Luxembourg National Research Fund

Swiss Metabolomics Society

UNIVERSITÉ DU LUXEMBOURG

Funded by the European Union

Navigating Millions of Chemicals in Metabolomics and Exposomics Workflows

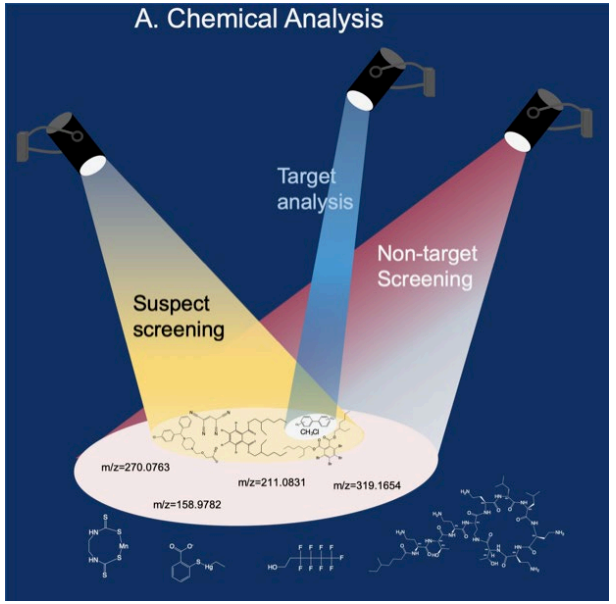


Image mod. from DOI: 10.1126/science.aay6636

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 / @ESchymanski / @schymane@mstdn.social
https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

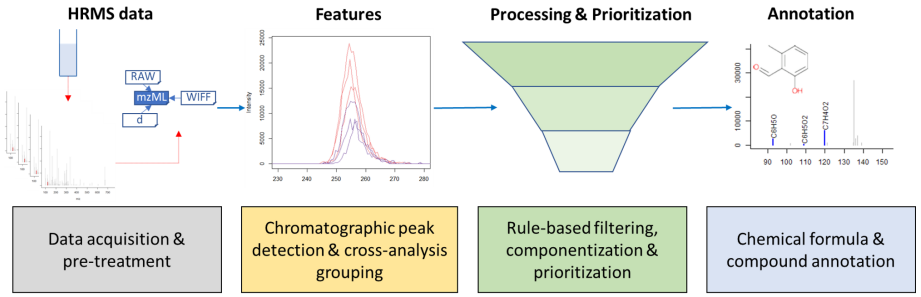
Swiss Metabolomics Society Keynote, 15 Sept. 2023, Zurich
https://swiss-metabolomics.ch/annual-meeting/

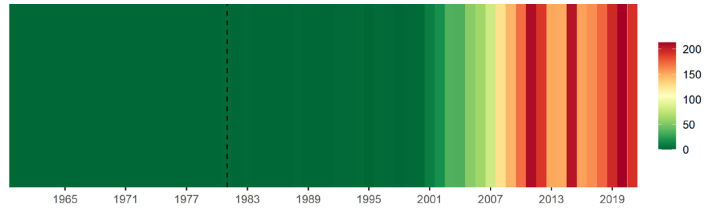
Slides available at DOI: 10.5281/zenodo.8343924

Today's slides zenodo

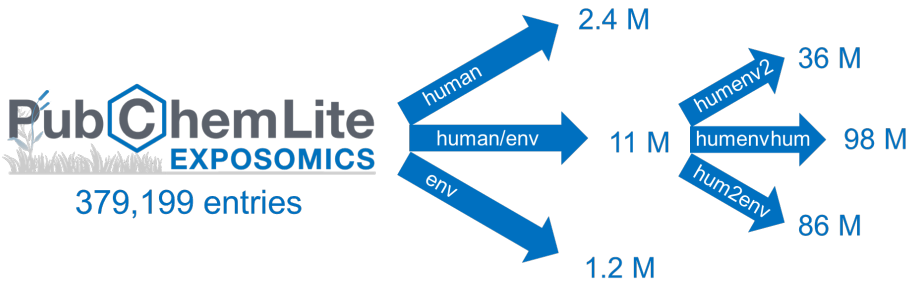
Outline

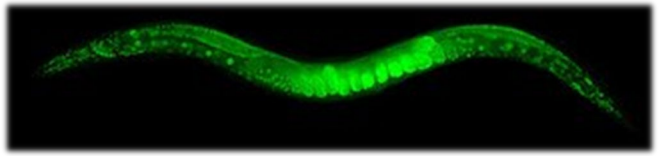
- Setting the Scene
  - NT-HRMS & Databases
  - The Great Data Challenge …
- Defining Chemical Space in Exposomics
- How to Tackle Transformations?
- Metabolomics / Exposomics-specific Workflows



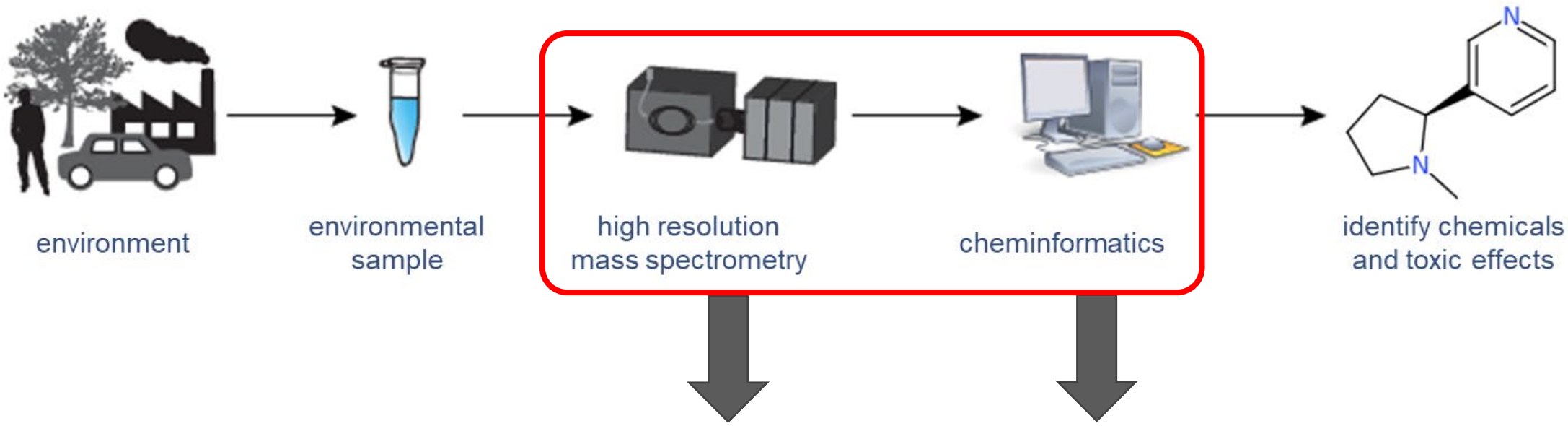


PubChemLite EXPOSOMICS



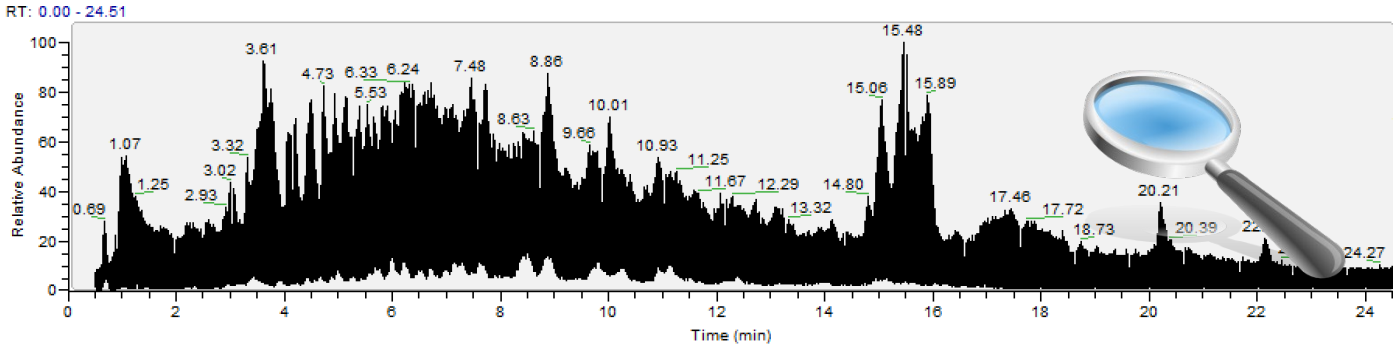


Environmental Cheminformatics & Non-target HR-MS



High resolution mass spectrometry

AND connecting chemical knowledge

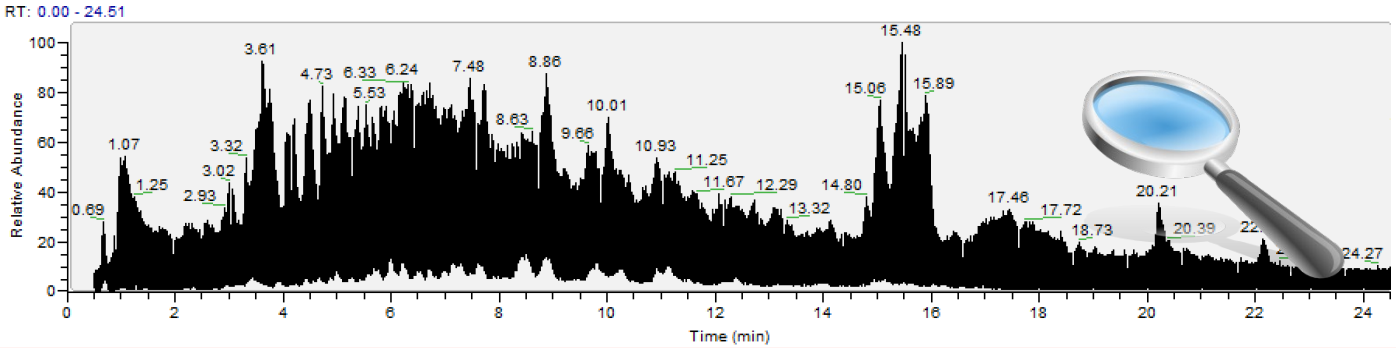


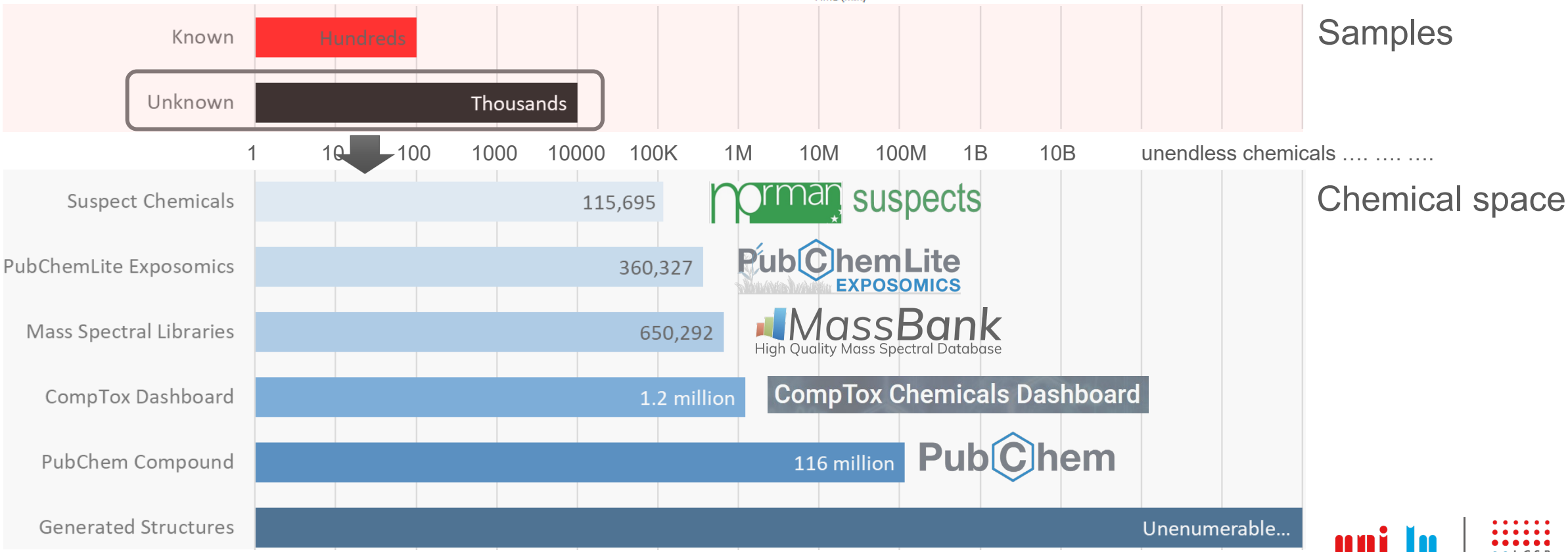
Samples

Environmental Cheminformatics & Non-target HR-MS

High resolution mass spectrometry

AND connecting chemical knowledge





Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 6 Sept. 2023.

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

https://massbank.eu/MassBank/ https://github.com/MassBank/MassBank-data/

MassBank Search Contents Download Accession Go More▼

MassBank Europe

MassBank
High Quality Mass Spectral Database

>> Search Spectra

Search for:

Basic Search | Peak List | Peaks

Compound Information

Compound name

Exact Mass

AND

Formula (e.g. C6H7N5, C5H*N5, C5*)

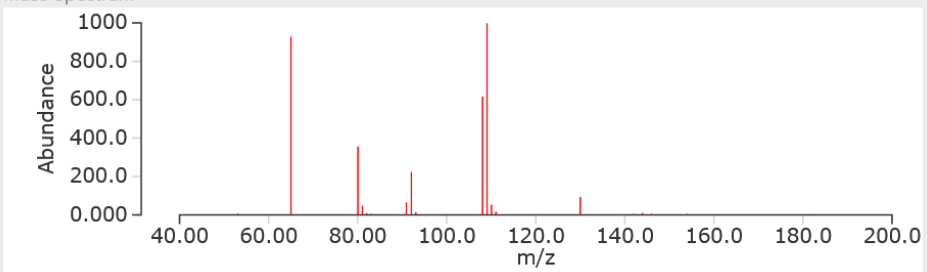
AND

Search

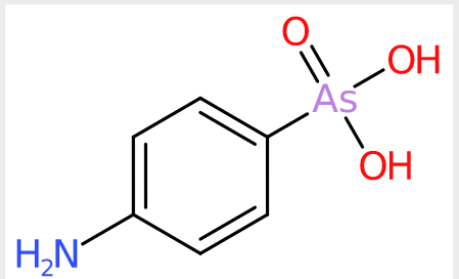
MassBank Record: LU040605

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

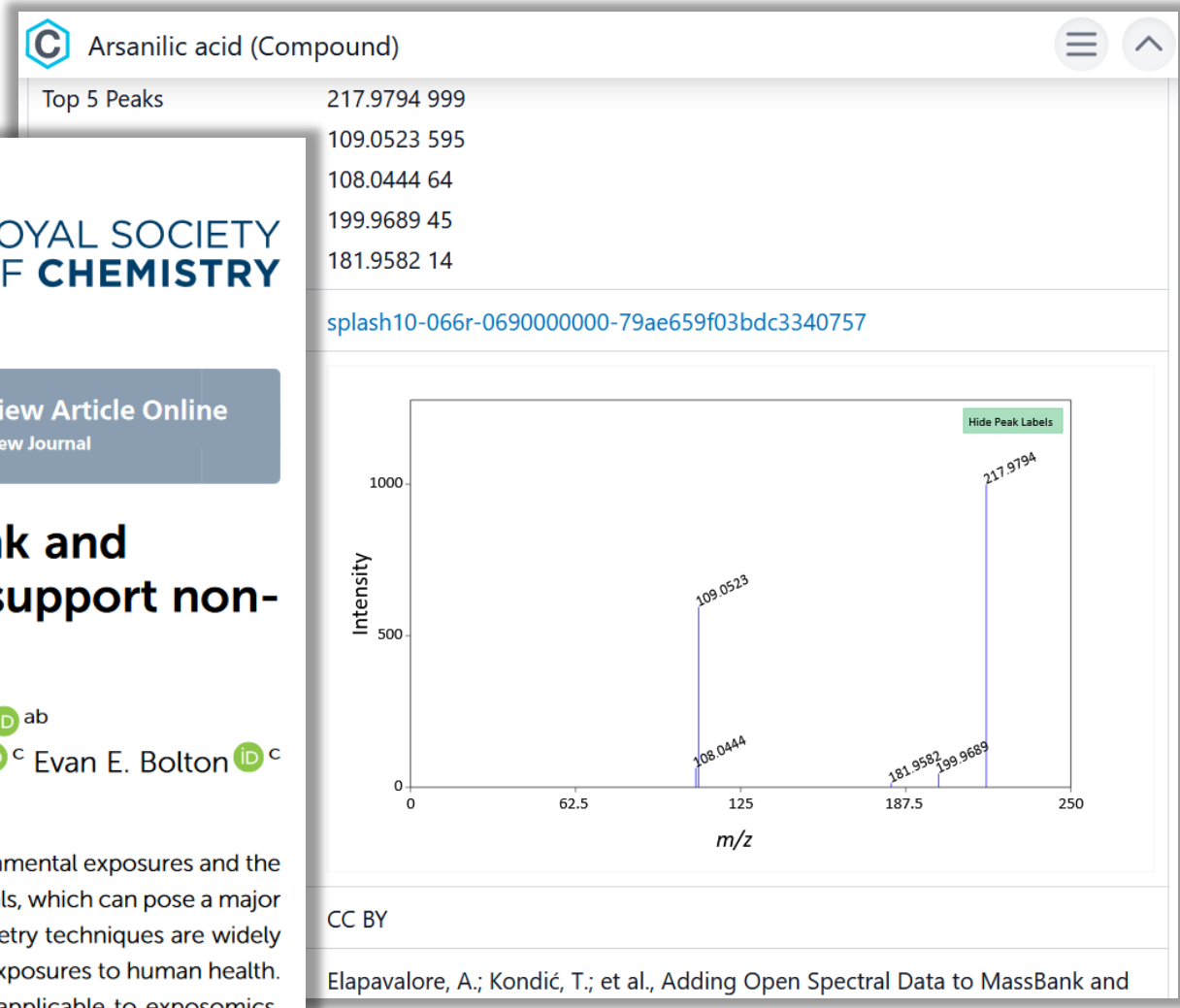
Mass Spectrum

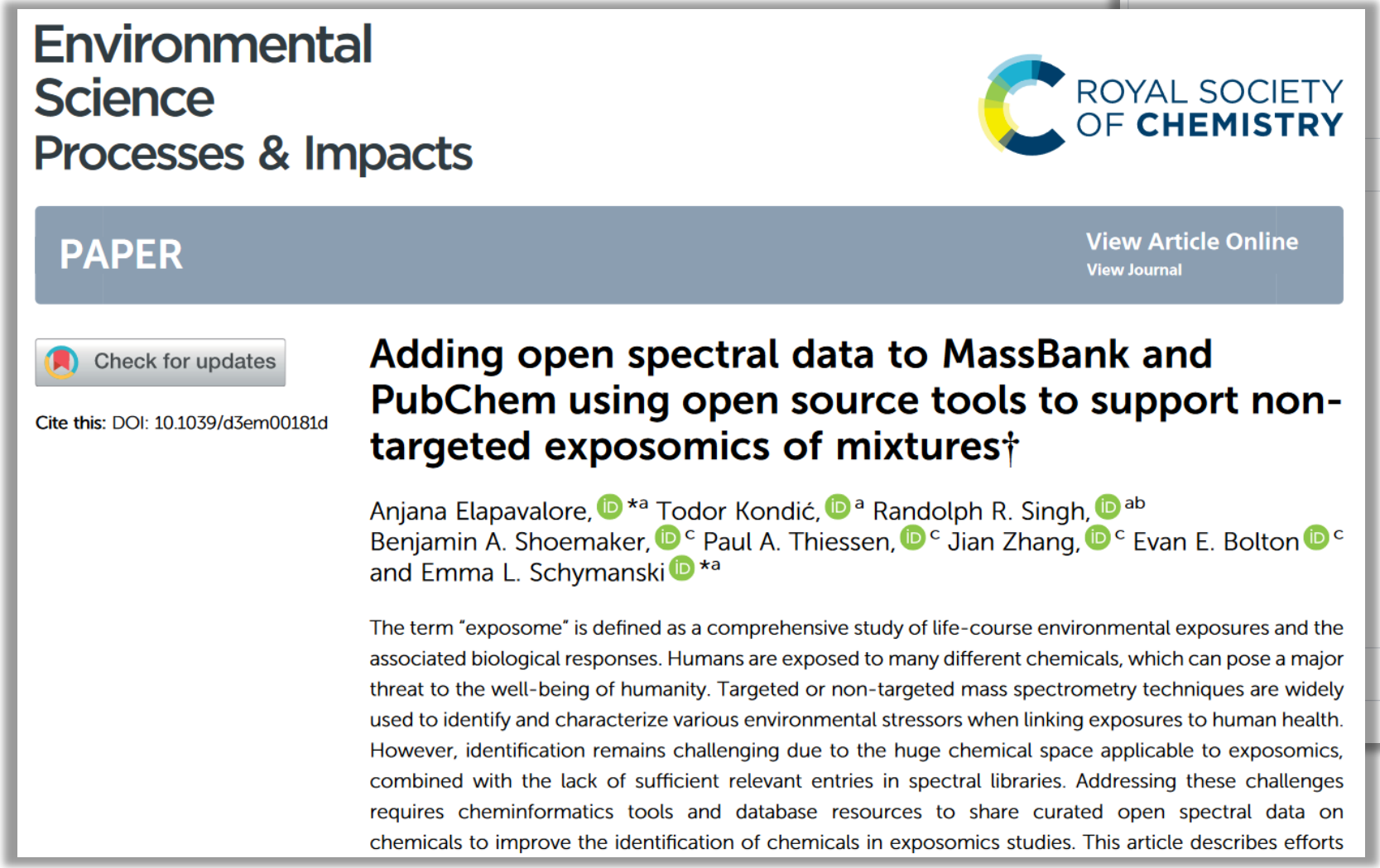


Chemical Structure



MassBank: Cross Resource Integration & New Article!





Environmental Science Processes & Impacts

ROYAL SOCIETY OF CHEMISTRY

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, *[a] Todor Kondić, [a] Randolph R. Singh, [ab] Benjamin A. Shoemaker, [c] Paul A. Thiessen, [c] Jian Zhang, [c] Evan E. Bolton [c] and Emma L. Schymanski *[a]

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

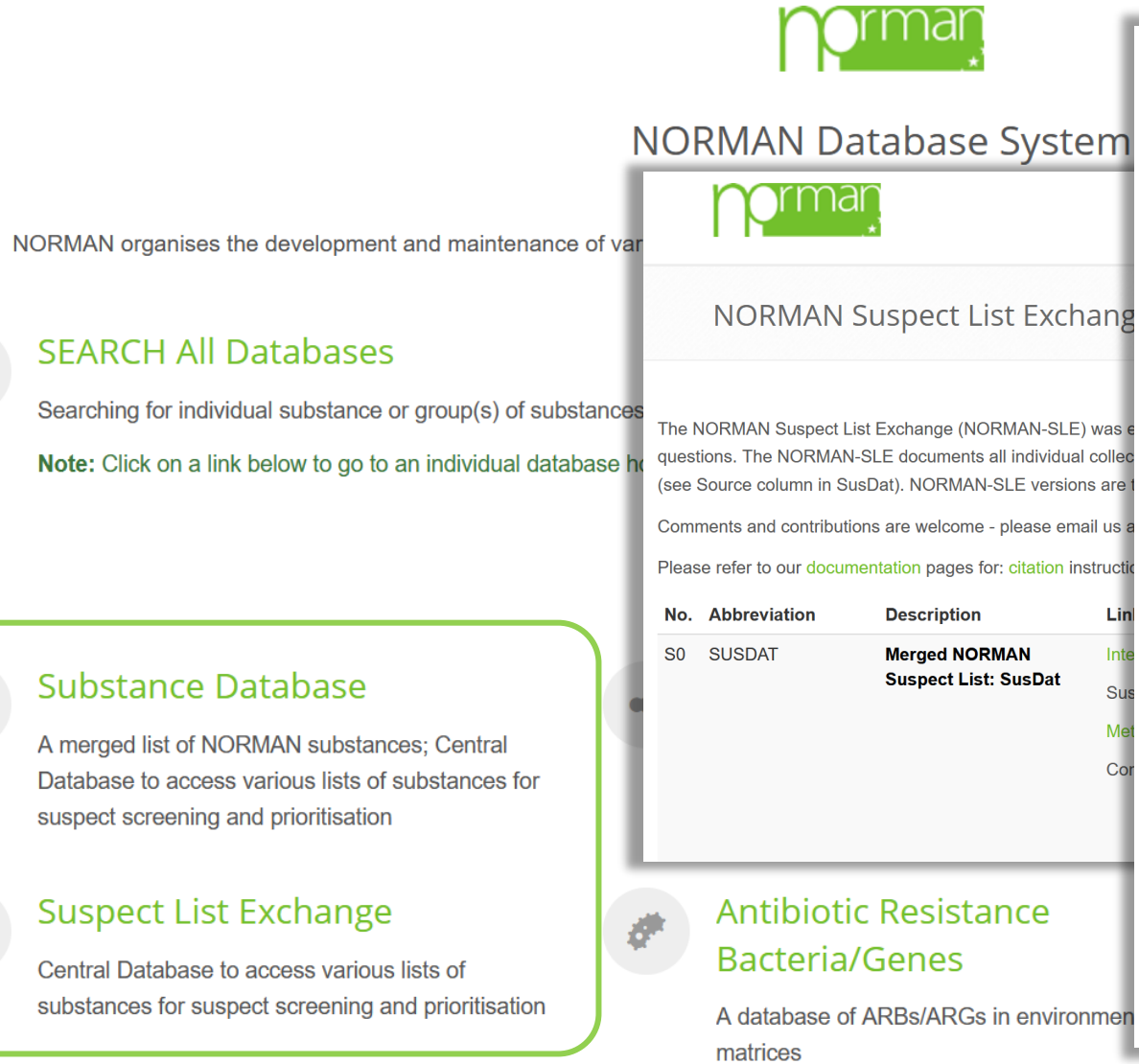
https://pubchem.ncbi.nlm.nih.gov/compound/7389#section=LC-MS

Elapavalore *et al*. (2023) DOI: 10.1039/D3EM00181D

Gathering Expert Knowledge: NORMAN-SLE

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

Mohammed Taha *et al*. (2022) DOI: 10.1186/s12302-022-00680-6

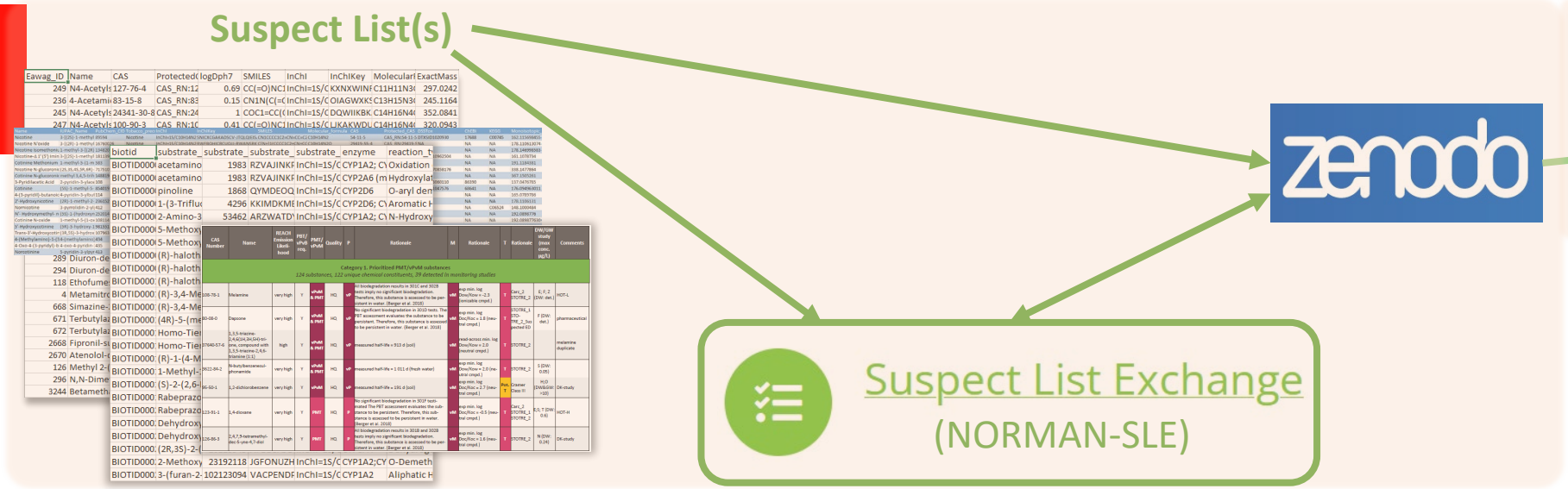


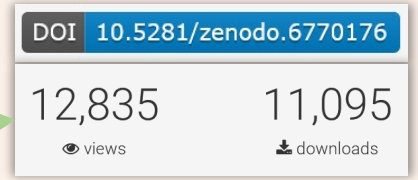
RESEARCH Open Access

Check for updates

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

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

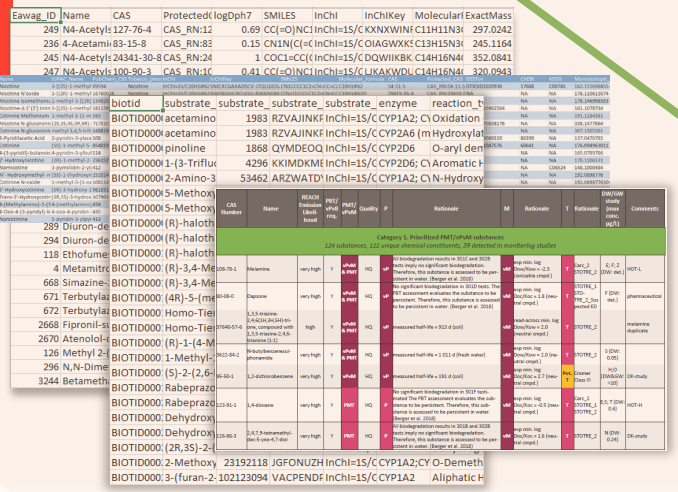


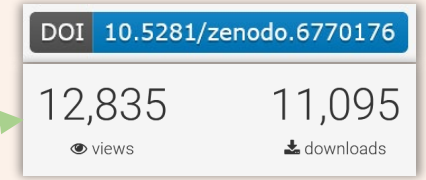


Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Suspect List(s)





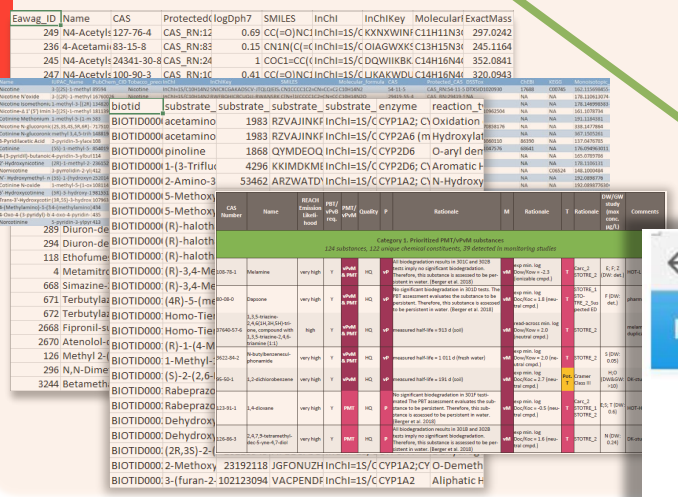
Suspect List Exchange (NORMAN-SLE)

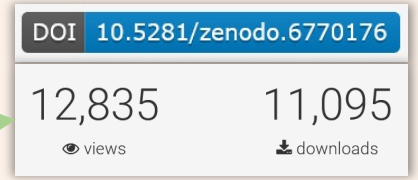
NORMAN Database System (NDS)

- Substance Database (NORMAN SusDat)
- Ecotoxicology
- 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

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

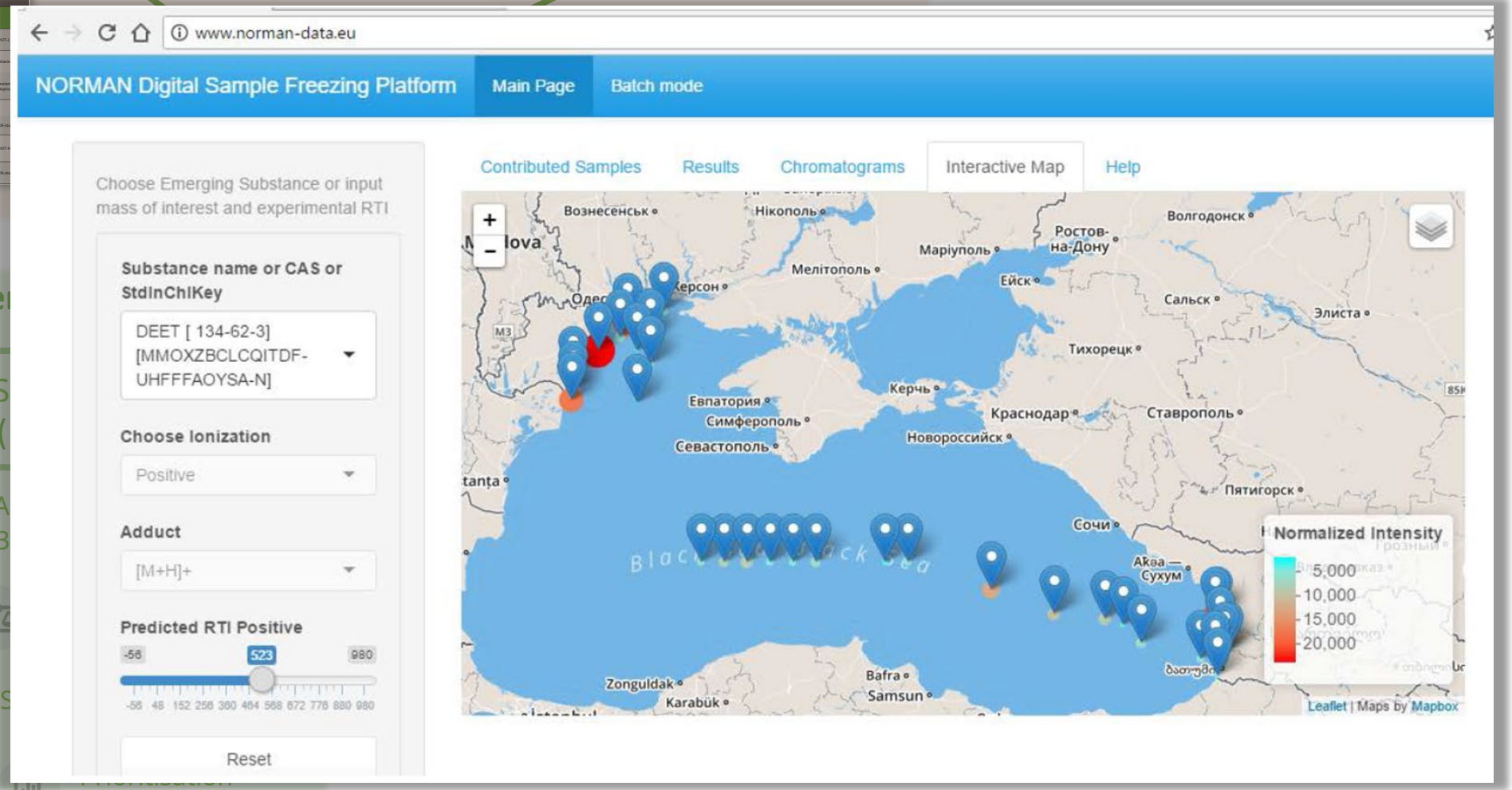
Suspect List(s)





NORMAN Database Syste

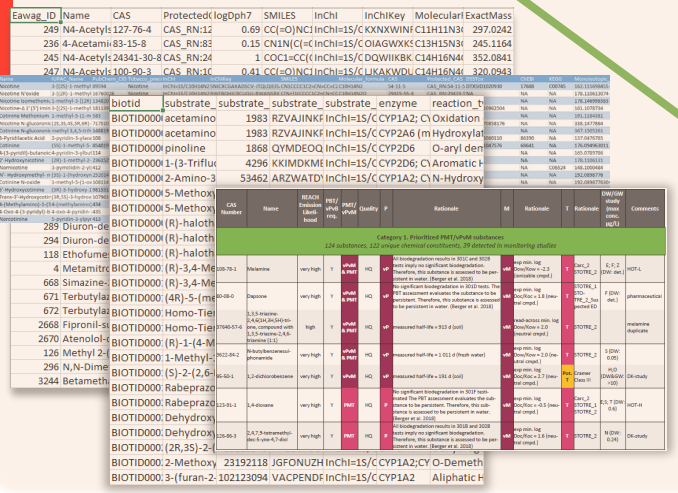
- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Indoor Environment
- Bioassays Monitoring Data
- Digital Sample Freezing Platform
- Chemical Occurrence Data



Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Suspect List(s)



zenodo

DOI 10.5281/zenodo.6770176

12,835 views

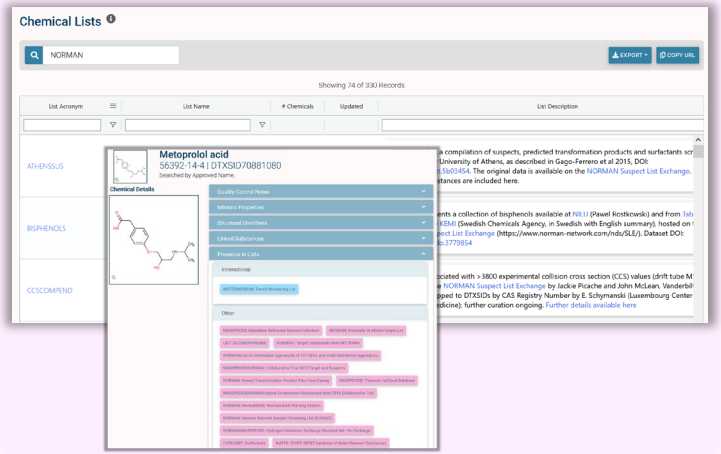
11,095 downloads

Suspect List Exchange (NORMAN-SLE)

NORMAN Database System (NDS)

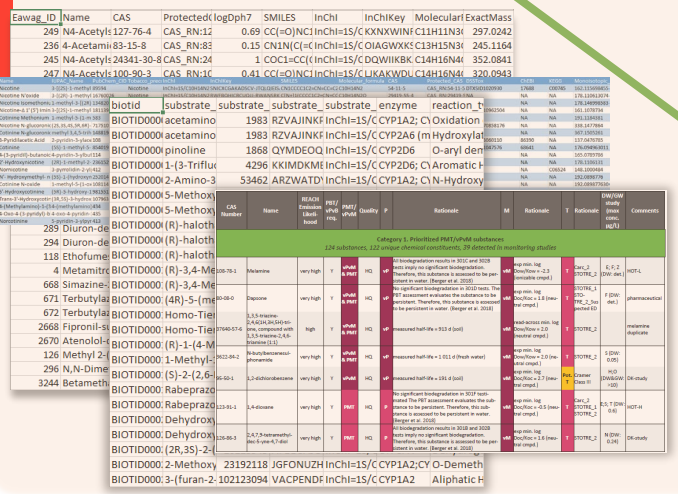
- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Indoor Environment
- Bioassays Monitoring Data
- Digital Sample Freezing Platform
- Chemical Occurrence Data
- Substance Database (NORMAN SusDat)
- Antibiotic Resistance Bacteria/Genes
- MassBank Europe
- Substance Factsheets
- Prioritisation

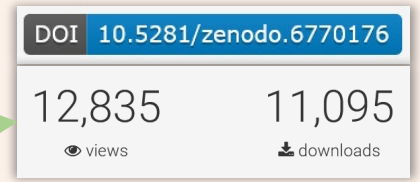
CompTox Chemicals Dashboard



Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

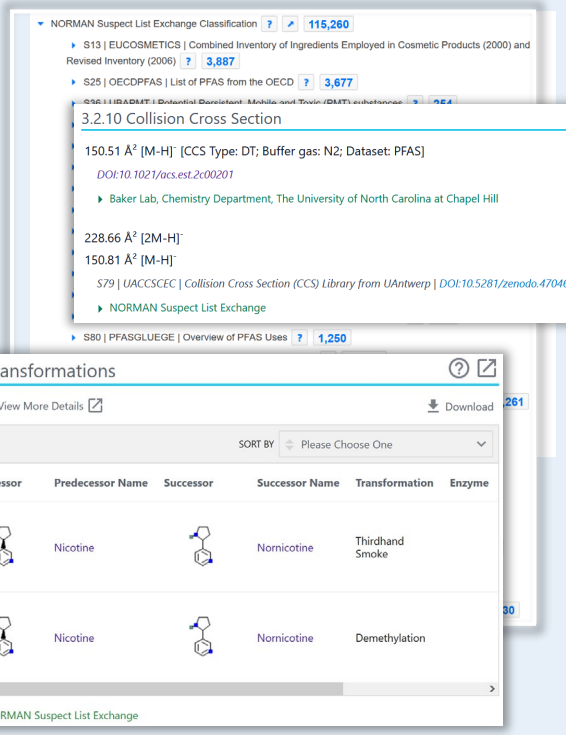
Suspect List(s)





Suspect List Exchange (NORMAN-SLE)

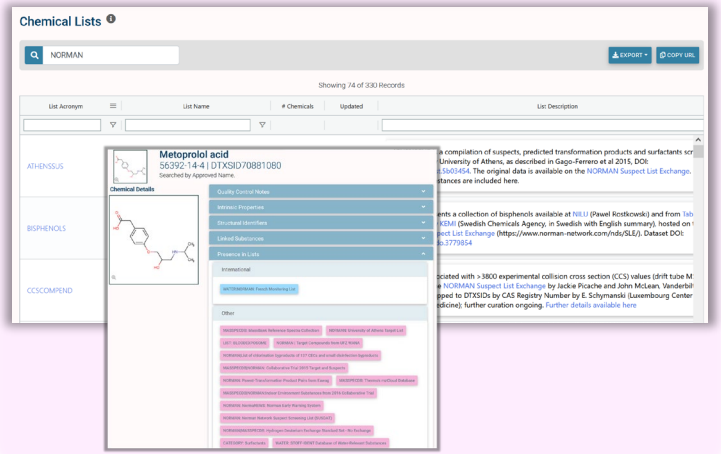
PubChem



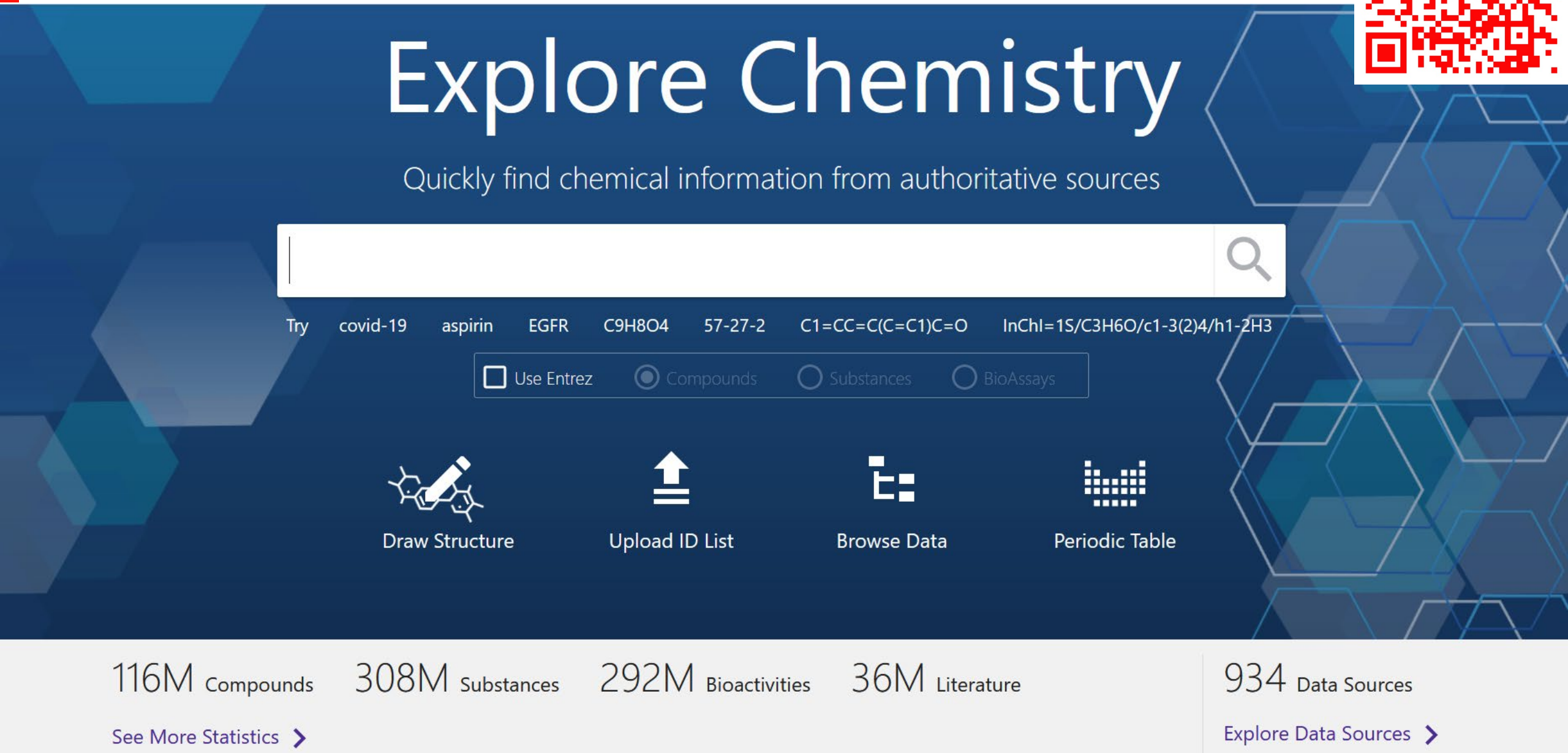
NORMAN Database System (NDS)

- 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



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

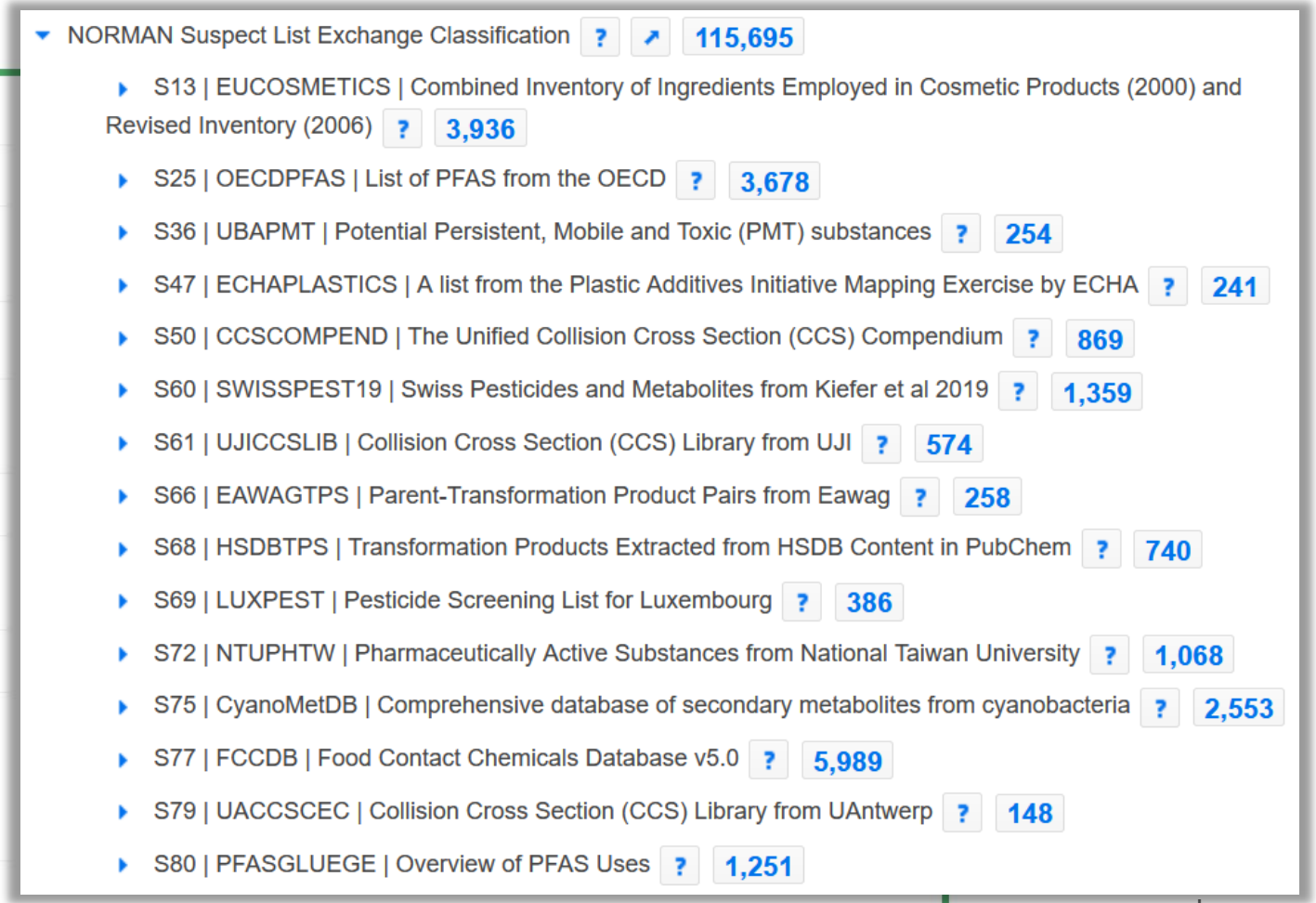


NORMAN-SLE in PubChem

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

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

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



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101 & https://pubchem.ncbi.nlm.nih.gov/source/23819

- NORMAN Suspect List Exchange Classification ? 115,248
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cos
    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 E
  - 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 Universit
  - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanoba
  - S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148
  - S80 | PFASGLUEGE | Overview of PFAS Uses ? 1,250

S00
S01
S02
S03
S04

ances ?

4

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

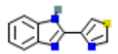
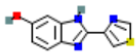
NORMAN Suspect List Exchange

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	Evidenc
	thiabendazole		5-hydroxythiabendazole	Aromatic hydroxylation of fused benzene ring / Human Phase I	CYP1A2	10.1186/

PubChem Tri-O-cresyl phosphate (Compound)

3.2.18 Collision Cross Section

192.43 Å² [M+Na]⁺
182.39 Å² [M+H]⁺
263.75 Å² [2M+Na]⁺

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

NORMAN Suspect List Exchange

PubChem 7-Desmethylmicrocystin RR (Compound)

7 Taxonomy

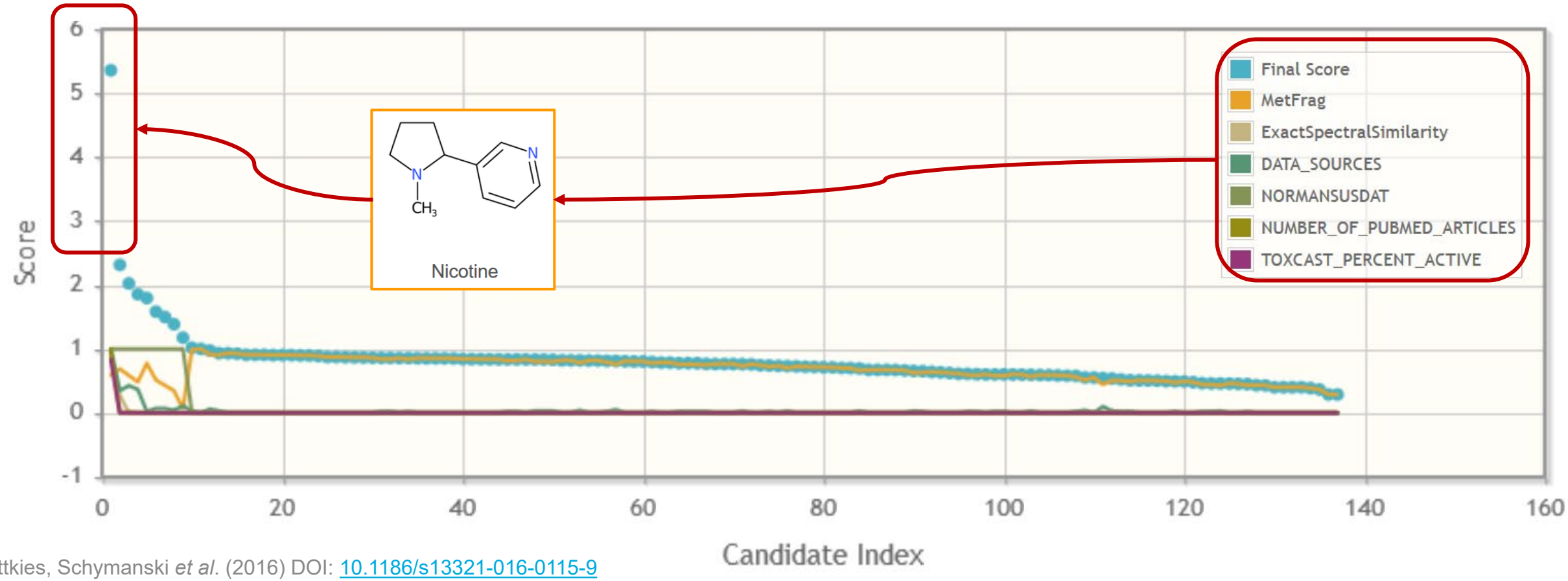
2 items Download

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

Mohammed Taha *et al*. (2022) DOI: 10.1186/s12302-022-00680-6

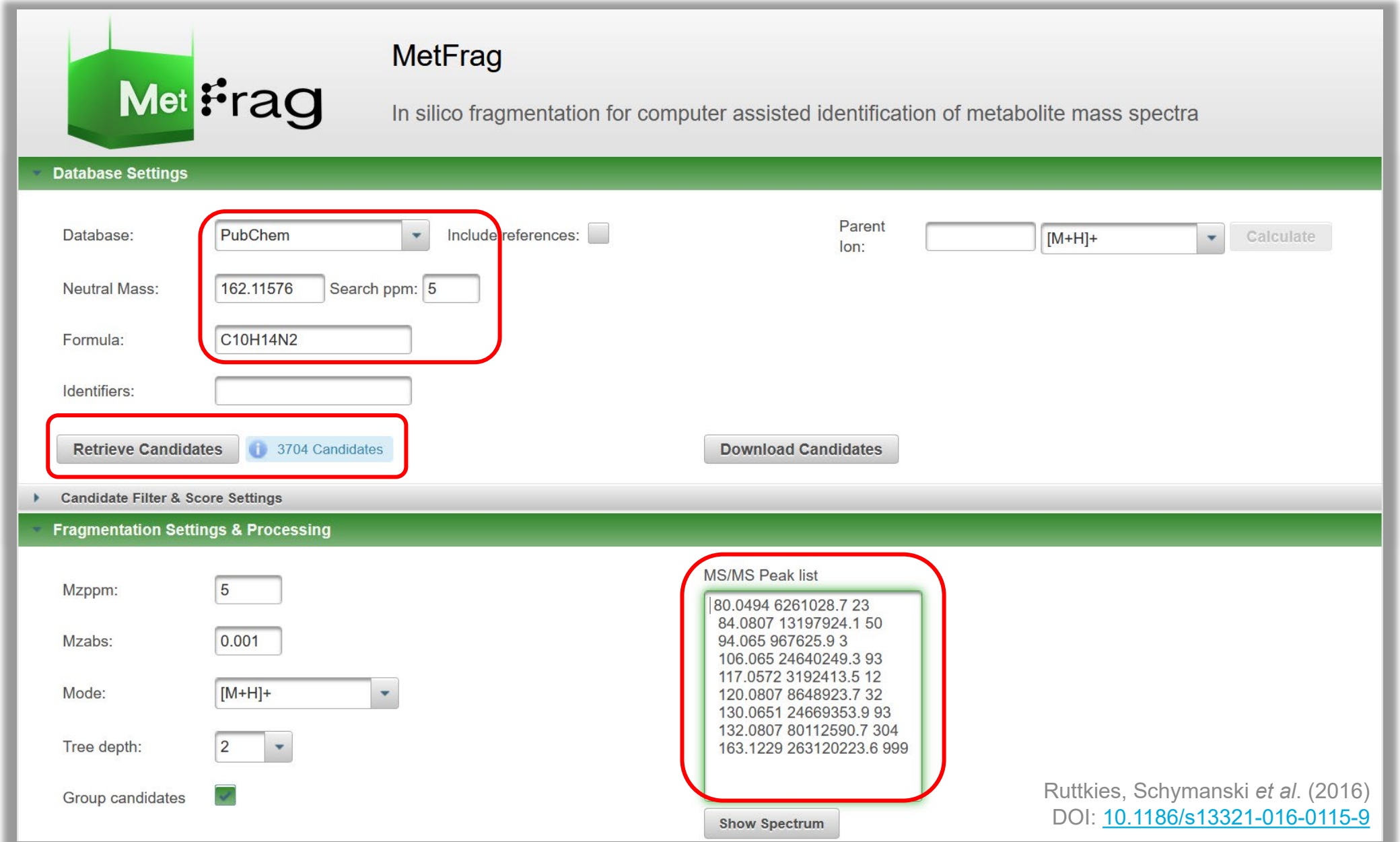
Connecting Knowledge for Chemical Identification: MetFrag

https://msbi.ipb-halle.de/MetFrag/ http://ipb-halle.github.io/MetFrag/



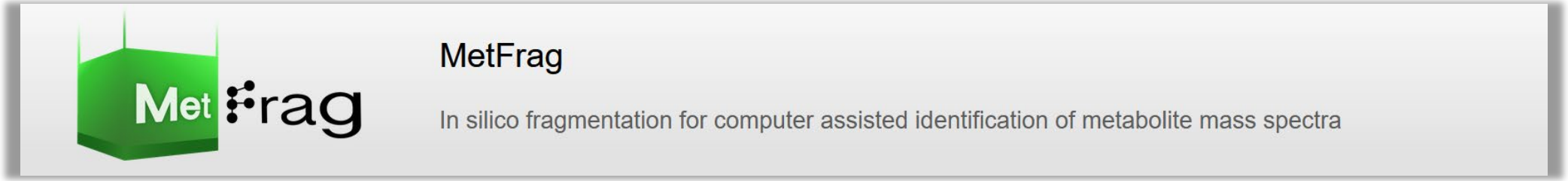
Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

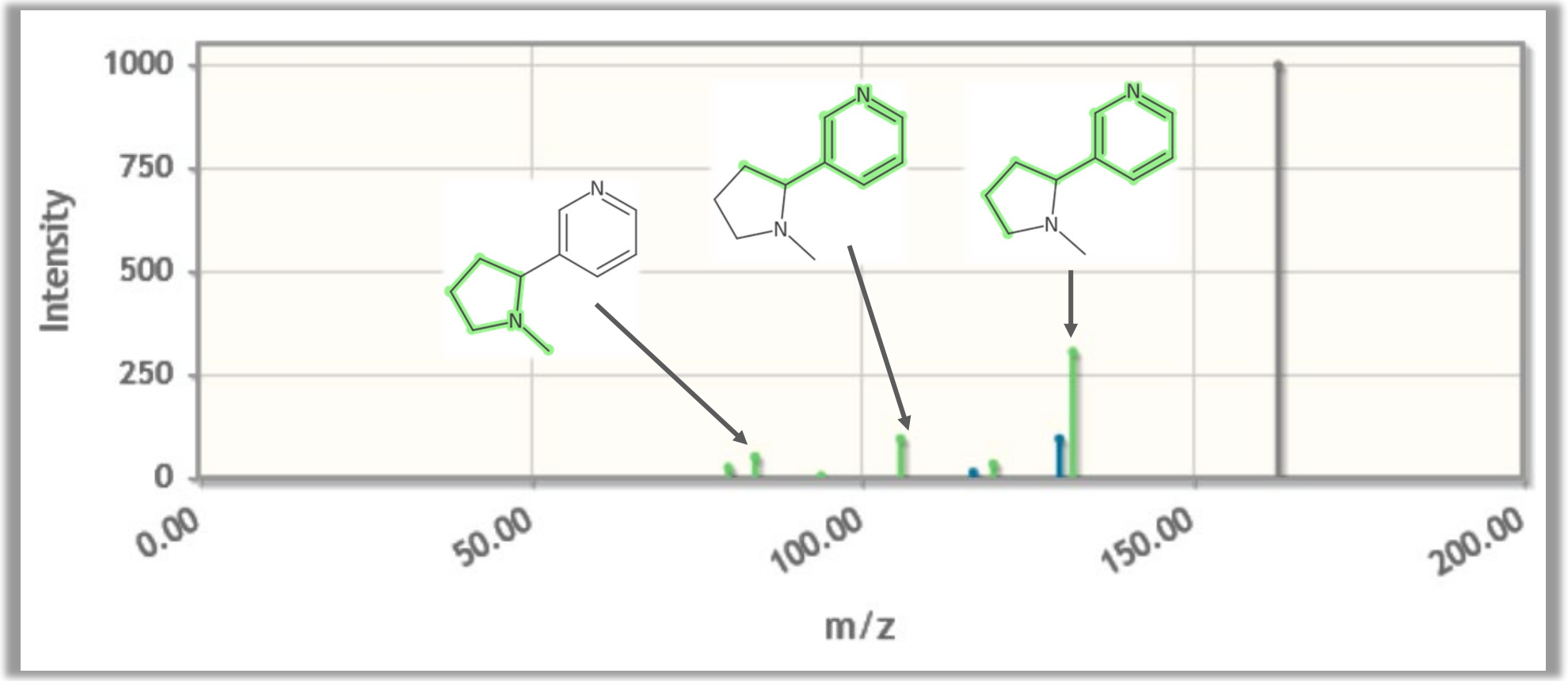
MetFrag: *In Silico* Fragmentation to Select Candidates



Ruttkies, Schymanski *et al*. (2016)
DOI: 10.1186/s13321-016-0115-9

MetFrag: *In Silico* Fragmentation to Select Candidates



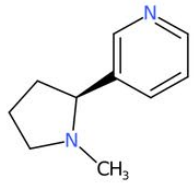
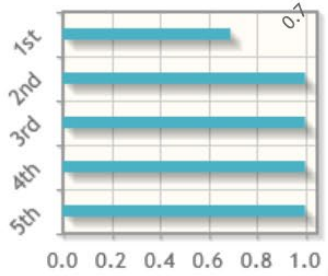
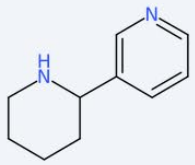
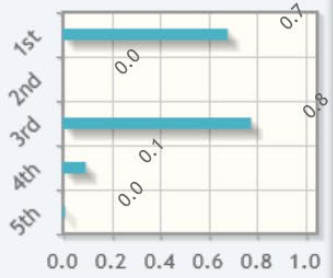
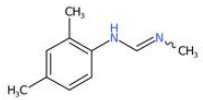
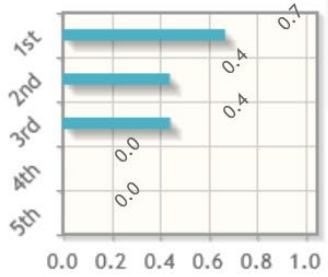


Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

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	 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine	89594 InChIKeyBlock1 = SNICXCGAKADSCV	162.1157	$C_{10}H_{14}N_2$		4.692	Peaks: 6 / 8 Fragments Scores Download
2	 3-piperidin-2-ylpyridine	2181 InChIKeyBlock1 = MTXSIJUGVMTTMU	162.1157	$C_{10}H_{14}N_2$			
3	 N-(2,4-dimethylphenyl)-N'-methylmethanimidamide	36326 InChIKeyBlock1 = JIIOLEGNERQDIP	162.1157	$C_{10}H_{14}N_2$		1.5586	Scores Download

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)

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

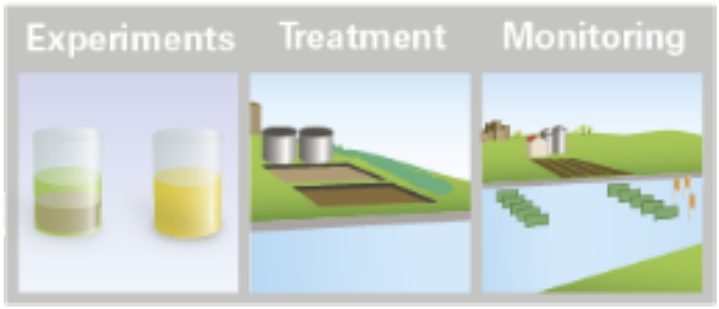
Review | Open Access | Published: 04 September 2023 | 35, Article number: 75 (2023)

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

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



Commercial / Open Software

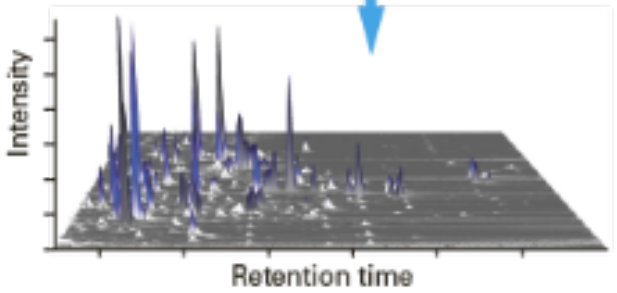
Databases

In Silico Prediction

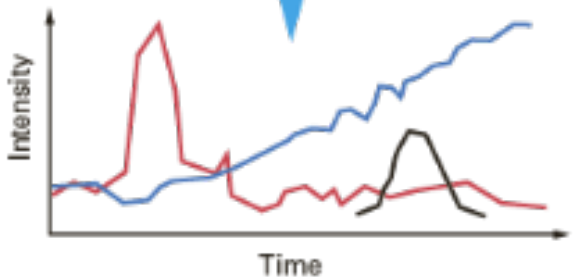
Sampling → Analysis → Data Pre-Processing → Prioritization → Identification

SAMPLE ARCHIVE

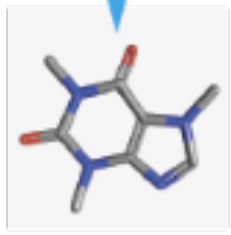
DIGITAL ARCHIVE



POLLUTION OVERVIEW



PATTERN



STRUCTURES

Modified from: Hollender, Schymanski *et al*. 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Open Source Workflows for NT-HRMS: patRoon

Software | Open Access | Published: 06 January 2021

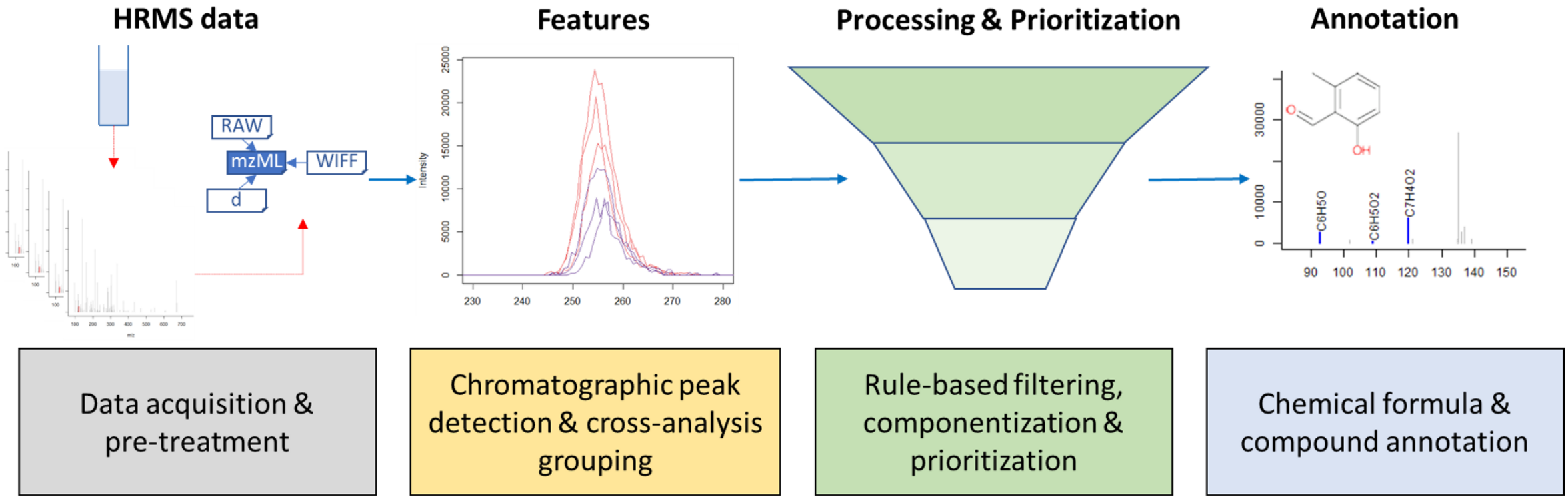
patRoon: open source software platform for environmental mass spectrometry based non-target screening

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

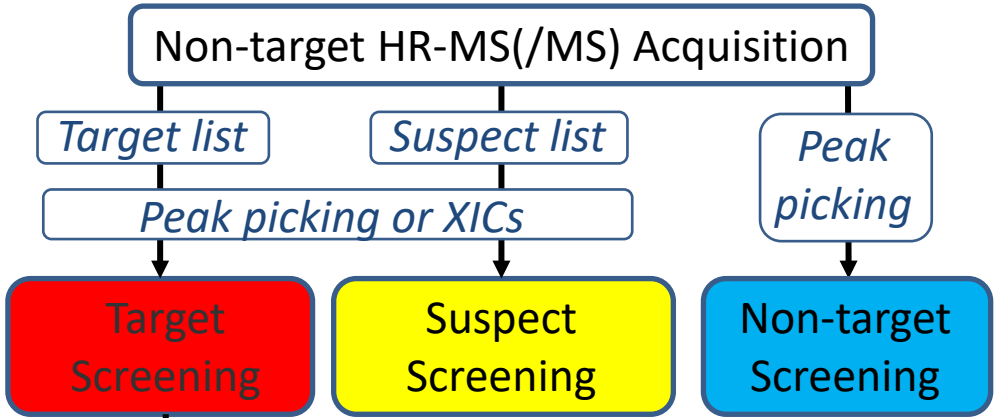
patRoon 2.0: Improved non-target analysis workflows including automated transformation product screening

Rick Helmus[1¶], Bas van de Velde[123], Andrea M. Brunner[24], Thomas L. ter Laak[12], Annemarie P. van Wezel[1], and Emma L. Schymanski[5]



Helmus *et al.* (2022) patRoon 2.0 *J. Open Source Software*, DOI: 10.21105/joss.04029
Helmus *et al.* (2021) patRoon. *J. Cheminformatics*. DOI: 10.1186/s13321-020-00477-w

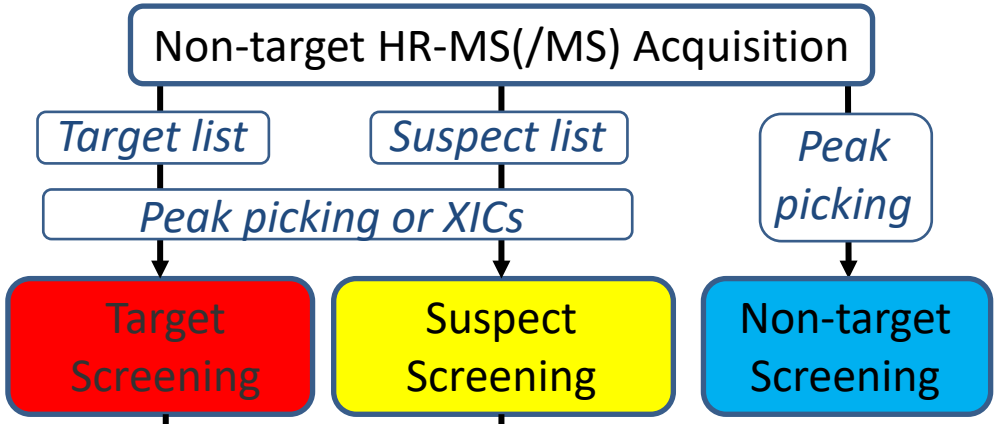
Identification Strategies and Confidence in NT-HRMS(/MS)



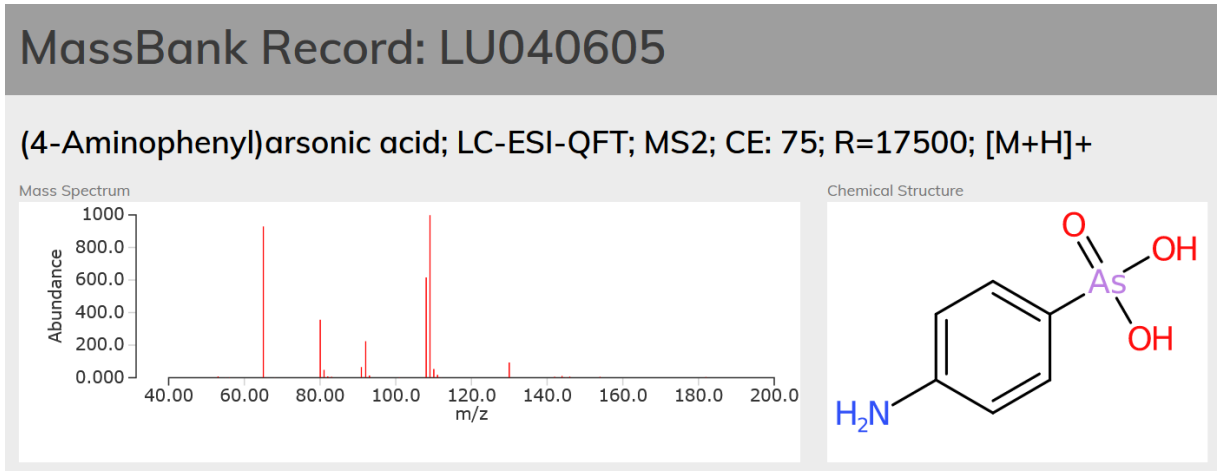
Limited by ref. std. availability …

Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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

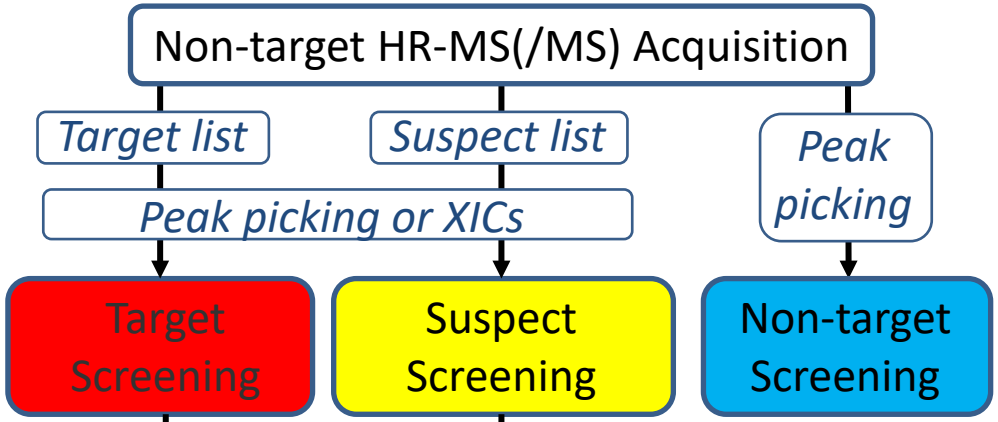


Level 1	**Confirmed Structure** by reference standard	*Limited by ref. std. availability …*
Level 2	**Probable Structure** by library/diagnostic evidence	MassBank High Quality Mass Spectral Database



Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105
Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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

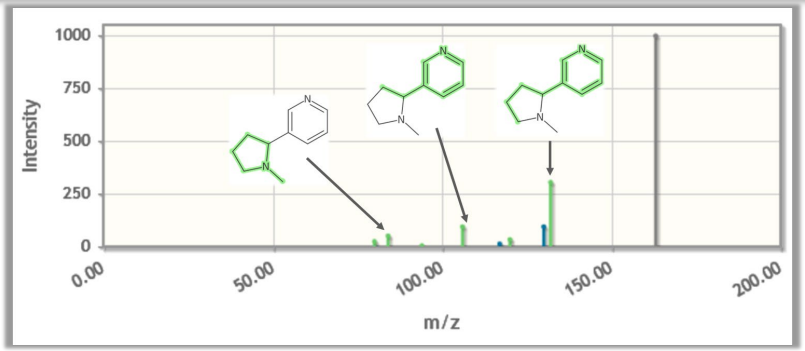


Start			**Level 1**	**Confirmed Structure** by reference standard
			Level 2	**Probable Structure** by library/diagnostic evidence
	Start		**Level 3**	**Tentative Candidate(s)** suspect, substructure, class

Limited by ref. std. availability …

MassBank
High Quality Mass Spectral Database

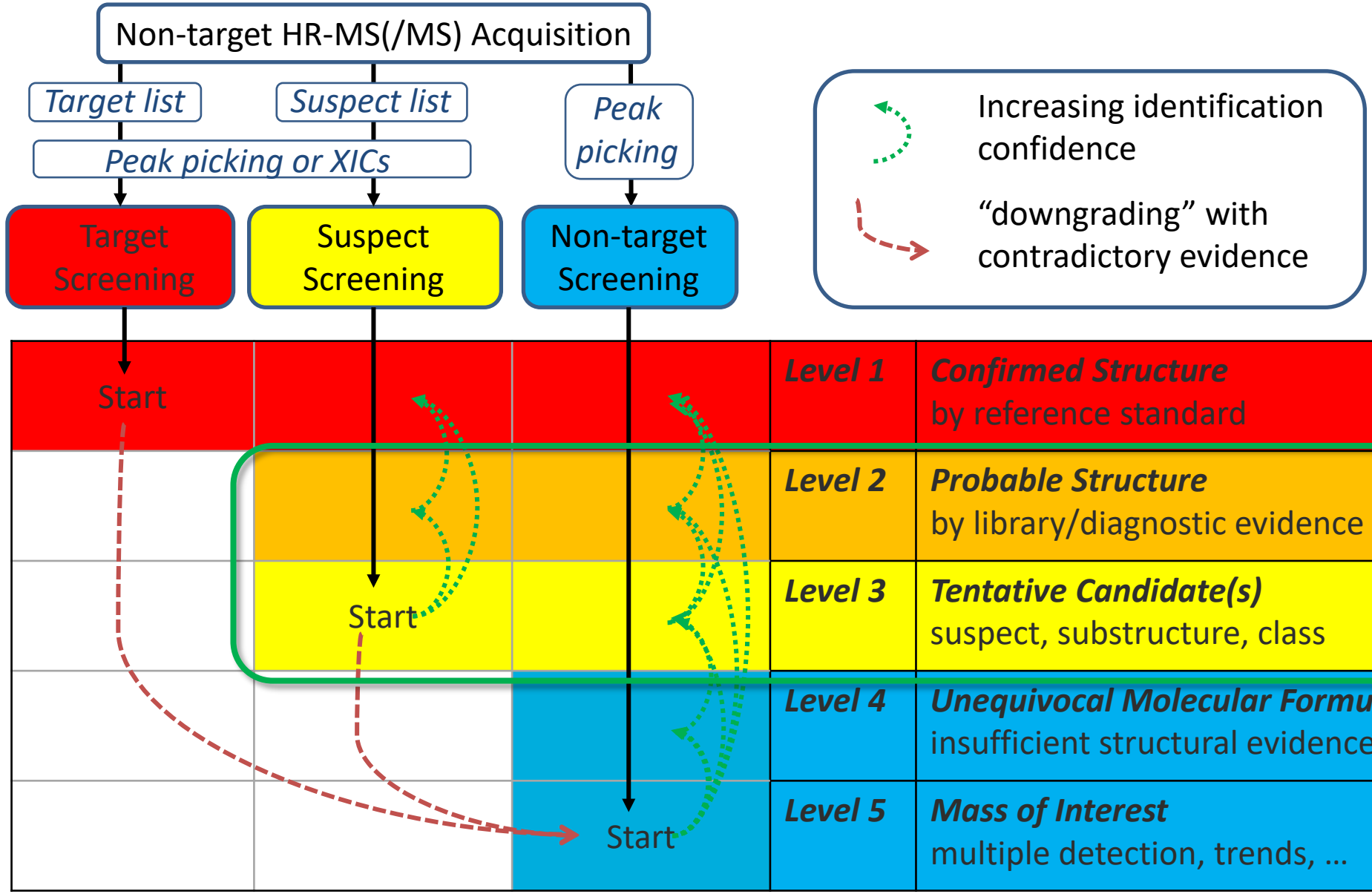
MetFrag



Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105
Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

eawag aquatic research

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



Limited by ref. std. availability …

MassBank
High Quality Mass Spectral Database

MetFrag

The "dark matter"

eawag aquatic research

Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

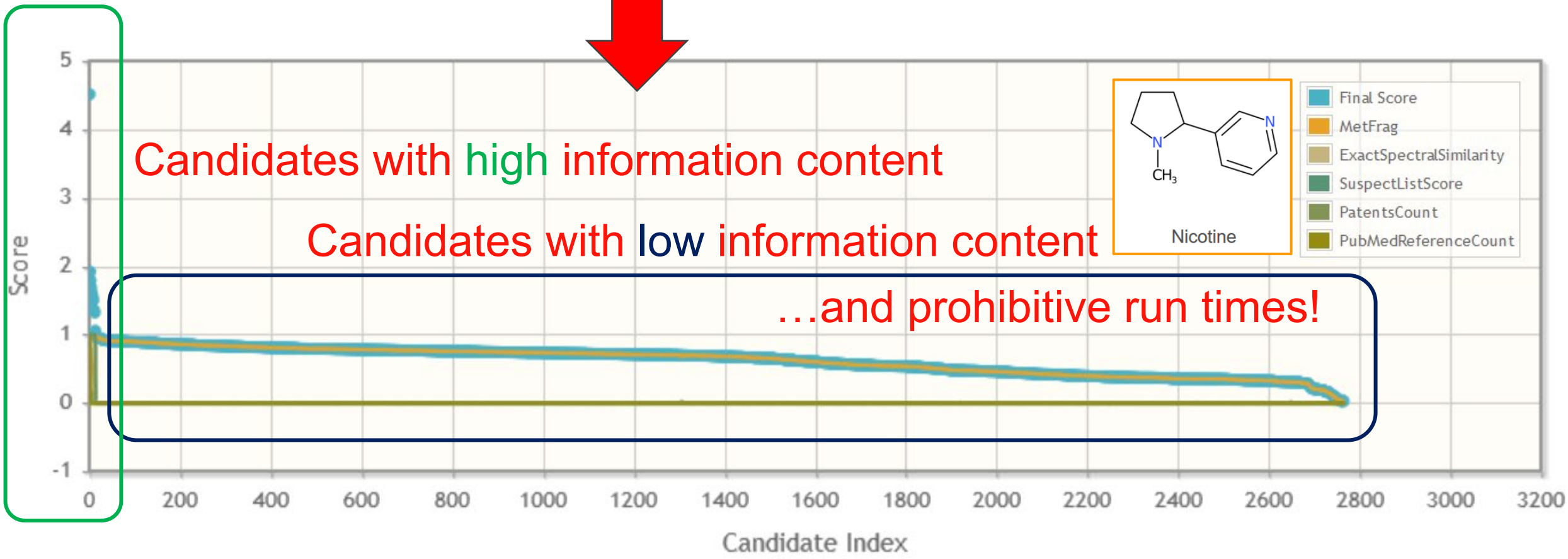
Grand Challenge: HR-MS “Chemical Space” is too Big!

