR: A Feature-Filtering Workflow for Untargeted LC–MS Based Metabolomics. *Anal. Chem.* **2020**, *92* (14), 9971–9981. https://doi.org/10.1021/acs.analchem.0c01594.

(87) Pirttilä, K.; Balgoma, D.; Rainer, J.; Pettersson, C.; Hedeland, M.; Brunius, C. Comprehensive Peak Characterization (CPC) in Untargeted LC–MS Analysis. *Metabolites* **2022**, *12* (2), 137. https://doi.org/10.3390/metabo12020137.

945 (88) Gloaguen, Y.; Kirwan, J. A.; Beule, D. Deep Learning-Assisted Peak Curation for Large-Scale LC-
946 MS Metabolomics. *Anal. Chem.* **2022**, *94* (12), 4930–4937.
947 https://doi.org/10.1021/acs.analchem.1c02220.
948 (89) Kantz, E. D.; Tiwari, S.; Watrous, J. D.; Cheng, S.; Jain, M. Deep Neural Networks for
949 Classification of LC-MS Spectral Peaks. *Anal. Chem.* **2019**, *91* (19), 12407–12413.
950 https://doi.org/10.1021/acs.analchem.9b02983.
951 (90) Chetnik, K.; Petrick, L.; Pandey, G. MetaClean: A Machine Learning-Based Classifier for
952 Reduced False Positive Peak Detection in Untargeted LC-MS Metabolomics Data. *Metabolomics*
953 **2020**, *16* (11), 117. https://doi.org/10.1007/s11306-020-01738-3.
954 (91) Albóniga, O. E.; González, O.; Alonso, R. M.; Xu, Y.; Goodacre, R. Optimization of XCMS
955 Parameters for LC–MS Metabolomics: An Assessment of Automated versus Manual Tuning and Its
956 Effect on the Final Results. *Metabolomics* **2020**, *16* (1), 14. https://doi.org/10.1007/s11306-020-
957 1636-9.
958 (92) Guo, J.; Yu, H.; Xing, S.; Huan, T. Addressing Big Data Challenges in Mass Spectrometry-Based
959 Metabolomics. *Chem. Commun.* **2022**, *58* (72), 9979–9990. https://doi.org/10.1039/D2CC03598G.
960 (93) Giné, R.; Capellades, J.; Badia, J. M.; Vughs, D.; Schwaiger-Haber, M.; Alexandrov, T.; Vinaixa,
961 M.; Brunner, A. M.; Patti, G. J.; Yanes, O. HERMES: A Molecular-Formula-Oriented Method to
962 Target the Metabolome. *Nat Methods* **2021**, *18* (11), 1370–1376. https://doi.org/10.1038/s41592-
963 021-01307-z.
964 (94) Fakouri Baygi, S.; Kumar, Y.; Barupal, D. K. IDSL.IPA Characterizes the Organic Chemical Space
965 in Untargeted LC/HRMS Data Sets. *J. Proteome Res.* **2022**, *21* (6), 1485–1494.
966 https://doi.org/10.1021/acs.jproteome.2c00120.
967 (95) Woldegebriel, M.; Derks, E. Artificial Neural Network for Probabilistic Feature Recognition in
968 Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. *Anal. Chem.* **2017**, *89* (2),
969 1212–1221. https://doi.org/10.1021/acs.analchem.6b03678.

(96) Wang, R.; Lu, M.; An, S.; Wang, J.; Yu, C. 3D-MSNet: A Point Cloud-Based Deep Learning Model for Untargeted Feature Detection and Quantification in Profile LC-HRMS Data. *Bioinformatics* **2023**, *39* (5), btad195. https://doi.org/10.1093/bioinformatics/btad195.

(97) Melnikov, A. D.; Tsentalovich, Y. P.; Yanshole, V. V. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. *Anal. Chem.* **2020**, *92* (1), 588–592. https://doi.org/10.1021/acs.analchem.9b04811.

(98) Bueschl, C.; Doppler, M.; Varga, E.; Seidl, B.; Flasch, M.; Warth, B.; Zanghellini, J. PeakBot: Machine-Learning-Based Chromatographic Peak Picking. *Bioinformatics* **2022**, *38* (13), 3422–3428. https://doi.org/10.1093/bioinformatics/btac344.

(99) Samanipour, S.; O'Brien, J. W.; Reid, M. J.; Thomas, K. V. Self Adjusting Algorithm for the Nontargeted Feature Detection of High Resolution Mass Spectrometry Coupled with Liquid Chromatography Profile Data. *Anal. Chem.* **2019**, *91* (16), 10800–10807. https://doi.org/10.1021/acs.analchem.9b02422.

(100) Woldegebriel, M.; Vivó-Truyols, G. Probabilistic Model for Untargeted Peak Detection in LC–MS Using Bayesian Statistics. *Anal. Chem.* **2015**, *87* (14), 7345–7355. https://doi.org/10.1021/acs.analchem.5b01521.

