

# Navigating Millions of Chemicals in Metabolomics and Exposomics Workflows

Prof. Dr. Emma L. Schymanski

*(plus many, many colleagues and collaborators!)*

Environmental Cheminformatics Group,

Luxembourg Centre for Systems Biomedicine, University of Luxembourg

[emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu) / [@ESchymanski](https://twitter.com/ESchymanski) / [@schymane@mstdn.social](https://www.instagram.com/schymane)

[https://wwwen.uni.lu/lcsb/research/environmental\\_cheminformatics/](https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/)

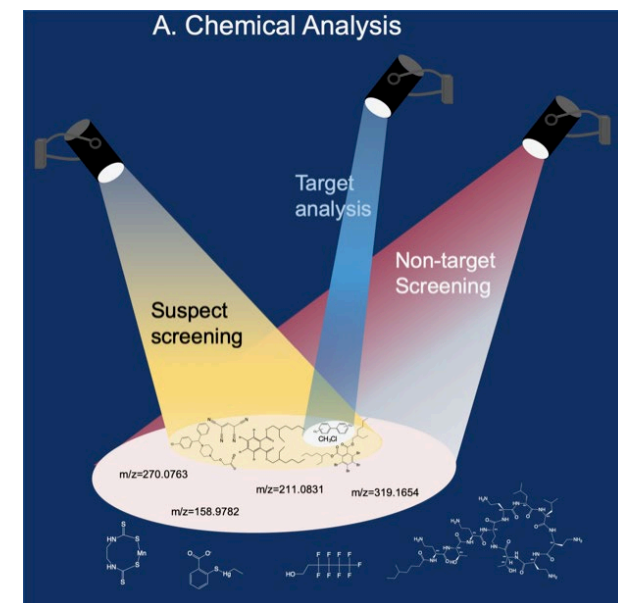


Image mod. from DOI:  
[10.1126/science.aay6636](https://doi.org/10.1126/science.aay6636)

Swiss Metabolomics Society Keynote, 15 Sept. 2023, Zurich

<https://swiss-metabolomics.ch/annual-meeting/>

Slides available at DOI:  
[10.5281/zenodo.8343924](https://doi.org/10.5281/zenodo.8343924)



# Outline

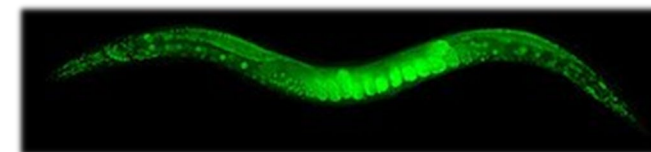
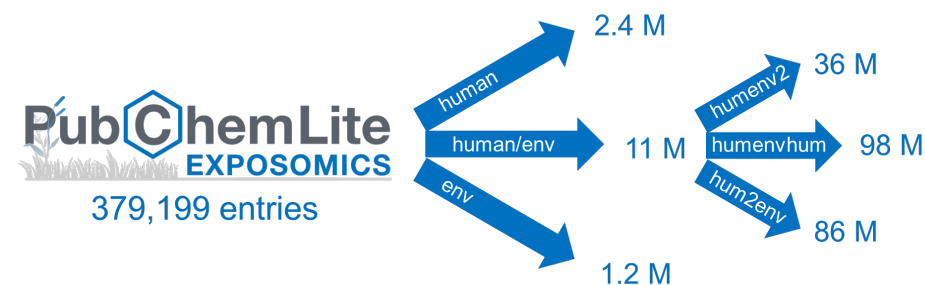
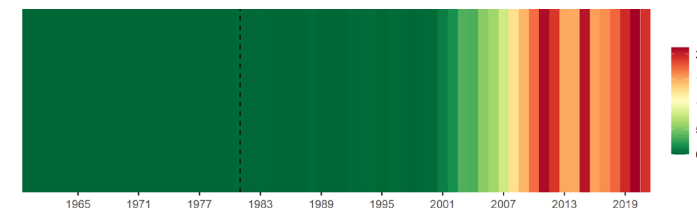
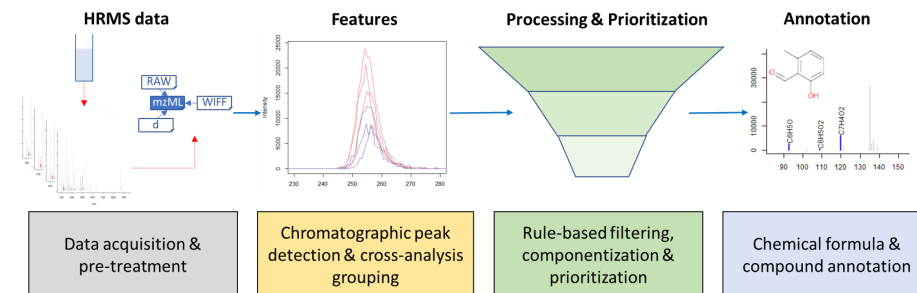
- Setting the Scene

- NT-HRMS & Databases
- The Great Data Challenge ...

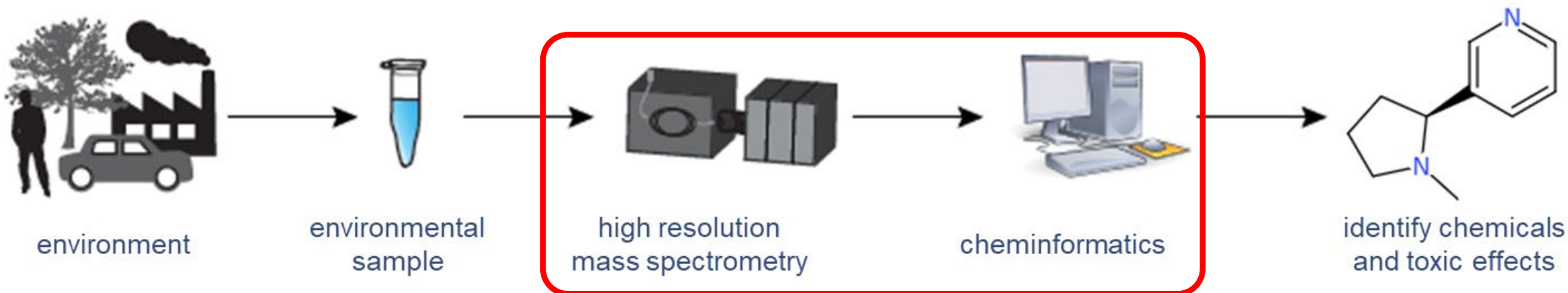
- Defining Chemical Space in Exposomics

- How to Tackle Transformations?

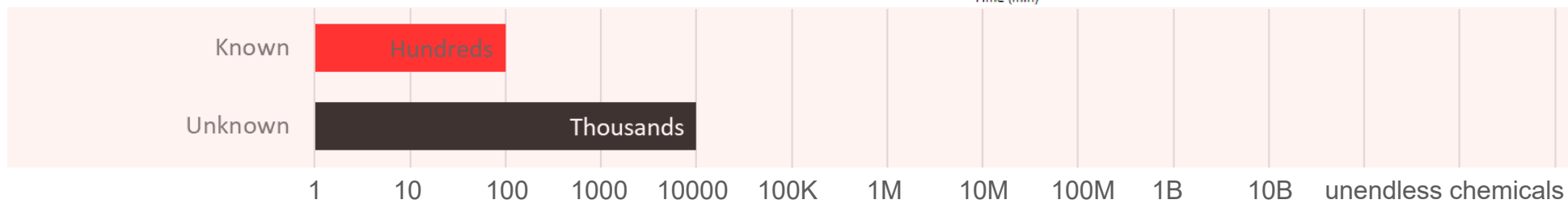
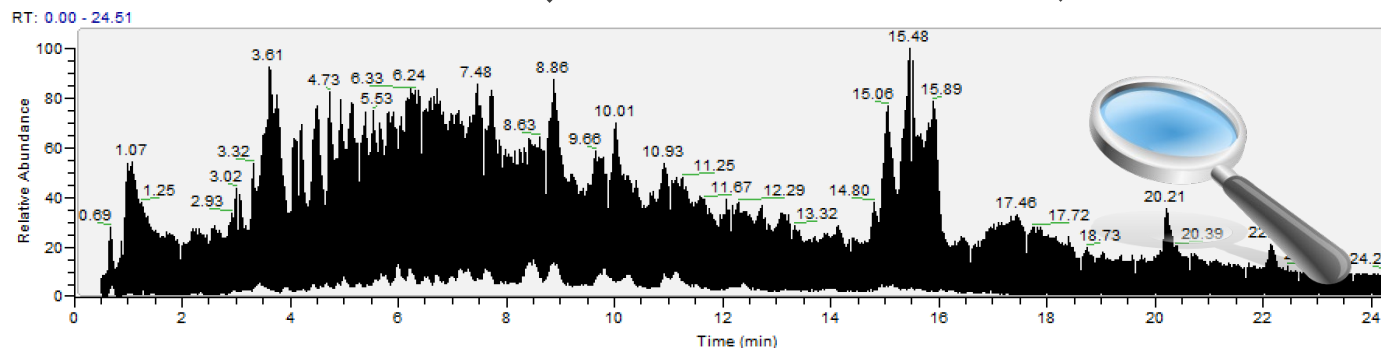
- Metabolomics / Exposomics-specific Workflows



# Environmental Cheminformatics & Non-target HR-MS



High resolution mass spectrometry  
AND connecting chemical knowledge



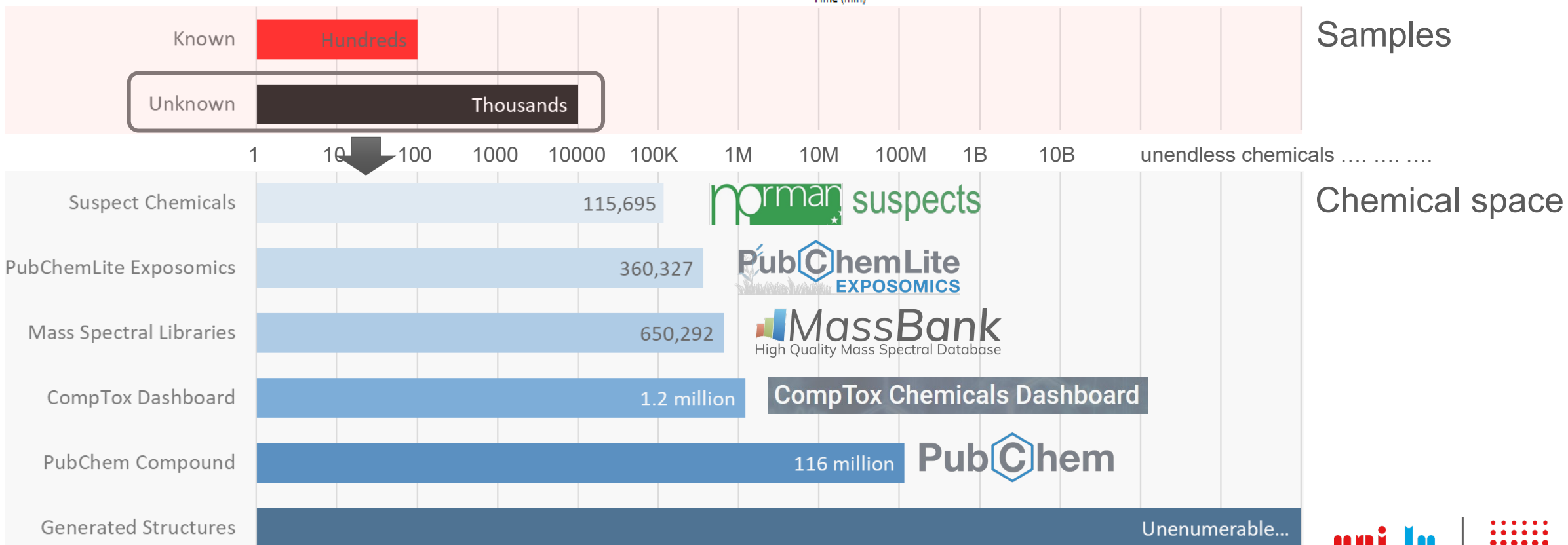
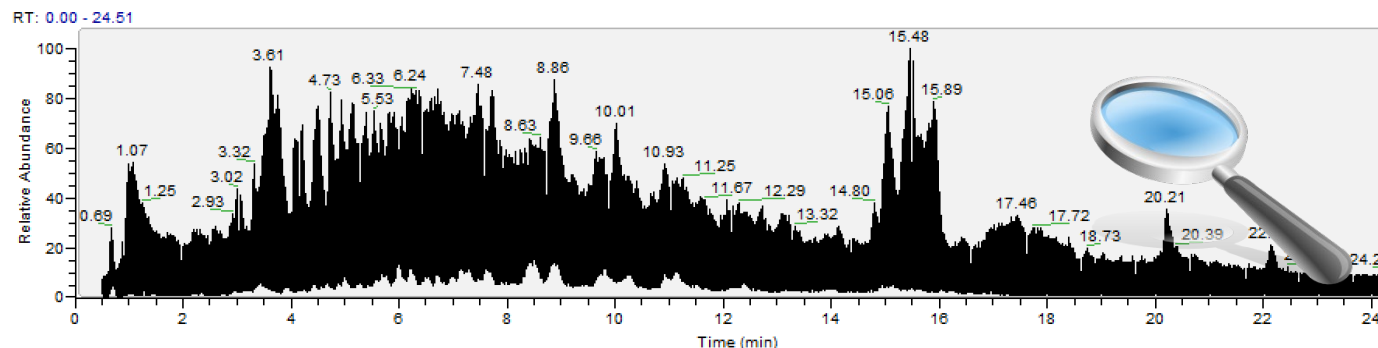
Samples



# Environmental Cheminformatics & Non-target HR-MS

High resolution  
mass spectrometry

AND connecting  
chemical knowledge



# Mass Spectral Libraries: MassBank (Open Source & Data!)

<https://massbank.eu/MassBank/>

<https://github.com/MassBank/MassBank-data/>

MassBank Europe

MassBank  
High Quality Mass Spectral Database

>> Search Spectra

MassBank Record: LU040605

(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]<sup>+</sup>

Search for:

Basic Search | Peak List | Peaks

Compound Information

Compound name

Exact Mass

AND

Formula ( e.g. C<sub>6</sub>H<sub>7</sub>N<sub>5</sub>, C<sub>5</sub>H<sup>\*</sup>N<sub>5</sub>, C<sub>5</sub><sup>\*</sup> )

AND

Search

Mass Spectrum

Abundance

m/z

Chemical Structure

Nc1ccc(cc1)S(=O)(=O)O

# MassBank: Cross Resource Integration & New Article!

## Environmental Science Processes & Impacts



PAPER

[View Article Online](#)  
[View Journal](#)

Check for updates

Cite this: DOI: 10.1039/d3em00181d

### Adding open spectral data to MassBank and PubChem using open source tools to support non-targeted exposomics of mixtures†

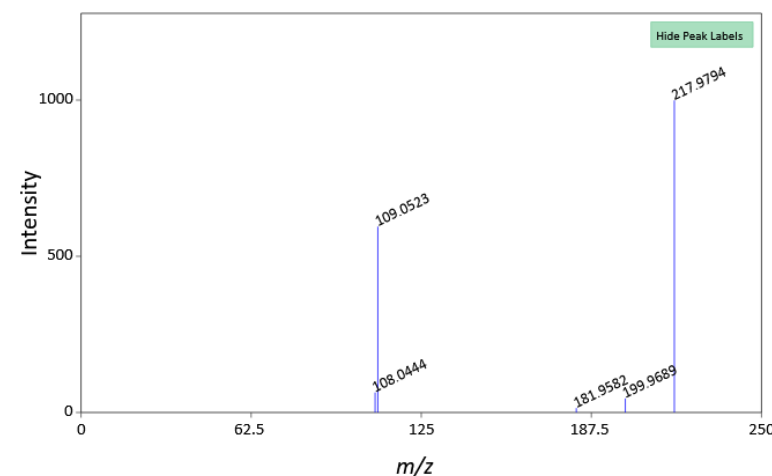
Anjana Elapavalore, \*<sup>a</sup> Todor Kondić, <sup>a</sup> Randolph R. Singh, <sup>ab</sup>  
Benjamin A. Shoemaker, <sup>c</sup> Paul A. Thiessen, <sup>c</sup> Jian Zhang, <sup>c</sup> Evan E. Bolton <sup>c</sup>  
and Emma L. Schymanski \*<sup>a</sup>

The term “exposome” is defined as a comprehensive study of life-course environmental exposures and the associated biological responses. Humans are exposed to many different chemicals, which can pose a major threat to the well-being of humanity. Targeted or non-targeted mass spectrometry techniques are widely used to identify and characterize various environmental stressors when linking exposures to human health. However, identification remains challenging due to the huge chemical space applicable to exposomics, combined with the lack of sufficient relevant entries in spectral libraries. Addressing these challenges requires cheminformatics tools and database resources to share curated open spectral data on chemicals to improve the identification of chemicals in exposomics studies. This article describes efforts

Arsanilic acid (Compound)

Top 5 Peaks	m/z
	217.9794 999
	109.0523 595
	108.0444 64
	199.9689 45
	181.9582 14

splash10-066r-0690000000-79ae659f03bdc3340757



CC BY

Elapavalore, A.; Kondić, T.; et al., Adding Open Spectral Data to MassBank and

<https://pubchem.ncbi.nlm.nih.gov/compound/7389#section=LC-MS>



# Gathering Expert Knowledge: NORMAN-SLE

<https://www.norman-network.com/nds/SLE/>



## NORMAN Database System



### NORMAN Suspect List Exchange

The NORMAN Suspect List Exchange (NORMAN-SLE) was established to facilitate the exchange of information on suspected substances. The NORMAN-SLE documents all individual collections of substances (see Source column in SusDat). NORMAN-SLE versions are regularly updated. Comments and contributions are welcome - please email us at [nds@norman-network.com](mailto:nds@norman-network.com). Please refer to our [documentation](#) pages for: [citation](#) instructions.

No.	Abbreviation	Description	Link
S0	SUSDAT	<b>Merged NORMAN Suspect List: SusDat</b>	<a href="#">Introduction</a>

### Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices

Mohammed Taha *et al.* (2022) DOI: [10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)

## RESEARCH

Open Access



# The NORMAN Suspect List Exchange (NORMAN-SLE): facilitating European and worldwide collaboration on suspect screening in high resolution mass spectrometry

Hiba Mohammed Taha<sup>1</sup>, Reza Aalizadeh<sup>2</sup>, Nikiforos Alygizakis<sup>3,2</sup>, Jean-Philippe Antignac<sup>4</sup>, Hans Peter H. Arp<sup>5,6</sup>, Richard Bade<sup>7</sup>, Nancy Baker<sup>8</sup>, Lidia Belova<sup>9</sup>, Lubertus Bijlsma<sup>10</sup>, Evan E. Bolton<sup>11</sup>, Werner Brack<sup>12,13</sup>, Alberto Celma<sup>10,14</sup>, Wen-Ling Chen<sup>15</sup>, Tiejun Cheng<sup>11</sup>, Parviel Chirsir<sup>1</sup>, Ľuboš Čirka<sup>16,3</sup>, Lisa A. D'Agostino<sup>17</sup>, Yannick Djoumbou Feunang<sup>18</sup>, Valeria Dulio<sup>19</sup>, Stellan Fischer<sup>20</sup>, Pablo Gago-Ferrero<sup>21</sup>, Aikaterini Galani<sup>2</sup>, Birgit Geueke<sup>22</sup>, Natalia Glowacka<sup>3</sup>, Juliane Glüge<sup>23</sup>, Ksenia Groh<sup>24</sup>, Sylvia Grosse<sup>25</sup>, Peter Haglund<sup>26</sup>, Pertti J. Hakkinen<sup>11</sup>, Sarah E. Hale<sup>5</sup>, Felix Hernandez<sup>10</sup>, Elisabeth M.-L. Janssen<sup>24</sup>, Tim Jonkers<sup>27</sup>, Karin Kiefer<sup>24</sup>, Michal Kirchner<sup>28</sup>, Jan Koschorreck<sup>29</sup>, Martin Krauss<sup>12</sup>, Jessy Krier<sup>1</sup>, Marja H. Lamoree<sup>27</sup>, Marion Letzel<sup>30</sup>, Thomas Letzel<sup>31</sup>, Qingliang Li<sup>11</sup>, James Little<sup>32</sup>, Yanna Liu<sup>33</sup>, David M. Lunderberg<sup>34,35</sup>, Jonathan W. Martin<sup>17</sup>, Andrew D. McEachran<sup>36</sup>, John A. McLean<sup>37</sup>, Christiane Meier<sup>29</sup>, Jeroen Meijer<sup>38</sup>, Frank Menger<sup>14</sup>, Carla Merino<sup>39,40</sup>, Jane Muncke<sup>22</sup>, Matthias Muschket<sup>12</sup>, Michael Neumann<sup>29</sup>, Vanessa Neveu<sup>41</sup>, Kelsey Ng<sup>3,42</sup>, Herbert Oberacher<sup>43</sup>, Jake O'Brien<sup>7</sup>, Peter Oswald<sup>3</sup>, Martina Oswaldova<sup>3</sup>, Jaqueline A. Picache<sup>37</sup>, Cristina Postigo<sup>44,14</sup>, Noelia Ramirez<sup>45,39</sup>, Thorsten Reemtsma<sup>12</sup>, Justin Renaud<sup>46</sup>, Pawel Rostkowski<sup>47</sup>, Heinz Rüdell<sup>48</sup>, Reza M. Salek<sup>41</sup>, Saer Samanipour<sup>49</sup>, Martin Scherlinger<sup>23,42</sup>, Ivo Schliebner<sup>29</sup>, Wolfgang Schulz<sup>50</sup>, Tobias Schulze<sup>12</sup>, Manfred Sengl<sup>30</sup>, Benjamin A. Shoemaker<sup>11</sup>, Kerry Sims<sup>51</sup>, Heinz Singer<sup>24</sup>, Randolph R. Singh<sup>1,52</sup>, Mark Sumarah<sup>46</sup>, Paul A. Thiessen<sup>11</sup>, Kevin V. Thomas<sup>7</sup>, Sonia Torres<sup>39</sup>, Xenia Trier<sup>53</sup>, Annemarie P. van Wezel<sup>54</sup>, Roel C. H. Vermeulen<sup>38</sup>, Jelle J. Vlaanderen<sup>38</sup>, Peter C. von der Ohe<sup>29</sup>, Zhanyun Wang<sup>55</sup>, Antony J. Williams<sup>56</sup>, Egon L. Willighagen<sup>57</sup>, David S. Wishart<sup>58</sup>, Jian Zhang<sup>11</sup>, Nikolaos S. Thomaidis<sup>2</sup>, Juliane Hollender<sup>23,24</sup>, Jaroslav Slobodnik<sup>3</sup> and Emma L. Schymanski<sup>1</sup>



### SEARCH All Databases

Searching for individual substance or group(s) of substances

**Note:** Click on a link below to go to an individual database



### Substance Database

A merged list of NORMAN substances; Central Database to access various lists of substances for suspect screening and prioritisation



### Suspect List Exchange

Central Database to access various lists of substances for suspect screening and prioritisation



### Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices

# Suspect List(s)

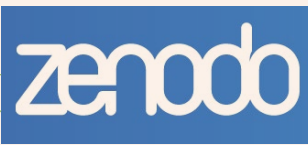
Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)Nc1ccc(C)cc1	InChi=1S/C6H7NO	KXNNXWJN	C1H11N3	297.0242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN(C)C(=O)c1ccc(C)cc1	InChi=1S/C6H9N	QOAGWXXK	C13H15N3	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1ccc(C(=O)N)cc1	InChi=1S/C7H7NO	QDQWIKBK	C14H16N4	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(C)cc1	InChi=1S/C6H7NO	UKAKWVDU	C14H16N4	320.0943

Substrate	Substrate	Substrate	Substrate	Enzyme	Reaction
BIOTID0001	acetaminofen	1983	RZVAJNKF	InChi=1S/C9H9NO2	CYP2A6; CYP2E1
BIOTID0001	propylthiouracil	1958	QVMDECO	InChi=1S/C5H5N3S	O-aryl de
BIOTID0001	3-Trifluoromethyl-4-nitrophenol	4296	KKIMDKMI	InChi=1S/C7H4F3NO2	CYP2A6; CYP2E1
BIOTID0001	2-Amino-3-thiophenol	53462	ARZWATD	InChi=1S/C4H5NS	CYP1A2; CYP2E1

CAS Number	Name	SMILES	InChi	InChiKey	Molecular	ExactMass
283	Diuron-de	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
294	Diuron-de	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
118	Ethofume	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
4	Metamit	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
668	Simazine	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
671	Terbutyl	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
672	Terbutyl	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
268	Figronil	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
2670	Atenolol	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
126	Methyl	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
296	N,N-Dime	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640
3244	Betameth	108-90-7	InChi=1S/C10H8N2O	QVWVWVWV	C10H8N2O	196.0640

## Suspect List Exchange

### (NORMAN-SLE)



DOI
10.5281/zenodo.6770176

# 12,835

views

# 11,095

downloads



# Suspect List(s)

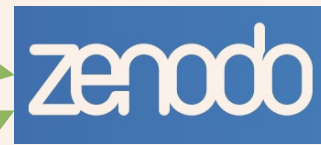
Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)Nc1cccnc1	CC(=O)Nc1cccnc1	133.07	297.0242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1C=CC(=O)C=C1	CN1C=CC(=O)C=C1	CN1C=CC(=O)C=C1	135.07	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1ccc(cc1)C(=O)N	COc1ccc(cc1)C(=O)N	COc1ccc(cc1)C(=O)N	153.07	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	320.0943

Substrate	Substrate	Substrate	Enzyme	Reaction
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP1A2; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2A6; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C8; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C9; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C19; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C18; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C17; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C16; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C15; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C14; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C13; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C12; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C11; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C10; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C9; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C8; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C7; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C6; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C5; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C4; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C3; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C2; CYP2D6
BIOTID0001	acetaminol	1983	RZVAJNKI	InChi=1S/CYP2C1; CYP2D6

CAS Number	Name	SMILES	InChi	InChiKey	Molecular	ExactMass	Priority	Category
283	Diuron-de	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
294	Diuron-de	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
118	Ethofume	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
4	Metamitr	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
668	Simazine	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
671	Terbutylal	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
672	Terbutylal	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
2668	Flpromil	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
2670	Atenolol	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
126	Methyl 2	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
296	N,N-Dime	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High
3244	Betameth	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	CC(=O)Nc1ccc(cc1)C	133.07	297.0242	1	High



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

Suspect List Exchange (NORMAN-SLE)

## NORMAN Database System (NDS)

### Ecotoxicology

Substance Database (NORMAN SusDat)

SARS-CoV-2 in sewage

Passive Sampling

Antibiotic Resistance Bacteria/Genes

Indoor Environment

MassBank Europe

Bioassays Monitoring Data

Substance Factsheets

Digital Sample Freezing Platform

Prioritisation

Chemical Occurrence Data



# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChI	InChIKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	179.0742	297.0242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CC(=O)C1=CC=CC=C1	CC(=O)C1=CC=CC=C1	CC(=O)C1=CC=CC=C1	135.0441	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	CC(=O)NC1=CC=C(C=C1)C	CC(=O)NC1=CC=C(C=C1)C	CC(=O)NC1=CC=C(C=C1)C	191.0841	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)NC1=CC=C(C=C1)C	CC(=O)NC1=CC=C(C=C1)C	CC(=O)NC1=CC=C(C=C1)C	191.0841	352.0841

BIOTID0001	acetamino	1983	RZVAJNKF	InChI=1S/C	CYP2A2	CYP2A2	Oxidation
BIOTID0002 <td>paracetamol</td> <td>1983</td> <td>QVMDCQ</td> <td>InChI=1S/C</td> <td>CYP2A6</td> <td>CYP2A6</td> <td>(m Hydroxylation)</td>	paracetamol	1983	QVMDCQ	InChI=1S/C	CYP2A6	CYP2A6	(m Hydroxylation)
BIOTID0003 <td>paracetamol</td> <td>1983</td> <td>QVMDCQ</td> <td>InChI=1S/C</td> <td>CYP2D6</td> <td>CYP2D6</td> <td>O-aryl de</td>	paracetamol	1983	QVMDCQ	InChI=1S/C	CYP2D6	CYP2D6	O-aryl de
BIOTID0004 <td>3-Trifluoromethylacetamin</td> <td>4296</td> <td>KKIMDKME</td> <td>InChI=1S/C</td> <td>CYP2D6</td> <td>CYP2D6</td> <td>C Aromatic</td>	3-Trifluoromethylacetamin	4296	KKIMDKME	InChI=1S/C	CYP2D6	CYP2D6	C Aromatic
BIOTID0005 <td>2-Amino-3-methylbutanoic acid</td> <td>53462</td> <td>ARZWATD</td> <td>InChI=1S/C</td> <td>CYP1A2</td> <td>CYP1A2</td> <td>C/N-Hydroxy</td>	2-Amino-3-methylbutanoic acid	53462	ARZWATD	InChI=1S/C	CYP1A2	CYP1A2	C/N-Hydroxy

CAS Number	Name	RT	Retention	Quality	Reference	M	Retention	Reference	Retention	Reference	Retention	Reference
283	Diuron-de	10.0	10.0	10.0	Diuron-de	10.0	10.0	10.0	10.0	10.0	10.0	10.0
294	Diuron-de	10.0	10.0	10.0	Diuron-de	10.0	10.0	10.0	10.0	10.0	10.0	10.0
118	Ethofomec	10.0	10.0	10.0	Ethofomec	10.0	10.0	10.0	10.0	10.0	10.0	10.0
4	Metamitri	10.0	10.0	10.0	Metamitri	10.0	10.0	10.0	10.0	10.0	10.0	10.0
668	Simazine	10.0	10.0	10.0	Simazine	10.0	10.0	10.0	10.0	10.0	10.0	10.0
671	Terbutylal	10.0	10.0	10.0	Terbutylal	10.0	10.0	10.0	10.0	10.0	10.0	10.0
672	Terbutylal	10.0	10.0	10.0	Terbutylal	10.0	10.0	10.0	10.0	10.0	10.0	10.0
2668	Pipromidil	10.0	10.0	10.0	Pipromidil	10.0	10.0	10.0	10.0	10.0	10.0	10.0
2670	Atenolol	10.0	10.0	10.0	Atenolol	10.0	10.0	10.0	10.0	10.0	10.0	10.0
126	Methyl 2	10.0	10.0	10.0	Methyl 2	10.0	10.0	10.0	10.0	10.0	10.0	10.0
296	N,N-Dime	10.0	10.0	10.0	N,N-Dime	10.0	10.0	10.0	10.0	10.0	10.0	10.0
3244	Betameth	10.0	10.0	10.0	Betameth	10.0	10.0	10.0	10.0	10.0	10.0	10.0



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

www.norman-data.eu

### NORMAN Digital Sample Freezing Platform

Main Page Batch mode

Contributed Samples Results Chromatograms Interactive Map Help

Choose Emerging Substance or input mass of interest and experimental RTI

Substance name or CAS or StdInChIKey

DEET [ 134-62-3 ]  
[MMOZXZBCLCQITDF-UHFFFAOYSA-N]

Choose Ionization

Positive

Adduct

[M+H]<sup>+</sup>

Predicted RTI Positive

-58 523 980

Reset

## NORMAN Database System

Ecotoxicology

SARS-CoV-2 in sewage

Passive Sampling

Indoor Environment

Bioassays Monitoring Data

Digital Sample Freezing Platform

Chemical Occurrence Data

# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logP <sub>ph7</sub>	SMILES	InChI	InChIKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)Nc1cccnc1	CN1C=CC=CC=C1	123.0742	297.0242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CC(=O)C1=CC=CC=C1	CC(=O)c1cccnc1	CC(=O)N1C=CC=C1	123.0742	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	CC(=O)NC1=CC=C(C=C1)C2=CC=CC=C2	CC(=O)Nc1cccnc1	CC(=O)N1C=CC=C1	352.0841	
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)NC1=CC=C(C=C1)C2=CC=CC=C2	CC(=O)Nc1cccnc1	CC(=O)N1C=CC=C1	320.0943	

Substrate	substrate	substrate	substrate	enzyme	reaction	InChI	InChIKey	Molecular	ExactMass
BIOTID0001	acetaminol	1983	RZVAJINK	InChI=1S/C11H13NO2	CYP2A6	CC(=O)Nc1cccnc1	CN1C=CC=CC=C1	151.1494	151.1494
BIOTID0001	acetaminol	1983	RZVAJINK	InChI=1S/C11H13NO2	CYP2A6	CC(=O)Nc1cccnc1	CN1C=CC=CC=C1	151.1494	151.1494
BIOTID0001	acetaminol	1983	RZVAJINK	InChI=1S/C11H13NO2	CYP2A6	CC(=O)Nc1cccnc1	CN1C=CC=CC=C1	151.1494	151.1494

CAS Number	Name	Priority	Substance	SMILES	InChI	InChIKey	Molecular	ExactMass	Accession
283	Diuron-de	High	Diuron	NC(=O)Nc1cccnc1	NC(=O)Nc1cccnc1	CN1C=CC=CC=C1	123.0742	123.0742	BIOTID0001
294	Diuron-de	High	Diuron	NC(=O)Nc1cccnc1	NC(=O)Nc1cccnc1	CN1C=CC=CC=C1	123.0742	123.0742	BIOTID0001
118	Ethofume	High	Ethofume	CC(=O)Nc1cccnc1	CC(=O)Nc1cccnc1	CN1C=CC=CC=C1	123.0742	123.0742	BIOTID0001



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

## Suspect List Exchange (NORMAN-SLE)

## NORMAN Database System (NDS)

- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Indoor Environment
- Bioassays Monitoring Data
- Digital Sample Freezing Platform
- Chemical Occurrence Data
- Substance Database (NORMAN SusDat)
- Antibiotic Resistance Bacteria/Genes
- MassBank Europe
- Substance Factsheets
- Prioritisation

## CompTox Chemicals Dashboard

**Metoprolol acid**  
56592-18-4 | DTXSID70881080

Chemical Details: Quality Control Notes, Internal Properties, Structural Identifiers, Linked Substances, Presence in Life, International, Other.