CAS® A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 204 million

PubChem — 116 million

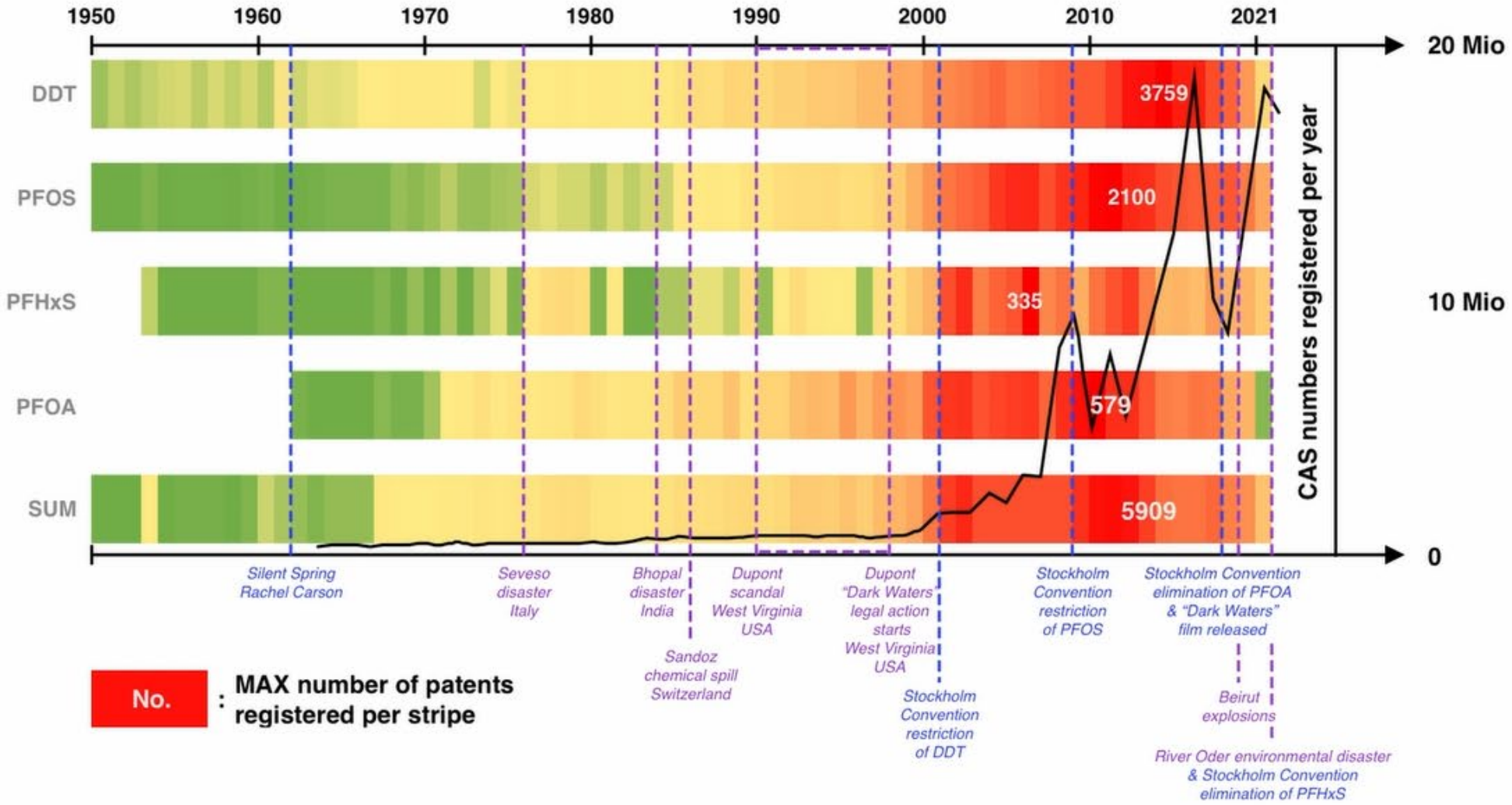
ChemSpider Search and share chemistry — 128 million

CompTox Chemicals Dashboard — 1.2 million



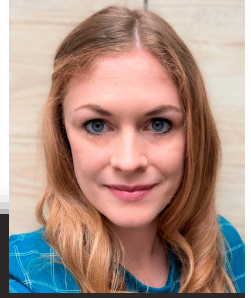
Our Chemical Past, Present, and Future

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



Composition & Animation: Jamie Perera; Figure: Dagny Aurich; mod. from Arp et al. (2023) DOI: 10.1021/acs.est.3c01735. Credits: Jamie Perera, Dagny Aurich, Paul Thiessen, Emma Schymanski

Chemical Stripes in R - for Patents & Literature

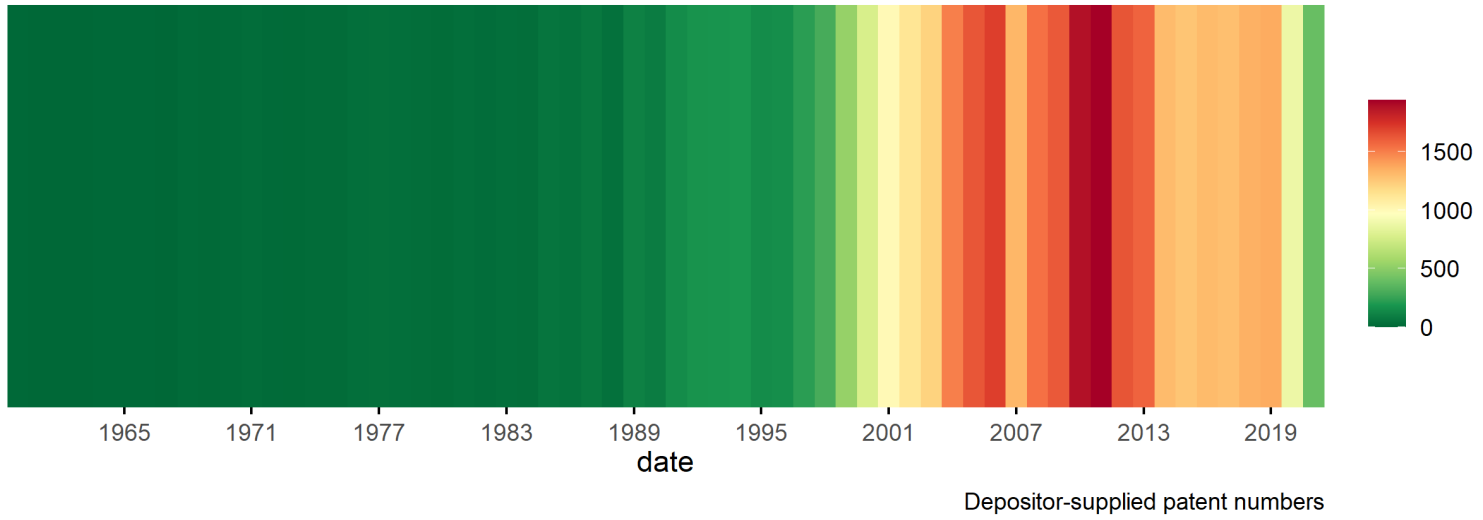


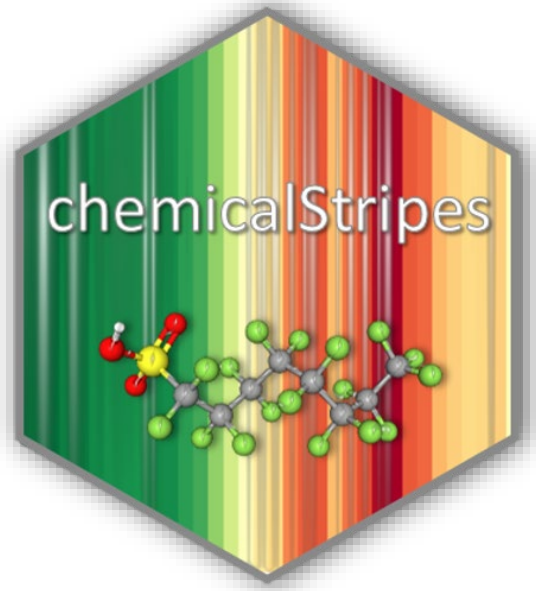
```
> chemical_stripes(74483)
Getting compound information
A total of 32461 patents were found for CID 74483
[=========================================================>----------------------------------------------------------------------]
Downloading patent data...
[=================================================================================================>-----------------------------]  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: C8HF17O3S
Exact Mass: 499.9374938

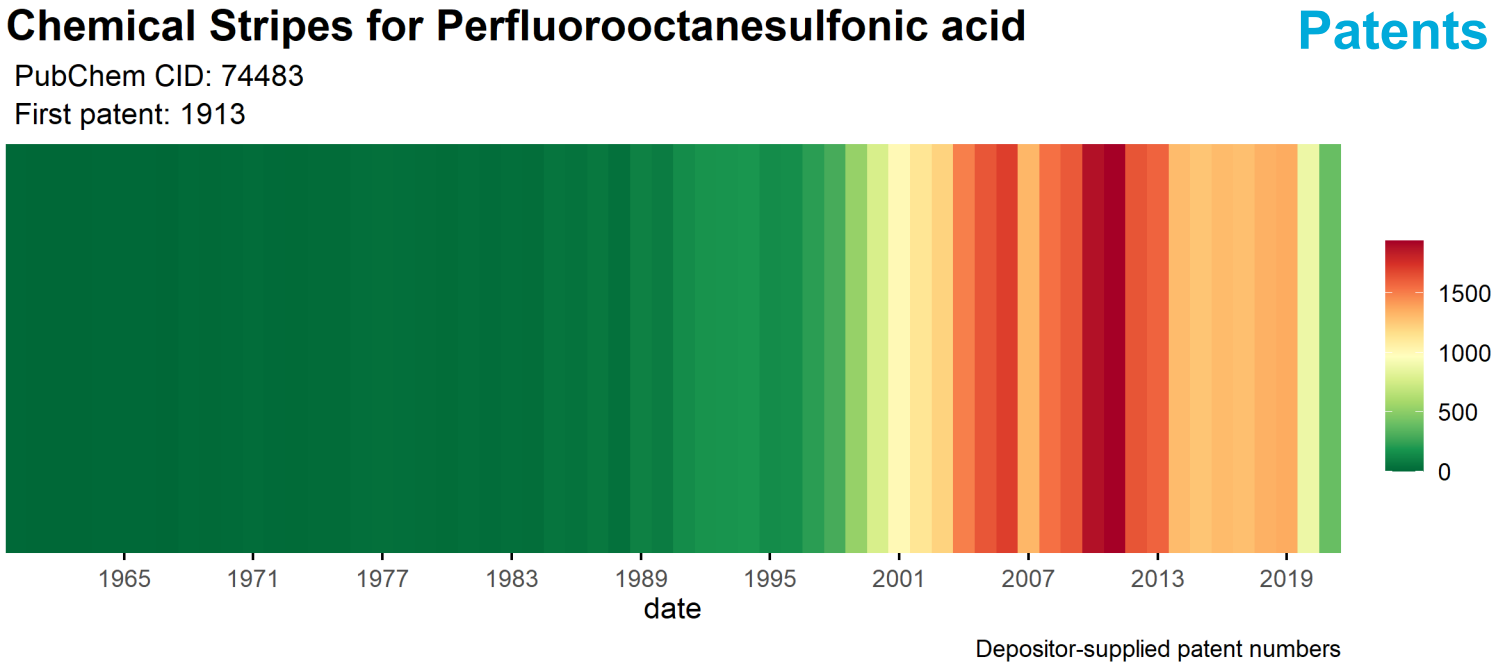


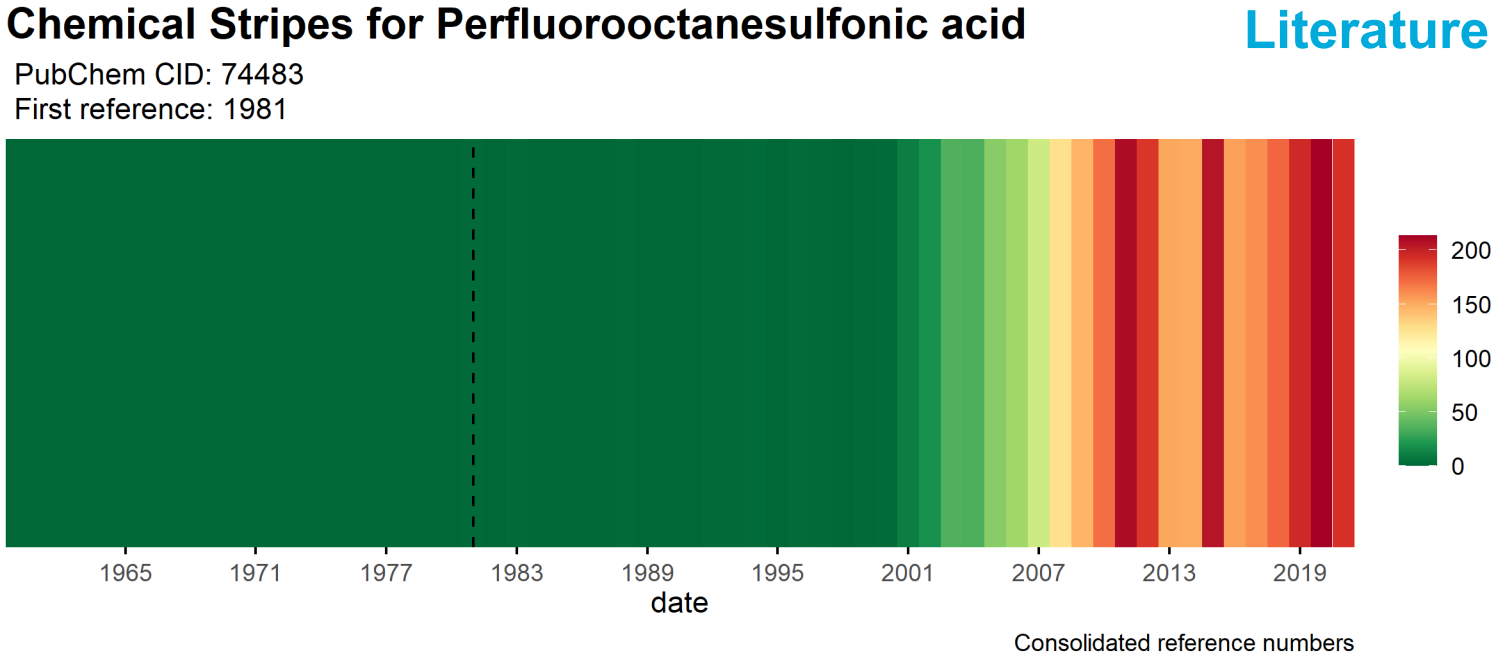


Aurich *et al* (2023). SETAC Presentation, DOI: 10.5281/zenodo.7885031 and https://gitlab.lcsb.uni.lu/eci/chemicalstripes

Chemical Stripes in R

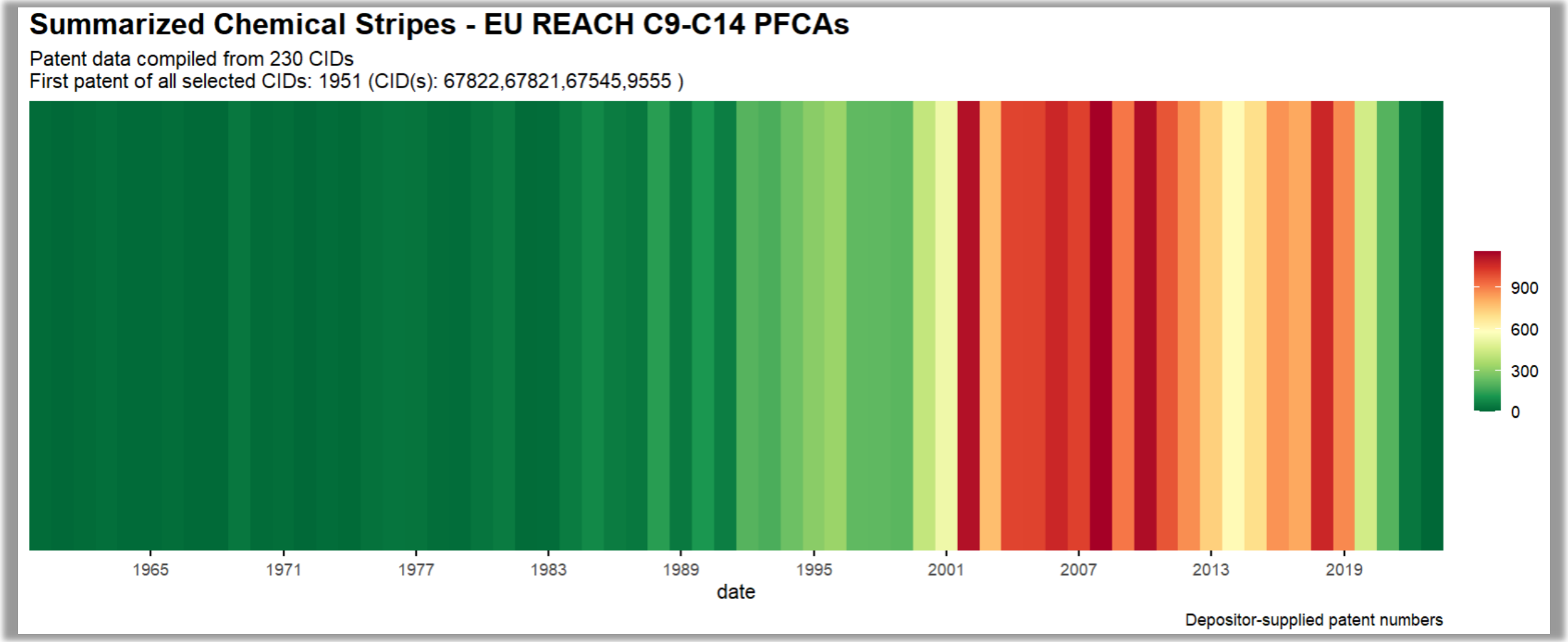
Patents & Literature





Schymanski *et al.*(2023) DOI: 10.26434/chemrxiv-2023-j823z
https://gitlab.lcsb.uni.lu/eci/chemicalstripes

“Summarized” Chemical Stripes in R

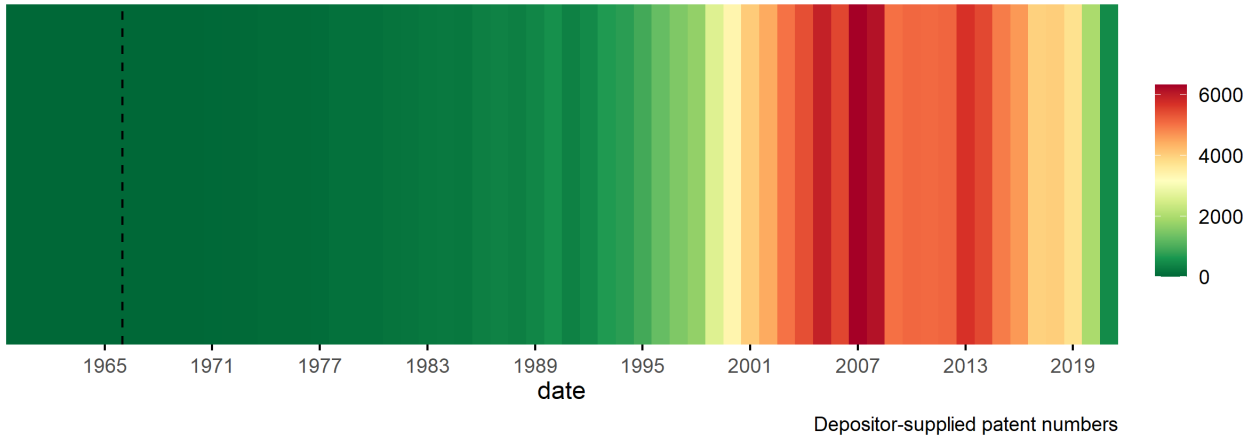


https://gitlab.lcsb.uni.lu/eci/chemicalstripes

Variety of Stripe Patterns: Patents

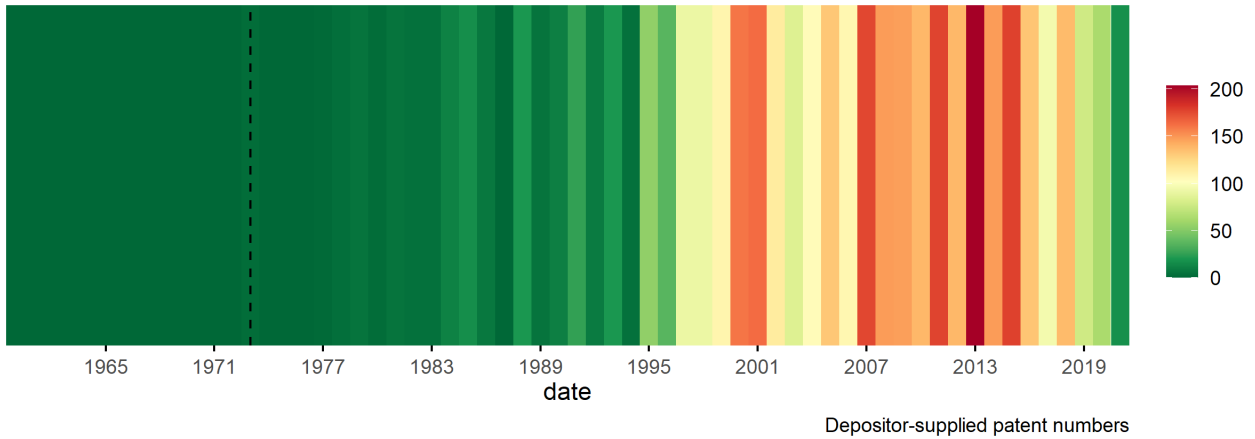
Chemical Stripes for Caffeine

PubChem CID: 2519
First patent: 1966



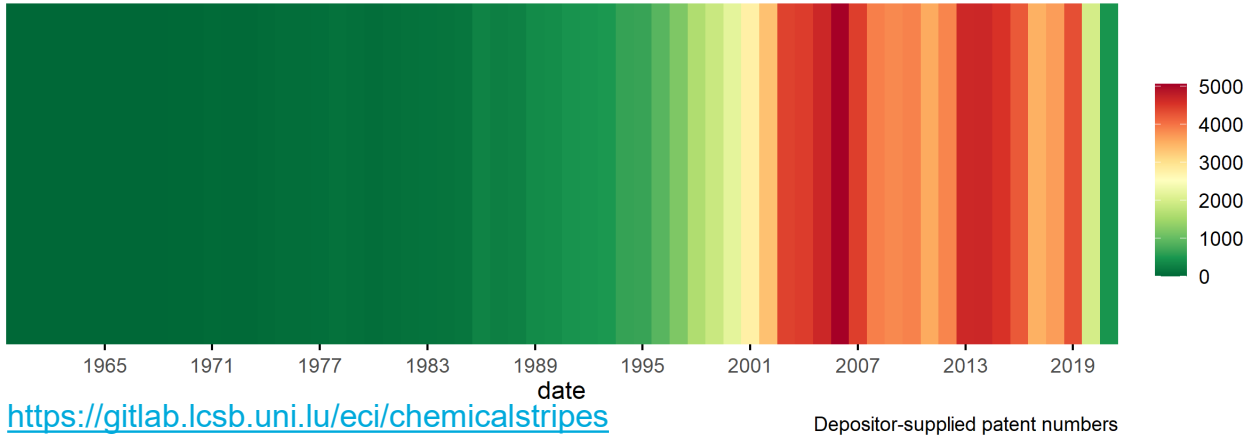
Chemical Stripes for Acetylcarnitine

PubChem CID: 1
First patent: 1973



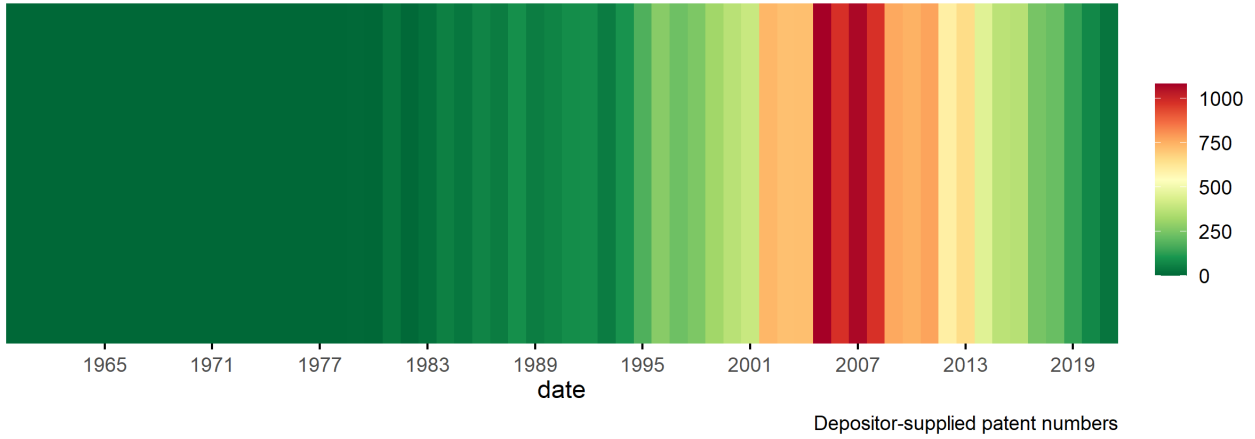
Chemical Stripes for Nicotine

PubChem CID: 89594
First patent: 1954



Chemical Stripes for Gallopamil

PubChem CID: 1234
First patent: 1951

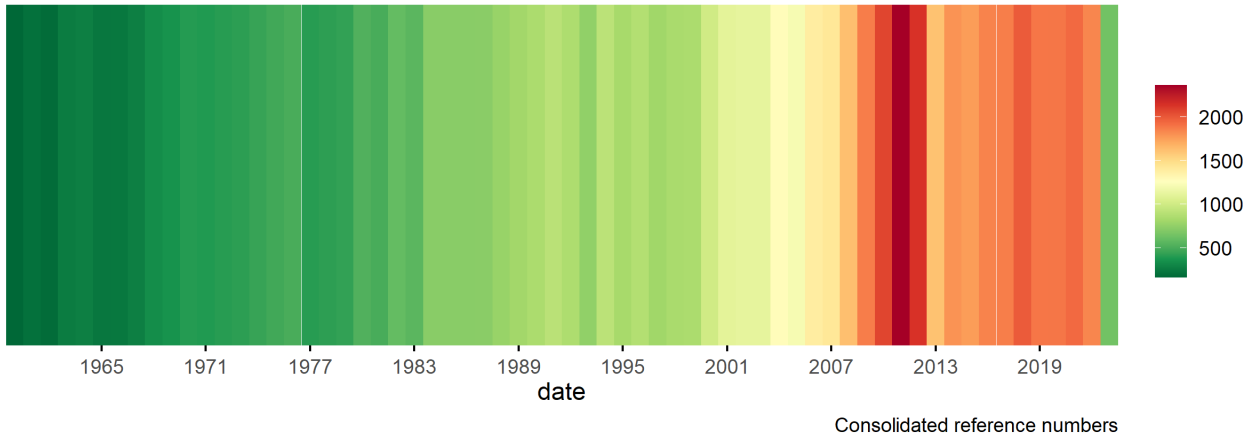


https://gitlab.lcsb.uni.lu/eci/chemicalstripes

Variety of Stripe Patterns: Literature

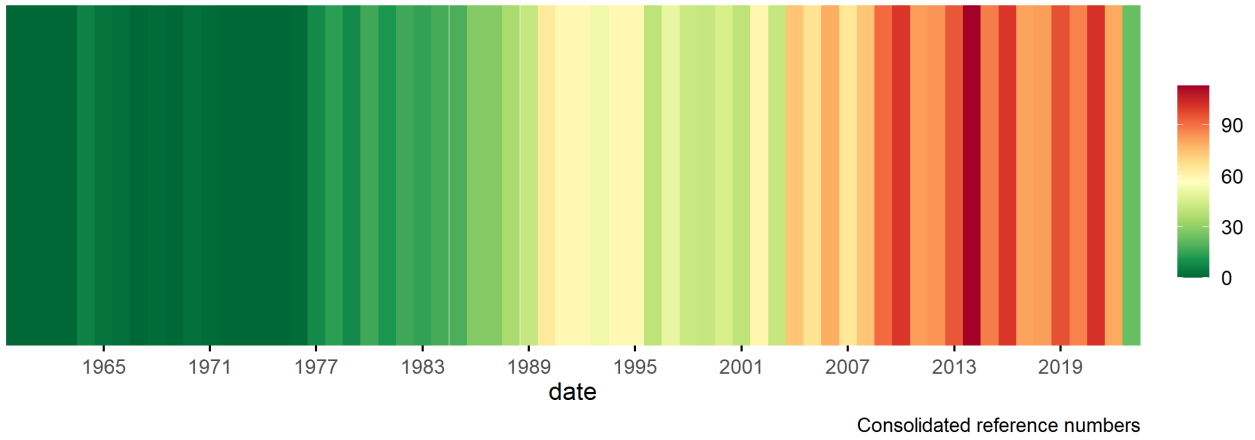
Chemical Stripes for Caffeine

PubChem CID: 2519
First reference: 1851



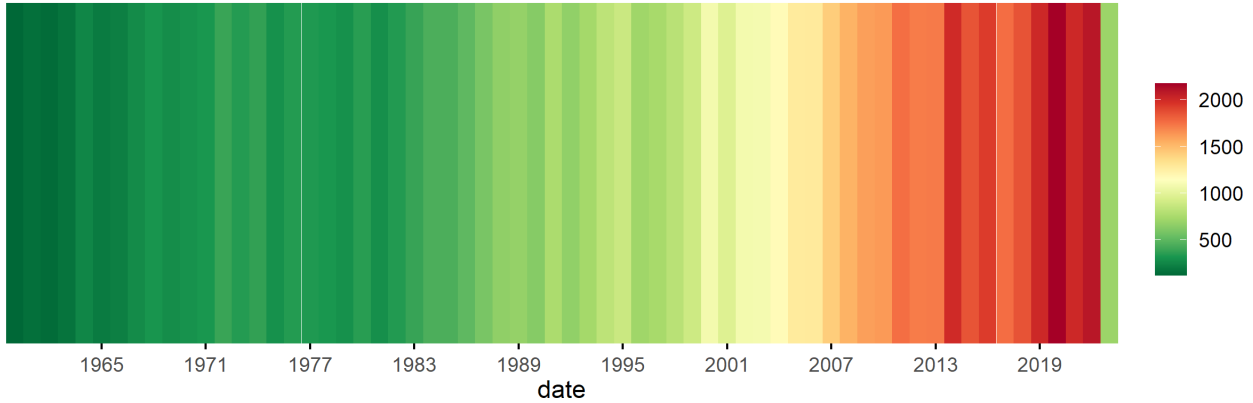
Chemical Stripes for Acetylcarnitine

PubChem CID: 1
First reference: 1959



Chemical Stripes for Nicotine

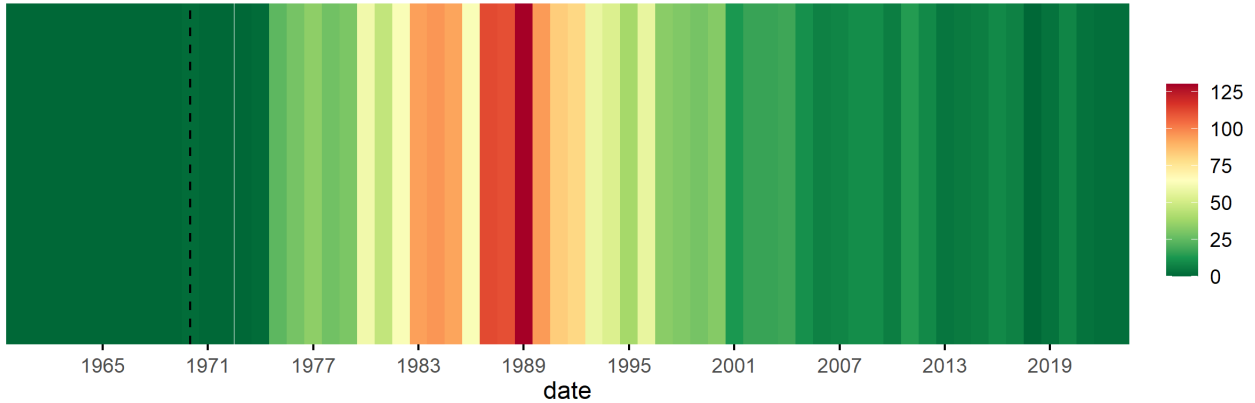
PubChem CID: 89594
First reference: 1854



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

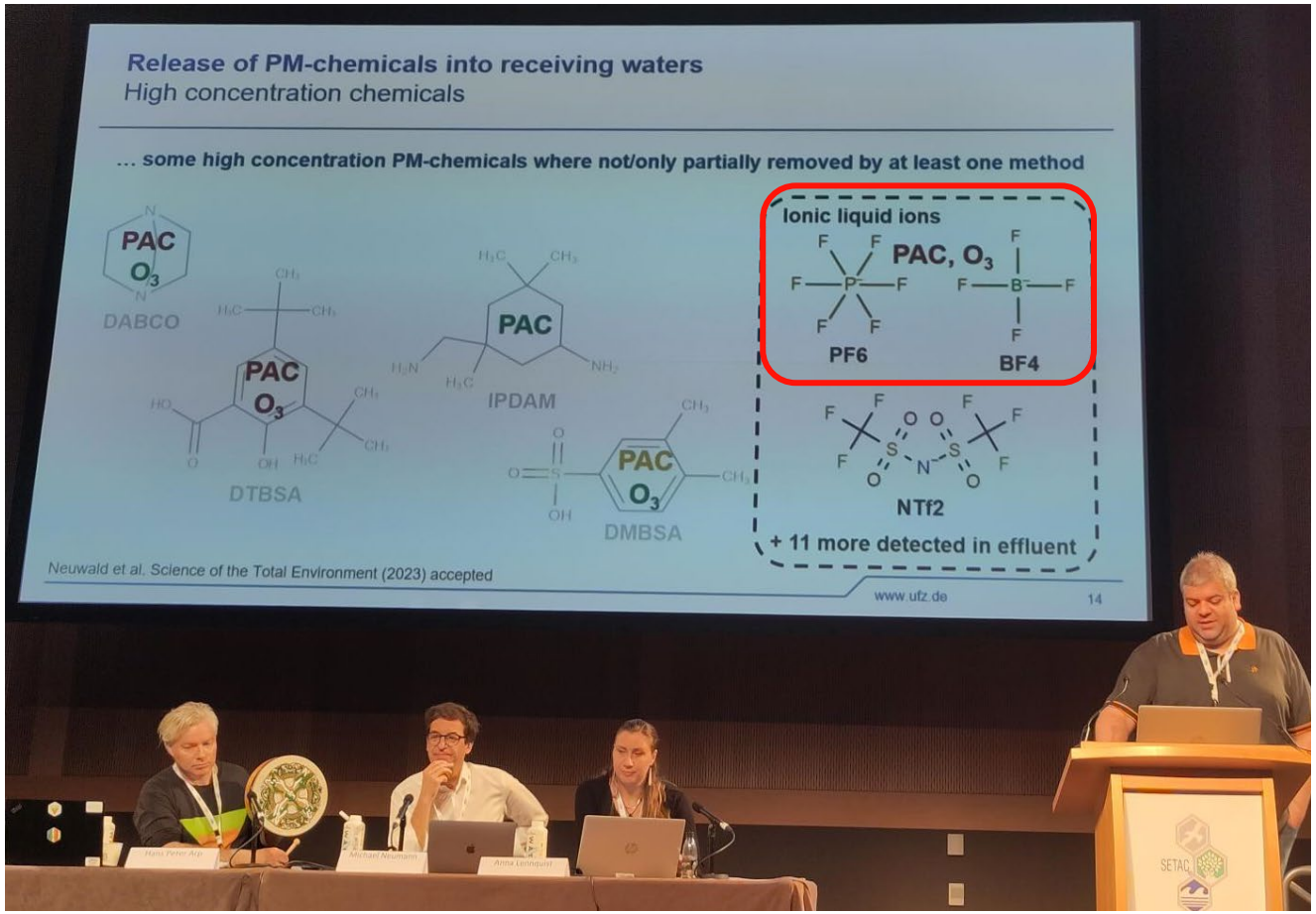
Chemical Stripes for Gallopamil

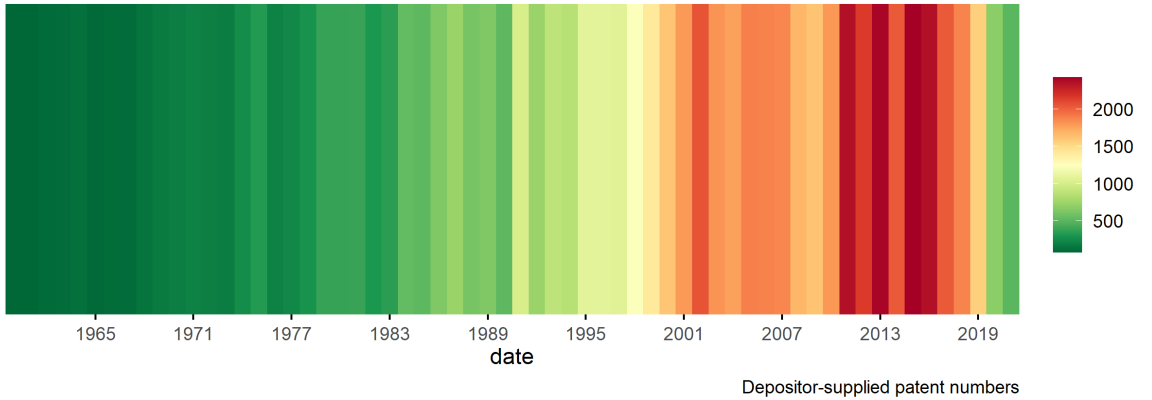
PubChem CID: 1234
First reference: 1970

