



**HAL**  
open science

# Scannotation: A Suspect Screening Tool for the Rapid Pre-Annotation of the Human LC-HRMS-Based Chemical Exposome

Jade Chaker, Erwann Gilles, Christine Monfort, Cécile Chevrier, Sarah Lennon, Arthur David

## ► To cite this version:

Jade Chaker, Erwann Gilles, Christine Monfort, Cécile Chevrier, Sarah Lennon, et al.. Scannotation: A Suspect Screening Tool for the Rapid Pre-Annotation of the Human LC-HRMS-Based Chemical Exposome. *Environmental Science and Technology*, 2023, 57 (48), pp.19253-19262. 10.1021/acs.est.3c04764 . hal-04297092

**HAL Id: hal-04297092**

**<https://ehesp.hal.science/hal-04297092v1>**

Submitted on 7 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Scannotation: a suspect screening tool for the rapid  
2 pre-annotation of the human LC-HRMS-based  
3 chemical exposome

4

5

6 Jade Chaker<sup>‡</sup>, Erwann Gilles<sup>‡</sup>, Christine Monfort, Cécile Chevrier, Sarah Lennon, Arthur David\*

7

8

9 Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) –

10 UMR\_S 1085, F-35000 Rennes, France

11

12 <sup>‡</sup>Shared authorship (co-first authors)

13 \*To whom correspondence should be addressed:

14 Tel: +33 299022885

15 email: [arthur.david@ehesp.fr](mailto:arthur.david@ehesp.fr)

16

17

18

19

20

21

22

23

24

## 25 Abstract

26 In an increasingly chemically polluted environment, rapidly characterizing the human chemical  
27 exposome (i.e. chemical mixtures accumulating in humans) at the population scale is critical to  
28 understand its impact on health. High-resolution mass spectrometry (HRMS) profiling of complex  
29 biological matrices can theoretically provide a comprehensive picture of chemical exposures. However,  
30 annotating the detected chemical features, particularly low-abundant ones, remains a significant  
31 obstacle to implementing such approaches at a large scale. We present Scannotation  
32 ([https://github.com/scannotation/Scannotation\\_software](https://github.com/scannotation/Scannotation_software)), an automated and user-friendly suspect  
33 screening tool for the rapid pre-annotation of HRMS preprocessed datasets. This software tool  
34 combines several MS1 chemical predictors, i.e., m/z, experimental and predicted retention times,  
35 isotopic patterns and neutral loss patterns, to score the proximity between features and suspects, thus  
36 efficiently prioritizing tentative annotations to verify. Scannotation and MS-DIAL4 were used to  
37 annotate blood serum samples of 75 Breton adolescents. Scannotation's combination of MS1-based  
38 chemical predictors allowed annotating 89 chemically diverse environmental compounds with high  
39 confidence (confirmed by MS2 when available). These compounds included 62% of emerging and  
40 unknown molecules, for which no toxicological and/or human biomonitoring data is reported in the  
41 literature. The complementarity observed with MS-DIAL4 results demonstrates the relevance of  
42 Scannotation for the efficient pre-annotation of large-scale exposomics datasets.

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

## 60 Introduction

61 Chemical pollution is a great and growing global problem, with more than tens of thousands of very  
62 diverse chemicals currently present on the market<sup>1</sup>. Human chemical exposure and toxicological data  
63 are only available for a few hundreds of these chemicals, meaning that a great share of chemicals  
64 potentially associated with deleterious health outcomes have not been investigated so far<sup>1</sup>.  
65 Nevertheless, the emergence of the exposome paradigm<sup>2</sup> as well as technological advances such as  
66 hyphenated high-resolution mass spectrometry techniques (e.g., LC-HRMS) have paved the way for  
67 the use of suspect screening (SS) and non-targeted screening (NTS) approaches, and therefore offer  
68 great promises for a comprehensive characterization of exogenous substances mixtures accumulating  
69 in humans<sup>3-7</sup>. It is however crucial to overcome the remaining methodological obstacles before  
70 implementing large-scale non-targeted exposomics studies to population-based studies. Indeed, the  
71 annotation of the tens of thousands of signals present in HRMS datasets remains one of the main  
72 bottlenecks, as only a few percent of signals are usually annotated<sup>4</sup>.

73 Annotation of complex HRMS data can be performed using NTS, which relies on the structure  
74 elucidation of features, prioritized as differential between two (or more) groups, or SS, which relies on  
75 the annotation of features prioritized for their similarity to compounds listed in a suspect library. This  
76 second methodology is particularly promising, in part because it has a strong potential for automation  
77 and allows for a very rapid prioritization of signals of interest. Furthermore, there is no restriction  
78 regarding the number of suspects that can be included, as well as the forms they are searched as (i.e.,  
79 parent or metabolite, adduct). The comparison of experimental features and suspects on characteristic  
80 properties such as their mass/charge ratio or their MS<sup>2</sup> fragmentation profile can be automatically  
81 performed, before being manually validated. Even though some bioinformatics solutions already exist  
82 to carry out MS<sup>2</sup>-based spectral library searches, allowing the automatized comparison of reference  
83 and experimental MS<sup>2</sup> spectra (e.g., xMSannotator, msPurity, MZmine2, MS-DIAL, patRoan,

84 CAMERA)<sup>8-13</sup>, the number of software tools is still limited and not necessarily suited for exposomics  
85 applications<sup>14</sup>, which present specific challenges detailed hereafter.

86 One of the main challenges of using biological matrices to characterize the human internal chemical  
87 exposome is the wide dynamic range of concentrations for compounds present in the samples.  
88 Compounds of interest in an exposomics context are often present at much lower levels (i.e., tens of  
89 pg/mL<sup>15</sup>) than many endogenous metabolites (i.e., up to a few mg/mL). These low-abundant  
90 xenobiotics do not systematically trigger MS2 acquisition<sup>16</sup>. This strongly limits the annotation's  
91 confidence level according to Schymanski's scale, which is the current reference<sup>17</sup>. Despite this fact,  
92 other factors accessible through MS1 data, such as retention time (Rt), distinctive isotope profiles  
93 based on halogen contents (often present in exogenous compounds such as pesticides), or detection  
94 of other phase I/II metabolites could already provide reliable indications on the annotation's  
95 plausibility. Moreover, while reference data, such as Rt from a standard or exhaustive knowledge of  
96 metabolism pathway, may not be available for each suspect compound, various models exist to  
97 provide predicted values for Rt<sup>18-20</sup> and structures for plausible metabolites<sup>21</sup>. Hence, these relevant  
98 MS1-based chemical predictors (experimental or predicted) can be combined into a substantial body  
99 of evidence pointing towards specific chemical identities for features of interest in exposomics  
100 applications. However, manually comparing these predictors between tens of thousands of suspects  
101 and features is a highly time-consuming task, so there is a need to develop tools that could speed up  
102 this process.

103 In this context, we developed the Scannotation software. It automatically scores the comparison  
104 between suspect compounds and features obtained from any pre-processing software<sup>10,11,22</sup> on three  
105 MS1-based chemical predictors: mass/charge ratio, isotopic ratios, and experimental or predicted  
106 retention time. It also generates common phase II metabolites and computes theoretical isotopic  
107 profiles for parent compounds and metabolites. Scannotation calculates and displays proximity scores  
108 between features and suspects called "confidence indices" for these three chemical predictors, as well

109 as an overall confidence index that effectively evaluates each automatized pre-annotation's reliability.  
110 This software allows the efficient prioritization of pre-annotations, which can then be validated  
111 manually. This prioritization saves considerable time, as it allows a fast prioritization of features to be  
112 further investigated to gain a higher confidence, decreases the amount of false positive annotations,  
113 thus contributing to the wider and faster pre-annotation of HRMS datasets. The list of suspects can  
114 easily be adjusted to the study since Scannotation can rapidly generate new predictors.

115 The efficiency of Scannotation was demonstrated in 75 serum samples from Breton adolescents  
116 (Pélagie cohort)<sup>23</sup>. These annotations were compared to those of MS-DIAL<sup>11</sup>, which performs MS2-  
117 based annotations and was previously demonstrated to adequately and efficiently annotate low-  
118 abundant compounds in serum samples<sup>16</sup>. The use of these two tools allowed the annotation of 89  
119 compounds from the internal chemical exposome of Breton adolescents, and demonstrated that the  
120 use of MS1 predictors was relevant and complementary to an MS2-based approach in an exposomic  
121 application.

## 122 Experimental section

### 123 1. Biological samples

124 Serum samples (n=75) were obtained from 12-year-old male children from the PELAGIE cohort. This  
125 population-based mother-child cohort included 3,421 women from Brittany (France) enrolled during  
126 early pregnancy (before 19 weeks of gestation) between 2002 and 2006<sup>23</sup>.

### 127 2. Sample preparation

128 Samples were prepared using a dual sample preparation method, previously optimized to widen the  
129 visible chemical space<sup>5</sup>. Briefly, protein precipitation (PPT) was performed on all 75 samples using a  
130 4:1 (v:v) ratio of cold methanol to matrix. After centrifugation, half of the supernatants was  
131 resuspended in 40  $\mu$ L of injection phase (i.e., 90:10 (v:v) ultrapure water to acetonitrile ratio), while  
132 the second half was further cleaned up using a Phree (Phenomenex) protein and phospholipid removal

133 plate. Extracts were then evaporated to dryness under vacuum, and recovered in 40  $\mu$ L of injection  
134 phase.

### 135 3. Data acquisition

136 Samples were analyzed using an AB SCIEX X500R QTOF-MS (Resolution > 30,000) interfaced with an  
137 AB SCIEX ExionLC AD UHPLC system. An Acquity UPLC HSS T3 C18 column (1.8  $\mu$ m, 1.0  $\times$  150 mm,  
138 Waters Corporation) maintained at 40°C was used to perform reverse phase chromatographic  
139 separation as described in Chaker *et al.* (2021)<sup>16</sup>. Samples were analyzed in full scan experiments in  
140 electrospray ionization (ESI) (–) and (+) modes. MS2 fragmentation data for chemical elucidation was  
141 obtained by analysis of selected samples in data-dependent acquisition experiments. Additional  
142 information regarding the chromatographic separation and ESI source parameters are available in the  
143 Supporting Information.

### 144 4. MS1 data pre-processing

145 Raw data acquired in full scan mode were pre-processed using vendor software MarkerView v.1.3 (AB  
146 SCIEX) and MSDIAL4 with parametrization previously optimized to detect low-abundant chemicals in  
147 blood serum samples<sup>16</sup>. Briefly, critical parameters values were set as: noise threshold of 10, mass  
148 tolerance of 10 ppm, retention time (Rt) tolerance of 1 min, average peak width of 12s, no isotope  
149 filtering. Feature areas in solvent blanks were systematically subtracted from corresponding feature  
150 areas in all samples.

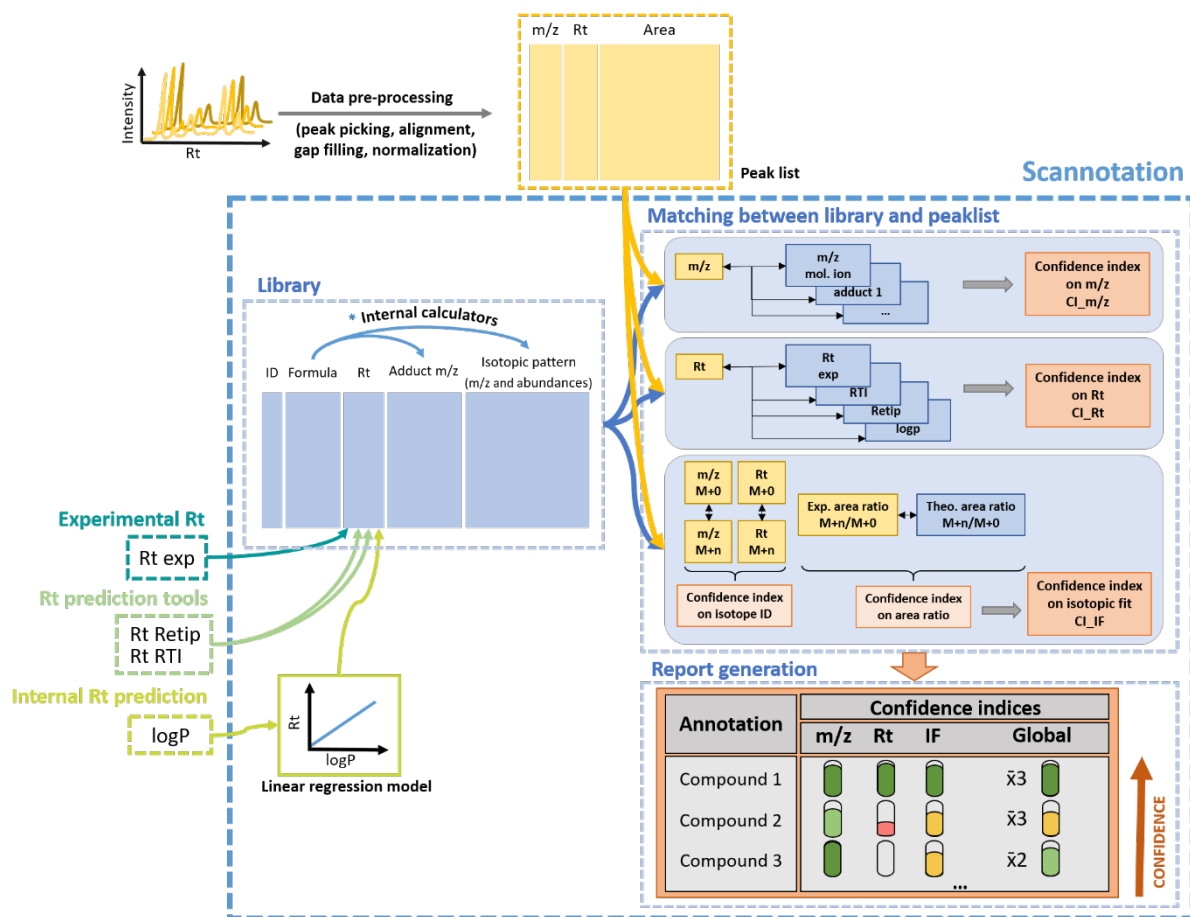
### 151 5. Suspect screening

#### 152 5.1. MS1-based suspect screening: Scannotation

153 Scannotation is a two-part Python program aimed at providing a prioritized list of scored pre-  
154 annotations available at [https://github.com/scannotation/Scannotation\\_software](https://github.com/scannotation/Scannotation_software). It was developed  
155 on Windows and tested on both Windows and on a Mac computer in a Windows virtual machine.  
156 Scannotation can process peak lists in .csv generated by any pre-processing software. The score is  
157 established by comparing the proximity between experimental features and suspects based on three

158 MS1-based chemical predictors (i.e., m/z, Rt, and isotopic pattern). Its operating principle is described  
 159 in Figure 1.

160



161

162 Fig. 1 - Scannotation annotation workflow relies on comparing a user-built library to a list of features.  
 163 Compounds' identifiers (name and SMILES), molecular formula, experimental, predicted retention time  
 164 (Rt) and logP values (when not listed by the user), allowing the software to compute molecular ion and  
 165 adduct masses, theoretical isotopic pattern, and a logP-predicted Rt. Any other predicted Rt can also  
 166 be added to the library and will be used by the software. The software then successively compares  
 167 experimental features to the suspect library data for three predictors: m/z, Rt and isotopic fit. Scores  
 168 are generated for each predictor, and combined into a global score.

169 The first module of this SS tool (the library) generates theoretical predictor data for a list of suspects,  
 170 based on user-provided data, as illustrated in Fig. 1 by the "Internal calculators" arrows in the "Library"



171 box. More precisely, molecular formulas (i.e. atomic contents) and atomic masses (from MIDAS  
172 <https://www.ncbi.nlm.nih.gov/CBBresearch/Yu/midas/index.html>) are used to compute m/z values  
173 for molecular ions and various common adducts in both ESI (-) and (+) modes (i.e., [M-H]<sup>-</sup>, [M-H<sub>2</sub>O-H]<sup>-</sup>  
174 , [M+Cl]<sup>-</sup>, [M+FA-H]<sup>-</sup> for negatively charged ions, and ([M+H]<sup>+</sup>, [M+NH<sub>4</sub>]<sup>+</sup>, [M+Na]<sup>+</sup>, [M+K]<sup>+</sup> for positively  
175 charged ones). Moreover, these m/z values can be computed for common metabolites (i.e.,  
176 glucuronide [+176.0321 m/z], sulfate [+79.9568 m/z], mercapturate [+145.0198 m/z] and cysteine  
177 conjugates [+103.0092 m/z]) of the library compounds by adding the corresponding expected mass  
178 increments<sup>24</sup>. The sulfate conjugate [+SO<sub>3</sub>] is added if O is present in the molecular formula and the  
179 glucuronide [+C<sub>6</sub>H<sub>8</sub>O<sub>6</sub>] conjugate is added if either N or O are present in the molecular formula.  
180 Molecular formulas are also used to generate theoretical isotopologues' m/z and relative abundance  
181 values by internally implementing the previously described MIDAS polynomial algorithm<sup>25</sup>. Users can  
182 provide experimental Rt values (if available), as well as predicted values from existing algorithms<sup>18,19</sup>.  
183 Moreover, octanol-water partition coefficient (logP) values provided by the user may be used to  
184 predict a logP-based Rt, under the condition that a sufficient number of compounds (i.e., over 20) have  
185 both an experimental Rt and a logP value. More details on this model are available in the SI. For library  
186 compounds without logP information, as well as for Scannotation-predicted metabolites, logP values  
187 are predicted through atomic composition-based model. This multiple linear regression model,  
188 adapted from Mannhold et al. (2009)<sup>26</sup>, was built on close to 1,800 compounds with available  
189 experimental logP values ranging from -5.08 to 9.29 (list available in SI Table A1). This model was  
190 trained by randomly choosing 80 % of the dataset, then validated on the remaining 20%. The process  
191 was repeated 1,000 times. Median coefficients were used to determine the contribution of each atom.  
192 Model training determined that the most sensitive parameters to predict logP from atomic  
193 composition were the numbers of atoms of carbon, halogens, and sulfur (logP increasing with number  
194 increase), as well as nitrogen and oxygen (logP decreasing with number increase). This slightly more  
195 complex model (i.e., the original model only considers the number of carbons and the number of  
196 heteroatoms) performed better than the original model, with a root mean square error (RMSE) of 1.32

197 compared to 2.04. Similarly, the Bayesian Information Criteria (BIC) value was lower (i.e., better  
198 adjustment) for the more complex model, with a value of 766 compared to 4553 for the original model.  
199 Details on this model are available in SI Table A1.

200 A human blood library, including 6 000 compounds, was constructed using data from the literature<sup>27</sup>  
201 and online databases such as the Blood Exposome Database<sup>28</sup>, Human Metabolome Database<sup>29</sup>,  
202 Exposome Explorer<sup>30</sup>, FoodBall<sup>31</sup>, and the NORMAN Network Suspect List Exchange<sup>32</sup>. This library is  
203 mainly comprised of food intake biomarkers, pesticides (and their metabolites), industrial pollutants,  
204 cosmetic ingredients, and pharmaceuticals/drugs (and their metabolites). Scannotation's library  
205 module was used to add potential metabolites, predict logP-based  $R_t$ , and compute theoretical  $m/z$   
206 values and isotopic patterns. Custom-made libraries can also be used.

207 Scannotation's second module performs the matching between the library and the peaklist. It  
208 computes confidence indices (CI), which score the proximity between suspects and features for each  
209 of the three predictors between 0 (no match) and 1 (perfect match). These scores are built as functions  
210 of the absolute difference between theoretical and experimental values of each predictor (i.e.,  $m/z$ ,  
211  $R_t$ , mass difference between isotopologues and area ratio between isotopologues), and of a tolerance  
212 associated with each of these parameters<sup>16</sup>. These tolerances, called  $\Delta$ , are either determined based  
213 on instrumental uncertainty (i.e., for  $m/z$  and  $m/z$  differences) or analytical variability (i.e., for  $R_t$  and  
214 area ratios). Additionally, for  $m/z$  and  $R_t$ , values of associated  $\Delta$  (i.e.,  $\Delta_{m/z}$  and  $\Delta_{R_t}$ ) vary depending on  
215 the chemical descriptor value. Indeed, higher  $m/z$  deviations (in ppm) are expected for lower  $m/z$   
216 values.  $\Delta_{m/z}$  is set to 15 ppm for masses strictly lower than 200 Da, and 10 ppm for masses over 200  
217 Da. Moreover, Scannotation allows the use of experimental  $R_t$  as well as  $R_t$  predicted through various  
218 tools, which all present different accuracies.  $R_t$  is also expected to vary in a non-linear manner  
219 throughout the chromatogram. Therefore, the used  $\Delta_{R_t}$  value depends on both the type of  $R_t$  (i.e.,  
220 experimental, predicted through RTI, predicted through Retip, or predicted through logP) and the  
221 value of the tentative annotation's theoretical  $R_t$ . In particular,  $\Delta_{R_t}$  values for experimental  $R_t$  can be

222 computed by Scannotation from a user-provided file which includes repeated (at least four) Rt  
223 measurements for any list of compounds (e.g., internal standards). All  $\Delta$  values used for this work are  
224 available in SI Table A2.

225 CI values for each predictor are combined into an overall score, named the global CI (Clg), as a mean  
226 of CI values to efficiently rank the pre-annotations generated by Scannotation. Moreover, the CI of  
227 annotated compounds, for which metabolites and/or neutral losses are detected, are associated to a  
228 letter from a to d (details available in SI Table A3). This leads to a better ranking in the result table and  
229 highlights the additional confidence in the tentative identification. The calculation of CI values and  
230 their adequation for suspect screening were previously described by Chaker et al. (2021)<sup>16</sup>.

231 5.2. Scannotation was used to perform suspect screening on the feature tables obtained  
232 from the aforementioned MS1 data pre-processing step (see paragraph 4. MS1 data  
233 pre-processing”). Manual curation on MS1 and MS2 data (when available) was  
234 performed to confirm pre-annotations. This process included verification of absence  
235 or significantly lower presence in the blank (i.e., area ratio sample/blank > 10), signal-  
236 to-noise ratio > 10, visual examination of peak shape, and verification of accuracy of  
237 m/z, Rt, and isotopic pattern. When available, experimental MS2 spectra was  
238 compared to MS2 spectra in databases (e.g. HMDB, MassBank<sup>29,33</sup>) or in-silico  
239 predictions (e.g., CFM-ID, MetFrag<sup>34,35</sup>). When a standard was available, experimental  
240 parameters of features were compared to those of the standard. This manual curation  
241 was performed on all compounds presenting a Clg value calculated with three  
242 predictors, as well as all compounds presenting a Clg value over 70% if only two  
243 predictors were available. MS2-based suspect screening: MS-DIAL

244 In addition to Scannotation’s screening, an MS2-based suspect screening approach was performed by  
245 processing raw data obtained from MS2 data dependent acquisitions with MS-DIAL (v.4.70)<sup>11</sup>. Critical  
246 parameters values were set as: minimum peak height of 10, mass tolerances of 0.0025 Da (10 ppm for  
247 a m/z of 250) in MS1 and MS2, Rt tolerance of 1 min, minimum peak width of 5 scans, and

248 consideration for Cl and Br elements enabled (for isotope recognition). Spectral MSP databases “All  
249 Public” available online (experimental spectra for 12,879 compounds in ESI (–) mode and 13,303  
250 compounds in ESI (+) mode) were used for suspect screening. Manual curation (i.e., manual  
251 verifications of peak shape, m/z, Rt, isotopic pattern and MS2 spectra if available) on MS1 and MS2  
252 data was performed to confirm annotations suggested by the software.

## 253 6. Quality assurance and quality control procedures

254 Several quality assurance and quality control procedures were implemented, including the systematic  
255 use of instrumental and extraction blanks, composite quality control samples, and ISTD. Details are  
256 available in SI.

# 257 Results and discussion

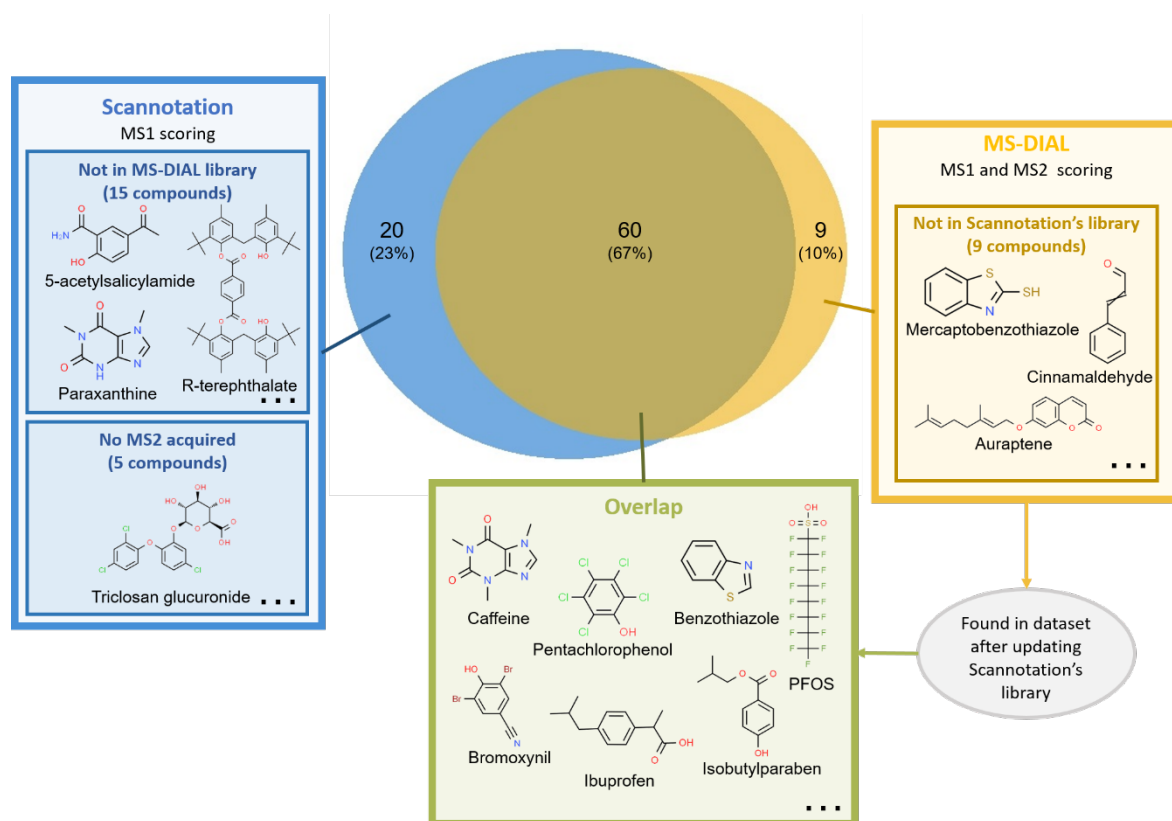
## 258 1. Comparison of MS1 and MS2-based suspect screening workflows

259 In total, 75 serum samples from a mother-child cohort were prepared with two different sample  
260 preparations<sup>5</sup> and injected on a UHPLC-ESI-QTOF. Quality control criteria were met (results presented  
261 in Supporting information table A6). Two suspect screening approaches were then implemented and  
262 compared to annotate the HRMS pre-treated datasets (i.e., approximately 50,000 and 93,000 features  
263 for ESI (–) and (+) modes respectively). Both software tools provided comparable numbers of raw pre-  
264 annotations. In total, Scannotation provided close to 33,000 pre-annotations while MS-DIAL provided  
265 close to 36,000. However, establishing a cut-off score of 70% for both tools (i.e., selecting pre-  
266 annotations with Scannotation Clg > 0.7 and MS-DIAL identification score > 70%) allowed reducing the  
267 number of tentative annotations by 78% using Scannotation while it only resulted in an 8% reduction  
268 for MS-DIAL. Reaching a similar level of prioritization for MS-DIAL can only be achieved by choosing  
269 perfect matches (i.e., scores of 100%). The discrepancy in scores is inherent to the algorithms used by  
270 both tools. Firstly, MS-DIAL incorporates a score based on MS2 spectral similarity, and Scannotation is  
271 currently only MS1-based. Secondly, MS-DIAL’s Rt and m/z score calculations are significantly less

272 restrictive than Scannotation's, as an error of half the tolerance margin will provide an MS-DIAL score  
273 of 88%<sup>11</sup> and a Scannotation CI of 50%. This is because MS-DIAL considers the distribution of errors  
274 (presumed gaussian) whereas Scannotation's CI formula is linear to represent the gap between  
275 theoretical and experimental values as directly and accurately as possible. Thirdly, MS-DIAL's isotopic  
276 fit score is based on five ratios between six isotopologues, instead of one ratio between two  
277 isotopologues for Scannotation, which is arguably less stringent. These last two points clarify why  
278 Scannotation's CI values decrease more rapidly compared to MS-DIALs', thus explaining a more  
279 pronounced discriminating effect of scoring. As there are often limited resources (particularly time)  
280 that can be dedicated to manual curation, an efficient prioritization tool is essential to focus on  
281 plausible tentative annotations first. For this work, manual verifications were conducted on 8000 pre-  
282 annotations from Scannotation and 3000 pre-annotations from MS-DIAL, which approximately  
283 required 35h and 15h of work, respectively (i.e., comparable rates).

284 After manual curation (as detailed in the Experimental section, paragraphs 5.1. and 5.2.), 89 prioritized  
285 annotations were proposed with a confidence level ranging from 1 to 4 according to Schymanski's  
286 scale<sup>17</sup>, with an overlap of 60 compounds for both software tools, as shown in Figure 2. The detailed  
287 prioritization process is available in SI Table A4 and the list of annotated compounds is available in SI  
288 Table A5. More specifically, of the 89 annotated compounds, 13 were attributed a confidence level of  
289 1 (15%), 60 were attributed a level 2a (67%), 6 were attributed a level 2b (7%), and 10 were attributed  
290 a level 4 (11%) according to Schymanski's scale<sup>17</sup>. Despite the heterogeneity in the levels of confidence  
291 reported here (i.e., between 1 and 4 on a scale of 1 to 5), these 89 annotations are supported by  
292 multiple orthogonal elements of proof. For instance, the level 4 does not accurately reflect the  
293 confidence that can be put in these annotations since they were suggested based on a combination of  
294 several MS1 predictors, including a match with Rt from a standard (when available), identification of  
295 neutral loss, or the presence of additional metabolite from the same parent compound. This is the case  
296 for the herbicide bromoxynil, which could not be fragmented in the samples during MS2 acquisitions,  
297 but for which the match between the experimental Rt of the annotated ion and of the standard's was

298 scored at 88% by Scannotation (Clg = 0.84), or for the triclosan glucuronide and sulfate as discussed  
 299 below. It should be noted that features for these 89 compounds were correctly peak picked by both  
 300 data processing software tools (i.e. checked after manual observations); observed differences in  
 301 compound lists can therefore be attributed to the annotation process.



302  
 303 *Fig. 2 - Overview of the data generated by two suspect screening tools, based on either MS1 or both*  
 304 *MS1 and MS2 predictors (Scannotation and MS-DIAL respectively).*

305 Scannotation was able to annotate 90% of these 89 compounds, including 22% that were exclusively  
 306 identified through this software. Among the compounds only annotated by Scannotation, five (e.g.,  
 307 triclosan sulfate and glucuronide or 2-chlorophenol) were either too low abundant or not fragmented,  
 308 and did not present reliable MS2 spectra. Since MS-DIAL mainly relies on matching MS2 spectra, these  
 309 compounds could therefore be missed and/or not prioritized by this solution. Even without MS2 data,  
 310 Scannotation was able to provide solid annotation for triclosan sulfate (Clg = 0.93) and glucuronide  
 311 (Clg = 0.92) based on Rt match with pure standards, specific isotopic profiles match (3 atoms of Cl),

312 identification of neutral loss (i.e., triclosan glucuronide and sulfate conjugates) and the presence of 2  
313 phase II conjugates from the same parent with a coherent elution order. These elements of proof are  
314 presented in SI Fig.S1. Moreover, for the 15 remaining compounds not annotated by MS-DIAL, there  
315 was no adequate reference spectra in the library. No adequate reference spectra may either mean no  
316 spectra at all, as for 5-acetylsalicylamide, or spectra acquired in the other ionization mode, or with  
317 different MS2 acquisition modes and/or collision energies, leading to very poor spectra matching (and  
318 thus elimination from the pool of suggested annotations), as for paraxanthine.

319 MS-DIAL annotated 77% of the 89 compounds, including 10% of them solely identified through this  
320 software. This was explained by the absence of these compounds in Scannotation's initial library.  
321 Subsequent addition of these 9 compounds to the library (i.e., molecular formula, identifiers, predicted  
322 Rt and logP values when available) allowed their annotation by MS1-based predictors, resulting in a  
323 mean Clg of 0.86 (all above 0.79). Many other highly scored putative annotations were suggested by  
324 MS-DIAL; however, they were mainly endogenous compounds. This was expected since a large  
325 proportion of the available MS2 spectra in MS-DIAL's databases are of endogenous compounds that  
326 were out of the scope of this work.

327 The Schymanski's scale is an undeniably useful tool to efficiently communicate confidence of  
328 annotations in a harmonized and easy-to-read way. However, in the context of exposomics  
329 applications with low-abundant compounds (particularly in complex matrices such as biological  
330 matrices), using MS2 predictors may face various critical obstacles. For instance, it is fairly frequent  
331 that there is either no MS2 data acquisition (i.e., acquisition not triggered due to low abundance) or  
332 no reliable MS2 spectra (i.e., acquisition triggered, but collision energy too low or compound too  
333 diluted to produce useful fragmentation data) for ions of interest. This lack of reliable MS2 data leads  
334 to the attribution of low confidence levels to compounds that may be relevant in a public health  
335 context due to their high toxicity and/or large prevalence in the population. When these issues arise,  
336 the use of other predictors based on MS1 could be relevant to differentiate between level 4

337 compounds ranging from bare formula matches to annotations supported by other elements such as  
338 coherent Rt and presence of related biotransformation products, and therefore to efficiently prioritize  
339 the massive number of suggested annotations for manual curation. This can also be helpful to create  
340 inclusion lists of ions of interest for possible further MS2 acquisitions.

341

342 In the end, we demonstrate that Scannotation is complementary to existing MS2-based suspect  
343 screening since both Scannotation and MS-DIAL were able to annotate a large number of exogenous  
344 chemicals in these 75 serum samples with a large majority (67%) being detected by both. We also  
345 demonstrate the importance of having a relevant library dedicated to HRMS-based exposomics  
346 studies.

## 347 2. Description of chemical exposure profiles in the Pélagie cohort

### 348 2.1. Environmental chemical exposures in the cohort

349 The data collected on the 75 samples injected in both ESI (-) and ESI (+) modes allowed annotating 89  
350 compounds from the internal chemical exposome. In these chemicals, several “everyday pollutants”  
351 commonly found in human biological samples (at varying detection frequencies and levels) were  
352 annotated, such as phthalates, paraben derivatives, pesticides, and per- and poly-fluorinated alkyl  
353 substances<sup>36,37</sup>. Interestingly, bromoxynil, a well-characterized herbicide and established endocrine  
354 disruptor<sup>38</sup> found in 64% of samples, had previously been reported in the urine of 22% of pregnant  
355 women from the same cohort (i.e., during the prenatal period of these 75 teenagers)<sup>39</sup>. This suggests  
356 that some individuals have been chronically exposed (or at repeated occasions) to this compound,  
357 including during this specific period of vulnerability. Moreover, lidocaine (local anesthetic) was  
358 surprisingly found in more than 90% of samples. After further investigation, it was determined that  
359 anesthetic patches containing lidocaine were used 1hour prior to the blood draw, thus confirming the  
360 relevance of the presented workflow to detect environmental exposures.



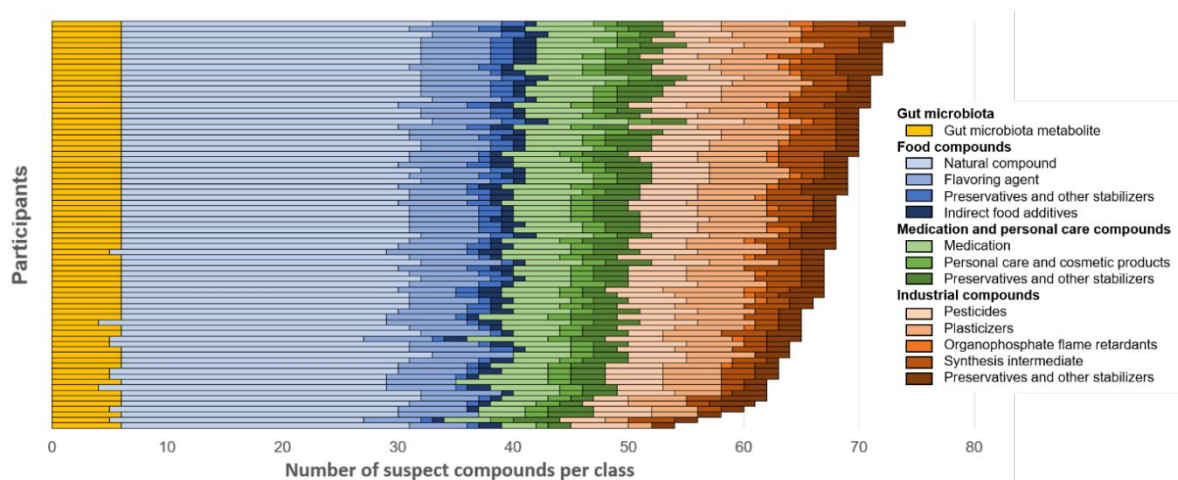
361 These annotations of commonly detected chemicals also demonstrate the sensitivity and the relevance  
362 of this HRMS-based profiling method and SS strategy to identify chemicals usually detected at trace  
363 levels using conventional targeted MS2 method. Levels of insecticide metabolite fipronil sulfone  
364 (detected in 35% of samples) were previously reported in human blood from the general population  
365 at concentrations varying between 0.1 and 4 ng/mL<sup>40</sup>. Similarly, bromoxynil levels in plasma samples  
366 from teenagers residing in rural areas were previously reported from trace levels to 140 ng/mL<sup>41</sup>.  
367 Likewise, previously reported levels of perfluorooctanesulfonic acid (PFOS) and  
368 perfluorohexanesulfonic acid (PFHxS) (detection frequencies of 100% and 95% respectively) in the  
369 German general population from 2009 to 2019 ranged from 0.9-9.9 ng/mL and from 0-4.6 ng/mL  
370 respectively<sup>42</sup>. As the MS1-based suspect screening approach allowed the annotation of all of these  
371 compounds, it appears that exogenous chemicals present at low levels in complex matrices may still  
372 be identified with a suspect screening strategy using Scannotation.

373 To provide a global overview of the exposure profiles, the 89 detected compounds were classified in  
374 four general categories: gut microbiota metabolites (including those from exogenous dietary  
375 substrates), food compounds (natural and artificial), medication and personal care compounds (e.g.,  
376 pain management, surfactants), and industrial compounds (e.g., synthesis intermediates used in the  
377 manufacturing of dyes, pesticides or plasticizers). Most of the 89 annotated compounds have multiple  
378 sources, such as ferulic acid, which is both a natural compound and a food preservative, or  
379 di(ethylhexyl)phthalate, which is a plasticizer present in many plastic products including flooring and  
380 upholstery, everyday household items, and food packaging. However, for illustrating purposes,  
381 primary uses (according to production volume) were considered for the proposed classifications. Gut  
382 microbiota, food compounds, medication and personal care compounds, and industrial compounds  
383 represented, respectively, 7%, 46%, 16%, and 31% of the overall number of identified compounds.  
384 Taken together, food compounds and medical and personal care products represented close to two-  
385 thirds (62%) of annotated compounds. The first category includes natural compounds (e.g. caffeine,  
386 piperine), flavoring agents (e.g. sweeteners aspartame and sucralose), and food contact chemicals (e.g.

387 1,3,5-tris(2,2-dimethylpropionylamino)benzene), while the second includes medication and  
388 metabolites (e.g. acetaminophen phase II metabolites), additives from personal care products (e.g.  
389 shampoo and shower gel surfactant cocamidopropyl betaine), and preservatives (e.g. isobutyl- and  
390 isopropylparaben). Detecting many compounds from these categories was expected, as their  
391 concentrations in blood can be up to  $10^6$  times higher than some industrial pollutants (such as  
392 pesticides or plasticizers)<sup>15</sup>, thus leading to easier detection and characterization (in particular, easier  
393 acquisition of MS2 spectra).

## 394 2.2. Inter-individual chemical exposure variability

395 In order to study the inter-individual variability that could be observed in terms of chemical exposure  
396 profiles, the presence or absence of each annotated compound in analyzed samples was assessed. The  
397 number of annotated compounds from each category and sub-category is represented in Figure 3.



398  
399 *Fig. 3 - Detection of suspect compounds in each participant. Compounds were classified in four main*  
400 *categories: gut microbiota metabolites, food compounds, medication and personal care compounds,*  
401 *and industrial compounds.*

402 It should be noted that most compounds have multiple possible uses. The classification presented here  
403 is based on primary use according to PubChem<sup>43</sup>. Secondary uses for each compound are also  
404 presented in the SI, table A4.

405 A median of 66 compounds were detected per sample. At the scale of the four main categories, food  
406 compounds are the most represented with a median of 52% (CV=4%) of annotated compounds per  
407 individual. The less represented category was gut microbiota-related chemicals, with a median of 9%  
408 of annotated compounds per individual (CV=8%). Industrial compounds represented a median of a  
409 quarter of all annotated compounds per individual, with a CV of 10%.

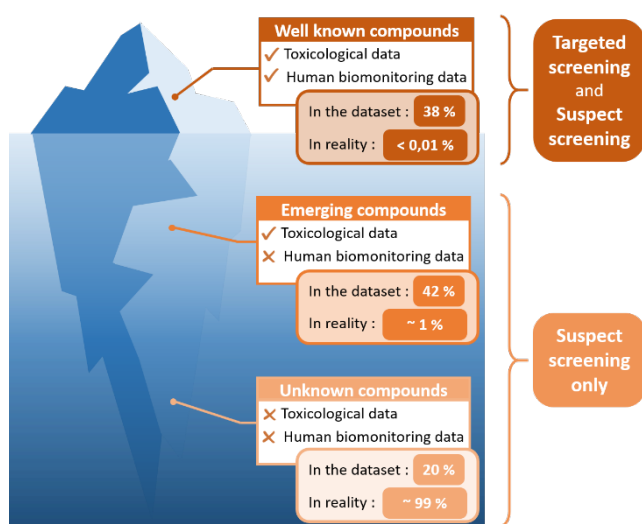
410 At the scale of subcategories, natural food compounds regrouped the largest proportion of annotated  
411 compounds (median of 38% of annotated compounds per individual). On the other hand, the less  
412 represented chemical class is organophosphate flame retardants (e.g., tris(2-butoxyethyl)phosphate)  
413 (median of 0%, average of 1%). Exposure profiles (i.e., combining exposures to different compounds)  
414 may be indicative of an individual's lifestyle. For example, 12 participants presented a co-exposure to  
415 acesulfame, aspartame and sucralose (all artificial sweeteners), which may indicate an overall more  
416 processed diet. It should however be noted that the high number of annotated compounds in  
417 comparison to the number of participants significantly limits the statistical power necessary to  
418 establish such profiles. It is also impossible to identify the source of these exposures with certainty, as  
419 this would require analyses of environmental samples (i.e., indoor air, items from diet, personal care  
420 products, etc.) and/or data obtained from questionnaires in addition to the data acquired from these  
421 biological samples.

422 Our results also highlight the capacity of suspect screening approaches to discover new relevant  
423 biomarkers of exposure to known toxicants. Bromoxynil metabolite 3,5-dibromo-4-hydroxybenzoic  
424 acid was detected in 97% of samples with areas 3 to 8 times higher than areas for bromoxynil (detected  
425 in 64% of samples). However, despite its apparent easier detection, this metabolite was not reported  
426 (as a biomarker of bromoxynil exposure or otherwise) in human biomonitoring studies in blood or  
427 urine before. These new biomarkers may subsequently be used either to retrospectively assess  
428 exposure if non-targeted data was acquired or may be included in new lists of targeted compounds of  
429 interest.

430 These exploratory approaches allow to significantly increase our knowledge of the chemical exposome,  
431 and particularly to investigate emerging and unknown compounds (without toxicological and/or  
432 human biomonitoring data available). Detection frequencies for all compounds were computed.  
433 Detailed results are available in SI Table A5. Overall, 66% of all annotated compounds were found in  
434 more than 80% of samples. Of these ubiquitous compounds, 17% (10 compounds) are not documented  
435 in the extensive NORMAN Network's SUSDAT list, which combines close to 110,000 structures from 98  
436 suspect lists provided by the scientific community<sup>32</sup>. These compounds include 3 phase II metabolites  
437 (sulfated forms), highlighting the need to include biotransformation products (known or predicted) in  
438 suspect lists. Furthermore, 20% (11 compounds) of these ubiquitous compounds have no reported  
439 toxicological data according to the CompTox chemistry dashboard<sup>44</sup> (Figure 4). For instance, a  
440 phthalate found in 92% of samples (i.e., Bis(2-(tert-butyl)-6-(3-(tert-butyl)-2-hydroxy-5-methylbenzyl)-  
441 4-methylphenyl) terephthalate) and only annotated with Scannotation has no reported toxicological  
442 data, even though phthalates are known endocrine and metabolic disruptors<sup>45</sup>. This again  
443 demonstrates that MS1-based suspect screening approaches can be of great use to uncover previously  
444 unknown or poorly known exposures to chemicals of potential concern. These compounds, while  
445 largely predominant in number in reality (more than 110 million compounds registered on PubChem<sup>43</sup>  
446 compared to less than a thousand biomonitored chemical species in some of the biggest human  
447 biomonitoring initiatives<sup>46,47</sup>), are not investigated by targeted approaches. The use of exploratory  
448 approaches focused on low-abundant chemicals, in terms of data acquisition (i.e., non-targeted  
449 analyses), data processing (i.e., adequate peak picking parameters), and annotation (i.e., annotation  
450 even without MS2 acquisitions), may help starting to bridge this gap in knowledge by uncovering  
451 emerging and unknown compounds in the studied population (Figure 4).

452

453



454

455 *Fig. 4 – Contribution of suspect screening approaches to the number of exogenous compounds detected*  
 456 *in human serum samples from the Pélagie cohort. Proportion of compounds in reality are based on the*  
 457 *total number of Pubchem entries (>110M), the number of CompTox chemistry dashboard entries (~1M),*  
 458 *and the number of chemicals usually biomonitored by the largest human biomonitoring initiatives to*  
 459 *date (<1000).*

## 460 Conclusion

461 Here, we demonstrate the efficiency of Scannotation to investigate the internal chemical exposome of  
 462 75 Breton adolescents using a MS1-based strategy to score the proximity between features obtained  
 463 from any pre-processing software and suspects, and providing an easy-to-read indicator of each pre-  
 464 annotation's reliability. This strategy was complementary to the one used by MS-DIAL based on MS2,  
 465 and Scannotation provided thousands of scored pre-annotations that led to the annotation of 89  
 466 environmental chemical compounds (confirmed with manual curation) with various uses including  
 467 pesticides, medication, preservatives and synthesis intermediates. It has also uncovered the relevance  
 468 of Scannotation's SS strategy to identify low-abundant and/or less documented compounds (not  
 469 annotated by MS-DIAL) or to detect new metabolites of known contaminants. We demonstrate that  
 470 this approach will help bridging a gap in knowledge by documenting the prevalence of some emerging  
 471 and unknown compounds in a given population. The chemical fingerprints acquired, and the list of

472 annotated compounds could be further used in association to contextual data from the cohort, to  
473 further describe the chemical exposome of the Breton teenage population, and to investigate  
474 determinants of these exposures.

## 475 Acknowledgment

476 This research was supported by a research chair of excellence (2016-52/Idex Université of Sorbonne  
477 Paris Cité) awarded to AD and a grant from the Brittany council (SAD). J.C. was funded by the Réseau  
478 Doctoral en Santé Publique. J.C., S.L. and A.D. acknowledge the research infrastructure France  
479 Exposome.

## 480 Supporting information

481 The Supporting Information is available free of charge at <http://pubs.acs.org>.

482 Description of sample preparation method, data acquisition parameters, quality control procedures,  
483 and elements of proof supporting the annotation of triclosan glucuronide (.docx)

484 Predictive Rt model, parameter tolerance values, codification of global confidence index, and list of  
485 annotations performed on cohort samples (.xlsx)

## 486 References

- 487 1. Fuller R, Landrigan PJ, Balakrishnan K, et al. Pollution and health: a progress update. *The Lancet*  
488 *Planetary Health* 2022;6(6):e535–e547; doi: 10.1016/S2542-5196(22)00090-0.
- 489 2. Wild CP. Complementing the Genome with an “Exposome”: The Outstanding Challenge of  
490 Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology,*  
491 *Biomarkers & Prevention* 2005;14(8):1847–1850; doi: 10.1158/1055-9965.EPI-05-0456.
- 492 3. Tkalec Ž, Codling G, Klánová J, et al. LC-HRMS based method for suspect/non-targeted screening  
493 for biomarkers of chemical exposure in human urine. *Chemosphere* 2022;300:134550; doi:  
494 10.1016/j.chemosphere.2022.134550.

- 495 4. Panagopoulos Abrahamsson D, Wang A, Jiang T, et al. A Comprehensive Non-targeted Analysis  
496 Study of the Prenatal Exposome. *Environ Sci Technol* 2021;55(15):10542–10557; doi:  
497 10.1021/acs.est.1c01010.
- 498 5. Chaker J, Kristensen DM, Halldorsson TI, et al. Comprehensive Evaluation of Blood Plasma and  
499 Serum Sample Preparations for HRMS-Based Chemical Exposomics: Overlaps and Specificities.  
500 *Anal Chem* 2022;94(2):866–874; doi: 10.1021/acs.analchem.1c03638.
- 501 6. Al-Salhi R, Monfort C, Bonvallet N, et al. Analytical strategies to profile the internal chemical  
502 exposome and the metabolome of human placenta. *Analytica Chimica Acta* 2022;1219:339983;  
503 doi: 10.1016/j.aca.2022.339983.
- 504 7. Pourchet M, Debrauwer L, Klanova J, et al. Suspect and non-targeted screening of chemicals of  
505 emerging concern for human biomonitoring, environmental health studies and support to risk  
506 assessment: From promises to challenges and harmonisation issues. *Environment International*  
507 2020;139:105545; doi: 10.1016/j.envint.2020.105545.
- 508 8. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of  
509 High-Resolution Metabolomics Data. *Anal Chem* 2017;89(2):1063–1067; doi:  
510 10.1021/acs.analchem.6b01214.
- 511 9. Lawson TN, Weber RJM, Jones MR, et al. msPurity: Automated Evaluation of Precursor Ion Purity  
512 for Mass Spectrometry-Based Fragmentation in Metabolomics. *Anal Chem* 2017;89(4):2432–  
513 2439; doi: 10.1021/acs.analchem.6b04358.
- 514 10. Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: Modular framework for processing,  
515 visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*  
516 2010;11(1):395; doi: 10.1186/1471-2105-11-395.
- 517 11. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: data-independent MS/MS deconvolution for  
518 comprehensive metabolome analysis. *Nat Methods* 2015;12(6):523–526; doi:  
519 10.1038/nmeth.3393.
- 520 12. Helmus R, ter Laak TL, van Wezel AP, et al. patRoom: open source software platform for  
521 environmental mass spectrometry based non-target screening. *Journal of Cheminformatics*  
522 2021;13(1):1; doi: 10.1186/s13321-020-00477-w.
- 523 13. Kuhl C, Tautenhahn R, Böttcher C, et al. CAMERA: An integrated strategy for compound spectra  
524 extraction and annotation of LC/MS data sets. *Anal Chem* 2012;84(1):283–289; doi:  
525 10.1021/ac202450g.
- 526 14. Misra BB. New software tools, databases, and resources in metabolomics: updates from 2020.  
527 *Metabolomics* 2021;17(5):49; doi: 10.1007/s11306-021-01796-1.
- 528 15. David A, Chaker J, Price EJ, et al. Towards a comprehensive characterisation of the human  
529 internal chemical exposome: Challenges and perspectives. *Environment International*  
530 2021;156:106630; doi: 10.1016/j.envint.2021.106630.
- 531 16. Chaker J, Gilles E, Léger T, et al. From Metabolomics to HRMS-Based Exposomics: Adapting Peak  
532 Picking and Developing Scoring for MS1 Suspect Screening. *Analytical Chemistry*  
533 2021;93(3):1792–1800; doi: 10.1021/acs.analchem.0c04660.

- 534 17. Schymanski EL, Jeon J, Gulde R, et al. Identifying Small Molecules via High Resolution Mass  
535 Spectrometry: Communicating Confidence. *Environ Sci Technol* 2014;48(4):2097–2098; doi:  
536 10.1021/es5002105.
- 537 18. Bonini P, Kind T, Tsugawa H, et al. Retip: Retention Time Prediction for Compound Annotation in  
538 Untargeted Metabolomics. *Anal Chem* 2020;92(11):7515–7522; doi:  
539 10.1021/acs.analchem.9b05765.
- 540 19. Aalizadeh R, Thomaidis NS, Bletsou AA, et al. Quantitative Structure–Retention Relationship  
541 Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging  
542 Contaminants in Environmental Samples. *J Chem Inf Model* 2016;56(7):1384–1398; doi:  
543 10.1021/acs.jcim.5b00752.
- 544 20. Stanstrup J, Neumann S, Vrhovšek U. PredRet: Prediction of Retention Time by Direct Mapping  
545 between Multiple Chromatographic Systems. *Anal Chem* 2015;87(18):9421–9428; doi:  
546 10.1021/acs.analchem.5b02287.
- 547 21. Wishart DS, Tian S, Allen D, et al. BioTransformer 3.0—a web server for accurately predicting  
548 metabolic transformation products. *Nucleic Acids Research* 2022;50(W1):W115–W123; doi:  
549 10.1093/nar/gkac313.
- 550 22. Smith CA, Want EJ, O’Maille G, et al. XCMS: Processing Mass Spectrometry Data for Metabolite  
551 Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal Chem*  
552 2006;78(3):779–787; doi: 10.1021/ac051437y.
- 553 23. Garlantezec R, Monfort C, Rouget F, et al. Maternal occupational exposure to solvents and  
554 congenital malformations: a prospective study in the general population. *Occupational and*  
555 *Environmental Medicine* 2009;66(7):456–463; doi: 10.1136/oem.2008.041772.
- 556 24. Huber C, Nijssen R, Mol H, et al. A large scale multi-laboratory suspect screening of pesticide  
557 metabolites in human biomonitoring: From tentative annotations to verified occurrences.  
558 *Environment International* 2022;168:107452; doi: 10.1016/j.envint.2022.107452.
- 559 25. Alves G, Ogurtsov AY, Yu Y-K. Molecular Isotopic Distribution Analysis (MIDAs) with Adjustable  
560 Mass Accuracy. *J Am Soc Mass Spectrom* 2014;25(1):57–70; doi: 10.1007/s13361-013-0733-7.
- 561 26. Mannhold R, Poda GI, Ostermann C, et al. Calculation of molecular lipophilicity: State-of-the-art  
562 and comparison of log P methods on more than 96,000 compounds. *J Pharm Sci* 2009;98(3):861–  
563 893; doi: 10.1002/jps.21494.
- 564 27. Rappaport SM, Barupal DK, Wishart D, et al. The blood exposome and its role in discovering  
565 causes of disease. *Environ Health Perspect* 2014;122(8):769–774; doi: 10.1289/ehp.1308015.
- 566 28. Barupal DK, Fiehn O. Generating the Blood Exposome Database Using a Comprehensive Text  
567 Mining and Database Fusion Approach. *Environmental Health Perspectives* n.d.;127(9):097008;  
568 doi: 10.1289/EHP4713.
- 569 29. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018.  
570 *Nucleic Acids Research* 2018;46(D1):D608–D617; doi: 10.1093/nar/gkx1089.
- 571 30. Neveu V, Moussy A, Rouaix H, et al. Exposome-Explorer: a manually-curated database on  
572 biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res*  
573 2017;45(D1):D979–D984; doi: 10.1093/nar/gkw980.



- 574 31. Fiamoncini J, Weinert C, Dragsted LO, et al. The FoodBALL Online Resources to Support Discovery  
575 of Novel Dietary Biomarkers with Metabolomics. 2015.
- 576 32. Mohammed Taha H, Aalizadeh R, Alygizakis N, et al. The NORMAN Suspect List Exchange  
577 (NORMAN-SLE): facilitating European and worldwide collaboration on suspect screening in high  
578 resolution mass spectrometry. *Environmental Sciences Europe* 2022;34(1):104; doi:  
579 10.1186/s12302-022-00680-6.
- 580 33. Horai H, Arita M, Kanaya S, et al. MassBank: a public repository for sharing mass spectral data for  
581 life sciences. *Journal of Mass Spectrometry* 2010;45(7):703–714; doi: 10.1002/jms.1777.
- 582 34. Wang F, Allen D, Tian S, et al. CFM-ID 4.0 – a web server for accurate MS-based metabolite  
583 identification. *Nucleic Acids Research* 2022;50(W1):W165–W174; doi: 10.1093/nar/gkac383.
- 584 35. Ruttkies C, Schymanski EL, Wolf S, et al. MetFrag relaunched: incorporating strategies beyond in  
585 silico fragmentation. *Journal of Cheminformatics* 2016;8(1):3; doi: 10.1186/s13321-016-0115-9.
- 586 36. Richterová D, Govarts E, Fábelová L, et al. PFAS levels and determinants of variability in exposure  
587 in European teenagers – Results from the HBM4EU aligned studies (2014–2021). *International  
588 Journal of Hygiene and Environmental Health* 2023;247:114057; doi:  
589 10.1016/j.ijheh.2022.114057.
- 590 37. Hartmann C, Jamnik T, Weiss S, et al. Results of the Austrian Children’s Biomonitoring Survey  
591 2020 part A: Per- and polyfluorinated alkylated substances, bisphenols, parabens and other  
592 xenobiotics. *International Journal of Hygiene and Environmental Health* 2023;249:114123; doi:  
593 10.1016/j.ijheh.2023.114123.
- 594 38. Arena M, Auteri D, Barmaz S, et al. Peer review of the pesticide risk assessment of the active  
595 substance bromoxynil (variant evaluated bromoxynil octanoate). *EFSA J* 2017;15(6):e04790; doi:  
596 10.2903/j.efsa.2017.4790.
- 597 39. Bonvallot N, Jamin EL, Regnaut L, et al. Suspect screening and targeted analyses: Two  
598 complementary approaches to characterize human exposure to pesticides. *Science of The Total  
599 Environment* 2021;786:147499; doi: 10.1016/j.scitotenv.2021.147499.
- 600 40. McMahan RL, Strynar MJ, Dagnino S, et al. Identification of fipronil metabolites by time-of-flight  
601 mass spectrometry for application in a human exposure study. *Environ Int* 2015;78:16–23; doi:  
602 10.1016/j.envint.2015.01.016.
- 603 41. Semchuk K, McDuffie H, Senthilselvan A, et al. Body mass index and bromoxynil exposure in a  
604 sample of rural residents during spring herbicide application. *J Toxicol Environ Health A*  
605 2004;67(17):1321–1352; doi: 10.1080/15287390490471424.
- 606 42. Göckener B, Weber T, Rüdell H, et al. Human biomonitoring of per- and polyfluoroalkyl  
607 substances in German blood plasma samples from 1982 to 2019. *Environ Int* 2020;145:106123;  
608 doi: 10.1016/j.envint.2020.106123.
- 609 43. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. *Nucleic Acids Research*  
610 2023;51(D1):D1373–D1380; doi: 10.1093/nar/gkac956.
- 611 44. Williams AJ, Grulke CM, Edwards J, et al. The CompTox Chemistry Dashboard: a community data  
612 resource for environmental chemistry. *Journal of Cheminformatics* 2017;9(1):61; doi:  
613 10.1186/s13321-017-0247-6.

- 614 45. Hlisníková H, Petrovičová I, Kolena B, et al. Effects and Mechanisms of Phthalates' Action on  
615 Reproductive Processes and Reproductive Health: A Literature Review. *Int J Environ Res Public*  
616 *Health* 2020;17(18):6811; doi: 10.3390/ijerph17186811.
- 617 46. Vorkamp K, Castaño A, Antignac J-P, et al. Biomarkers, matrices and analytical methods targeting  
618 human exposure to chemicals selected for a European human biomonitoring initiative.  
619 *Environment International* 2021;146:106082; doi: 10.1016/j.envint.2020.106082.
- 620 47. CDC. Fourth National Report on Human Exposure to Environmental Chemicals. 2022.

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638 For Table of Contents Only

639

640