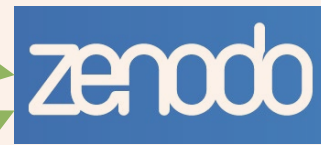
Showing 74 of 300 Records

# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	KXNNXWIFC1H1H1N3	297.0242	
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1C(C=C)C=CC1	CN1C(C=C)C=CC1	COAGWXXK13H15N3	245.1164	
245	N4-Acetyl	24341-30-8	CAS_RN24	1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	QDQWIKBK14H16N4	352.0841	
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	LUKAKWYD14H16N4	320.0943	

CAS Number	Name	SMILES	InChi	InChiKey	Molecular	ExactMass	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
1983	RZVAJNK	InChi=1S/CYP2A6	CYP2A6	YXGJYKXWYK13H15N3	297.0242			0.69	CC(=O)Nc1ccc(C)cc1	InChi=1S/CYP2A6	CYP2A6	YXGJYKXWYK13H15N3	297.0242
1983	QYMDXQ	InChi=1S/CYP2D6	CYP2D6	YXGJYKXWYK13H15N3	297.0242			0.15	CN1C(C=C)C=CC1	InChi=1S/CYP2D6	CYP2D6	YXGJYKXWYK13H15N3	297.0242
4296	KKIMDKM	InChi=1S/CYP2C9	CYP2C9	YXGJYKXWYK13H15N3	297.0242			1	CC(=O)Nc1ccc(C)cc1	InChi=1S/CYP2C9	CYP2C9	YXGJYKXWYK13H15N3	297.0242
53462	ARZWATD	InChi=1S/CYP1A2	CYP1A2	YXGJYKXWYK13H15N3	297.0242			0.41	CC(=O)Nc1ccc(C)cc1	InChi=1S/CYP1A2	CYP1A2	YXGJYKXWYK13H15N3	297.0242



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

## Suspect List Exchange (NORMAN-SLE)



NORMAN Suspect List Exchange Classification 115,260

- S13 | EU-COSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) 3,887
- S25 | OECD/PFAS | List of PFAS from the OECD 3,677

3.2.10 Collision Cross Section

150.51 Å<sup>2</sup> [M-H]<sup>-</sup> [CCS Type: DT; Buffer gas: N<sub>2</sub>; Dataset: PFAS]  
DOI:10.1021/acs.est.2c02001

Baker Lab, Chemistry Department, The University of North Carolina at Chapel Hill

228.66 Å<sup>2</sup> [2M-H]<sup>-</sup>  
150.81 Å<sup>2</sup> [M-H]<sup>-</sup>

S79 | UAACSC/CC | Collision Cross Section (CCS) Library from UAntwerp | DOI:10.5281/zenodo.4704648

NORMAN Suspect List Exchange

S80 | PFAS/LUEGIE | Overview of PFAS Uses 1,250

## NORMAN Database System (NDS)

- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Indoor Environment
- Bioassays Monitoring Data
- Digital Sample Freezing Platform
- Chemical Occurrence Data
- Substance Database (NORMAN SusDat)
- Antibiotic Resistance Bacteria/Genes
- MassBank Europe
- Substance Factsheets
- Prioritisation

## CompTox Chemicals Dashboard

Chemical Lists

Search: NORMAN

Showing 7/4 of 300 Records

List Acronym	List Name	Chemicals	Updated	List Description
ATHENSUS				
BSPHENOLS				
CCCMPEND				

**Metoprolol acid**  
56592-14-4 [TXSID7081080]  
Searched by Approved Name.

Quality Control Notes

- Internal Properties
- Structural Identifiers
- Linked Substances
- Presence in Life

Other

Metoprolol acid is a compilation of suspects, predicted transformation products and surfactants at University of Athens, as described in Gago-Ferreiro et al 2015, DOI: 10.1021/acs.est.5b01454. The original data is available on the NORMAN Suspect List Exchange. Names are included here.

Metoprolol acid is a collection of biogenics available at NIB (Pavel Kratochvíl) and from Lab K3M (Swedish Chemicals Agency, in Swedish with English summary), hosted on SciList (Erlang) (<https://www.norman-network.com/nds/SLE/>). Dataset DOI: 10.1186/s12302-022-00680-6

Metoprolol acid is included with > 3800 experimental collision cross section (CCS) values (drift tube M) in NORMAN Suspect List Exchange by Jackie Prusache and John McLean. Vendor listed as DTSDS by CAS Registry Number by G. Schymanski, Luxembourg Center for Science. Further curation ongoing. Further details available here.

7.5 Transformations

2 items View More Details

Sort By: Please Choose One

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	Nicotine		Norcotine	Thirdhand Smoke	
	Nicotine		Norcotine	Demethylation	

NORMAN Suspect List Exchange

What is PubChem? <https://pubchem.ncbi.nlm.nih.gov/>



# Explore Chemistry

Quickly find chemical information from authoritative sources

Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Use Entrez

Compounds

Substances

BioAssays



Draw Structure



Upload ID List



Browse Data



Periodic Table

116M Compounds

308M Substances

292M Bioactivities

36M Literature

934 Data Sources

[See More Statistics >](#)

[Explore Data Sources >](#)

# NORMAN-SLE in PubChem



The NORMAN network enhances the exchange of information on emerging environmental substances, and encourages the validation and harmonisation of common measurement methods and monitoring tools so that the requirements of risk assessors and risk managers can be better met. The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point to find suspect lists relevant for various environmental monitoring questions, described in DOI:10.1186/s12302-022-00680-6

<b>Organization</b>	NORMAN Network (c/o UniLu)
<b>Category</b>	Research and Development
<b>URL</b>	<a href="https://www.norman-network.com/nds/SLE/">https://www.norman-network.com/nds/SLE/</a>
<b>License Note</b>	Data: CC-BY 4.0; Code (hosted by ECI, LCSB): Artistic-2.0
<b>License URL</b>	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
<b>Contact Name</b>	Emma Schymanski
<b>Address</b>	6 avenue du Swing, Belvaux, Luxembourg, 4367
<b>Data Source ID</b>	23819
<b>Data in PubChem</b>	118,487 Live Substances 22,317 Annotations 1 Classification
<b>Last Updated</b>	2023/09/06

▼ NORMAN Suspect List Exchange Classification	?	↗	115,695
▶ S13   EUCOSMETICS   Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006)	?		3,936
▶ S25   OECDPFAS   List of PFAS from the OECD	?		3,678
▶ S36   UBAPMT   Potential Persistent, Mobile and Toxic (PMT) substances	?		254
▶ S47   ECHAPLASTICS   A list from the Plastic Additives Initiative Mapping Exercise by ECHA	?		241
▶ S50   CCSCOMPEND   The Unified Collision Cross Section (CCS) Compendium	?		869
▶ S60   SWISSPEST19   Swiss Pesticides and Metabolites from Kiefer et al 2019	?		1,359
▶ S61   UJICCSLIB   Collision Cross Section (CCS) Library from UJI	?		574
▶ S66   EAWAGTPS   Parent-Transformation Product Pairs from Eawag	?		258
▶ S68   HSDBTPS   Transformation Products Extracted from HSDB Content in PubChem	?		740
▶ S69   LUXPEST   Pesticide Screening List for Luxembourg	?		386
▶ S72   NTUPHTW   Pharmaceutically Active Substances from National Taiwan University	?		1,068
▶ S75   CyanoMetDB   Comprehensive database of secondary metabolites from cyanobacteria	?		2,553
▶ S77   FCCDB   Food Contact Chemicals Database v5.0	?		5,989
▶ S79   UACCSCEC   Collision Cross Section (CCS) Library from UAntwerp	?		148
▶ S80   PFASGLUEGE   Overview of PFAS Uses	?		1,251

▼ NORMAN Suspect List Exchange Classification ? ↗ 115,248

▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetics (2000) and Revised Inventory (2006) ? 3,887

▶ S25 | OECDPFAS | List of PFAS from the OECD ? 3,677

▶ S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances

▶ S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise

▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ?

▶ S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ?

▶ S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574

▶ S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258

▶ S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem

▶ S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386

▶ S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University

▶ S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria

▶ S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148

▶ S80 | PFASGLUECE | Overview of PFAS Uses ? 1,250

### PubChem Phloroglucinol (Compound)

#### 9.1.1 Use Classification

Cosmetics -> Hair dyeing

S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) | DOI:10.5281/zenodo.2624118

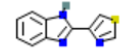
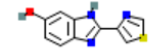
▶ NORMAN Suspect List Exchange

### PubChem Thiabendazole (Compound)

#### 10.10 Transformations

10 items Download

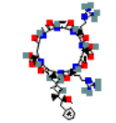
Search SORT BY Predecessor Name - A to Z

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence
	thiabendazole		5-hydroxythiabendazole	Aromatic hydroxylation of fused benzene ring / Human Phase I	CYP1A2	10.1186

### PubChem 7-Desmethylmicrocystin RR (Compound)

#### 7 Taxonomy

2 items Download

Structure	Taxonomy	Evidence IDs	Data Source
	Microcystis aeruginosa	PMID:1440646 DOI:10.1016/0041-0101(92)90054-9	NORMAN Suspect List Exchange

S00 PubChem Tri-O-cresyl phosphate (Compound)

S01 3.2.18 Collision Cross Section

S02 192.43 Å<sup>2</sup> [M+Na]<sup>+</sup>

S03 182.39 Å<sup>2</sup> [M+H]<sup>+</sup>

S04 263.75 Å<sup>2</sup> [2M+Na]<sup>+</sup>

S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp | DOI:10.5281/zenodo.4704648

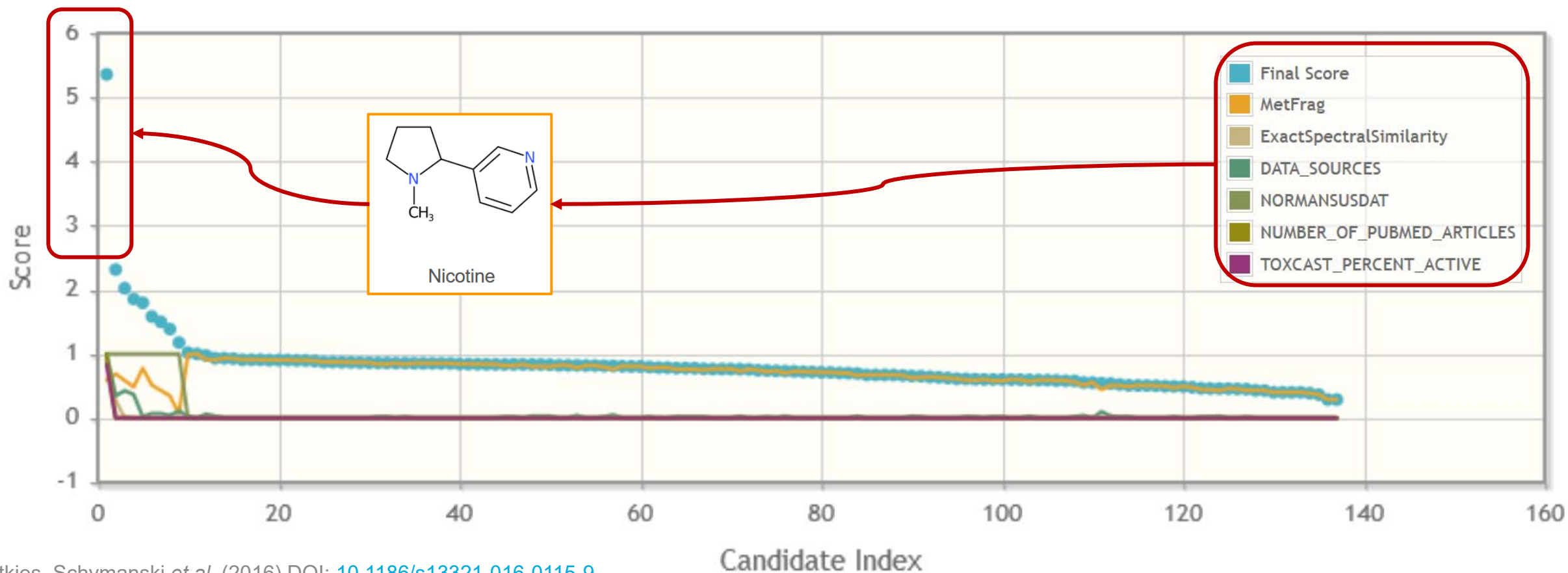
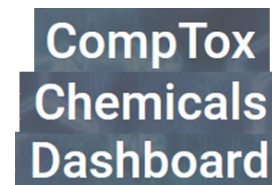
▶ NORMAN Suspect List Exchange

# Connecting Knowledge for Chemical Identification: MetFrag



<https://msbi.ipb-halle.de/MetFrag/>


<http://ipb-halle.github.io/MetFrag/>





# MetFrag: *In Silico* Fragmentation to Select Candidates





## MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

### Database Settings

Database:   Include references:

Parent Ion:

Neutral Mass:  Search ppm:

Formula:

Identifiers:

### Candidate Filter & Score Settings

### Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates

#### MS/MS Peak list

```
80.0494 6261028.7 23
84.0807 13197924.1 50
94.065 967625.9 3
106.065 24640249.3 93
117.0572 3192413.5 12
120.0807 8648923.7 32
130.0651 24669353.9 93
132.0807 80112590.7 304
163.1229 263120223.6 999
```

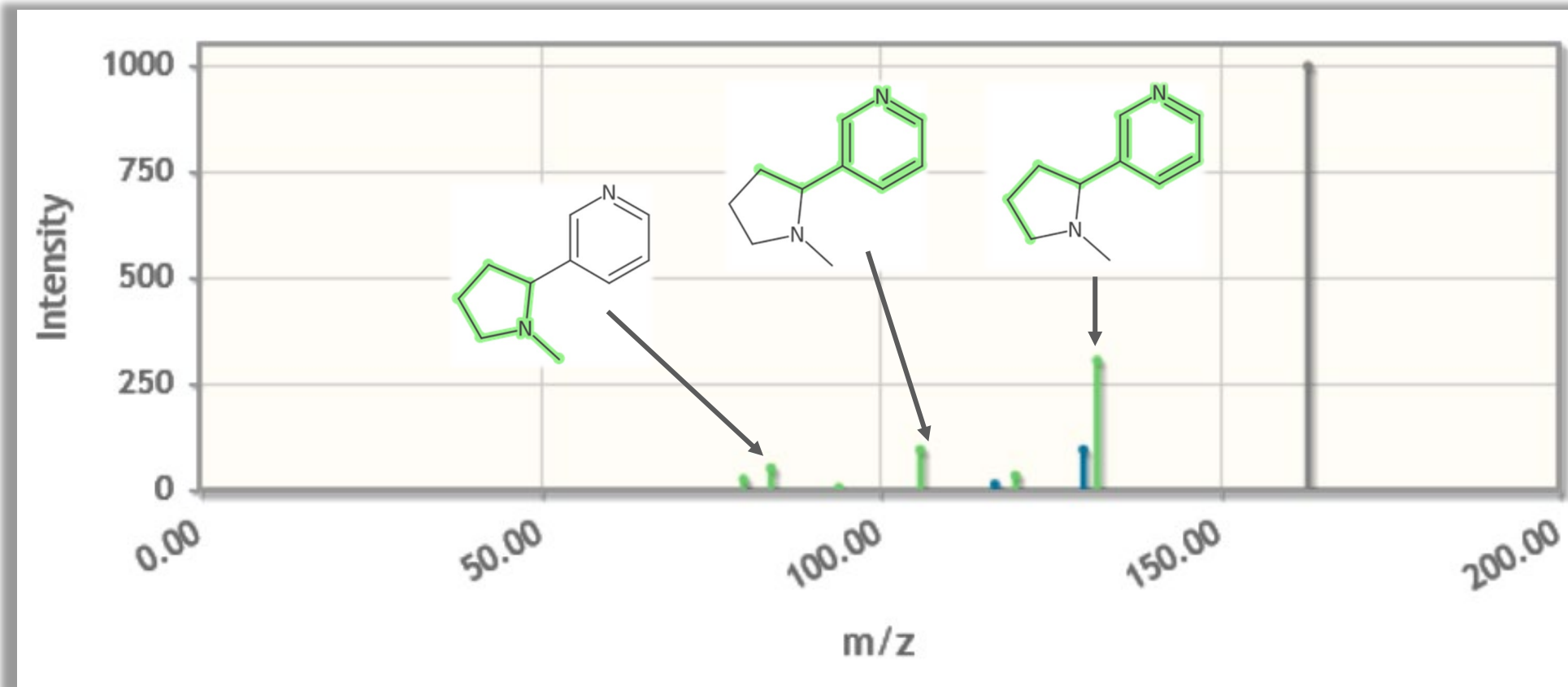


# MetFrag: *In Silico* Fragmentation to Select Candidates



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra




- matched
- not matched
- excluded



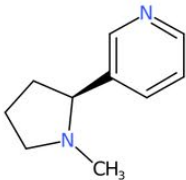
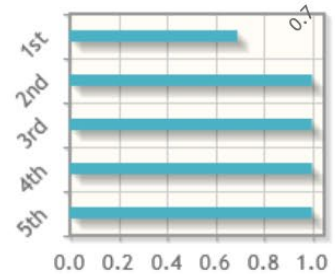
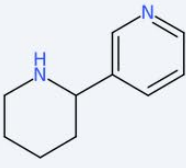
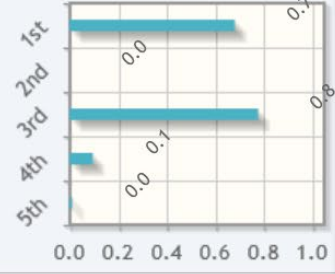
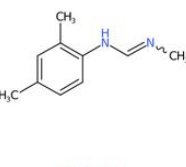
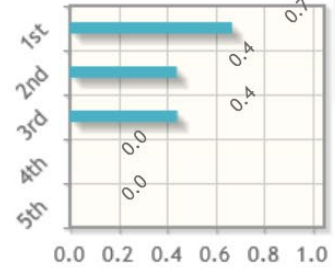
# MetFrag: ...Plus Metadata for Context





## MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 <p>3-[(2S)-1-methylpyrrolidin-2-yl]pyridine</p>	89594 InChIKeyBlock1 = SNICXCGAKADSCV	162.1157	C <sub>10</sub> H <sub>14</sub> N <sub>2</sub>		4.692	Peaks: 6 / 8 <a href="#">Fragments</a> <a href="#">Scores</a> <a href="#">Download</a>
2	 <p>3-piperidin-2-ylpyridine</p>	2181 InChIKeyBlock1 = MTXSIJUGVMTTMU	162.1157	C <sub>10</sub> H <sub>14</sub> N <sub>2</sub>			
3	 <p>N-(2,4-dimethylphenyl)-N'-methylmethanimidamide</p>	36326 InChIKeyBlock1 = JIIOLEGNERQDIP	162.1157	C <sub>10</sub> H <sub>14</sub> N <sub>2</sub>		1.5586	<a href="#">Scores</a> <a href="#">Download</a>

### Scores View

Candidate Name: 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine  
 Candidate Identifier: 89594|47445

	Name	Normalized Value	Raw Value
▶	MetFrag	0.692	246.1725
▶	ExactSpectralSimilarity	1.0	1.0
▶	AnnoTypeCount	1.0	9.0
▶	Patent_Count	1.0	102416.0
▶	PubMed_Count	1.0	38574.0

# Non-target High Resolution Mass Spectrometry (NT-HRMS)

[Home](#) > [Environmental Sciences Europe](#) > [Article](#)

## NORMAN guidance on suspect and non-target screening in environmental monitoring

Review | [Open Access](#) | [Published: 04 September 2023](#) | 35, Article number: 75 (2023)

[Download PDF](#) ↓

✓ You have full access to this [open access](#) article



[Environmental Sciences Europe](#)

[Submit manuscript](#) →

[Juliane Hollender](#) , [Emma L. Schymanski](#), [Lutz Ahrens](#), [Nikiforos Alygizakis](#), [Frederic Béen](#), [Lubertus Bijlsma](#), [Andrea M. Brunner](#), [Alberto Celma](#), [Aurelie Fildier](#), [Qiuguo Fu](#), [Pablo Gago-Ferrero](#), [Ruben Gil-Solsona](#), [Peter Haglund](#), [Martin Hansen](#), [Sarit Kaserzon](#), [Anneli Kruve](#), [Marja Lamoree](#), [Christelle Margoum](#), [Jeroen Meijer](#), [Sylvain Merel](#), [Cassandra Rauert](#), [Pawel Rostkowski](#), [Saer Samanipour](#), [Bastian Schulze](#), [Tobias Schulze](#), [Randolph R. Singh](#), [Jaroslav Slobodnik](#), [Teresa Steininger-Mairinger](#), [Nikolaos S. Thomaidis](#), [Anne Togola](#), [Katrin Vorkamp](#), [Emmanuelle Vulliet](#), [Linyan Zhu](#) & [Martin Krauss](#)

— [Show fewer authors](#)

Hollender *et al.* (2023) DOI: [10.1186/s12302-023-00779-4](https://doi.org/10.1186/s12302-023-00779-4)

[Sections](#)

[Figures](#)

[References](#)

[Abstract](#)

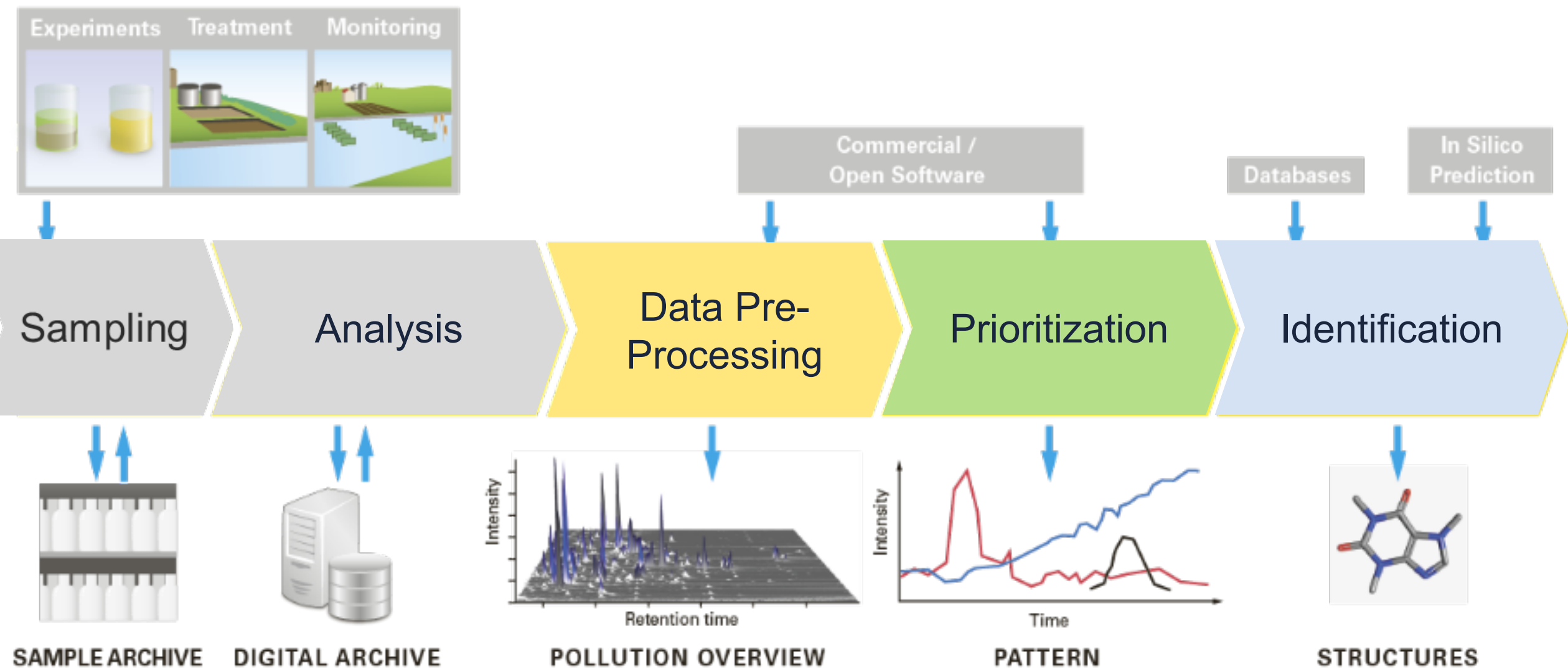
[Motivation for this guidance](#)

[Overview on analytical methods for NTS](#)

[Sampling and sample preparation for NTS](#)

[LC-HRMS/MS analysis](#)

# Non-target High Resolution Mass Spectrometry (NT-HRMS)



# Open Source Workflows for NT-HRMS: patRoan



Software | [Open Access](#) | [Published: 06 January 2021](#)

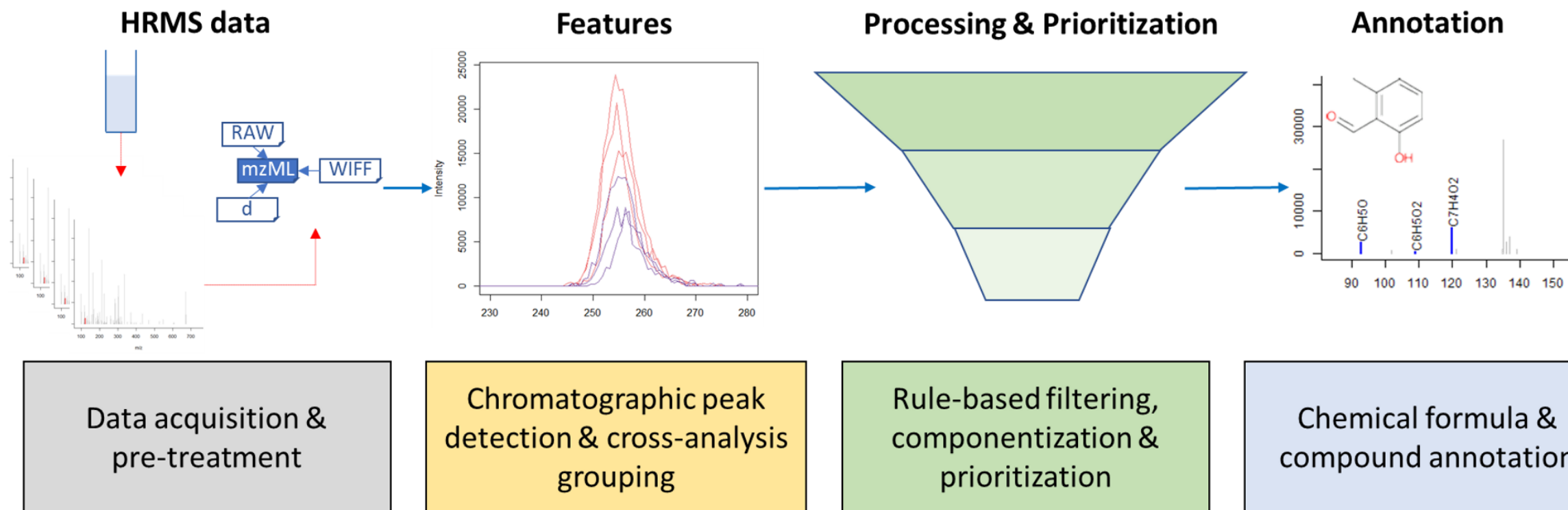
**patRoan: open source software platform for environmental mass spectrometry based non-target screening**

*Journal of Cheminformatics* **13**, Article number: 1 (2021) | [Cite this article](#)

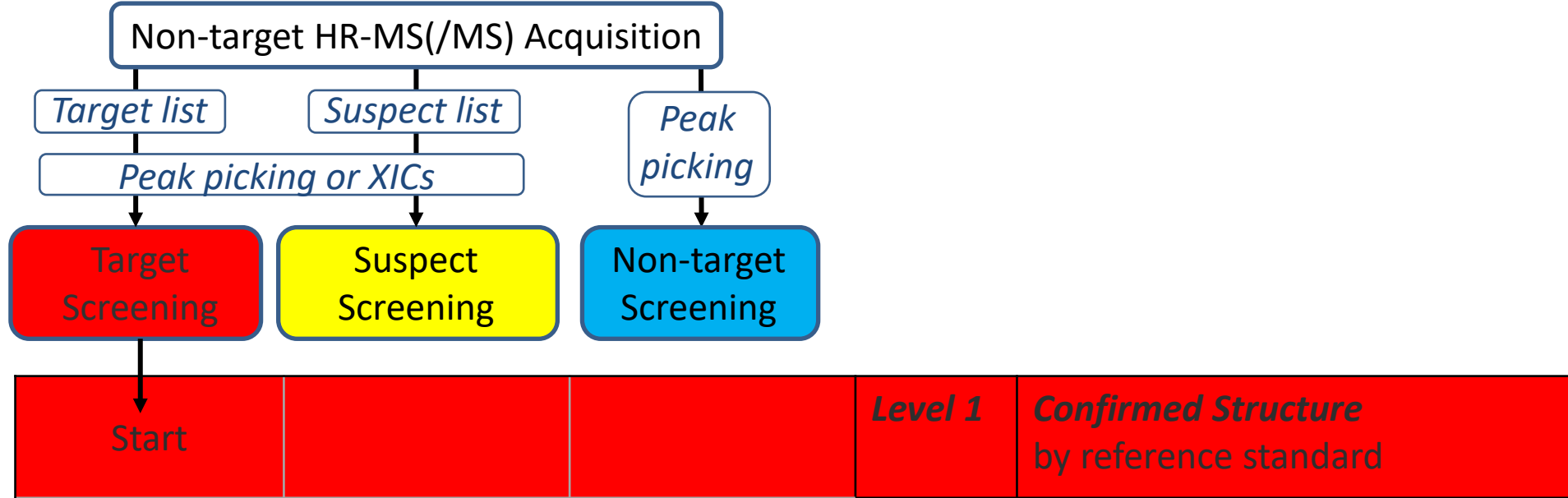
Rick Helmus , Thomas L. ter Laak, Annemarie P. van Wezel, Pim de Voogt & Emma L. Schymanski

patRoan 2.0: Improved non-target analysis workflows including automated transformation product screening

Rick Helmus<sup>1†</sup>, Bas van de Velde<sup>123</sup>, Andrea M. Brunner<sup>24</sup>, Thomas L. ter Laak<sup>12</sup>, Annemarie P. van Wezel<sup>1</sup>, and Emma L. Schymanski<sup>5</sup>

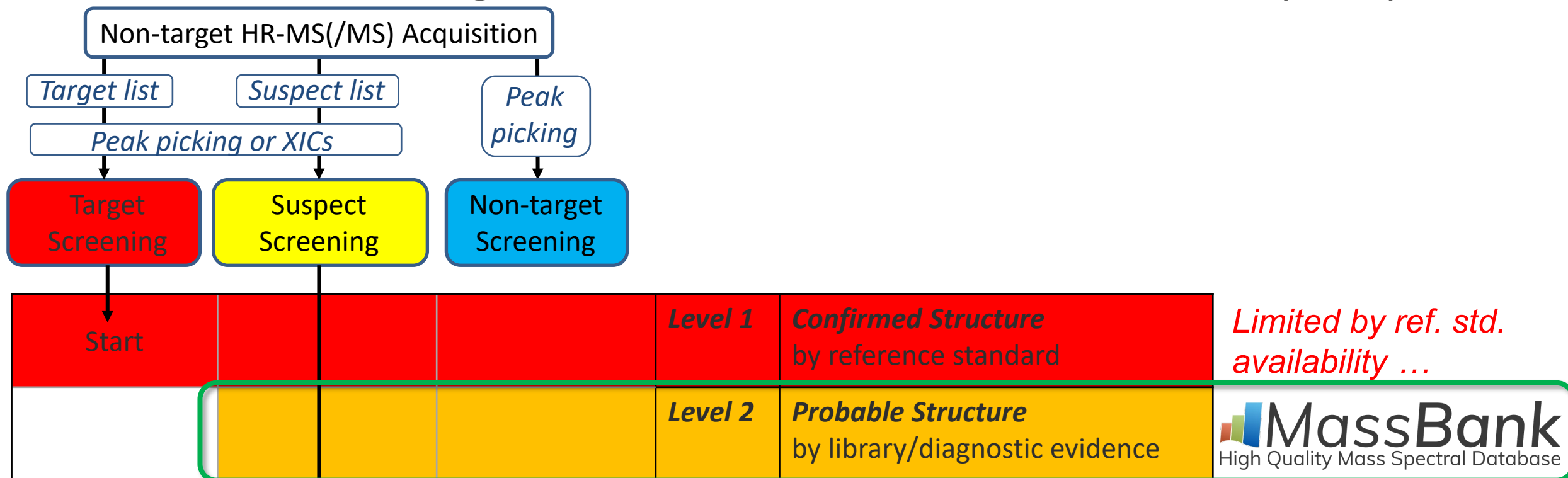


# Identification Strategies and Confidence in NT-HRMS(/MS)



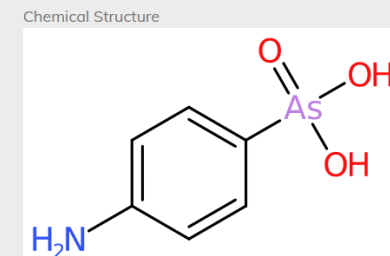
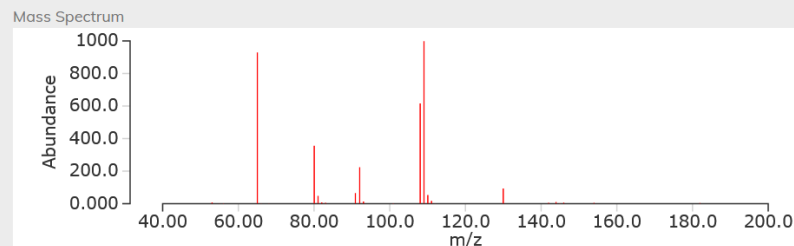
*Limited by ref. std. availability ...*

# Identification Strategies and Confidence in NT-HRMS(/MS)



## MassBank Record: LU040605

(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]<sup>+</sup>

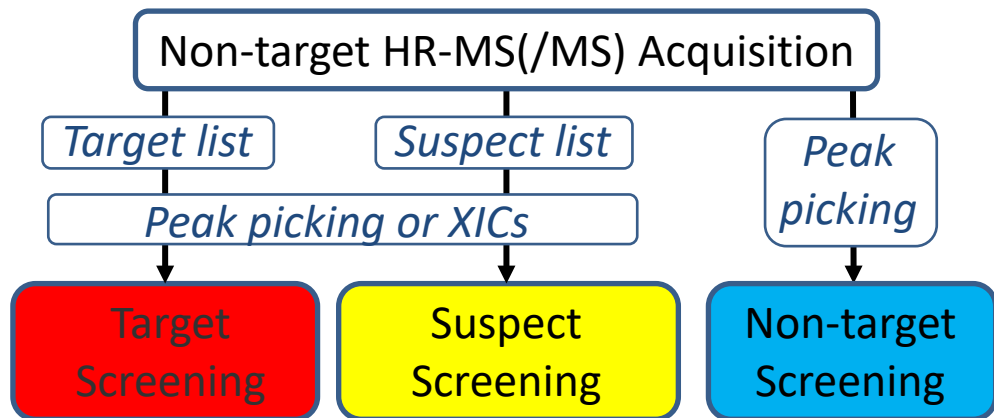


**eawag**  
aquatic research





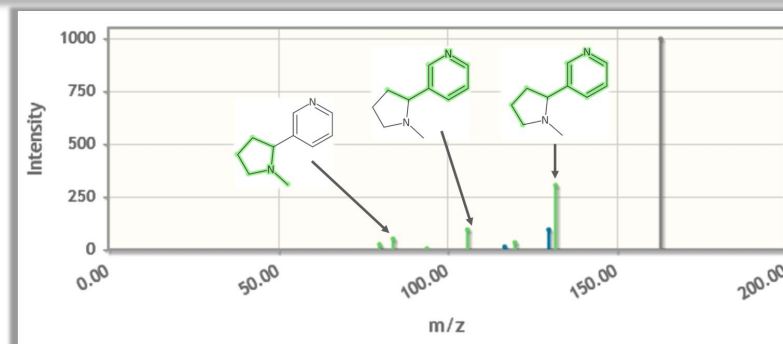
# Identification Strategies and Confidence in NT-HRMS(/MS)



Start			<b>Level 1</b>	<b>Confirmed Structure</b> by reference standard
			<b>Level 2</b>	<b>Probable Structure</b> by library/diagnostic evidence
	Start		<b>Level 3</b>	<b>Tentative Candidate(s)</b> suspect, substructure, class

Limited by ref. std.  
availability ...

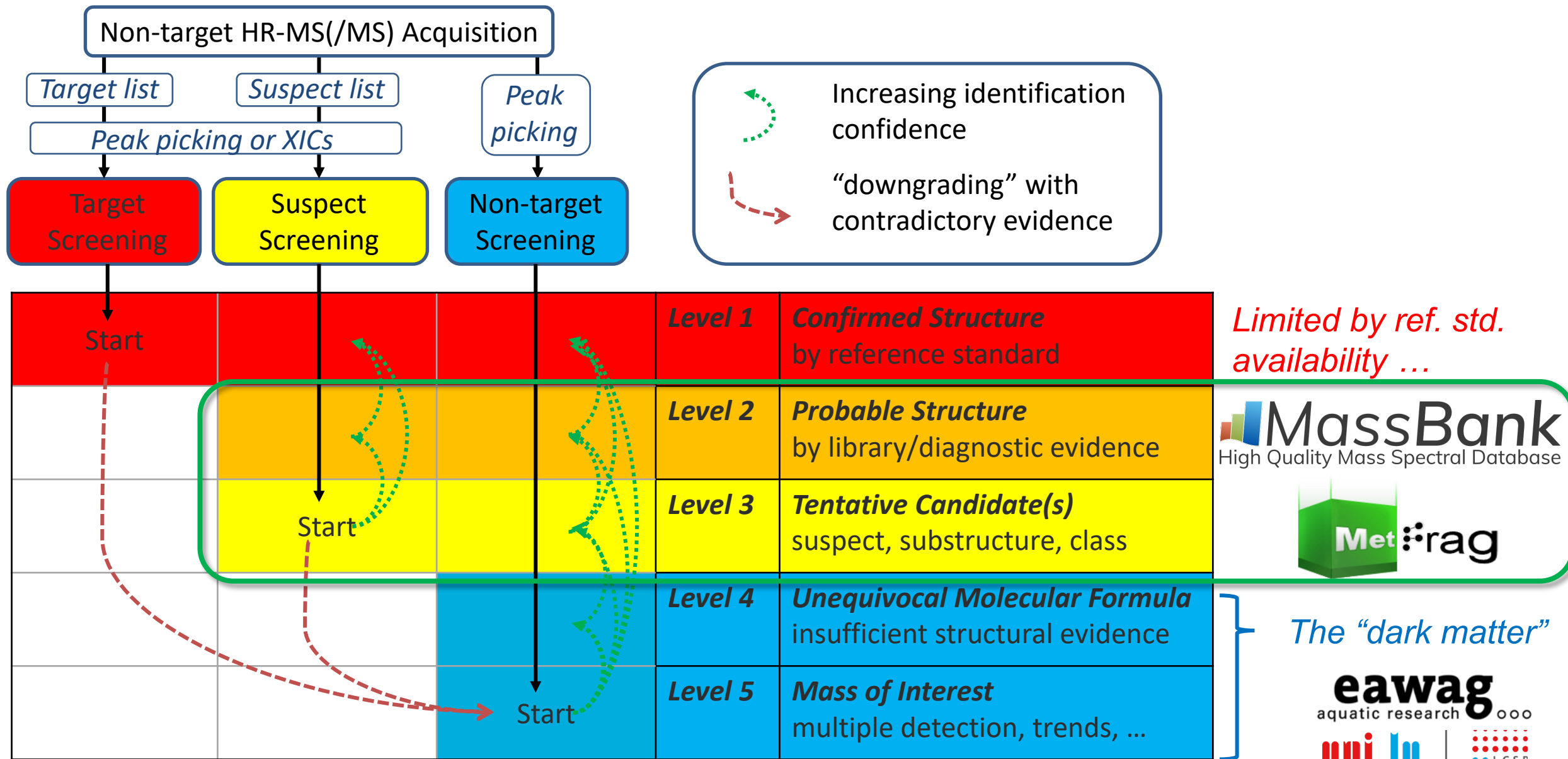
**MassBank**  
High Quality Mass Spectral Database



**eawag**  
aquatic research ooo



# Identification Strategies and Confidence in NT-HRMS(/MS)



# Grand Challenge: HR-MS “Chemical Space” is too Big!



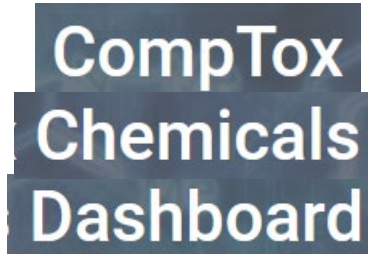
204 million



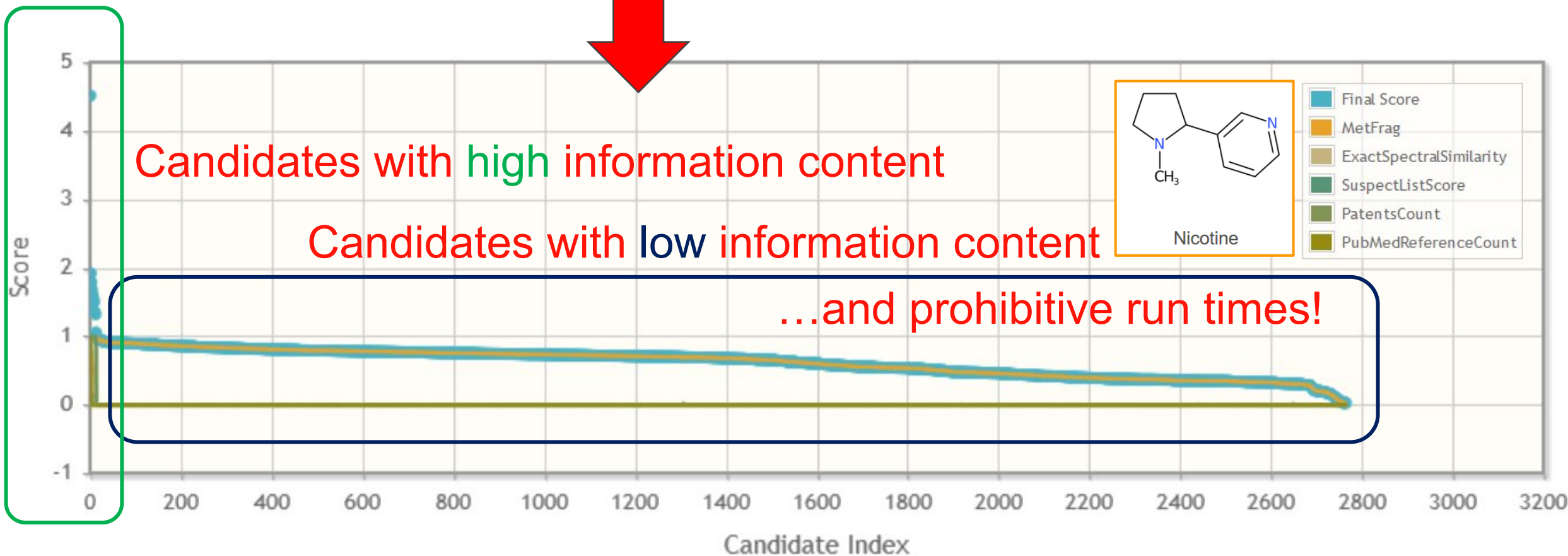
116 million



128 million



1.2 million



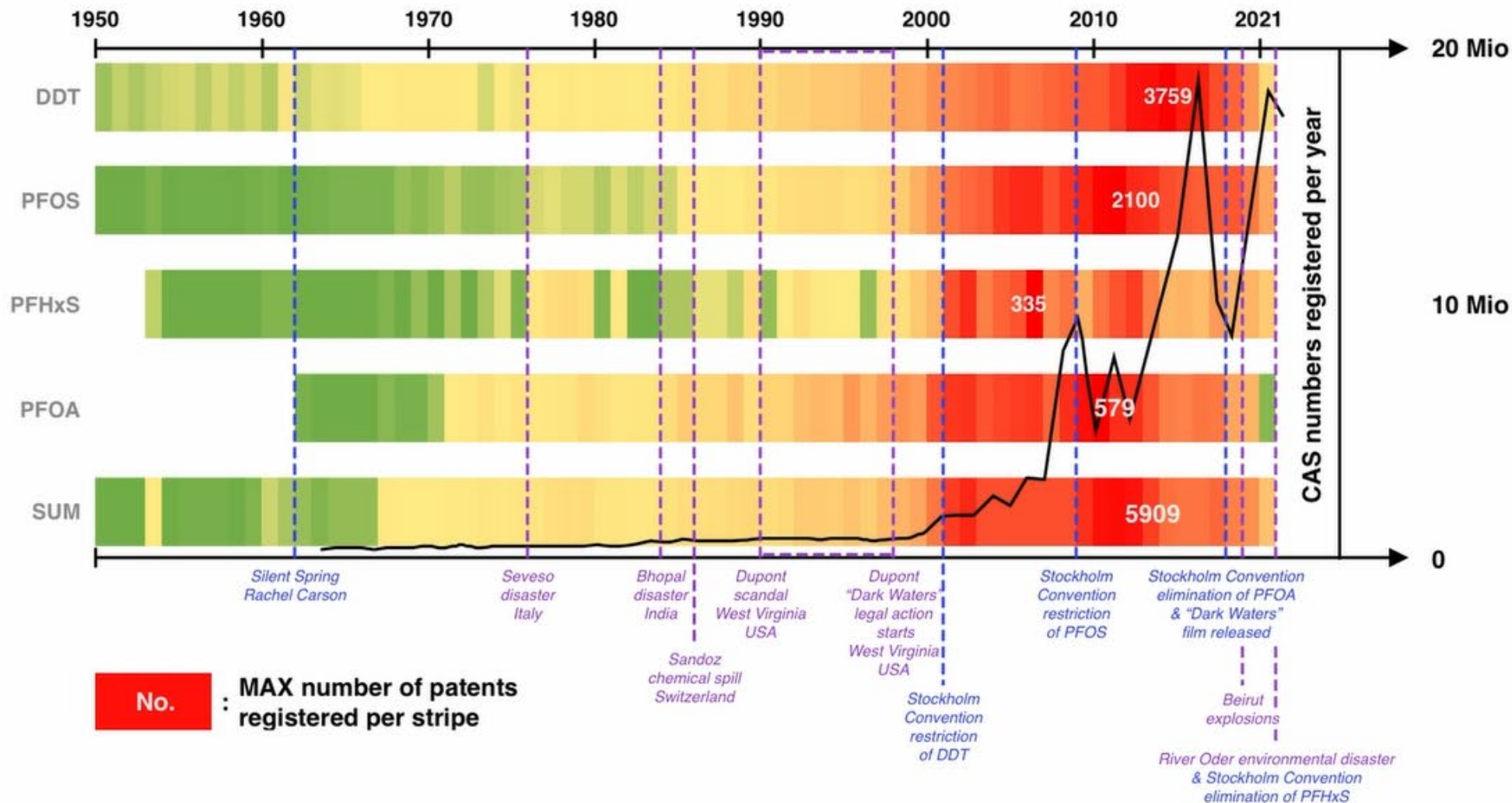


# Our Chemical Past, Present, and Future



JPML

low / medium / high number of patents registered per year (WIPO)



# Chemical Stripes in R - for Patents & Literature

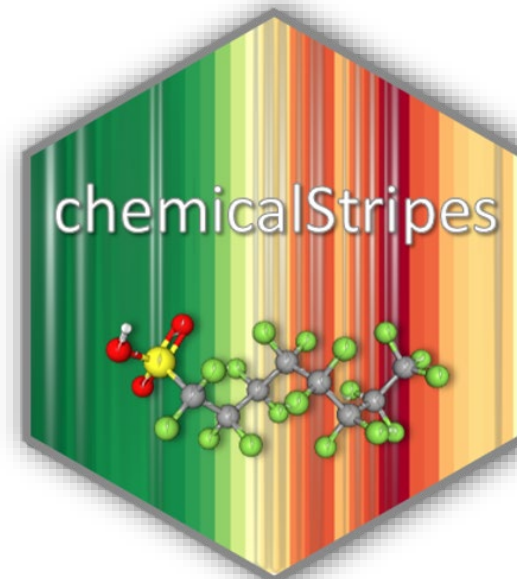
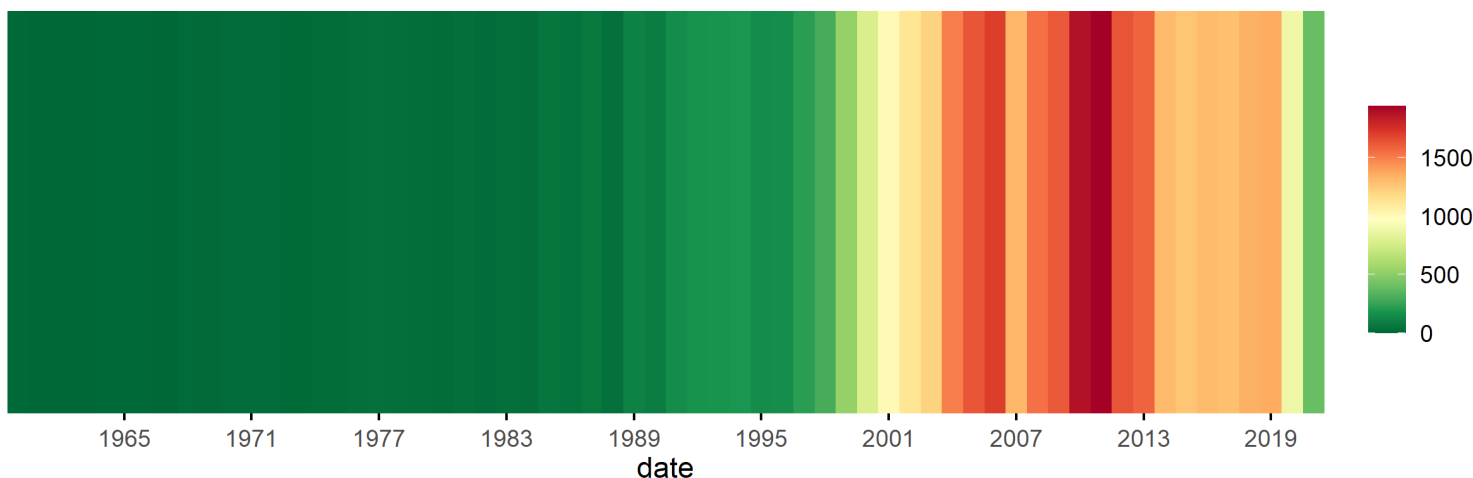


```
> chemical_stripes(74483)
Getting compound information
A total of 32461 patents were found for CID 74483
[=====]-----]
Downloading patent data...
[=====]-----] 60% 7s
Processing patent data
32460 patents were processed for CID 74483
[=====]-----] 80% 3s
Plotting chemical stripes for the years between 1960 and 2021

Your stripes have been saved as png_74483_1960_2021.png in your folder C:/Users/dagny.aurich/Documents/R_stripes/png_74483_1960_2021.png
```

## Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483  
IUPACName: 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluorooctane-1-sulfonic acid  
Molecular Formula: C<sub>8</sub>HF<sub>17</sub>O<sub>3</sub>S  
Exact Mass: 499.9374938



# Chemical Stripes in R

## Patents & Literature

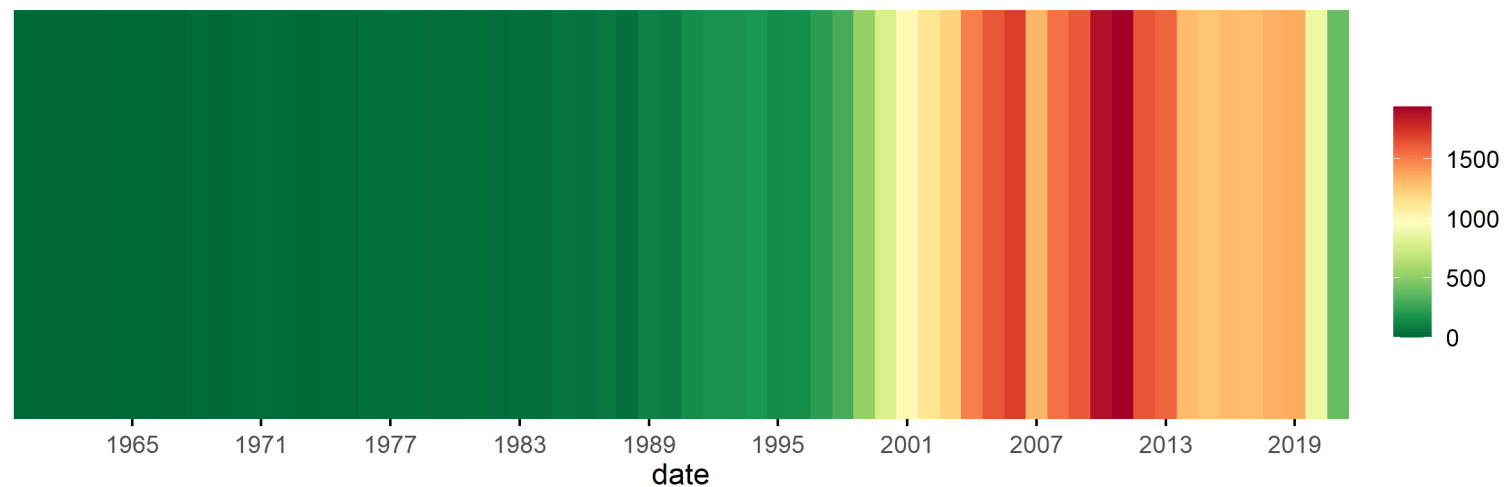


### Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First patent: 1913

Patents



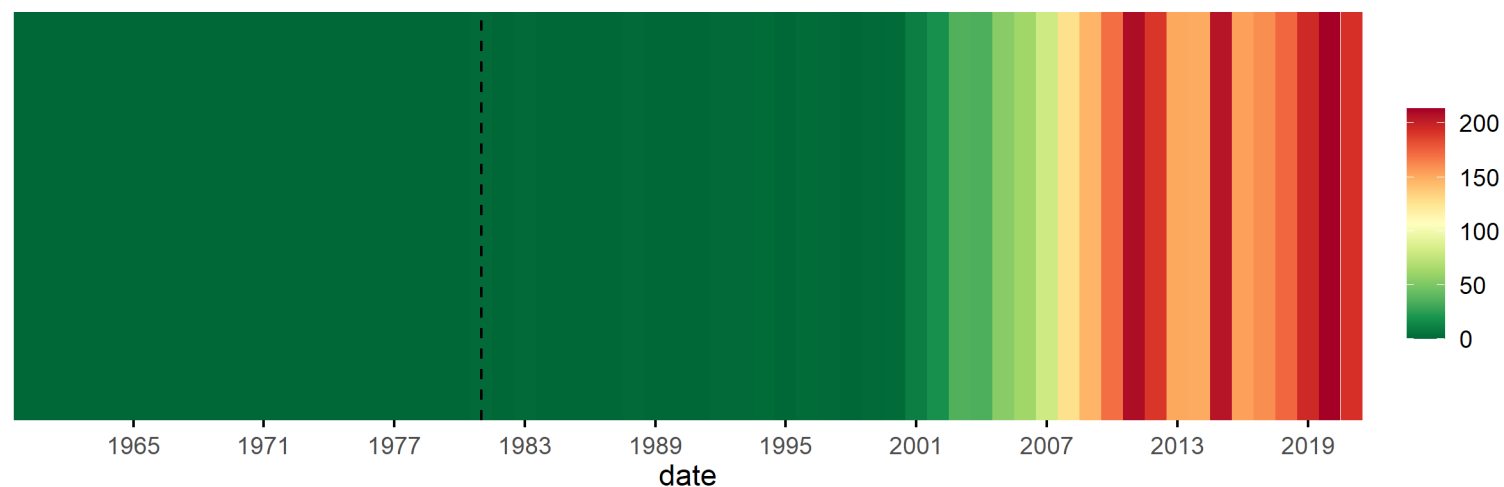
Depositor-supplied patent numbers

### Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First reference: 1981

Literature



Consolidated reference numbers

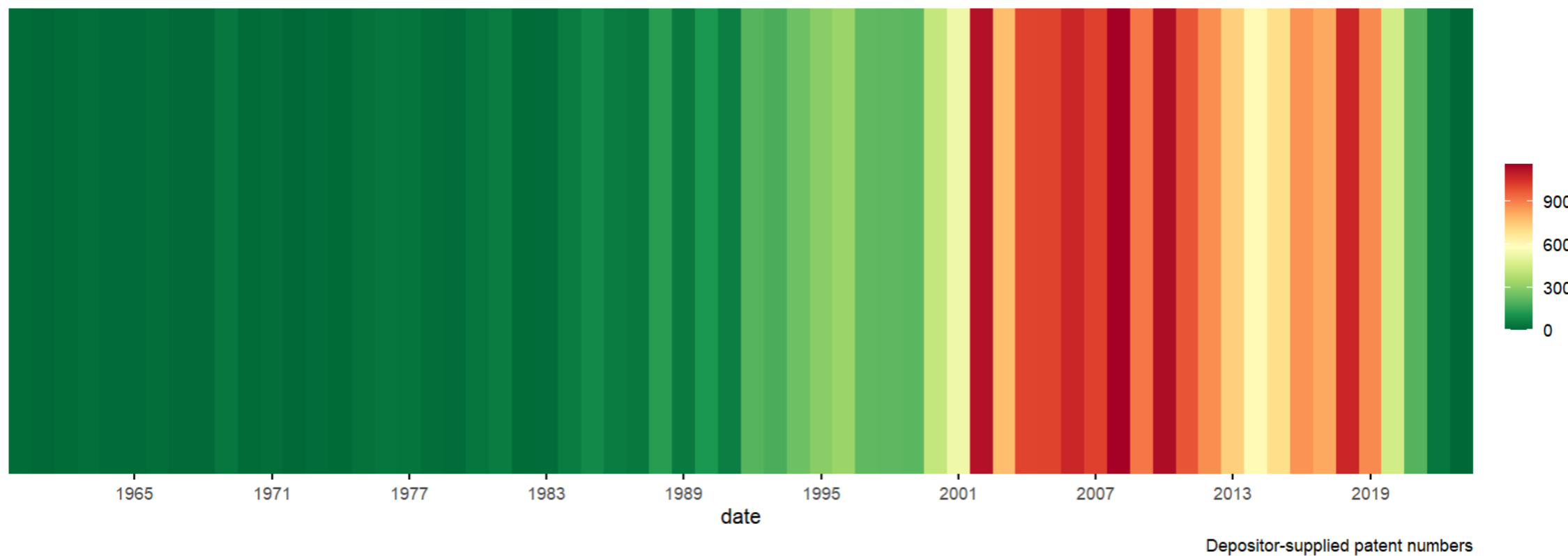
# “Summarized” Chemical Stripes in R



## Summarized Chemical Stripes - EU REACH C9-C14 PFCAs

Patent data compiled from 230 CIDs

First patent of all selected CIDs: 1951 (CID(s): 67822,67821,67545,9555 )

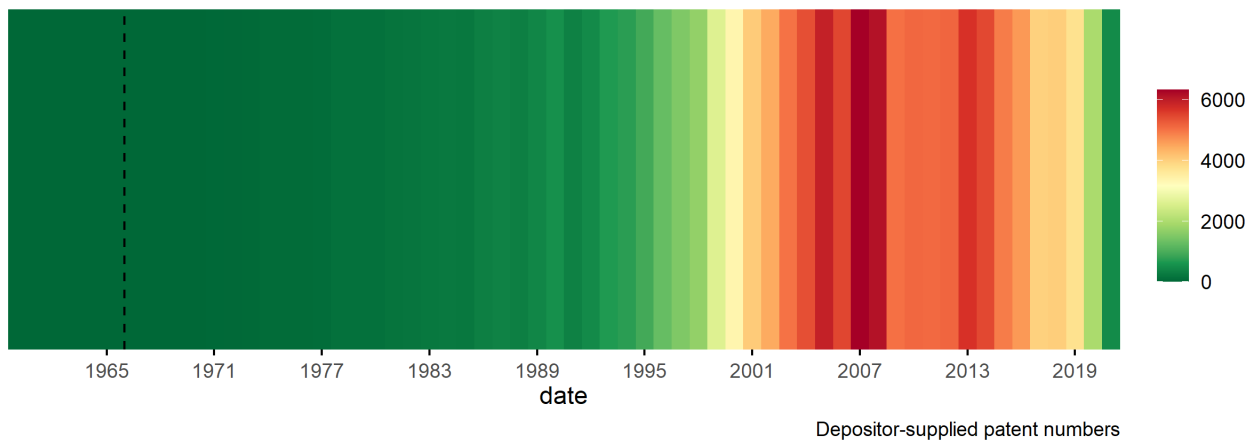


# Variety of Stripe Patterns: Patents



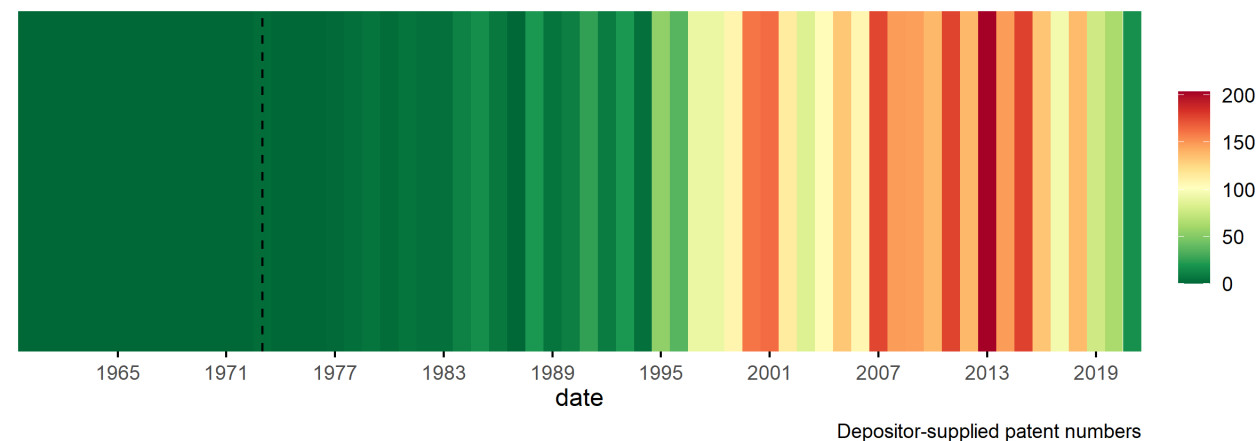
## Chemical Stripes for Caffeine

PubChem CID: 2519  
First patent: 1966



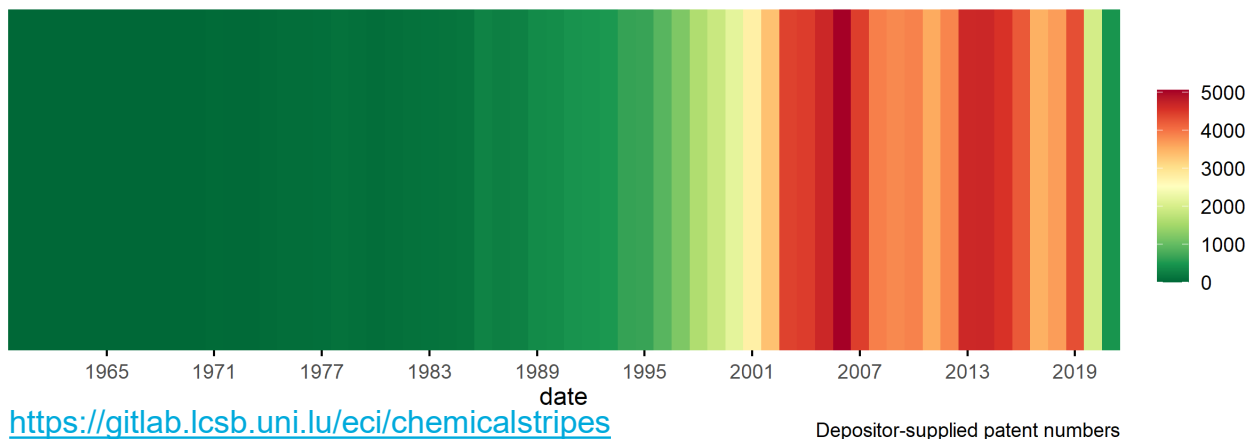
## Chemical Stripes for Acetylcarnitine

PubChem CID: 1  
First patent: 1973



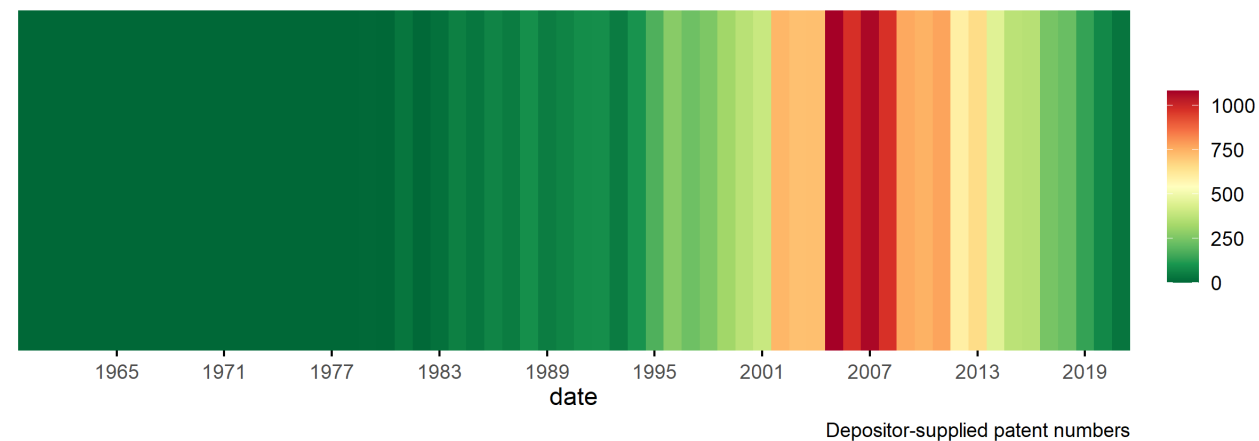
## Chemical Stripes for Nicotine

PubChem CID: 89594  
First patent: 1954



## Chemical Stripes for Gallopamil

PubChem CID: 1234  
First patent: 1951



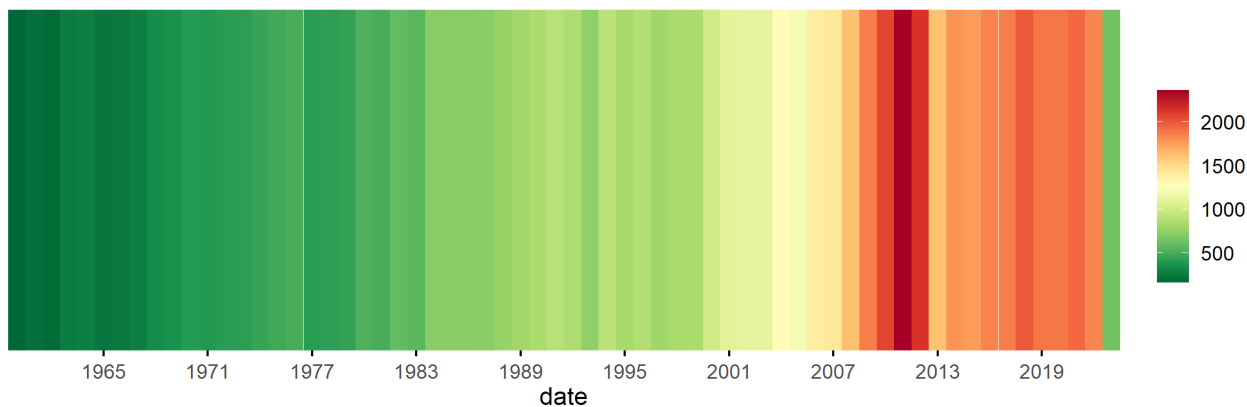


# Variety of Stripe Patterns: Literature



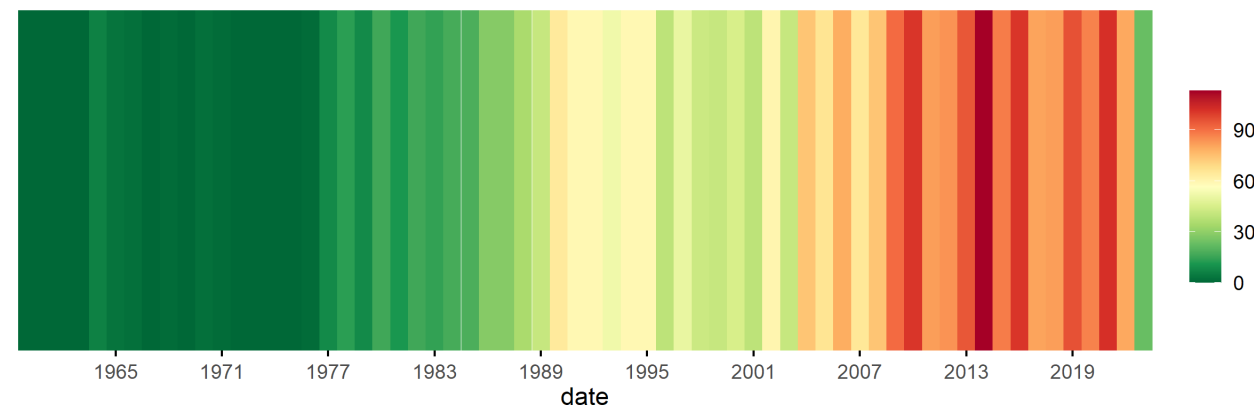
## Chemical Stripes for Caffeine

PubChem CID: 2519  
First reference: 1851



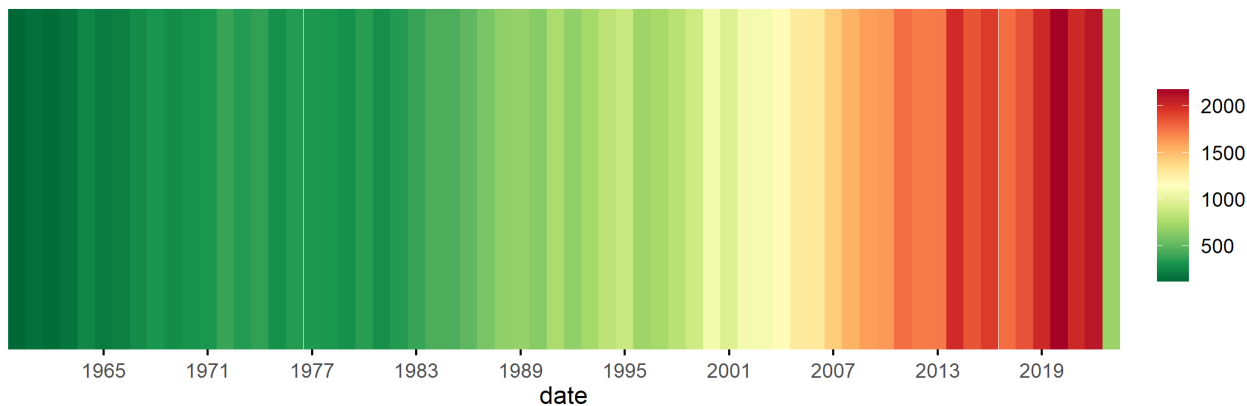
## Chemical Stripes for Acetylcarnitine

PubChem CID: 1  
First reference: 1959



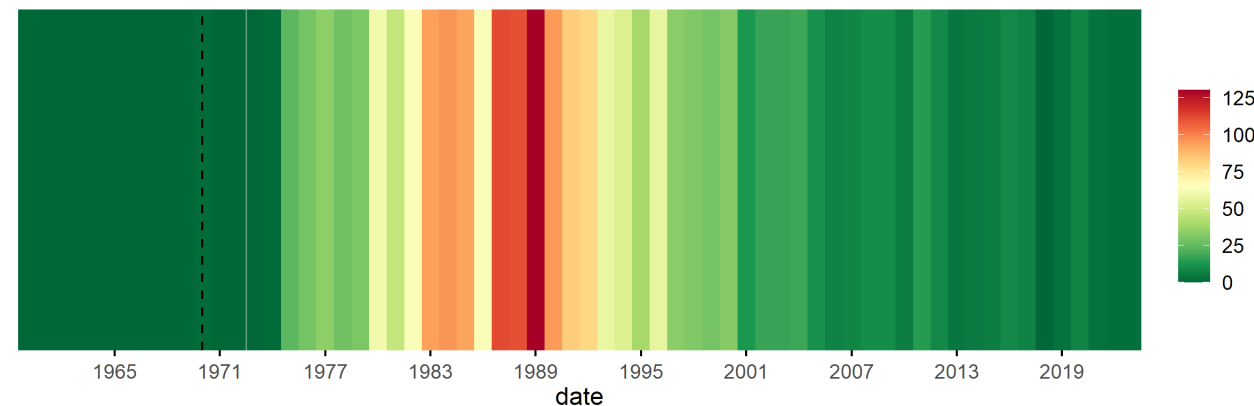
## Chemical Stripes for Nicotine

PubChem CID: 89594  
First reference: 1854



## Chemical Stripes for Gallopamil

PubChem CID: 1234  
First reference: 1970



<https://gitlab.lcsb.uni.lu/eci/chemicalstripes>

Consolidated reference numbers

Consolidated reference numbers

# The Chemical Stripes and Patent Data

(live stripe calculations during SETAC 2023)



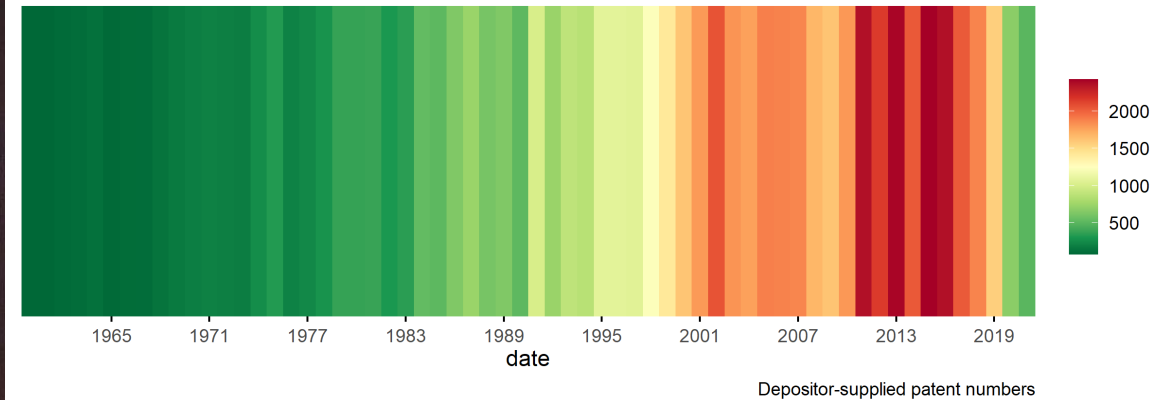
**Release of PM-chemicals into receiving waters**  
High concentration chemicals

... some high concentration PM-chemicals where not/only partially removed by at least one method

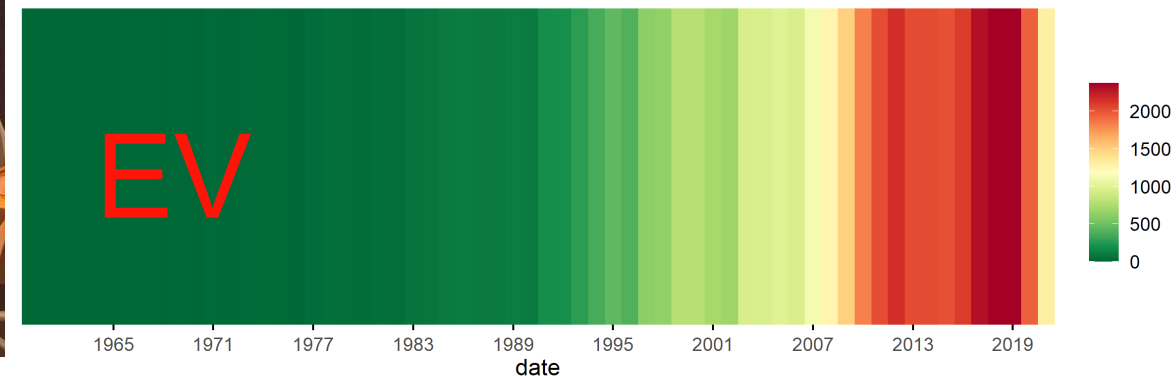
**Ionic liquid ions**  
PAC, O<sub>3</sub>  
PF<sub>6</sub> BF<sub>4</sub>  
NTf<sub>2</sub>  
+ 11 more detected in effluent

Neuwald et al. Science of the Total Environment (2023) accepted  
www.ufz.de 14

## Chemical Stripes for Tetrafluoroboric acid



## Chemical Stripes for Lithium tetrafluoroborate



Left: Neuwald et al, STOTEN, DOI: [10.1016/j.scitotenv.2023.163921](https://doi.org/10.1016/j.scitotenv.2023.163921)

Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

# Outline

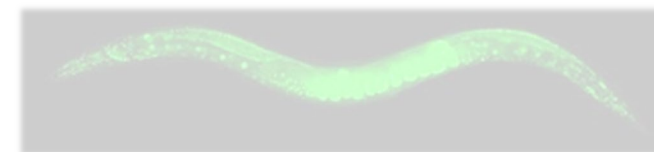
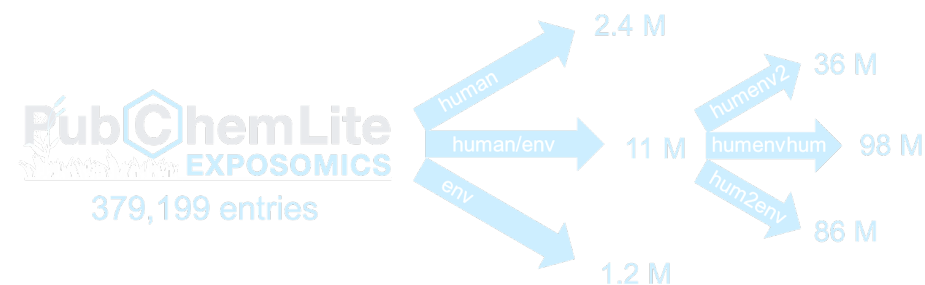
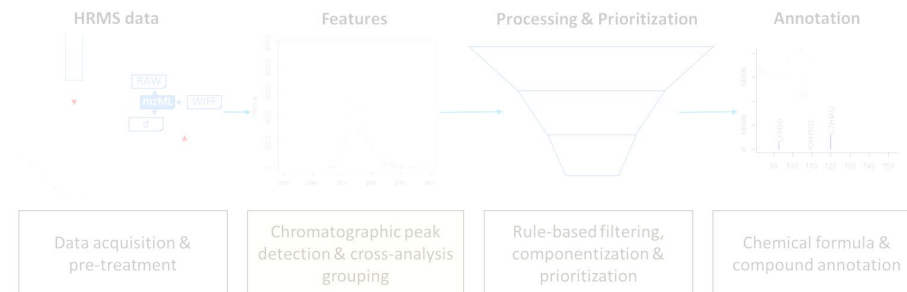
- Setting the Scene

- NT-HRMS & Databases
- The Great Data Challenge ...

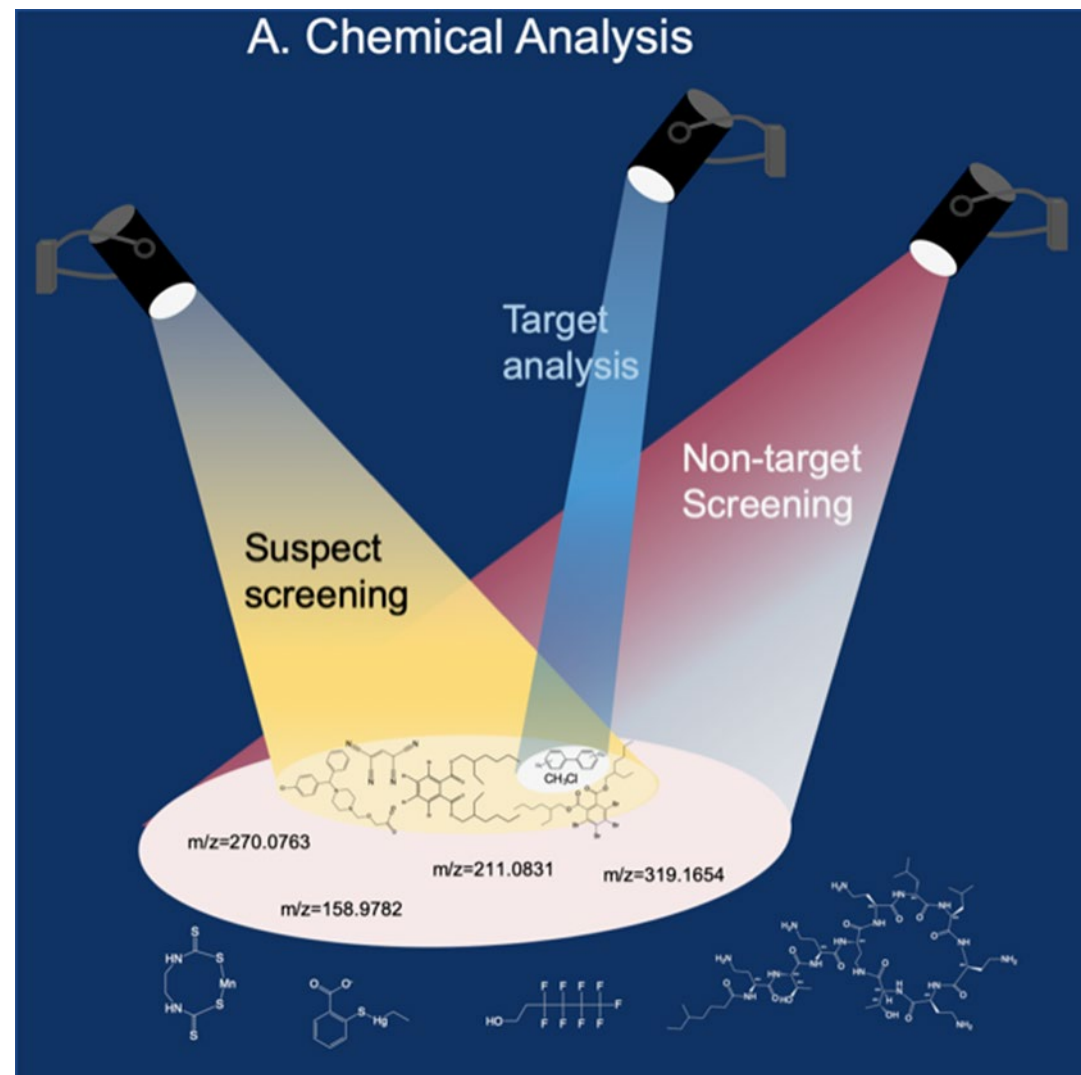
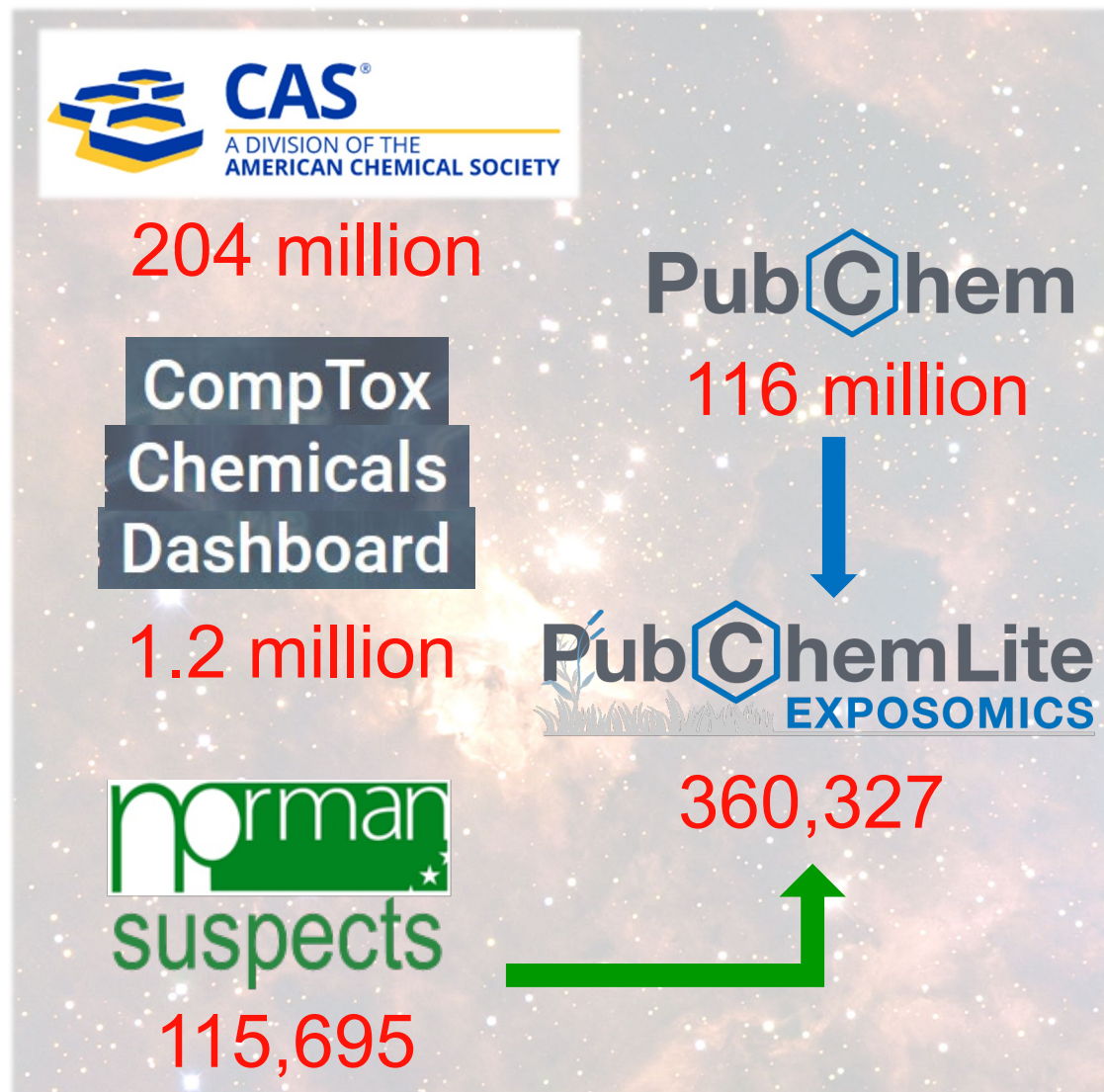
- Defining Chemical Space in Exposomics

- How to Tackle Transformations?

- Metabolomics / Exposomics-specific Workflows



# The Problem: Which chemicals are relevant? How to find them?



# Creating Informative Subsets of PubChem



PubChem Compound TOC ? 67,343,260

- ▶ Agrochemical Information ? 3,135
- ▶ Associated Disorders and Diseases ? 30,136
- ▶ Biologic Description ? 2,511,444
- ▶ Biological Test Results ? 4,567,078
- ▶ Chemical and Physical Properties ? 268,878
- ▶ Classification ? 22,965,005
- ▶ Drug and Medication Information ? 21,177
- ▶ Food Additives and Ingredients ? 7,627
- ▶ Identification ? 4,808
- ▶ Information Sources ? 47,725,078
- ▶ Interactions and Pathways ? 207,277
- ▶ Literature ? 4,076,955
- ▶ Names and Identifiers ? 7,021,765
- ▶ Patents ? 39,104,437
- ▶ Pharmacology and Biochemistry ? 114,060
- ▶ Related Records ? 13,282,616
- ▶ Safety and Hazards ? 184,712
- ▶ Spectral Information ? 1,576,070
- ▶ Structures ? 11,819,155
- ▶ Toxicity ? 118,115
- ▶ Use and Manufacturing ? 107,948

- ▼ Agrochemical Information ? 3,135
  - Agrochemical Category ? 1,977
  - Agrochemical Transformations ? 1,491
  - EU Pesticides Data ? 1,248
  - USDA Pesticide Data Program ? 663
- ▼ Associated Disorders and Diseases ? 30,136
  - Disease and References 13,386
- ▼ Interactions and Pathways ? 207,277
  - ▶ Molecular Imaging Information ? 1,910
  - ▶ Protein Bound 3D Structures ? 71,666
  - Chemical-Target Interactions ? 83,557
  - Drug-Drug Interactions ? 3,398
  - Drug-Food Interactions ? 1,246
- ▶ Pathways ? 62,525

PubChem Furathiocarb (Compound)

7.3 EU Pesticides Data

Active Substance	furathiocarb
Status	Not approved [Reg. (EC) No 1107/2009]
Legislation	2002/2076

PubChem Nicotine (Compound)

14 Associated Disorders and Diseases

Showing 11 to 15 of 252 items

Disease	Evidence Type	Evidence PMID
Airway Obstruction	marker/mechanism	33082140
Alzheimer Disease	therapeutic	16627626 17135361

PubChem Acetylcarnitine (Compound)

12.1 Pathways

4 items

Oxidation of Branched-Chain Fatty Acids

Source: PathBank External ID: SMP0000030

Taxonomy: Homo sapiens (human)

# Creating Informative Subsets of PubChem



## PubChemLite EXPOSOMICS

~350,000 entries "small"



- PubChem Compound TOC ? 67,343,260
  - Agrochemical Information ? 3,135
  - Associated Disorders and Diseases ? 30,136
  - Biologic Description ? 2,511,444
  - Biological Test Results ? 4,567,078
  - Chemical and Physical Properties ? 268,878
  - Classification ? 22,965,005
  - Drug and Medication Information ? 21,177
  - Food Additives and Ingredients ? 7,627
  - Identification ? 4,808
  - Information Sources ? 47,725,078
  - Interactions and Pathways ? 207,277
  - Literature ? 4,076,955
  - Names and Identifiers ? 7,021,765
  - Patents ? 39,104,437
  - Pharmacology and Biochemistry ? 114,060
  - Related Records ? 13,282,616
  - Safety and Hazards ? 184,712
  - Spectral Information ? 1,576,070
  - Structures ? 11,819,155
  - Toxicity ? 118,115
  - Use and Manufacturing ? 107,948



Updated PubChemLite to Aug 2023 version #127

Merged sneumann merged 2 commits into ipb-halle:master from schyman:master last week

Database Settings

Database: PubChemLite\_exposomics

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16CIN5

Identifiers:

Retrieve Candidates 4 Candidates

zenodo Search Upload Communities

August 26, 2023 Dataset Open Access

### PubChemLite for Exposomics

Bolton, Evan; Schyman, Emma; Kondic, Todor; Thiessen, Paul; Zhang, Jian (Jeff)

This is the repository for regular updates of the PubChemLite for Exposomics data collection. PubChemLite for Exposomics is a subset of PubChem selected from major categories of the Table of Contents page at the PubChem Classification Browser, described in DOI:10.1186/s13321-021-00489-0.

PubChemLite for Exposomics is compiled from 10 categories: AgroChemInfo, BioPathway, DrugMedicInfo, FoodRelated, PharmacolInfo, SafetyInfo, ToxicityInfo, KnownUse, DisorderDisease, Identification.

PubChemCIDs have been collapsed by InChIKey first block, reporting the structure from the most annotated CID, plus related CIDs. Entries that will be ignored by MetFrag (salts, disconnected substances) or cause errors (e.g. transition metals) have been removed. The Patent and PubMed ID counts are extracted from files on the PubChem FTP site. The 'AnnoTypeCount' term counts how many of the categories are represented, the subsequent column (named per category) counts the number of annotation categories available in the next sub-category of the TOC entry.

These files can be used 'as is' as localCSV for MetFrag Command Line.

Schyman et al. (2021) DOI: [10.1186/s13321-021-00489-0](https://doi.org/10.1186/s13321-021-00489-0)

# PubChemLite: Fewer and more relevant candidates – with context

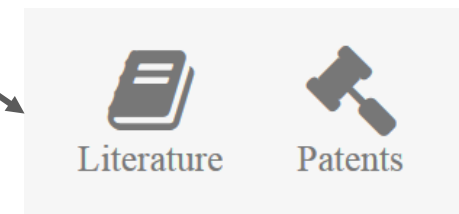
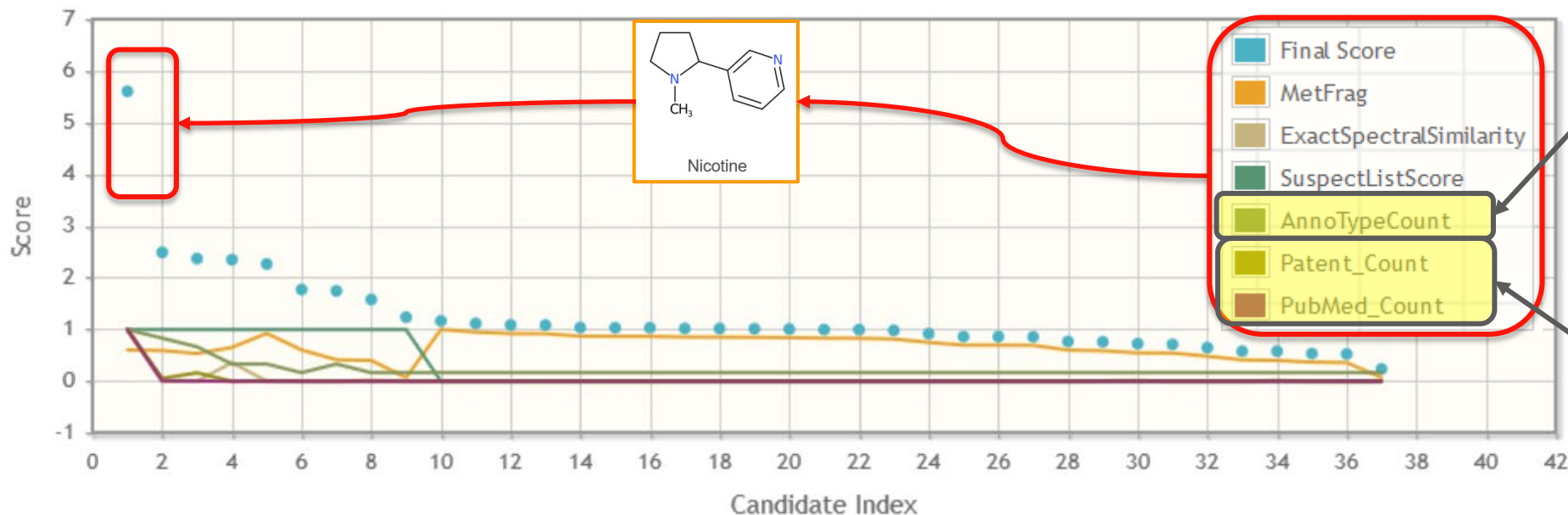


PubChem Compound TOC ? 67,343,260

- Agrochemical Information ? 3,135
- Associated Disorders and Diseases ? 30,136
- Biologic Description ? 2,511,444
- Biological Test Results ? 4,567,078
- Chemical and Physical Properties ? 268,878
- Classification ? 22,965,005
- Drug and Medication Information ? 21,177
- Food Additives and Ingredients ? 7,627
- Identification ? 4,808
- Information Sources ? 47,725,078
- Interactions and Pathways ? 207,277
- Literature ? 4,076,955
- Names and Identifiers ? 7,021,765
- Patents ? 39,104,437
- Pharmacology and Biochemistry ? 114,060
- Related Records ? 13,282,616
- Safety and Hazards ? 184,712
- Spectral Information ? 1,576,070
- Structures ? 11,819,155
- Toxicity ? 118,115
- Use and Manufacturing ? 107,948

## Statistics

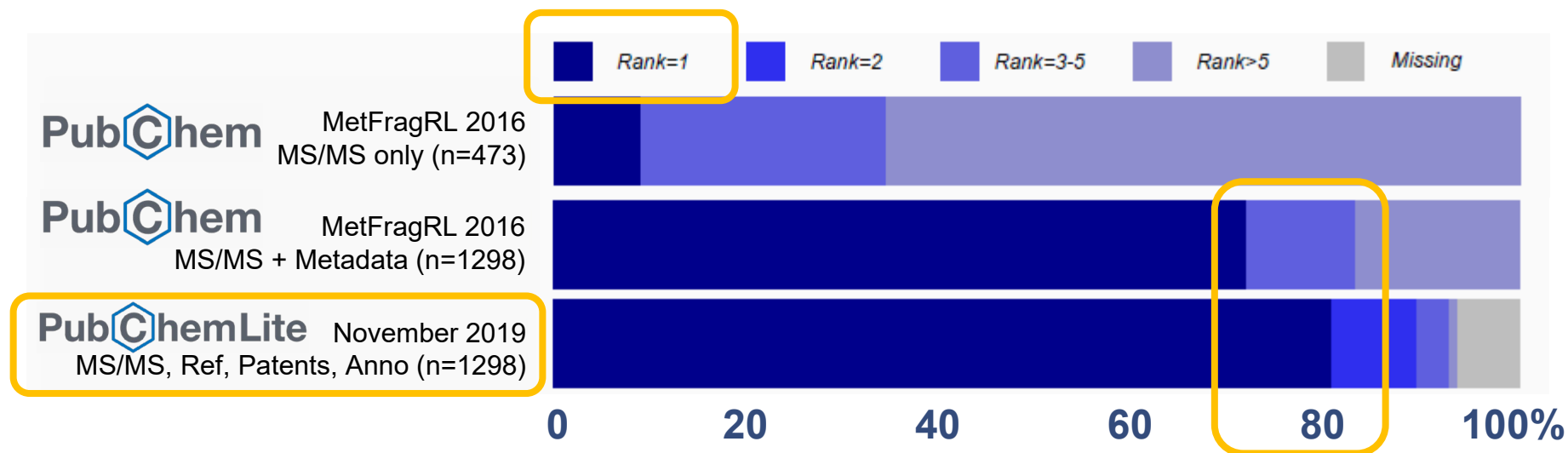
### Candidate Score Distribution



# How does PubChemLite perform?



- ~111 M => ~400 K ... how does this influence performance?

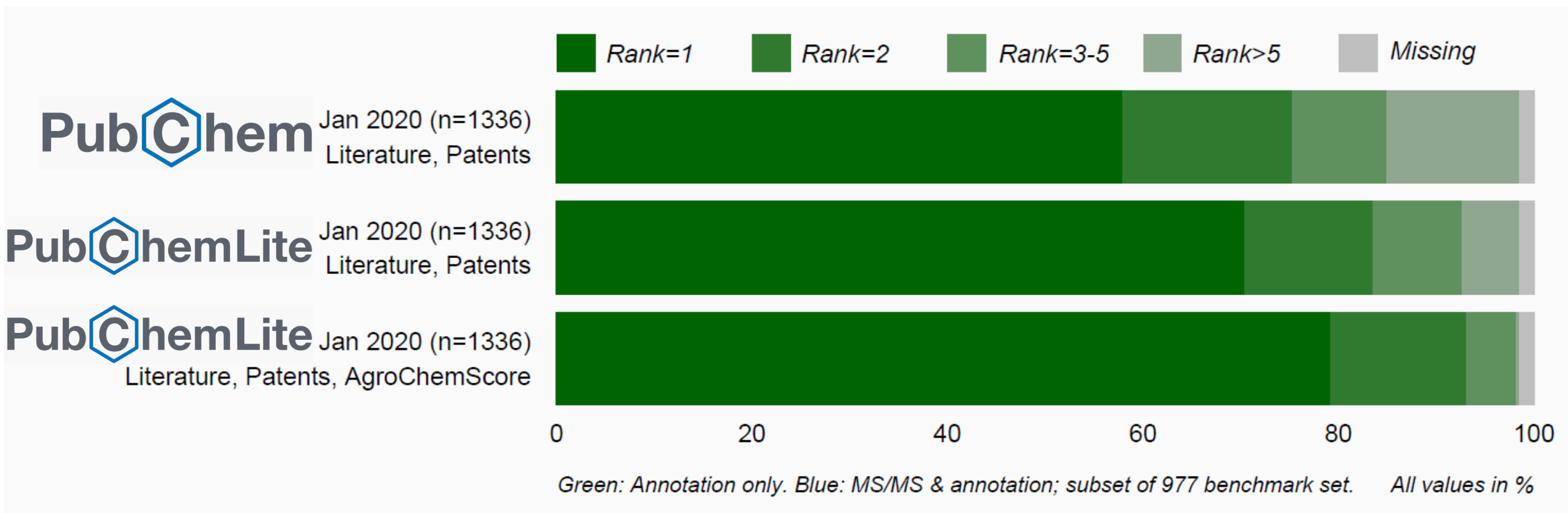


>80 % ranked in first place  
~90 % ranked first or second!

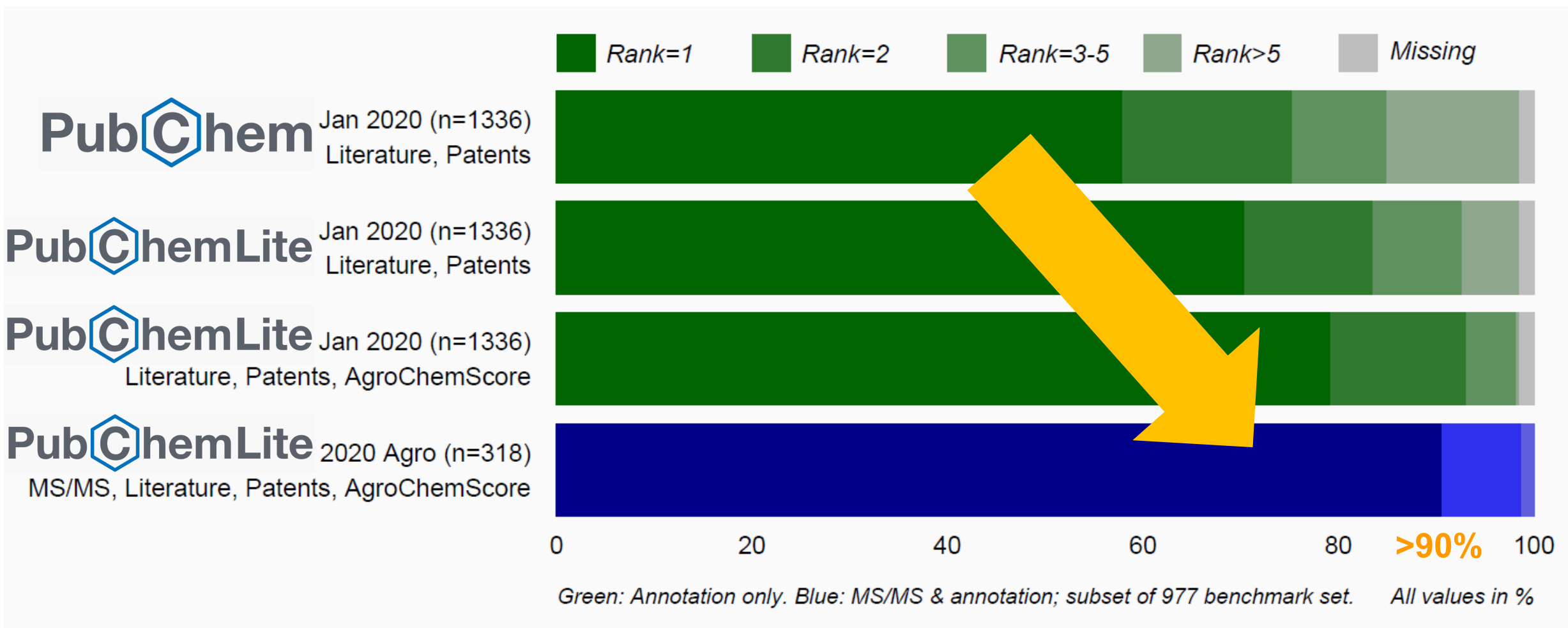
} high throughput identification



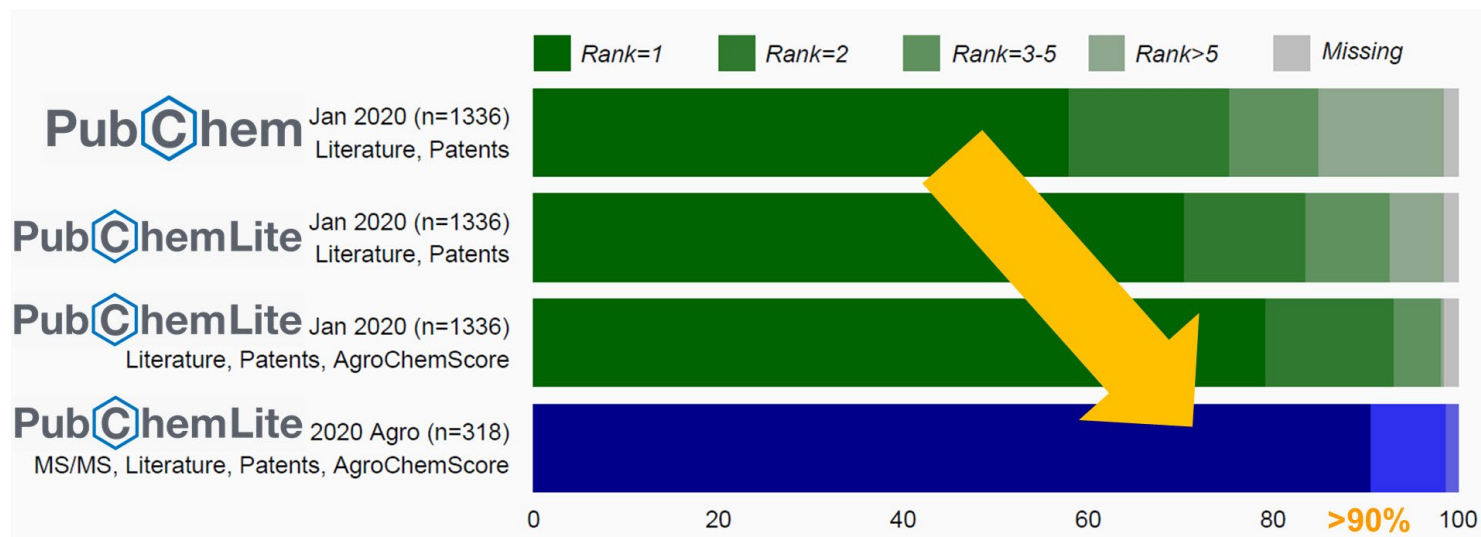
# Influence of the Annotation Content in PubChemLite



# Influence of the Annotation Content in PubChemLite



# Influence of the Annotation Content in PubChemLite



>90% of (well known) entries are Top 1 rank

>97% are Top 2 ... 100 % Top 3 ...

***Exposomics: Top 1-3 matches will cover a lot!  
(and quickly!)***

# PubChemLite is a Dynamic Collection ...



- ... built every week, updated online every month
- ... re-evaluated with every build for continuous quality control
- ... growing as knowledge grows!

**Emma Schymanski**  
@ESchymanski

Our PubChemLite for #Exposomics collection hit a milestone last week, cracking the 400,000 entry mark with updated safety annotation content in @pubchem Monthly (approx.) updates since mid 2021 are now available here (DOI redirects to latest version): [doi.org/10.5281/zenodo...](https://doi.org/10.5281/zenodo)

**PubChemLite EXPOSOMICS**  
~370,000 entries "small"

20220304	20220311	Diff	
AnnoTypeCount	392609	403507	10898
AgroChemInfo	2426	2426	0
BioPathway	119924	120099	175
DrugMedicInfo	8780	8781	1
FoodRelated	4671	4696	25
PharmacolInfo	90218	90223	5
SafetyInfo	125998	138796	12798
ToxicityInfo	101343	101343	0
KnownUse	77502	77519	17
DisorderDisease	19551	19551	0
Identification	3180	3180	0

## Versions

Version 1.25.0	Aug 26, 2023
10.5281/zenodo.8285969	
Version 1.24.0	Jul 28, 2023
10.5281/zenodo.8191746	
Version 1.23.0	Jun 30, 2023
10.5281/zenodo.8099155	
Version 1.22.0	May 26, 2023
10.5281/zenodo.7973696	
Version 1.21.0	Apr 29, 2023
10.5281/zenodo.7878471	

[View all 26 versions](#)

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.5995885](https://doi.org/10.5281/zenodo.5995885). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

DOI: [10.5281/zenodo.5995885](https://doi.org/10.5281/zenodo.5995885)

# Putting PubChemLite into Practice?



- Downloadable & fully integrated in patRoom & MetFrag

zenodo

Search [ ] Upload Communities emma.schymanski@uni.lu

Zenodo.org will be unavailable for 2 hours on September 29th from 06:00-08:00 UTC. See announcement.

August 26, 2023

PubChemLite for Exposomics

Dataset Open Access

Communities: LCSB Environmental Cheminformatics Group

2,087 views 1,778 downloads

See more details...

MetFragWeb Learn more Use MetFrag directly from your browser which makes it pretty easy for beginners

MetFrag CL Learn more The new MetFrag commandline tool providing additional scoring terms for scoring candidates with MS/MS spectra

<https://rickhelmus.github.io/patRoom/>

patRoom 1.2.0

Installation

Getting start

For a very quick start:

```
library(patRoom)
newProject()
```

https://github.com/rickhelmus/patRoom/issues

License: GPL-3

Citation: Citing patRoom

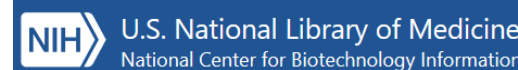
Dev status: PASSED, build passing, codecov 82%, image size 2.17 GB

DOI: [10.5281/zenodo.5995885](https://doi.org/10.5281/zenodo.5995885)

<https://ipb-halle.github.io/MetFrag/>



UNIVERSITY OF AMSTERDAM



U.S. National Library of Medicine  
National Center for Biotechnology Information



# Outline

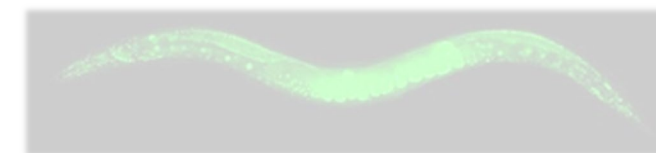
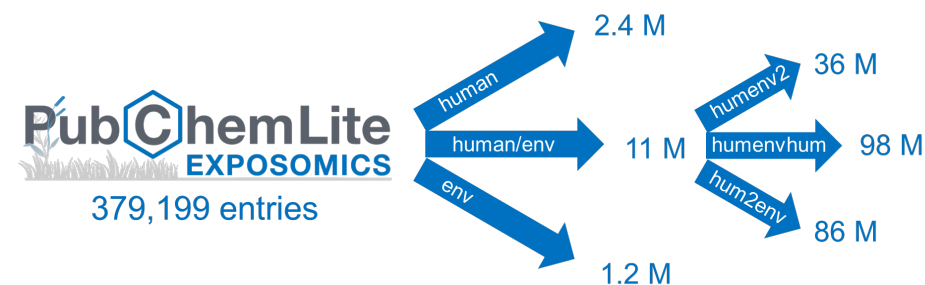
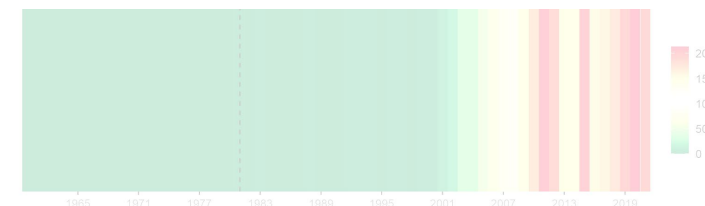
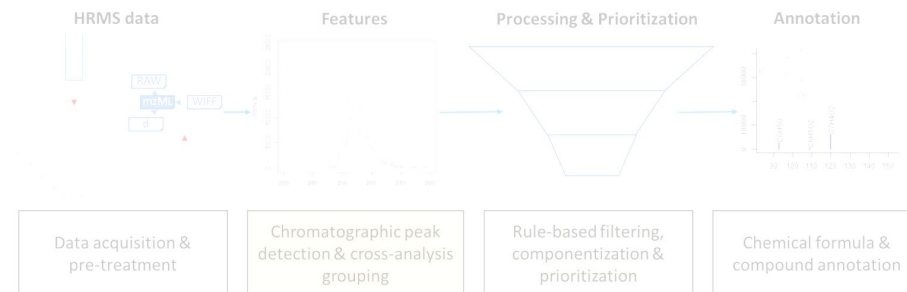
- Setting the Scene

- NT-HRMS & Databases
- The Great Data Challenge ...

- Defining Chemical Space in Exposomics

- How to Tackle Transformations?

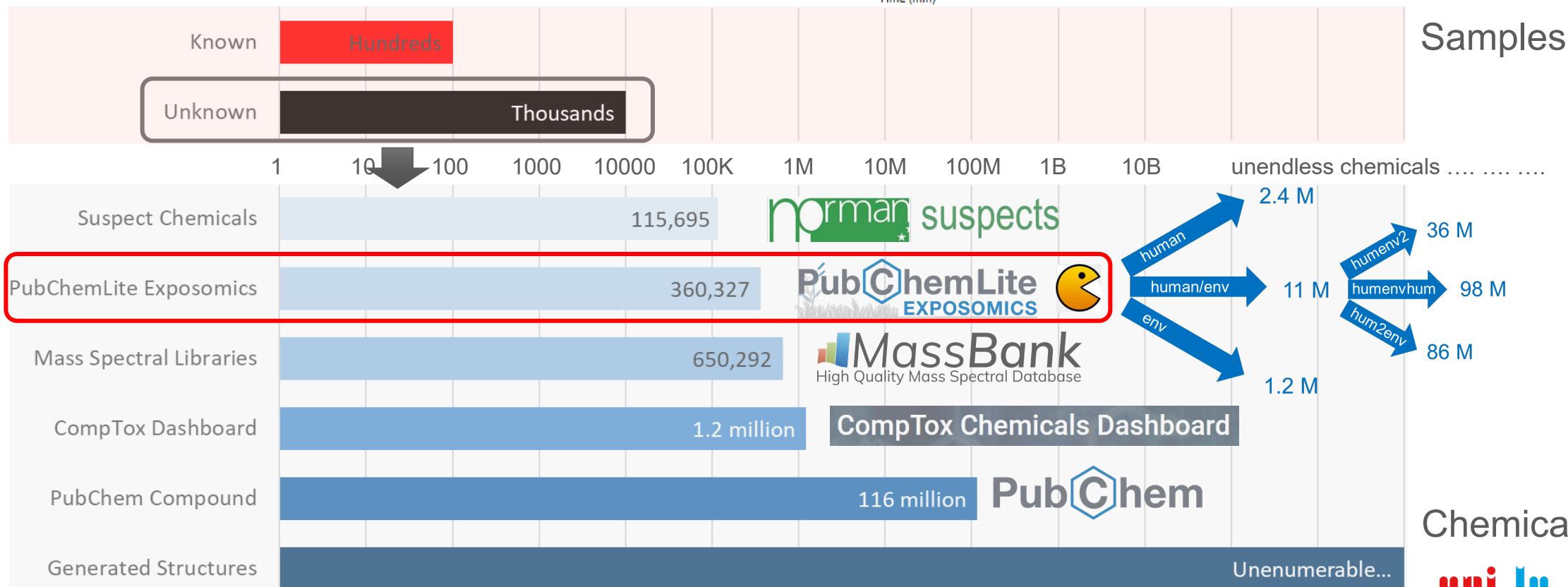
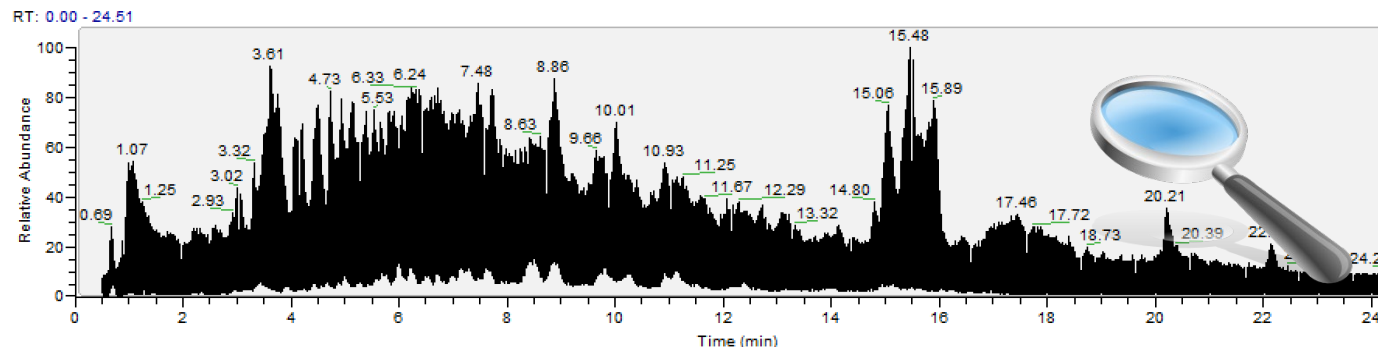
- Metabolomics / Exposomics-specific Workflows



# Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution  
mass spectrometry

AND connecting  
chemical knowledge



Samples

Chemical space

# Transforming PubChemLite with **BioTransformer 3.0**

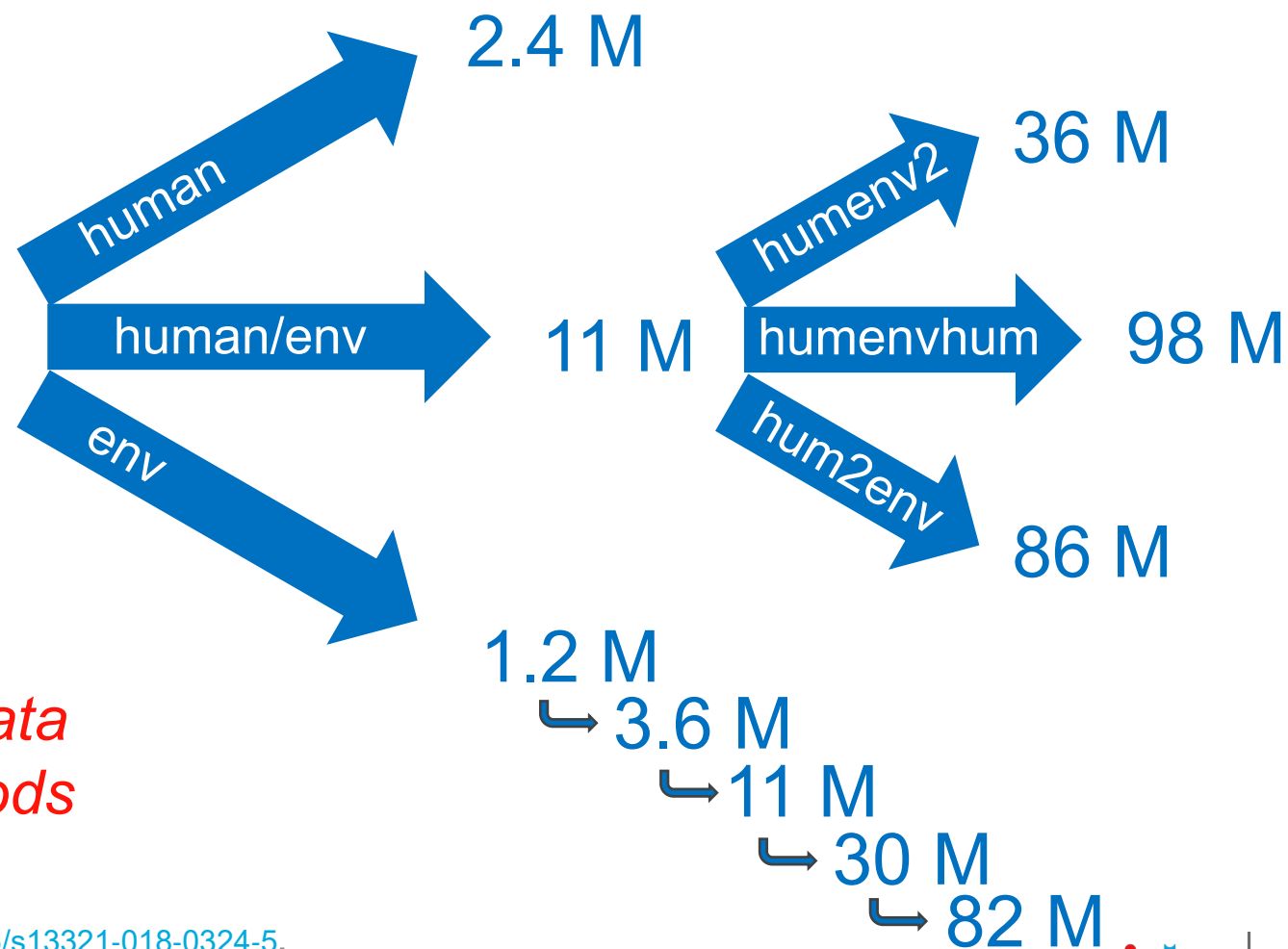
<http://biotransformer.ca/>

**PubChemLite**  
EXPOSOMICS

379,199 entries

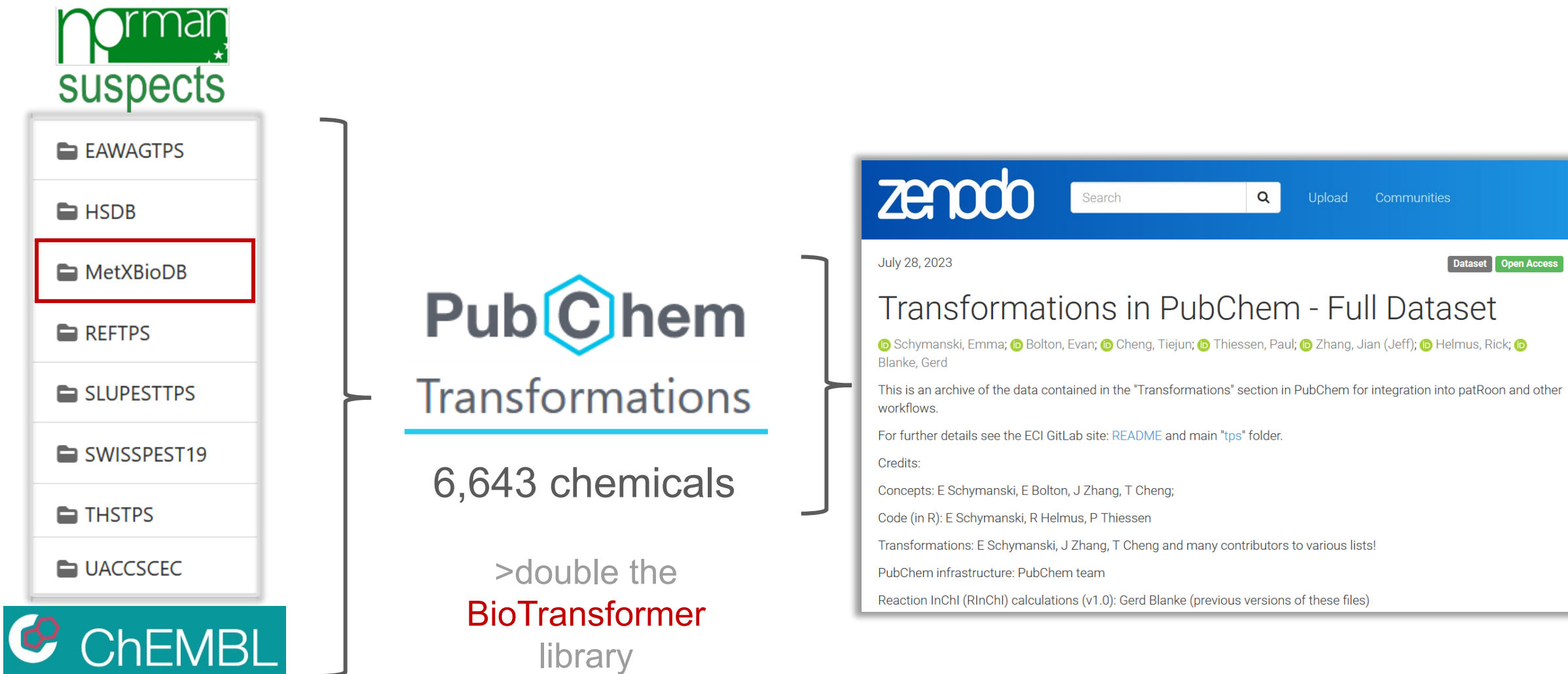
version 1.0.0; DOI: [10.5281/zenodo.5995886](https://doi.org/10.5281/zenodo.5995886)

*Combinatorial explosion – more data  
needed to predict reaction likelihoods*





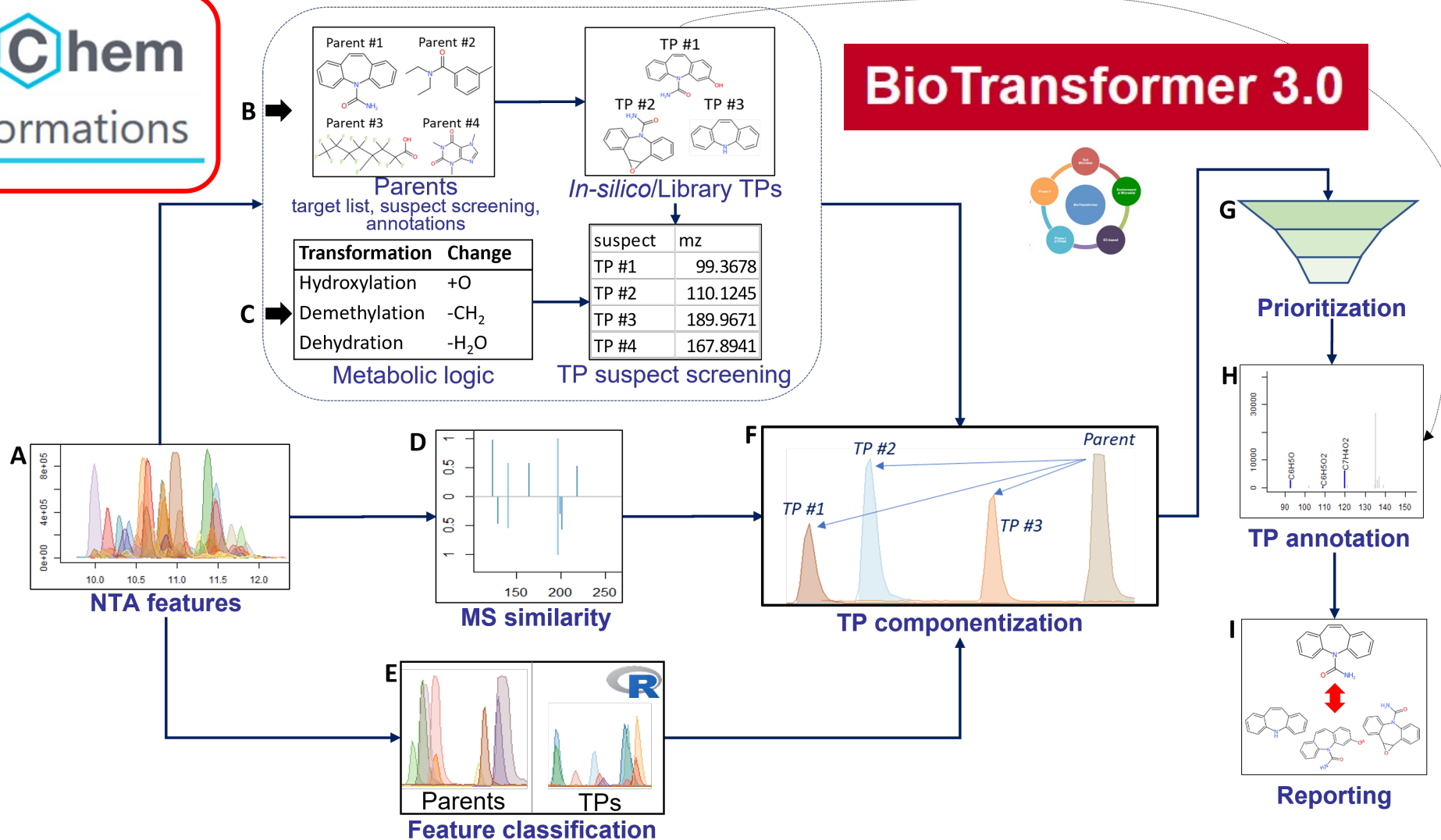
# FAIR Transformations in PubChem and NORMAN-SLE



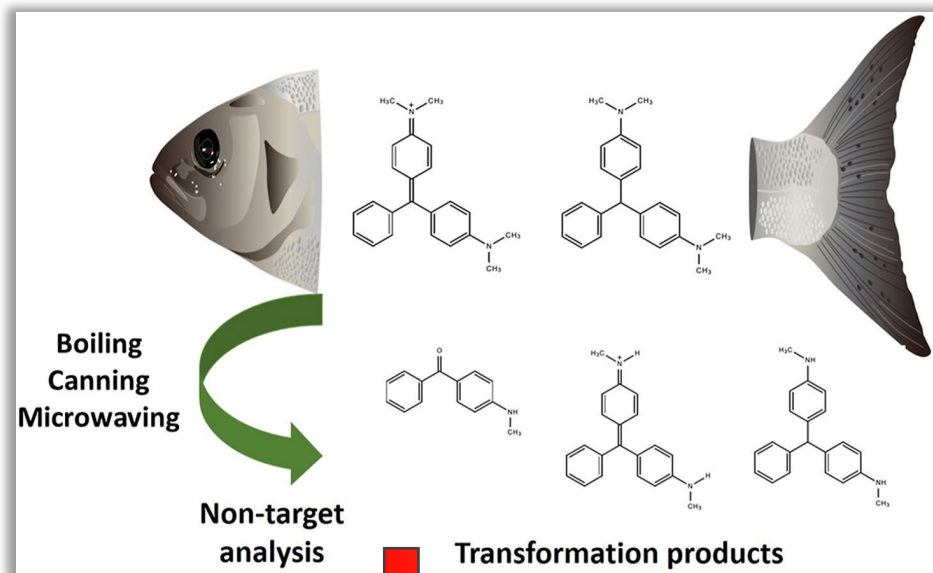
# Open Transformation Products Workflows in patRoou 2.0



PubChem  
Transformations



# What about *new* information? Help add it in!



zenodo

Search [ ] [Q] Upload Communities

March 8, 2023 Dataset Open Access

## S74 | REFTPS | Transformation Products and Reactions from Literature

Schymanski, Emma; Baesu, Anca; Chirsir, Parviel

This is the collection associated with list S74 REFTPS Transformation Products and Reactions from Literature on the NORMAN Suspect List Exchange.

<https://www.norman-network.com/nds/SLE/>

## PubChem (4-(Methylamino)phenyl(phenyl)methanone (Compound)

### 7.1 Transformations

2 items Download

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Evidence DO
	Malachite green		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation	10.1016/j.crf.
	Malachite green cation		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation	10.1016/j.crf.

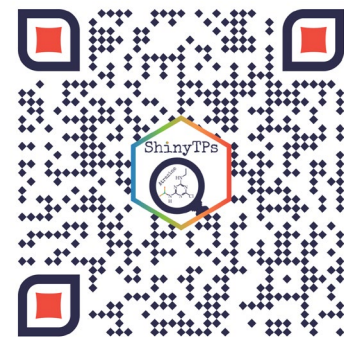
< >

▼ NORMAN Suspect List Exchange

Source	NORMAN Suspect List Exchange
Record Name	4-(Methylamino)benzophenone



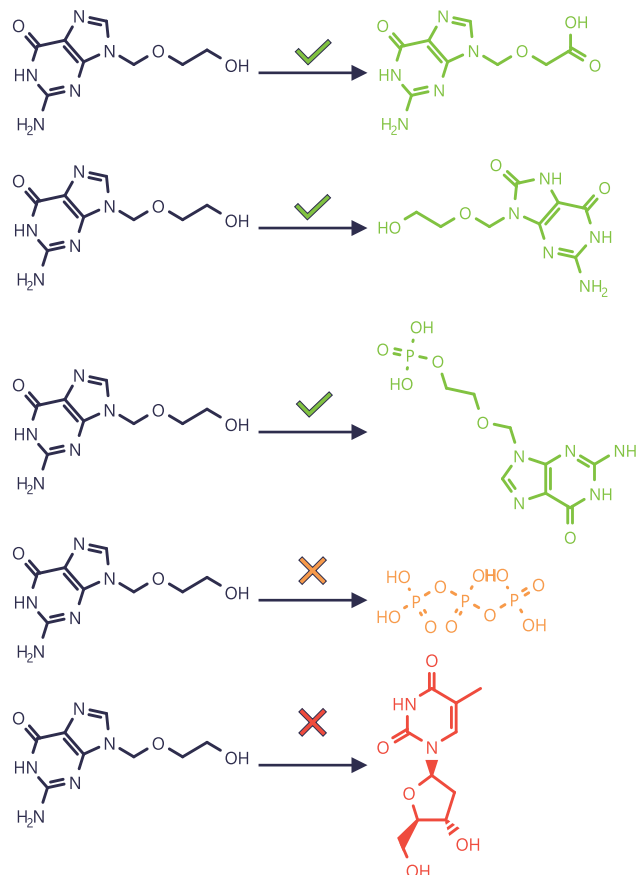
# Curating Text-mined Content in PubChem



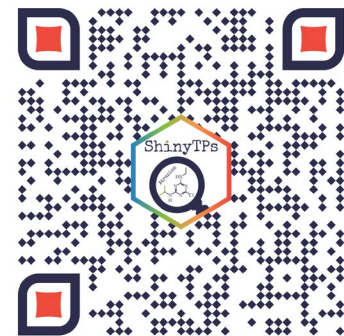
PubChem Acyclovir (Compound)

## 8.6 Metabolism/Metabolites

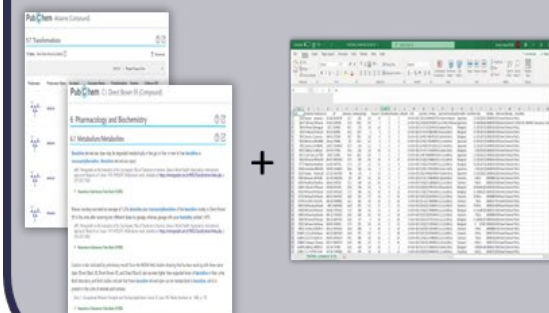
**Acyclovir** is metabolized partially to **9-carboxymethoxymethylguanine** and minimally to **8-hydroxy-9-(2-hydroxyethoxymethyl)guanine**. In vitro, **acyclovir** also is metabolized to **acyclovir monophosphate**, diphosphate, and **triphosphate** in cells infected with herpes viruses, principally by intracellular phosphorylation of the drug by virus coded **thymidine** kinase and several cellular enzymes.



# ShinyTPs: Curating TPs from Text Mining Results



## 1. Read in data



PubChem API interface showing search results and a data table with columns for Compound ID, Name, and SMILES.

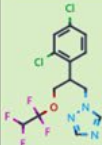

## 2. Launch ShinyTPs



## 3. Check available data

Select an input compound

Tetraconazole

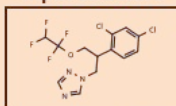
Parent	Parent CID	TP	TP CID
	80277		1810180

## 4. Curate reactions

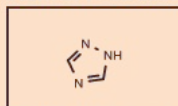
Select an input compound

Tetraconazole

Input structure



Potential TP structure



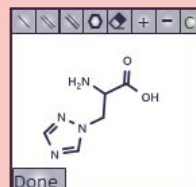
Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).

## 5. Add missing entries

Select an input compound

Tetraconazole

Draw a structure

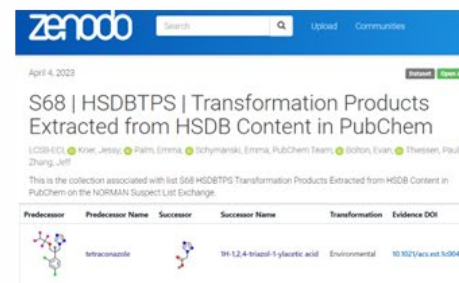


Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).



SMILES:

NC(CN1C=NC=N1)C(=O)O

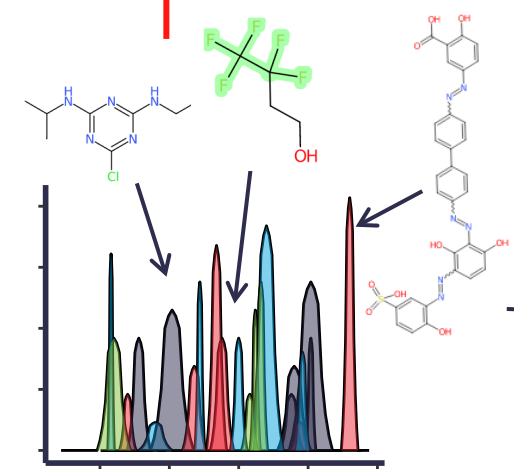
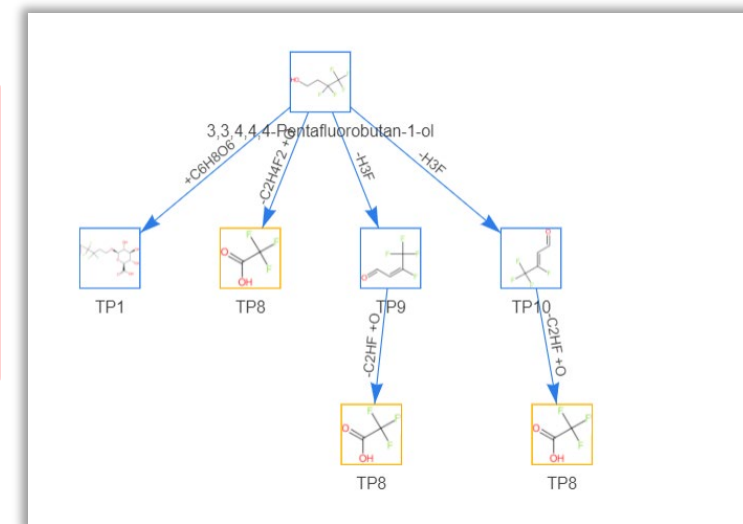
## 6. Export and upload



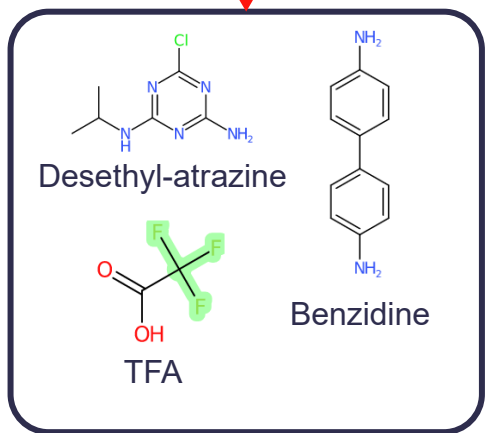
zenodo upload page showing the collection title: S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem. The page includes a search bar, a date (April 4, 2023), and a table of predecessor and successor compounds.

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Evidence DOI
	tetraconazole		1H-1,2,4-triazol-1-ylacetic acid	Environmental	10.1021/acs.est.5c0466

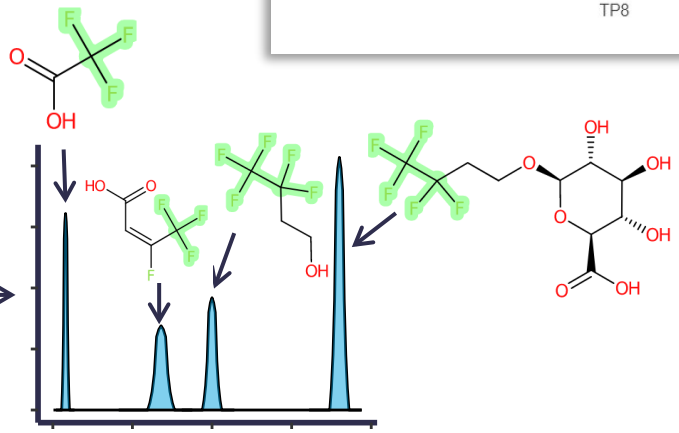
# Applying TP NT-HRMS Workflows with patRoan



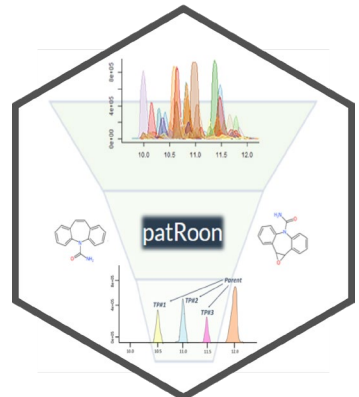
Suspect/Non-target screening



TP library



TP suspect screening



# Outline

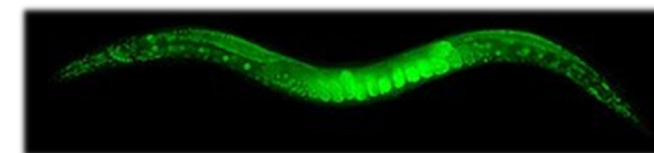
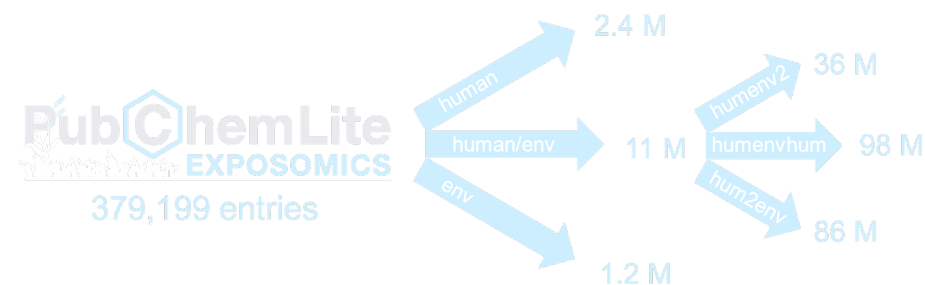
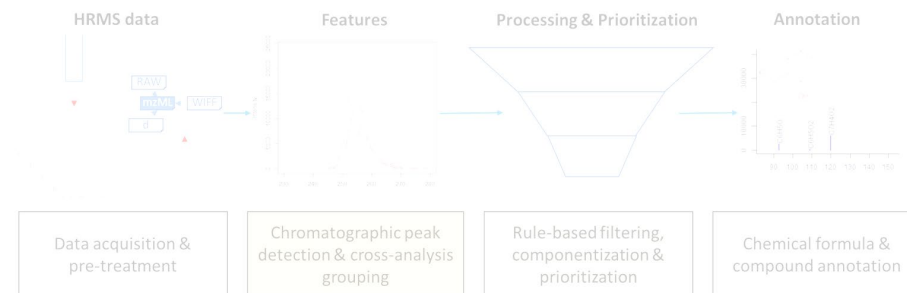
- Setting the Scene

- NT-HRMS & Databases
- The Great Data Challenge ...

- Defining Chemical Space in Exposomics

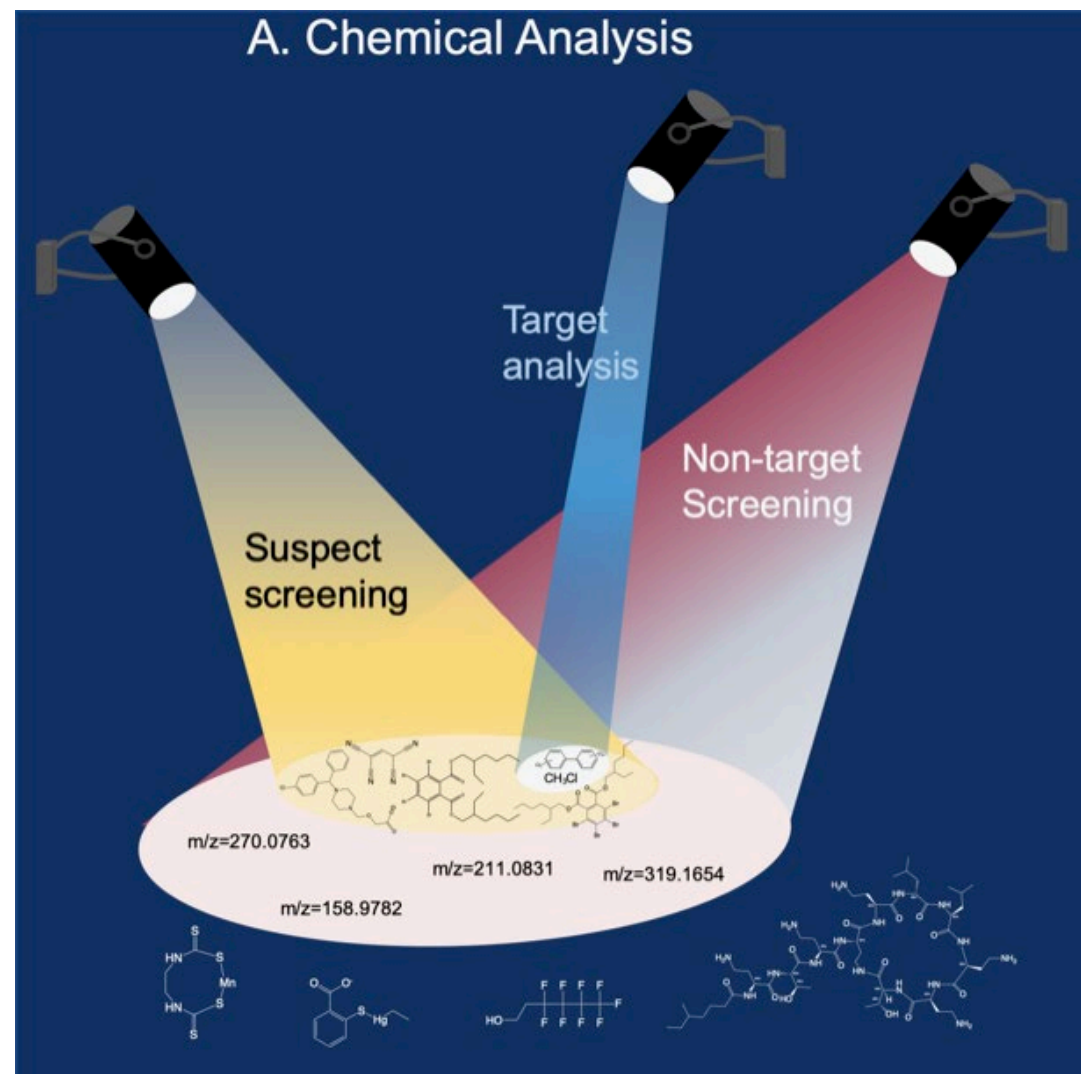
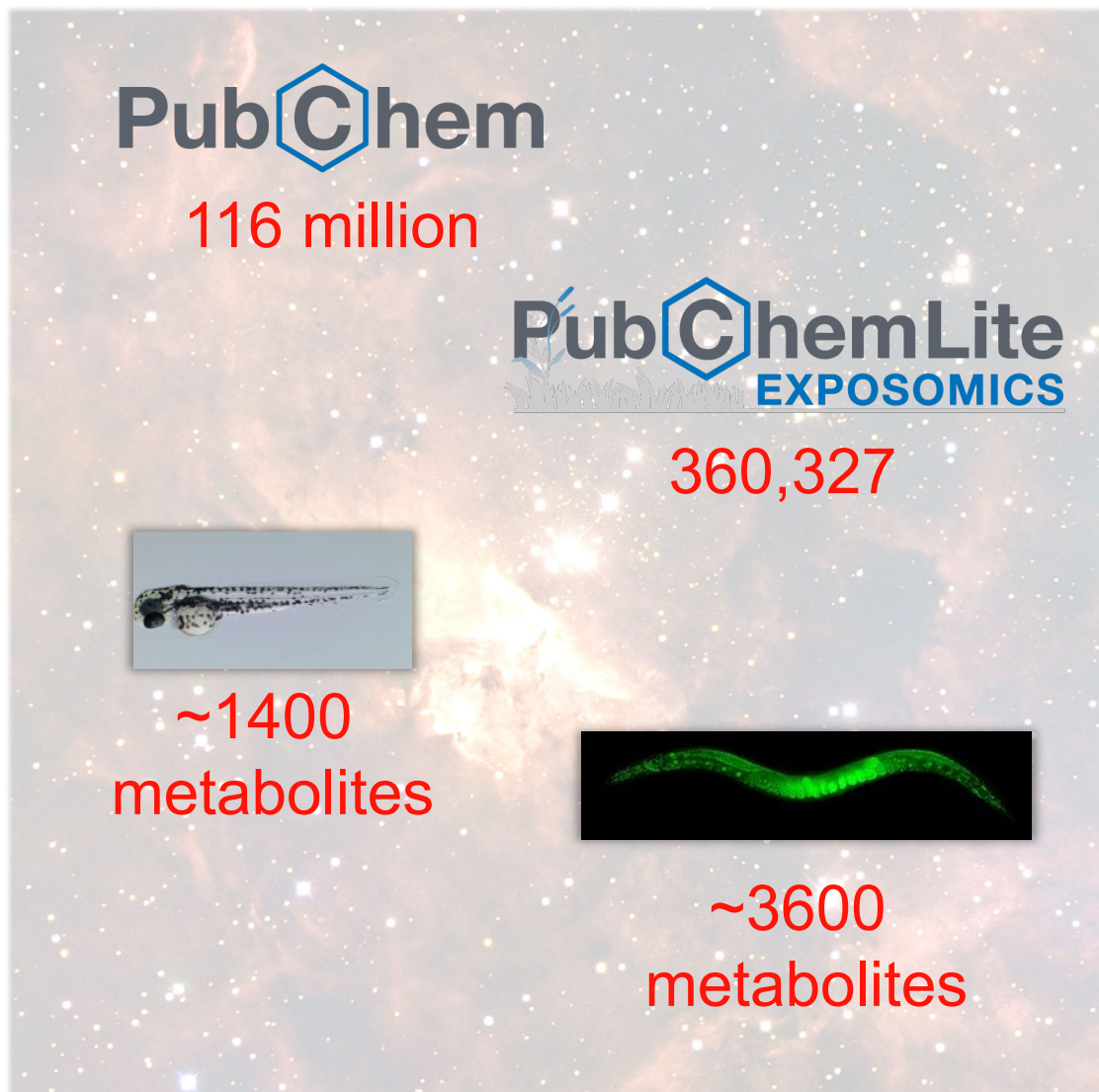
- How to Tackle Transformations?

- Metabolomics / Exposomics-specific Workflows

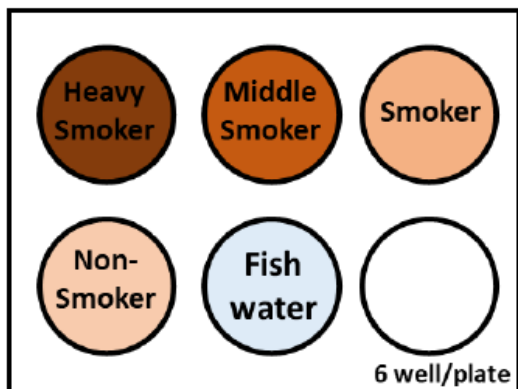
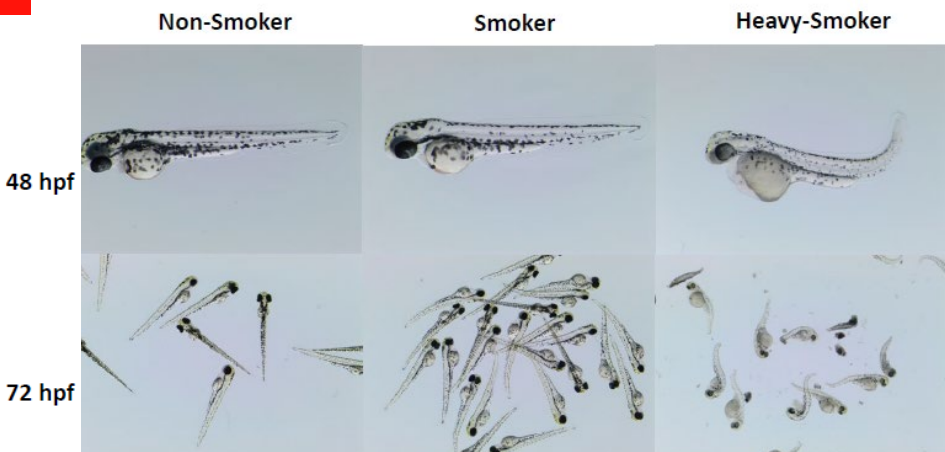




# Finally ... Do we always need the entire chemical space?



# Organism-specific Metabolomics?



PubChem Danio rerio (zebrafish) (Taxonomy)

2-Phosphoglyceric acid [WikiPathways:WP1356](#)

CONTENTS

- Title and Summary
- 1 Names and Identifiers
- 2 Related Taxonomies
- 3 Chemicals and Bioactivities
  - 3.1 Whole-Organism Bioactivities
  - 3.2 Pathway Compounds
  - 3.3 Metabolites
  - 3.4 Natural Products
  - 3.5 Glycans

423 items

Search  SORT BY

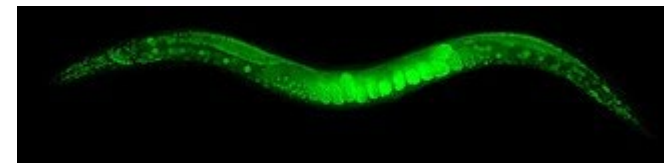
Structure	Compound CID	Compound	Evidence IDs	Data Source
	59	2-Phosphoglyceric acid	DOI:10.1093/nar/gkaa1024	ECI Group, LCSB, University of Luxembourg

<https://pubchem.ncbi.nlm.nih.gov/taxonomy/7955#section=Metabolites>

Taxonomy: Kim *et al.* (2022) DOI: [10.1016/j.jmb.2022.167514](https://doi.org/10.1016/j.jmb.2022.167514)

Image source: Carla Merino, Noelia Ramirez (URV). Merino *et al* (in prep), Torres *et al* (in prep).

# Organism-specific Metabolomics?



zenodo

Search [ ] Upload Communities

September 9, 2019 Dataset Open Access

## WormJam Metabolites Local CSV for MetFrag

Witting, Michael; Schymanski, Emma

This is a local CSV file of WormJam (https://msbi.ipb-halle.de/MetFrag/).

The text file provided by Michael (also headers for MetFrag import).

This CSV file is for users wanting to use MetFrag online; please use the file in the repository.

`C_elegans_metabolites.Rmd` 9.79 KB

```
title: "Finding_C. elegans_Metabolites on PubChem"
author: "Emma SCHYMANSKI"
date: "19/01/2022"
output: pdf_document
```

```
knitr::opts_chunk$set(echo = TRUE)
```

PubChem Caenorhabditis elegans (Taxonomy)

70 4-Methyl-2-oxopentanoic acid DOI:10.1080/21624054.2017.1373939 ECI Group, LCSB, University of Luxembourg

Page 1 of 233

ECI Group, LCSB, University of Luxembourg

### 3.4 Natural Products

712 items Download

Search SORT BY Compound CID - Increasing

Structure	Compound CID	Compound	Evidence IDs	Data Source
	47	3-Methyl-2-oxovaleric acid	DOI:10.3389/FMOLB.2018.00096	LOTUS - the natural products occurrence database

Topic	Documentation
Taxonomy	<a href="#">C. elegans as PDF or RMarkdown</a>



## Background

### WormJam

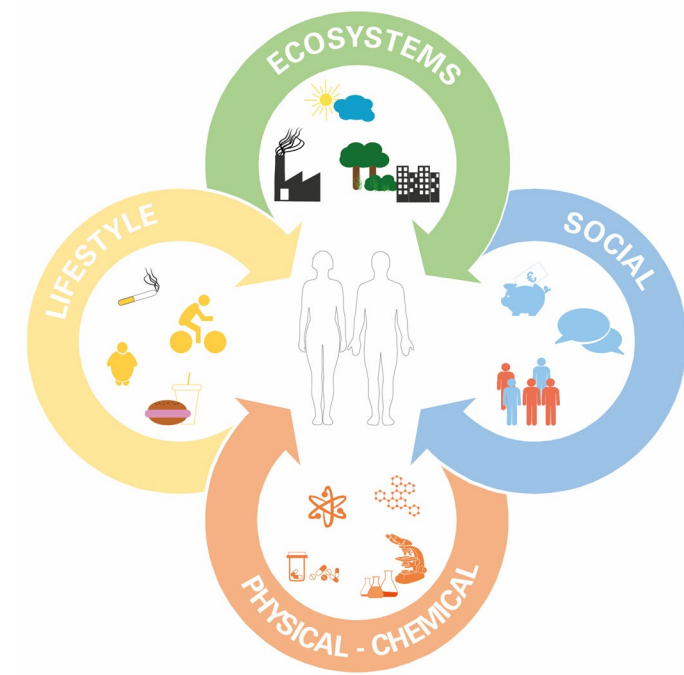
In 2019, Michael Witting provided a copy of 1203 WormJam (Worm Jamboree) metabolites from the multi-author WormJam study (DOI:10.1080/21624054.2017.1373939). This was turned into a MetFrag database and put on Zenodo by the Environmental Cheminformatics group (ECI) at LCSB (DOI:10.5281/zenodo.3403364) for integration into MetFrag Web, and for download for MetFragCL users.

<https://pubchem.ncbi.nlm.nih.gov/taxonomy/6239#section=Metabolites>

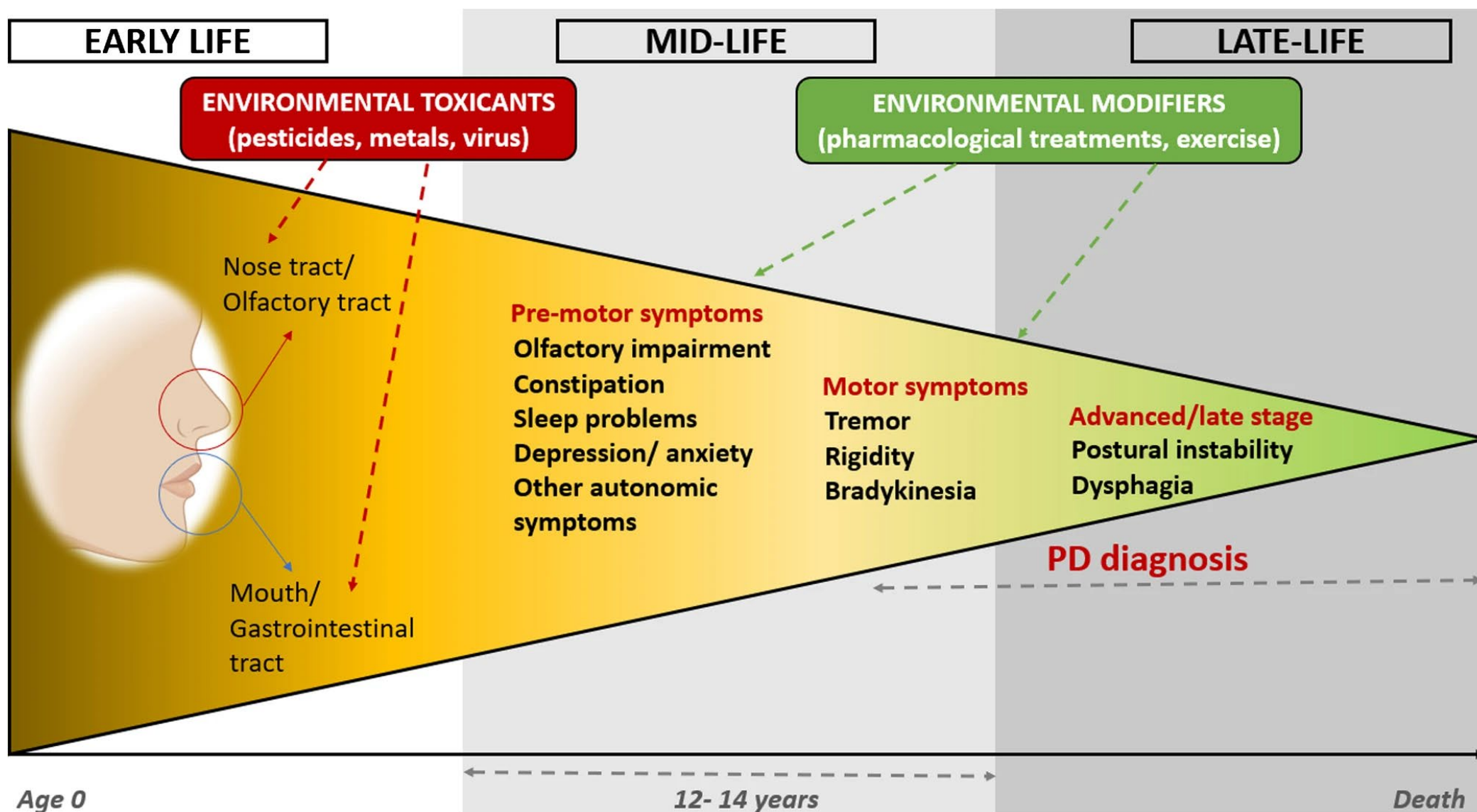


# The Exposome

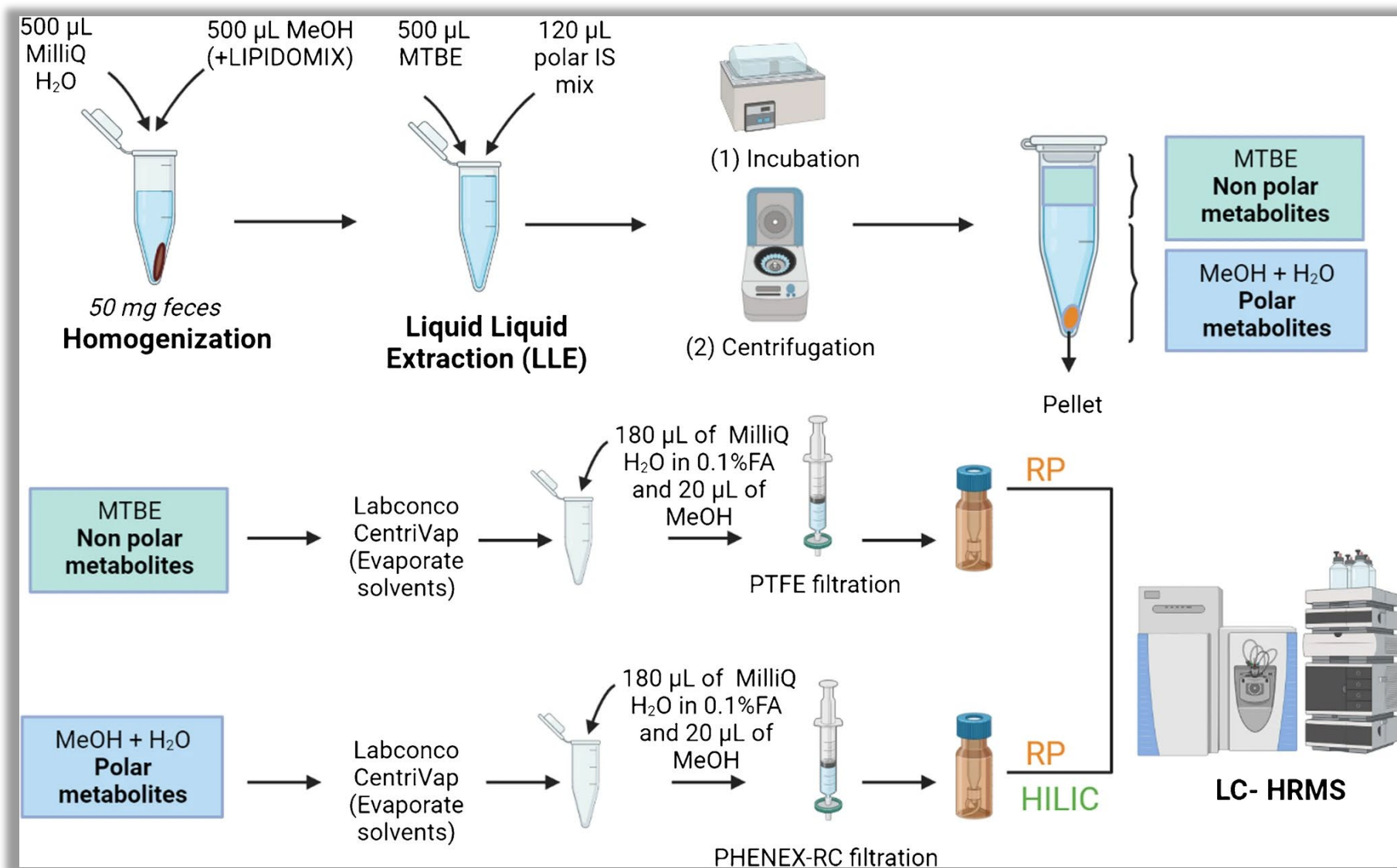
- Christopher Wild (2005)
  - “**All exposures** from **conception onwards**, including those from **lifestyle, diet and the environment**”
- Miller and Jones (2014) functionalized this to:
  - “The cumulative measure of **environmental influences** and **associated biological responses throughout the lifespan**, including exposures from the **environment, diet, behaviour and endogenous processes**”
- (Under)estimated **16 % (9 million) deaths per year** worldwide due to **environmental pollution** alone (Landrigan *et al.* 2018)
- Less than approx. **10 % of disease** can be explained by **genetic factors alone** (Vermeulen *et al.* 2020 & refs within)



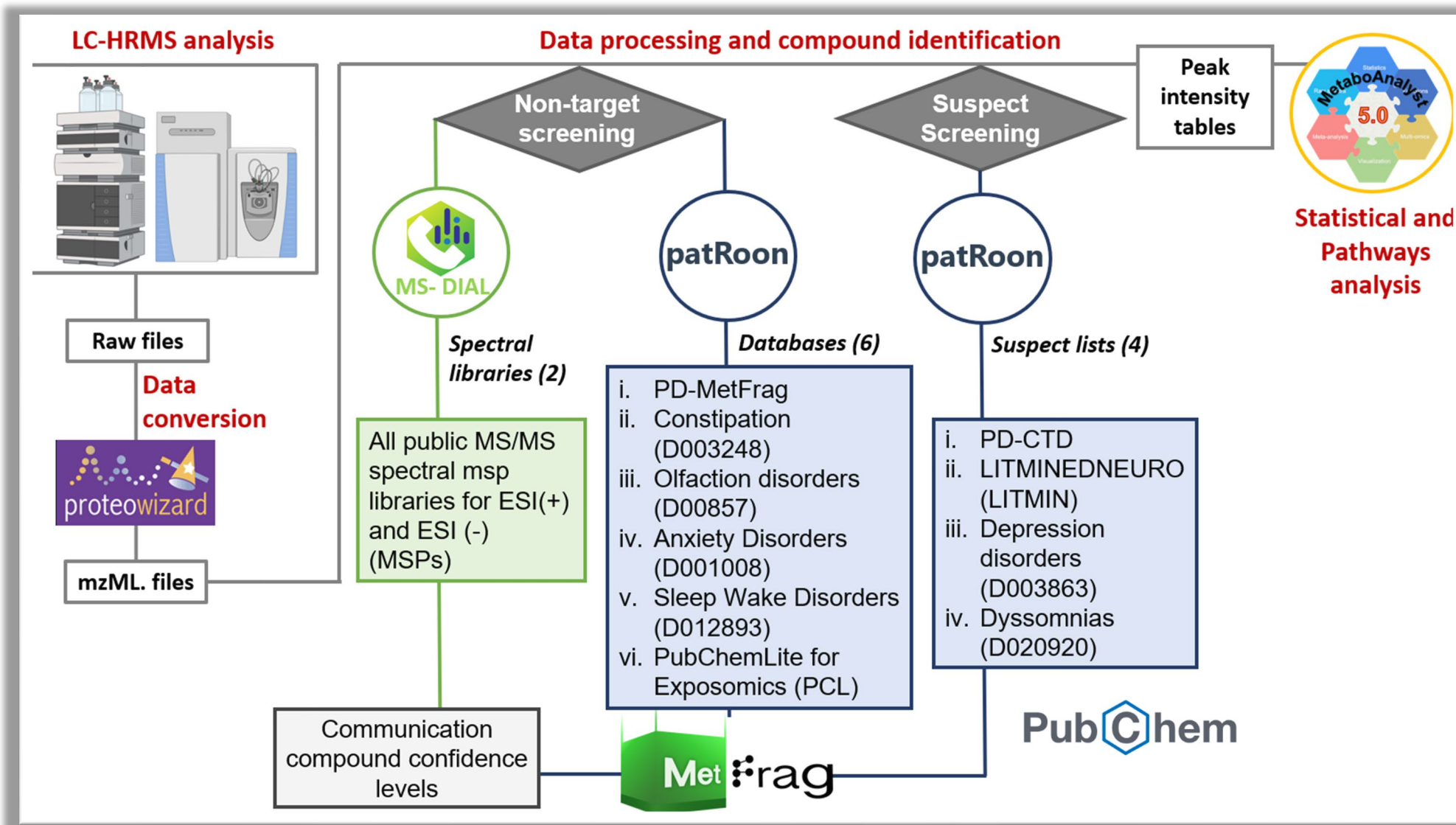
# Exposomics and Neurodegenerative Disease: Timeline!



# Exposomics: Sample Prep. for Broad Range of Analysis



# Exposomics: Disease-specific Databases and Interpretation



# Exposomics: Disease-specific Databases: LITMINEDNEURO

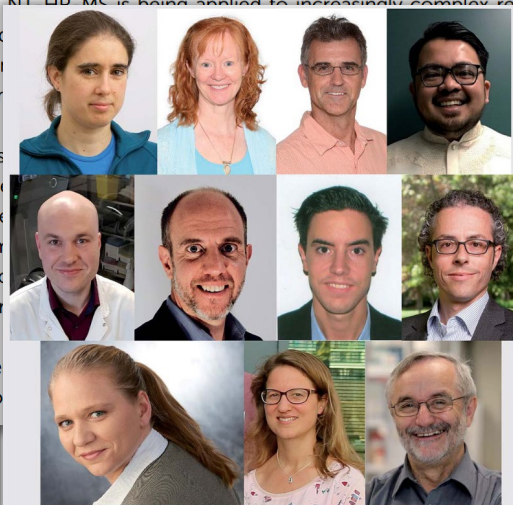
LC-HRMS analysis

Data processing and compound identification

## Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: potential and challenges†

Emma L. Schymanski, <sup>id</sup>\*<sup>a</sup> Nancy C. Baker, <sup>id</sup><sup>b</sup> Antony J. Williams, <sup>id</sup><sup>c</sup> Randolph R. Singh, <sup>id</sup><sup>ad</sup> Jean-Pierre Trezzi, <sup>id</sup><sup>ef</sup> Paul Wilmes, <sup>id</sup><sup>f</sup> Pierre L. Kolber, <sup>id</sup><sup>gh</sup> Rejko Kruger, <sup>id</sup><sup>gh</sup> Nicole Paczia, <sup>id</sup><sup>i</sup> Carole L. Linster <sup>id</sup><sup>j</sup> and Rudi Balling <sup>id</sup><sup>j</sup>

Connecting chemical exposures over a lifetime to complex chronic diseases with multifactorial causes such as neurodegenerative diseases is an immense challenge requiring a long-term, interdisciplinary approach. Rapid developments in analytical and data technologies, such as non-target high resolution mass spectrometry (NT-HR-MS), have opened up new possibilities to accomplish this, inconceivable 20 years ago. While NT-HR-MS is being applied to increasingly complex research questions, there are still many unidentified environmental chemicals and their interactions with biological systems. This presents a challenge to human health outcomes and the identification of environmental and biological factors at cross many scientific disciplines, such as toxicology, epidemiology, and genetics. The integration of data matrices as a form of exposomics, which briefly covers these chemical exposures, has the potential for sustained and exploratory neurodegeneration contributions.



MetaboAnalyst 5.0

Statistical and

Peak intensity tables

Suspect screening

Peak Room

Suspect lists (4)

i. PD-CTD  
ii. LITMINEDNEURO (LITMIN)  
iii. Depression

Chemical	Effect							
	Seizures	Nervous System Diseases	Peripheral Nervous System Diseases	Brain Diseases	Muscular Diseases	Basal Ganglia Diseases	Parkinson Disease, Secondary	Coma
Cisplatin	20	47	140	13	0	4	1	1
Ethanol	100	23	11	18	26	1	3	20
Lead	28	107	68	102	4	2	2	1
Lithium	30	50	9	22	5	36	13	25
Valproic Acid	32	10	3	65	6	10	18	45
Carbaryl	3	PubMed Article Count: Co-occurring chemical effect						
Vincristine	17							
Phenytoin	37							

### NEURO: Neurotoxins from PubMed

Search LITMINEDNEURO Chemicals

Identifier substring search

#### List Details

**Description:** This is a list of chemicals associated with neurotoxicity compiled through automated literature mining of PubMed using MeSH terms and associating these with single chemical substances (where possible). Articles were identified in which a nervous system disease was annotated with the MeSH node C10 with disease subheading "chemically induced" and subheading "toxicity", "poisoning", or "adverse effects". We identified nervous system diseases through a look-up in the MeSH tree. Nerve diseases caused by trauma and manually identified "common English terms" that could not be associated with any specific chemicals (e.g. "particulate matter", "contrast media") were omitted. In total 4,528 chemicals were identified; this list contains 1243 chemicals associated with 5 or more literature references, all of which have been registered in the Dashboard. The details of this work are reported in "Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: potential and challenges" by Schymanski et al, DOI: 10.1039/c9em00068b

**Number of Chemicals:** 1243

CompTox  
Chemicals  
Dashboard

Check for updates

Cite this: *Environ. Sci.: Processes Impacts*, 2019, 21, 1426

Received 12th February 2019  
Accepted 2nd July 2019

DOI: 10.1039/c9em00068b

rs.c.li/esp



# Exposomics: Disease-specific Information in PubChem



PubChem Rotenone (Compound)

## 14 Associated Disorders and Diseases

83 items [Download](#)

Search

Disease	Evidence Type	Evidence PMID
Motor Disorders	marker/mechanism	26770656
		27016191
		29649621

Database

## PubChem Protein, Gene, Pathway, and Taxonomy Data Collections: Bridging Biology and Chemistry through Target-Centric Views of PubChem Data


Sunghwan Kim<sup>†</sup>, Tiejun Cheng<sup>†</sup>, Siqian He, Paul A. Thiessen, Qingliang Li, Asta Gindulyte, Evan E. Bolton   



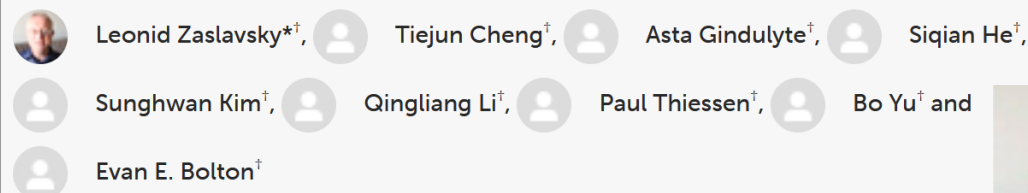
PubChem Rotenone (Compound)

## 15.10 Chemical-Disease Co-Occurrences in Literature

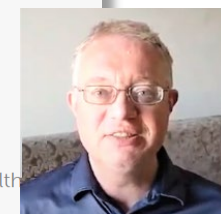
Showing 3 of 25 [View More Co-Occurrence and Evidence Data](#) [Download](#)

Disease	Selected evidence
Parkinson Disease	1,590 articles <a href="#">View All</a>  <a href="#">Rotenone-Induced Model of Parkinson's Disease: Beyond Mitochondrial Complex I Inhibition</a> PMID 36593435; DOI 10.1007/s12035-022-03193-8; Molecular neurobiology 2023 Jan; ?(?): (Review Article) Name matches: parkinson's disease <i>rotenone</i>
	<a href="#">HR LC-MS/MS metabolomic profiling of Yucca aloifolia fruit and the potential neuroprotective effect on rotenone-induced Parkinson's disease in rats</a> PMID 36854038; DOI 10.1371/journal.pone.0282246; PloS one 2023; 18(2):e0282246 Name matches: parkinson's disease <i>rotenone</i>

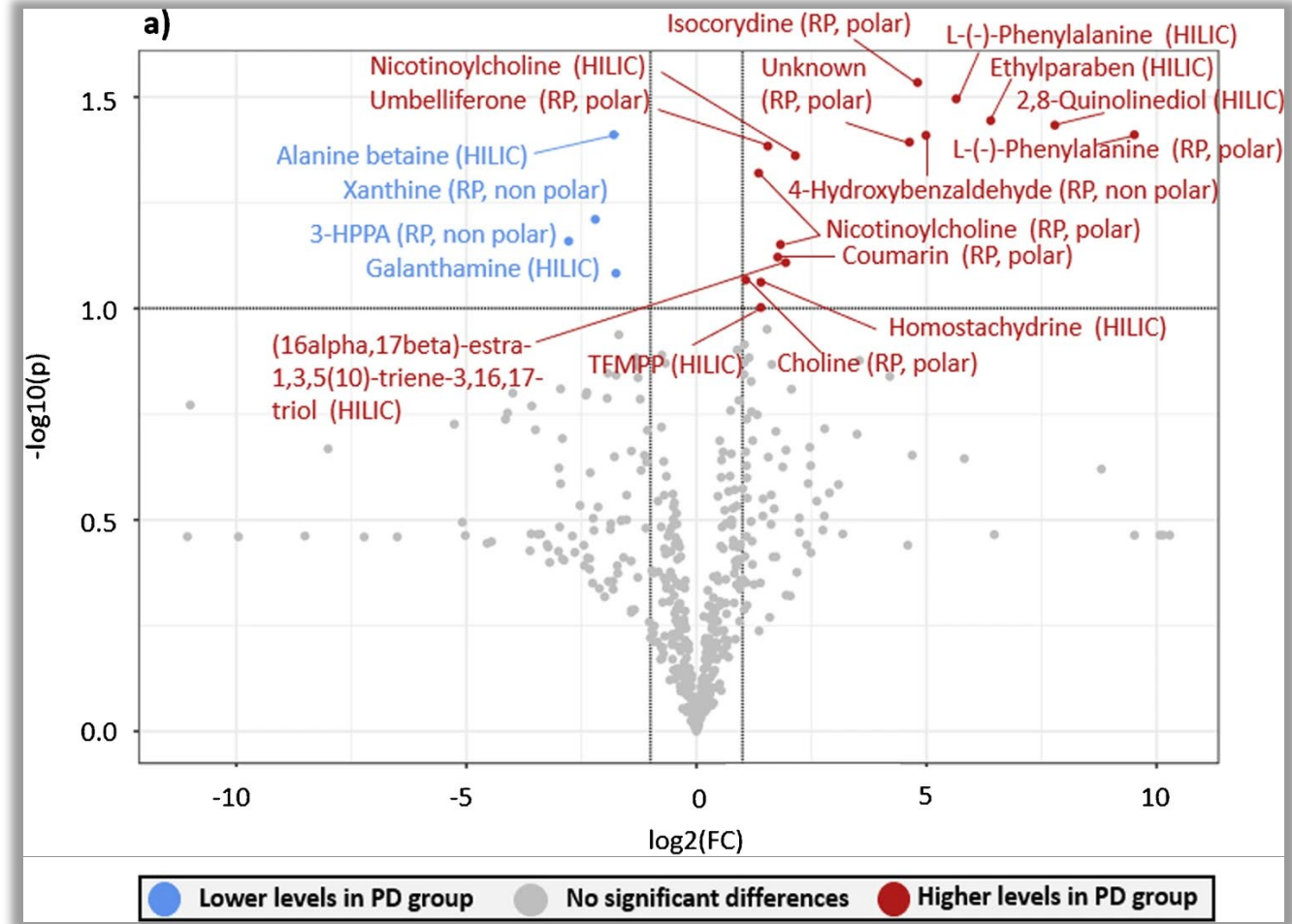
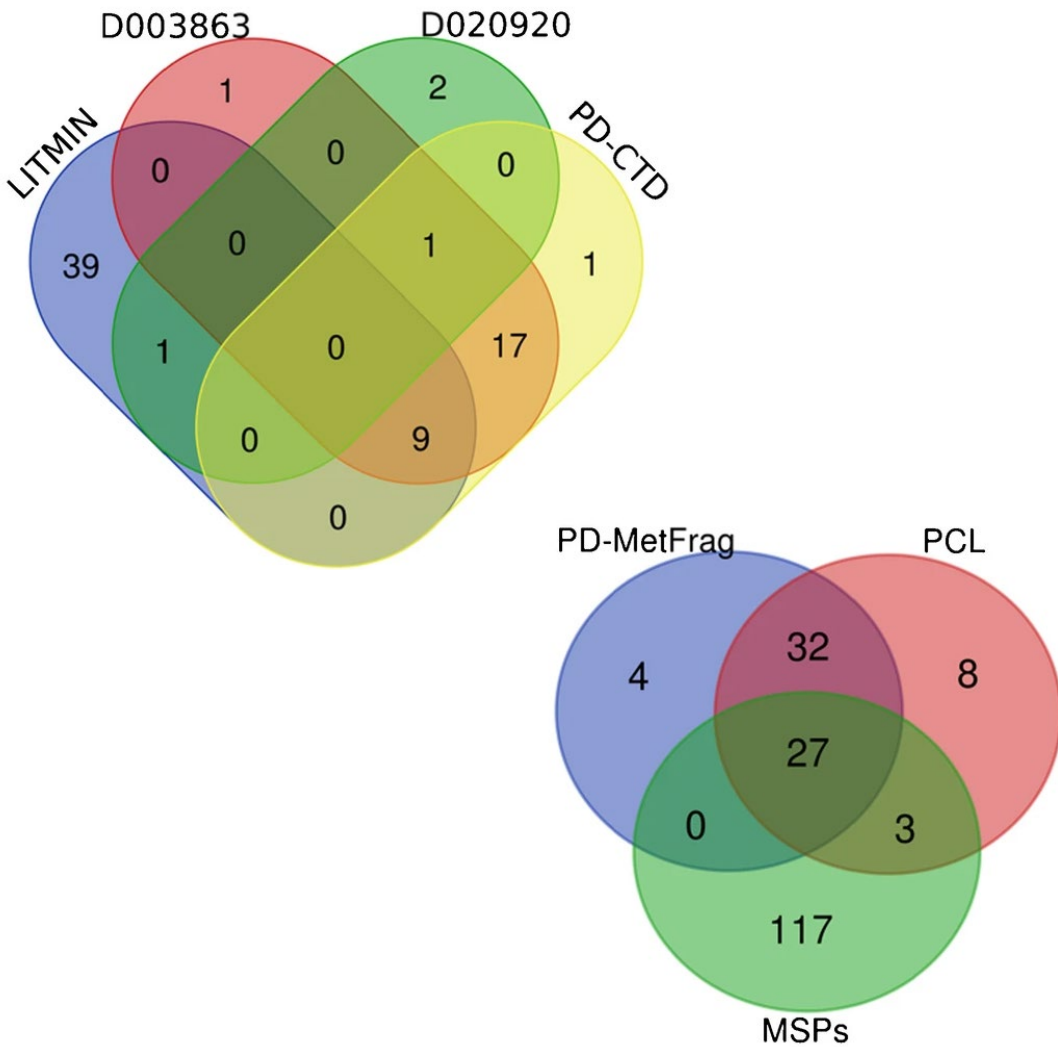
## Discovering and Summarizing Relationships Between Chemicals, Genes, Proteins, and Diseases in PubChem