(101) Liu, Y.; Yang, Y.; Chen, W.; Shen, F.; Xie, L.; Zhang, Y.; Zhai, Y.; He, F.; Zhu, Y.; Chang, C. DeepRTAlign: Toward Accurate Retention Time Alignment for Large Cohort Mass Spectrometry Data Analysis. *Nat Commun* **2023**, *14* (1), 8188. https://doi.org/10.1038/s41467-023-43909-5.

(102) Skoraczyński, G.; Gambin, A.; Miasojedow, B. Alignstein: Optimal Transport for Improved LC-MS Retention Time Alignment. *GigaScience* **2022**, *11*, giac101. https://doi.org/10.1093/gigascience/giac101.

(103) Vitale, C. M.; Lommen, A.; Huber, C.; Wagner, K.; Garlito Molina, B.; Nijssen, R.; Price, E. J.; Blokland, M.; van Tricht, F.; Mol, H. G. J.; Krauss, M.; Debrauwer, L.; Pardo, O.; Leon, N.; Klanova, J.; Antignac, J.-P. Harmonized Quality Assurance/Quality Control Provisions for Nontargeted

995 Measurement of Urinary Pesticide Biomarkers in the HBM4EU Multisite SPECIMEn Study. *Anal.*
996 *Chem.* **2022**, *94* (22), 7833–7843. https://doi.org/10.1021/acs.analchem.2c00061.
997 (104) Place, B. J.; Ulrich, E. M.; Challis, J. K.; Chao, A.; Du, B.; Favela, K.; Feng, Y.-L.; Fisher, C. M.;
998 Gardinali, P.; Hood, A.; Knolhoff, A. M.; McEachran, A. D.; Nason, S. L.; Newton, S. R.; Ng, B.;
999 Nuñez, J.; Peter, K. T.; Phillips, A. L.; Quinete, N.; Renslow, R.; Sobus, J. R.; Sussman, E. M.; Warth,
1000 B.; Wickramasekara, S.; Williams, A. J. An Introduction to the Benchmarking and Publications for
1001 Non-Targeted Analysis Working Group. *Anal. Chem.* **2021**, *93* (49), 16289–16296.
1002 https://doi.org/10.1021/acs.analchem.1c02660.
1003 (105) Knolhoff, A. M.; Premo, J. H.; Fisher, C. M. A Proposed Quality Control Standard Mixture and
1004 Its Uses for Evaluating Nontargeted and Suspect Screening LC/HR-MS Method Performance. *Anal.*
1005 *Chem.* **2021**, *93* (3), 1596–1603. https://doi.org/10.1021/acs.analchem.0c04036.
1006 (106) Adams, K. J.; Pratt, B.; Bose, N.; Dubois, L. G.; St. John-Williams, L.; Perrott, K. M.; Ky, K.;
1007 Kapahi, P.; Sharma, V.; MacCoss, M. J.; Moseley, M. A.; Colton, C. A.; MacLean, B. X.; Schilling, B.;
1008 Thompson, J. W. Skyline for Small Molecules: A Unifying Software Package for Quantitative
1009 Metabolomics. *J. Proteome Res.* **2020**, *19* (4), 1447–1458.
1010 https://doi.org/10.1021/acs.jproteome.9b00640.
1011 (107) Chaker, J.; Gilles, E.; Monfort, C.; Chevrier, C.; Lennon, S.; David, A. Scannotation: A Suspect
1012 Screening Tool for the Rapid Pre-Annotation of the Human LC-HRMS-Based Chemical Exposome.
1013 *Environ. Sci. Technol.* **2023**. https://doi.org/10.1021/acs.est.3c04764.
1014 (108) Celma, A.; Ahrens, L.; Gago-Ferrero, P.; Hernández, F.; López, F.; Lundqvist, J.; Pitarch, E.;
1015 Sancho, J. V.; Wiberg, K.; Bijlsma, L. The Relevant Role of Ion Mobility Separation in LC-HRMS
1016 Based Screening Strategies for Contaminants of Emerging Concern in the Aquatic Environment.
1017 *Chemosphere* **2021**, *280*, 130799. https://doi.org/10.1016/j.chemosphere.2021.130799.
1018 (109) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.;
1019 Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B.
1020 H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.;

1021 Hogan Jr, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.;
1022 Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach,
1023 T. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass*
1024 *Spectrometry Reviews* **2019**, *38* (3), 291–320. https://doi.org/10.1002/mas.21585.
1025

Highlights

- Preprocessing of raw data from suspect screening and non-targeted analysis by liquid chromatography coupled to high resolution mass spectrometry (SSA/NTA LC-HRMS) is affected by reproducibility and incomplete peak peaking
- Optimization tools and guidelines were developed to improve SSA/NTA LC-HRMS data preprocessing
- Quality assurance/Quality control provisions for SSA/NTS LC-HRMS data preprocessing are proposed to assess performance of preprocessing

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: