

## Adding Open Spectral Data to MassBank and PubChem Using Open Source Tools to Support Non-Targeted Exposomics of Mixtures

Anjana Elapavalore<sup>1\*</sup>, Todor Kondić<sup>1</sup>, Randolph R. Singh<sup>1,2</sup>, Benjamin A. Shoemaker<sup>3</sup>,  
Paul A. Thiessen<sup>3</sup>, Jian Zhang<sup>3</sup>, Evan E. Bolton<sup>3</sup>, Emma L. Schymanski<sup>1\*</sup>

<sup>1</sup> Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367, Belvaux, Luxembourg. ORCIDs AE: [0000-0002-0295-6618](#); TK: [0000-0001-6662-4375](#); ELS: [0000-0001-6868-8145](#).

<sup>2</sup> IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer), Laboratoire Biogéochimie des Contaminants Organiques, Rue de l'Île d'Yeu, BP 21105, Nantes Cedex 3, 44311, France. ORCID RRS: [0000-0003-4500-3400](#).

<sup>3</sup> National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, MD, 20894, USA. ORCIDs BAS: [0000-0002-1356-285X](#); PAT: [0000-0002-1992-2086](#); JZ: [0000-0002-6192-4632](#); EEB: [0000-0002-5959-6190](#).

\* Correspondence: AE: [anjana.elapavalore@uni.lu](mailto:anjana.elapavalore@uni.lu) and ELS: [emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu)

### Supplementary Information

This section contains three additional figures (Figures S1-S3) and a table (Table S1) in the following pages to support the article contents.

## Config

### Import

The list of knowns. Required columns: ID, SMILES, Name and RT (the last two can be empty). Remember to quote SMILES and Name entries!

The list of unknowns. Required columns: ID, mz and RT (RT can be empty).

Set table. Required columns ID and set.

### Raw Files in mzML Format

Use this file table to assign adduct modes and tags to the data files.

	Files	mode	set	tag
1	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE15.mzML	mH	mix499_M	15
2	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE30.mzML	mH	mix499_M	30
3	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE45.mzML	mH	mix499_M	45
4	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE60.mzML	mH	mix499_M	60
5	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE75.mzML	mH	mix499_M	75
6	ix499_neg/mzML's/20200303_ENTACT_RP_mix499_neg_CE90.mzML	mH	mix499_M	90

Figure S1: ShinyScreen Input Configuration.

## Spectra Extraction and Automatic QA

### A Extract Spectra

Number of processes:

Precursor m/z error (coarse) (+/-):  [Da]

Precursor m/z error (fine) (+/-):  [ppm]

EIC m/z error (+/-):  [Da]

Retention time tolerance (+/-):  [min]

Select set(s).

### Processed Sets

set	extracted	qa	ms2
mix499_M	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

### B Automatic Quality Control

Intensity threshold (MS1):

Intensity threshold (MS2):

Signal-to-noise ratio:

Retention time shift tolerance (+/-):  [min]

Figure S2: Spectral Extraction and Prescreening with ShinyScreen. A) Spectral Extraction Settings B) Automatic Quality Control Settings.

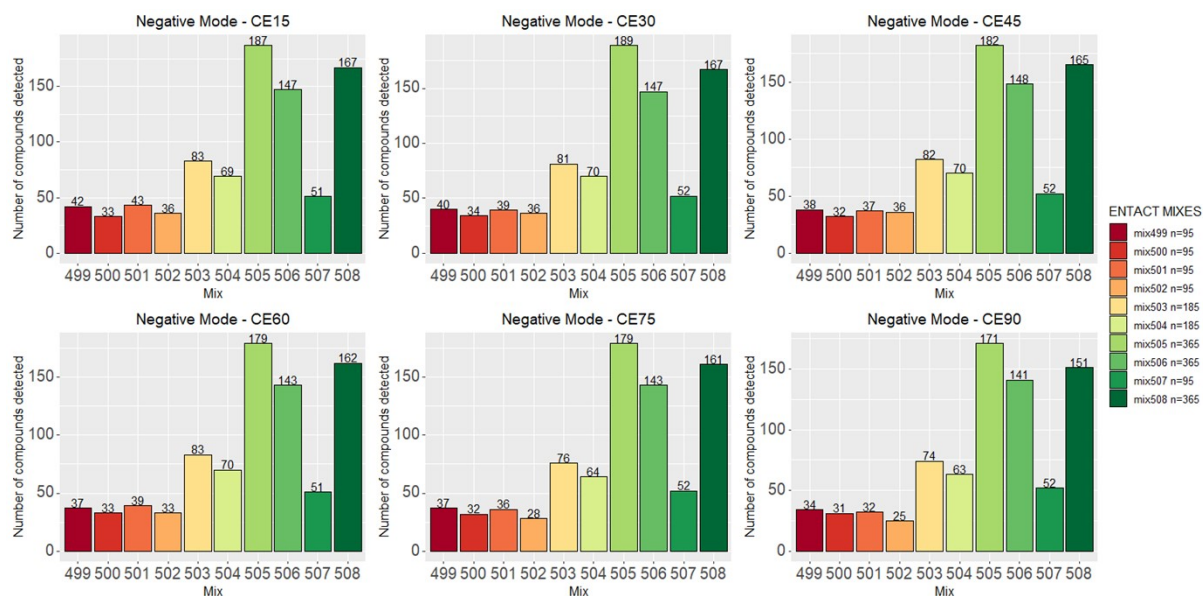


Figure S3: Prescreening results, negative mode.

Table S1: The number of compounds that passed quality control in ShinyScreen.

Mix	Total no. of compounds in each mix	No. of Compounds that passed the QC Positive mode						No. of Compounds that passed the QC Negative mode					
		CE15	CE30	CE45	CE60	CE75	CE90	CE15	CE30	CE45	CE60	CE75	CE90
499	95	44	47	46	45	44	46	42	40	38	37	37	34
500	95	36	37	39	38	38	39	33	34	32	33	32	31
501	95	51	46	49	46	45	43	43	39	37	39	36	32
502	95	34	38	32	33	32	33	36	36	36	33	28	25
503	185	95	97	93	89	96	95	83	81	82	83	76	74
504	185	95	95	92	89	93	87	69	70	70	70	64	63
505	365	199	197	197	191	196	192	187	189	182	179	179	171
506	365	206	197	200	203	202	204	147	147	148	143	143	141
507	95	48	47	45	46	46	45	51	52	52	51	52	52
508	365	203	198	197	192	193	195	167	167	165	162	161	151