National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health  
Bethesda, MD, United States



# Exposomics: Disease-specific Data Interpretation

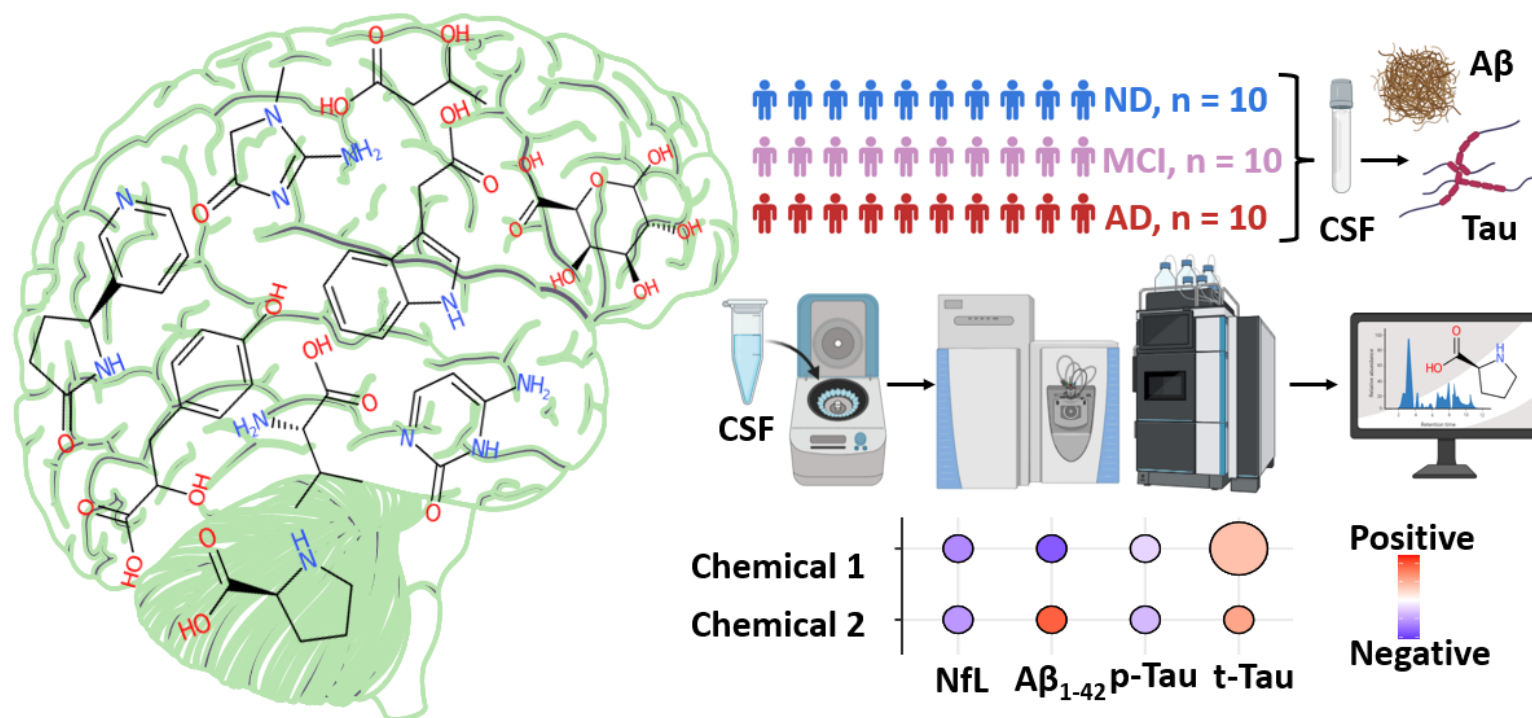


# Exposomics: Disease-specific Databases and Interpretation

Information available varies widely depending on disease and sample

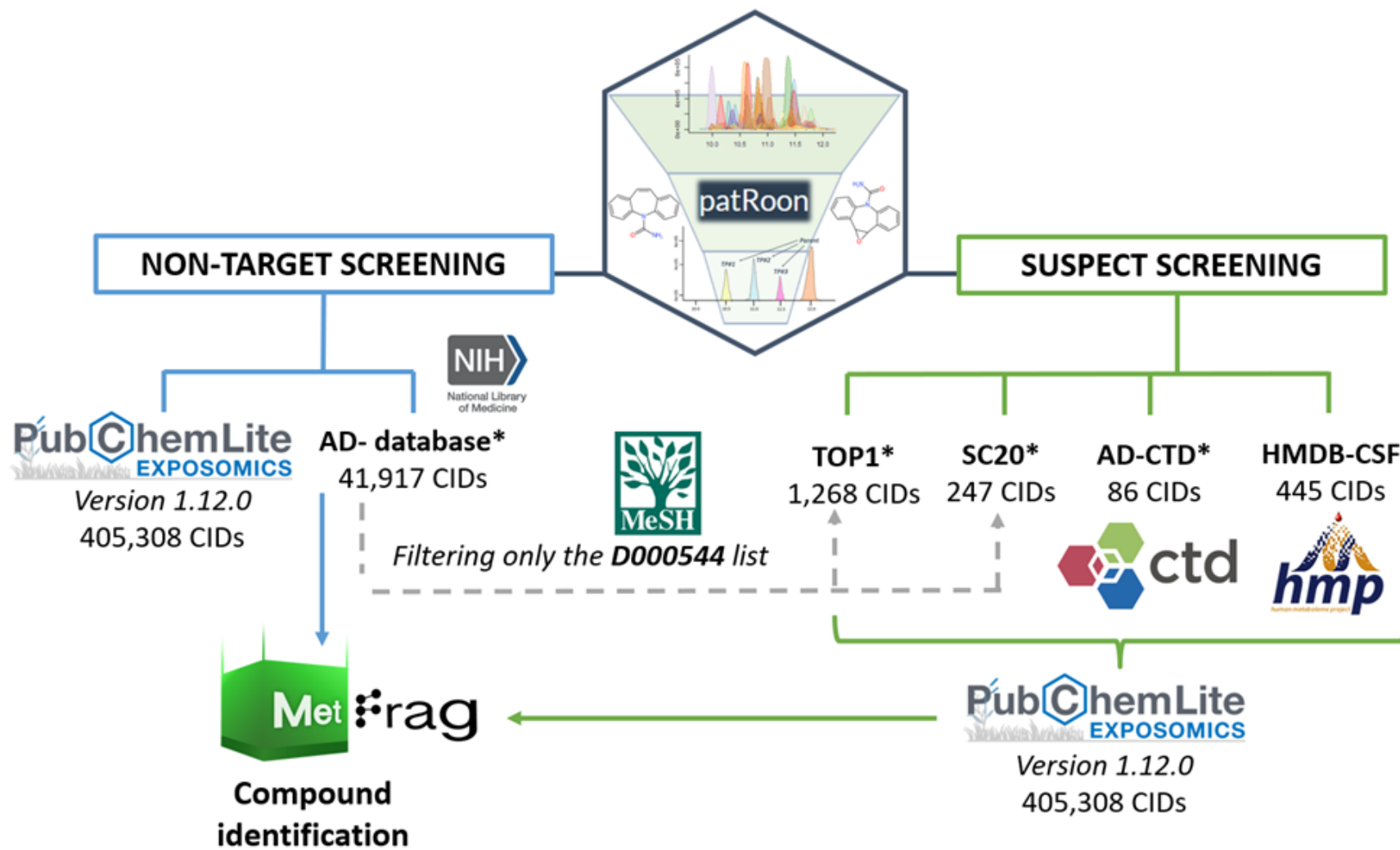


MICROH



# Exposomics: Disease-specific Databases and Interpretation

Information available varies widely depending on disease and sample



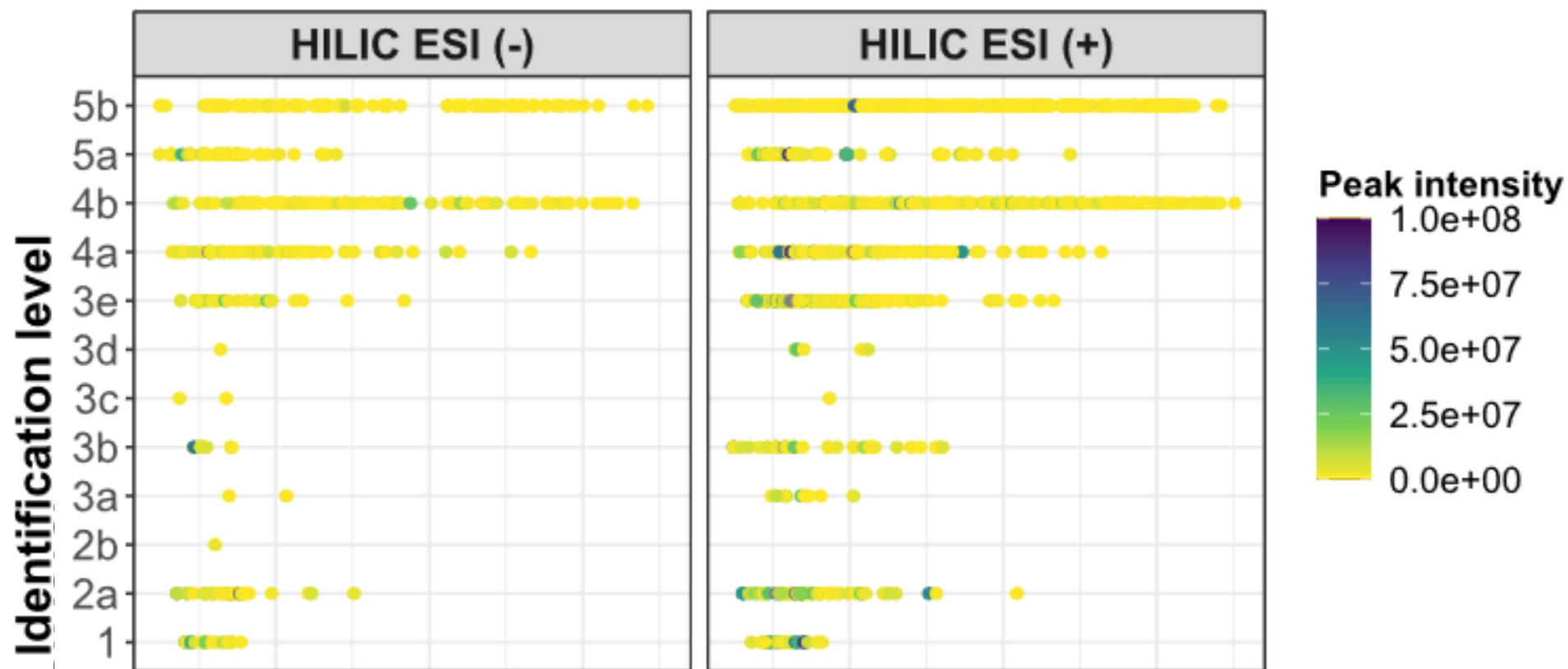
# Identification Evidence & Confidence Levels ...



	Minimum requirements	Means used for identification
<b>Level 1</b>	Confirmed structure with reference standard	Internal library
<b>Level 2a</b>	>= 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 70-100	
	Fragment presence 50-100	
<b>Level 2b</b>	>= 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 70-100	
	Fragment presence 50-100	
	Structure unknown in library	
<b>Level 3a</b>	>= 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 50-70	
	Fragment presence 50-100	
<b>Level 3b</b>	< 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 50-100	
	Fragment presence 50-100	
<b>Level 3c</b>	< 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 50-100	
	Fragment presence 50-100	
	Structure unknown in library	
<b>Level 3d</b>	>= 3 ion fragments matching	MS-FINDER
	Score >= 8	
<b>Level 3e</b>	>= 3 ion fragments matching	MS-FINDER
	Score > 5	

	Minimum requirements	Means used for identification
<b>Level 4a</b>	Dot product <50 and/or	MS-DIAL MSPs libraries
	Fragment presence < 50	
<b>Level 4b</b>	< 3 ion fragments matching and or/	MS-FINDER
	Score <5 or/	
	Only formula information	
<b>Level 5a</b>	Exact mass of interest and MS2 information but not library match	MS-DIAL MSPs libraries and/or MS-FINDER
<b>Level 5b</b>	Exact mass of interest without MS2 information	MS-DIAL MSPs libraries and/or MS-FINDER

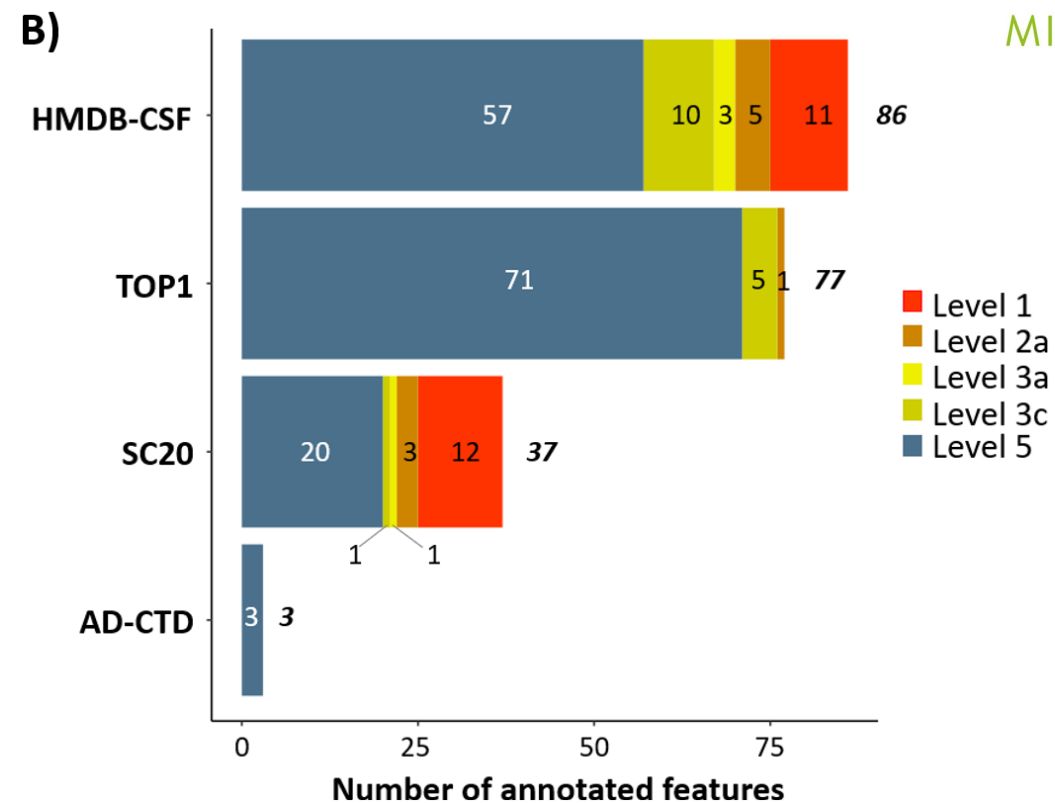
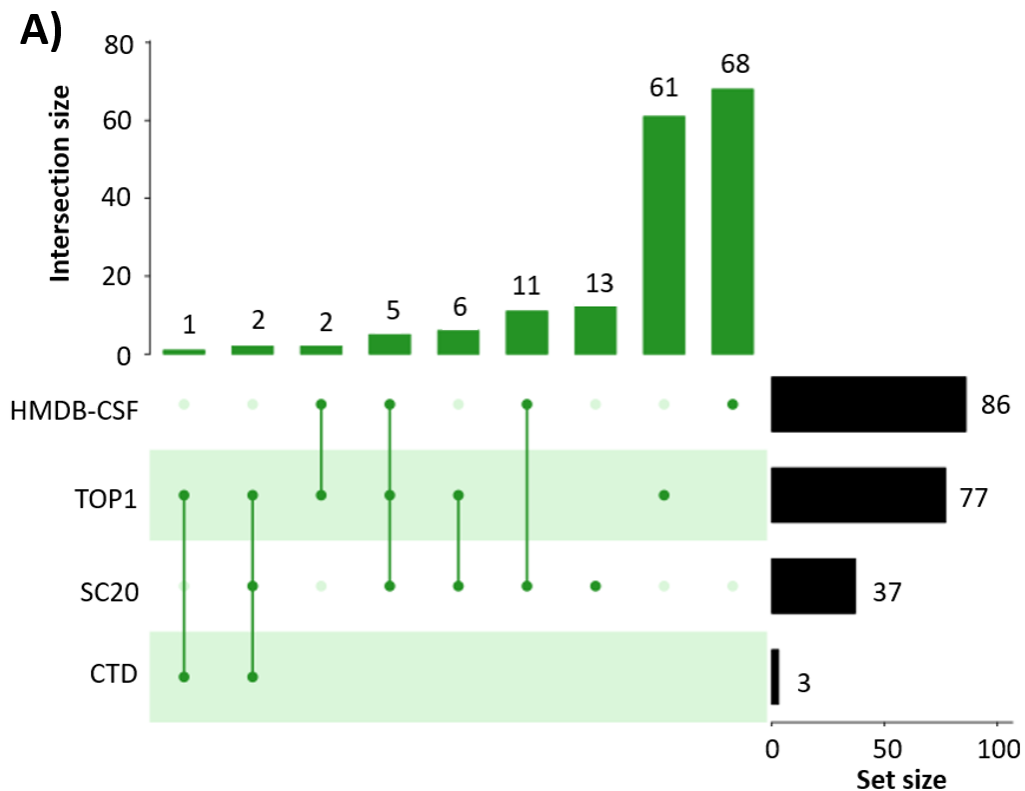
# $m/z$ & Intensity by Annotation Confidence Level & Method



# Annotation Level and Overlap between Databases

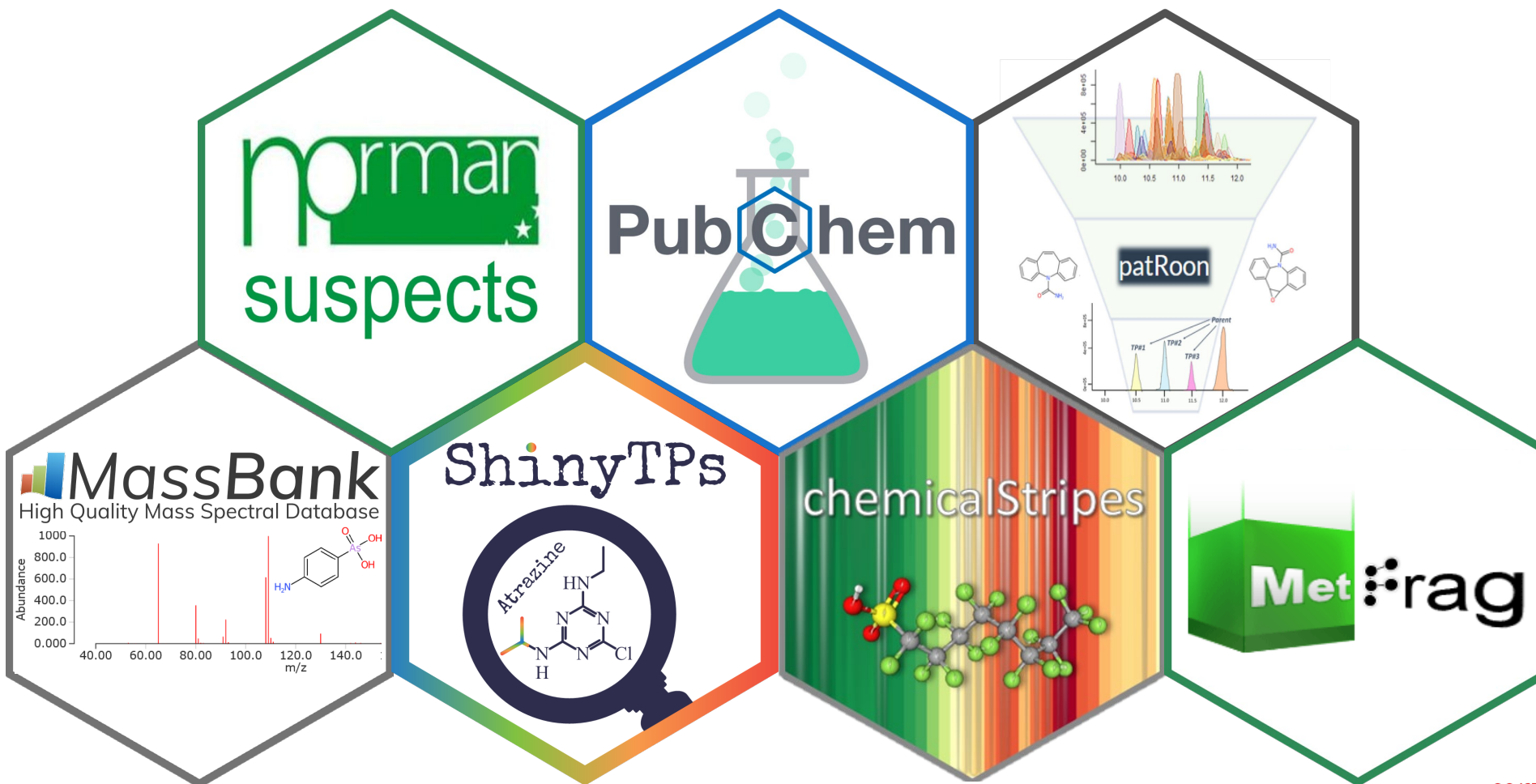


MICROH



# “Take home” Messages

- Open, FAIR data supports NT-HRMS workflows





# “Take home” Messages

- Dynamic, Open Collection: 

- Expert knowledge to fill the gaps
- Annotation content is extremely powerful

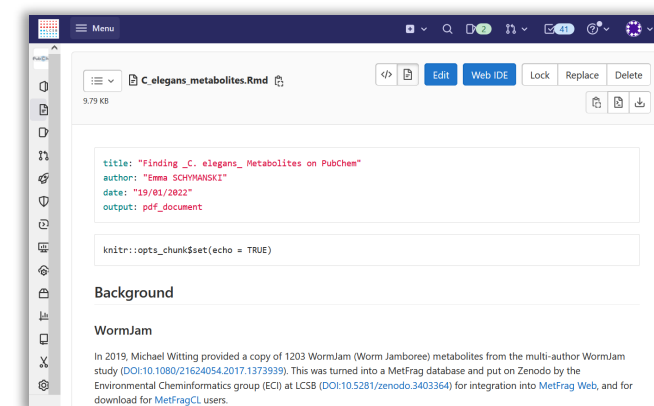
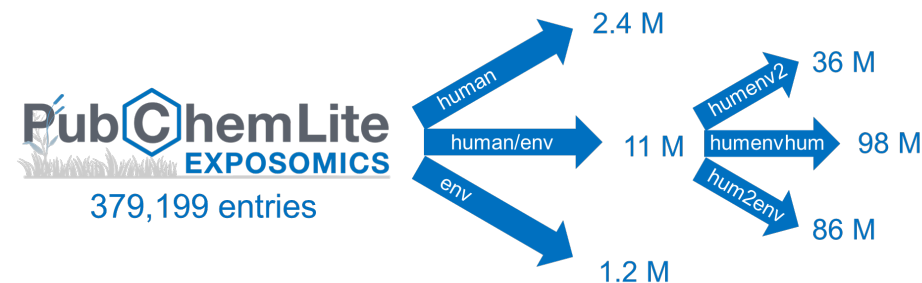
- Transformations and Metabolism

- More data needed to improve predictions

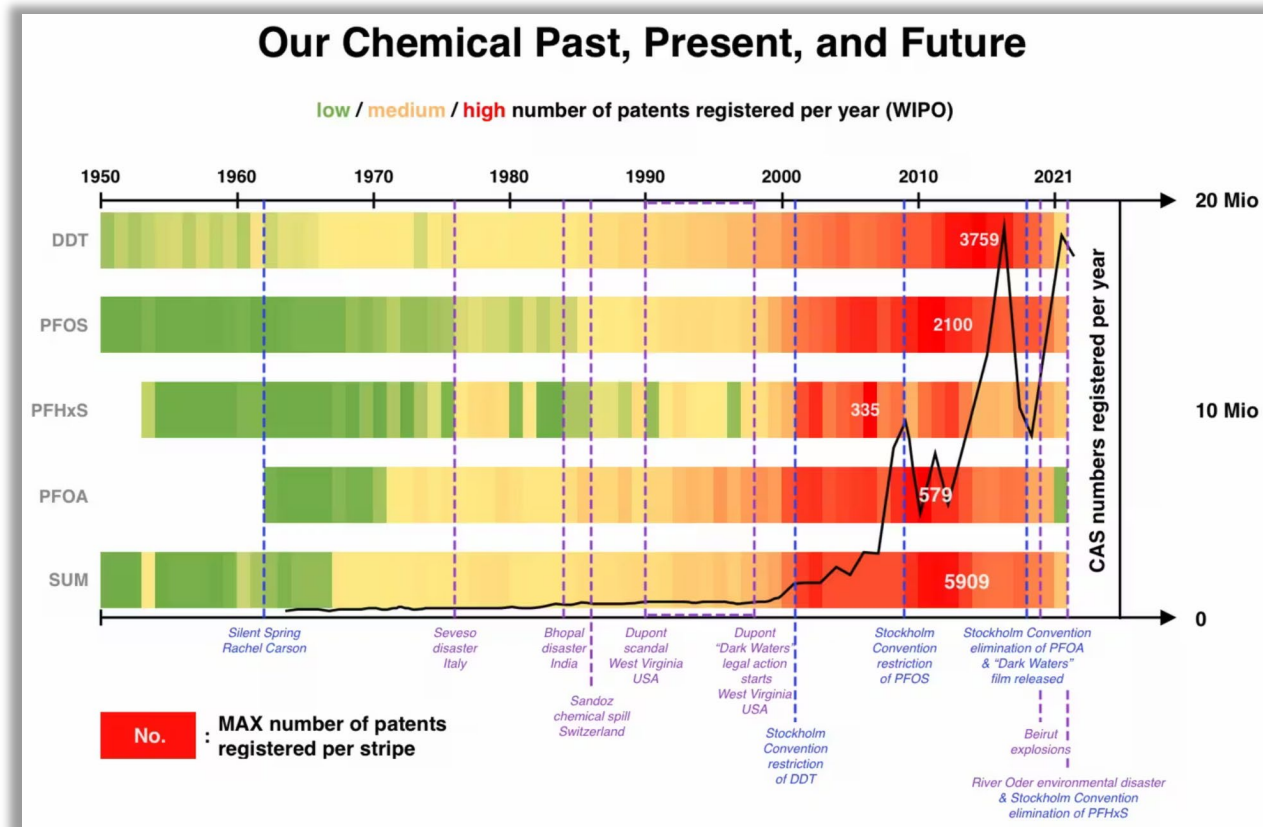
- Workflows available in patRoom

- Specific collections are easy to create for context

- Help contribute by adding your knowledge!



# Today's Video / Soundtrack



Video: <https://vimeo.com/jpmlmusic/ourchemicalpastpresentandfuture>

Soundtrack: <https://soundcloud.com/jamieperera/our-chemical-past-present-and-future>

# Acknowledgements!

Today's slides:

DOI: [10.5281/zenodo.8343924](https://doi.org/10.5281/zenodo.8343924)

Email: [emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu)

Twitter/X: [@ESchymanski](https://twitter.com/ESchymanski)

[@schymane@mstdn.social](https://social.mstdn.com/@schymane)

ORCID: [0000-0001-6868-8145](https://orcid.org/0000-0001-6868-8145)



# PubChem

UNIVERSITY OF AMSTERDAM



Evan Bolton, Jian (Jeff) Zhang, Paul Thiessen, PubChem team

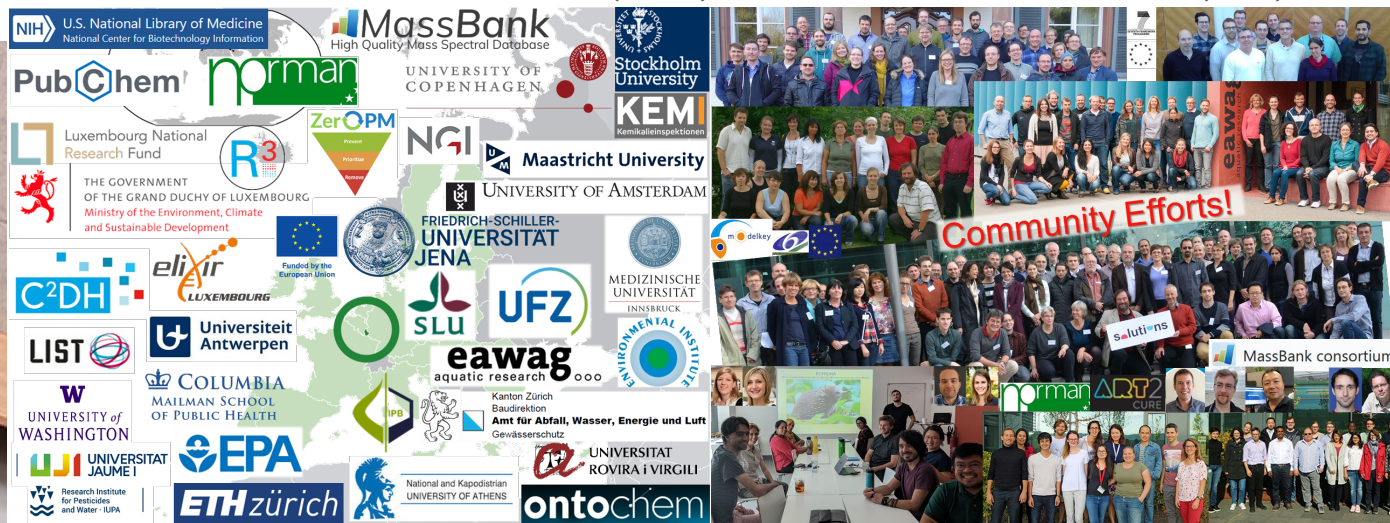


JPML

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information



This work was supported in part by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM), National Institutes of Health (NIH)



Funded by the European Union



H2020:  
101036756



Luxembourg  
National  
Research Fund