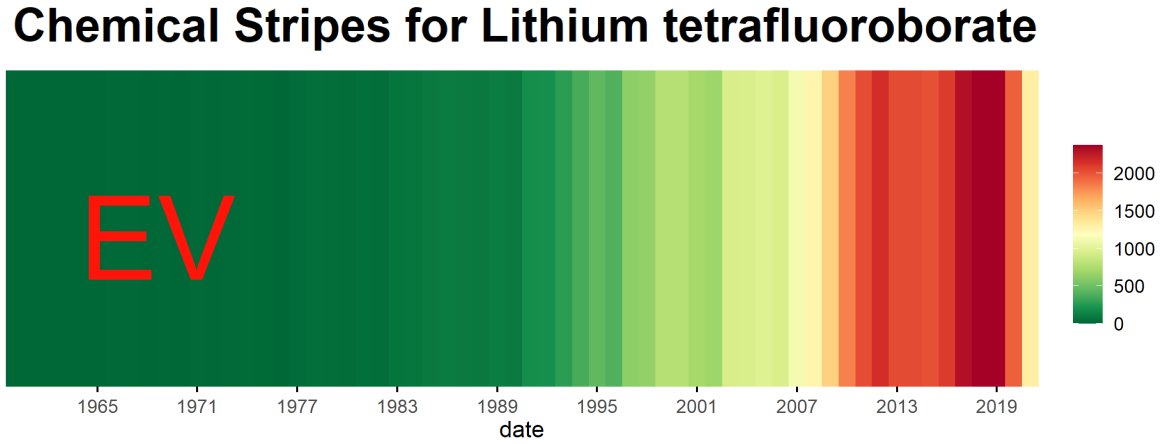


The Chemical Stripes and Patent Data

(live stripe calculations during SETAC 2023)



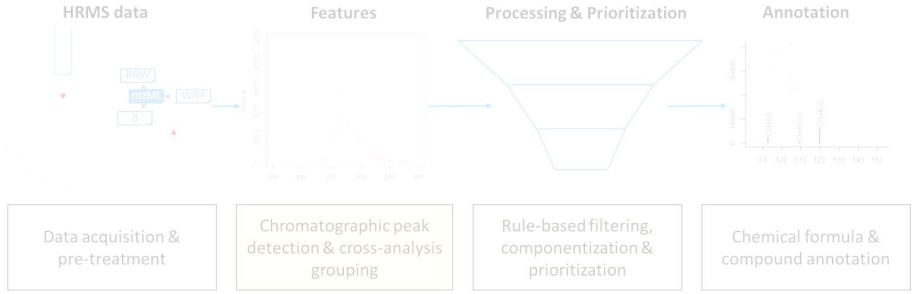




Left: Neuwald *et al, STOTEN*, DOI: 10.1016/j.scitotenv.2023.163921
Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

Outline

- Setting the Scene
  - NT-HRMS & Databases
  - The Great Data Challenge …
- Defining Chemical Space in Exposomics
- How to Tackle Transformations?
- Metabolomics / Exposomics-specific Workflows



PubChemLite EXPOSOMICS

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

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

204 million

PubChem

116 million

CompTox Chemicals Dashboard

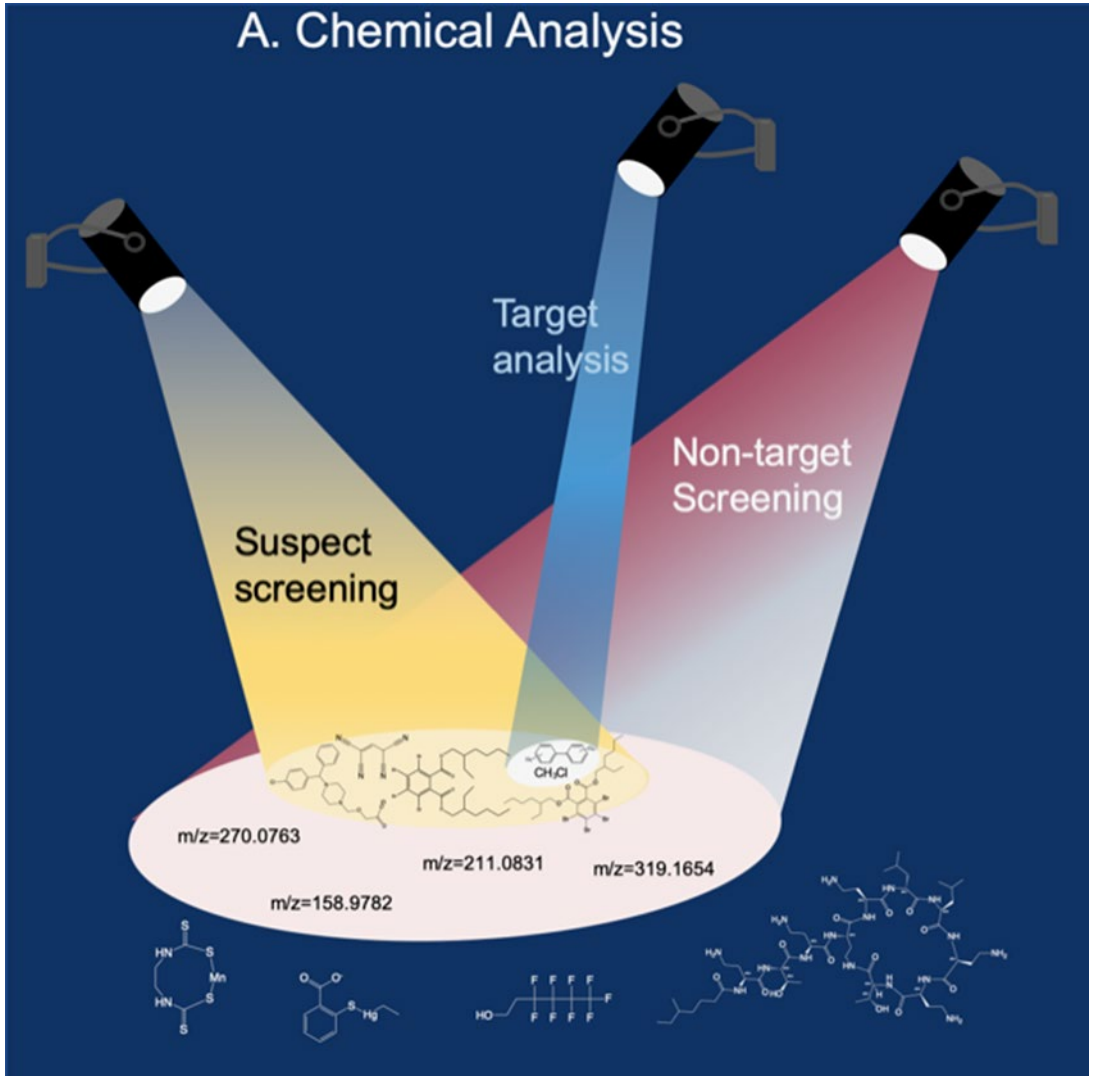
1.2 million

PubChemLite EXPOSOMICS

360,327

norman suspects

115,695



Numbers updated 9 Sept. 2023

Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

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 — Download

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

PubChem Acetylcarnitine (Compound)

12.1 Pathways

4 items — Download

Oxidation of Branched-Chain Fatty Acids
Source: PathBank External ID: SMP0000030
Taxonomy: Homo sapiens (human)

Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

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

PubChemLite EXPOSOMICS

~350,000 entries “small”

MetFrag

Updated PubChemLite to Aug 2023 version #127

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

Database Settings

Database: PubChemLite_exposomics

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16ClN5

Identifiers:

Retrieve Candidates 4 Candidates

zenodo Search Upload Communities

August 26, 2023 Dataset Open Access

PubChemLite for Exposomics

Bolton, Evan; Schymanski, 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, PharmacoInfo, 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.

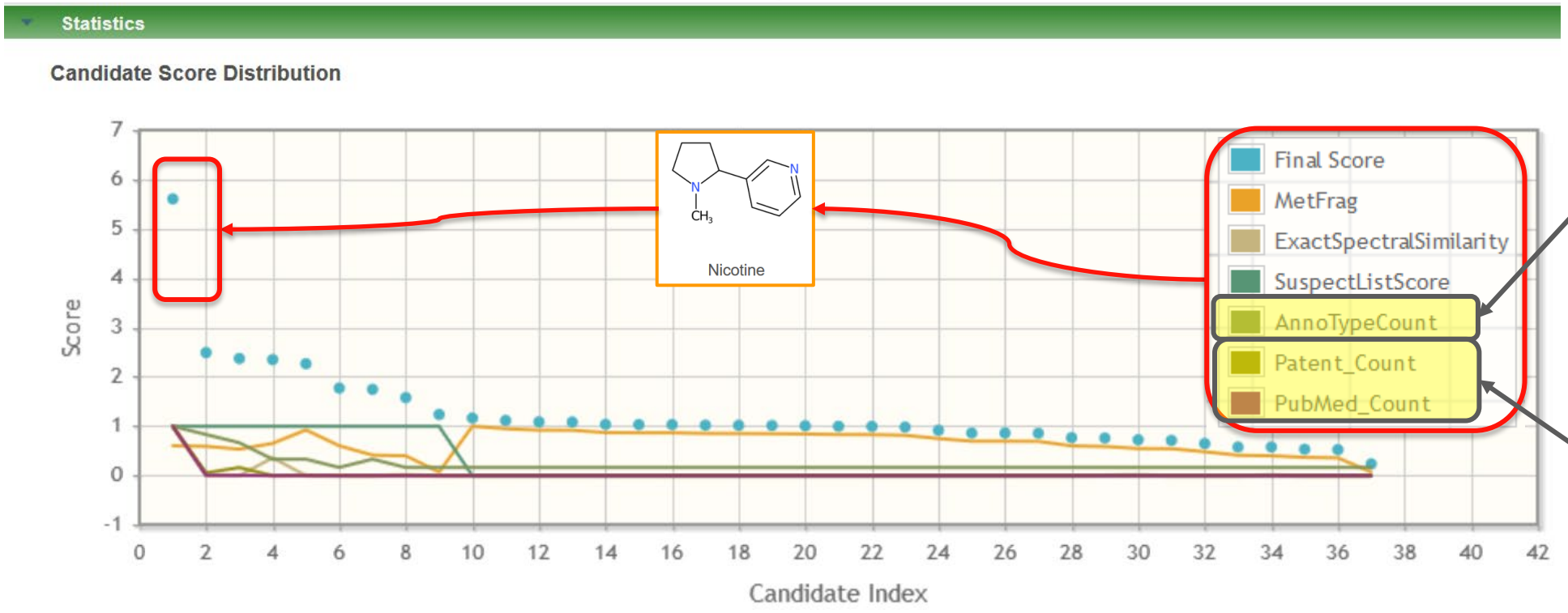
Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

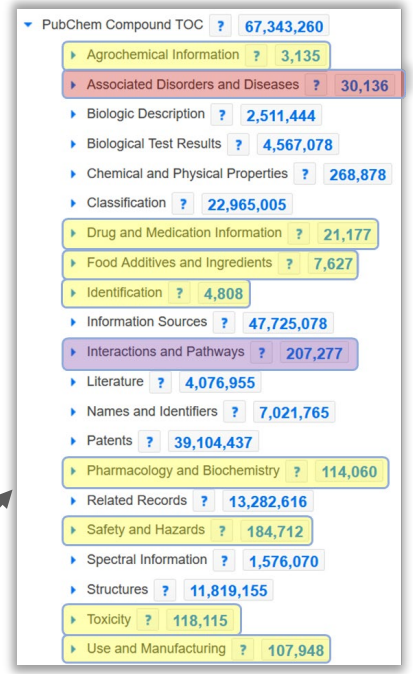
NIH U.S. National Library of Medicine National Center for Biotechnology Information

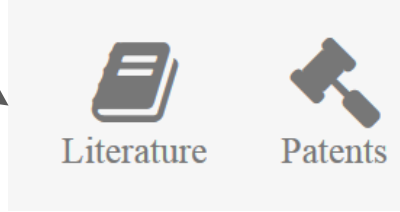
LCSB

PubChemLite: Fewer and more relevant candidates – with context

MetFrag · MassBank High Quality Mass Spectral Database · norman suspects · PubChemLite EXPOSOMICS



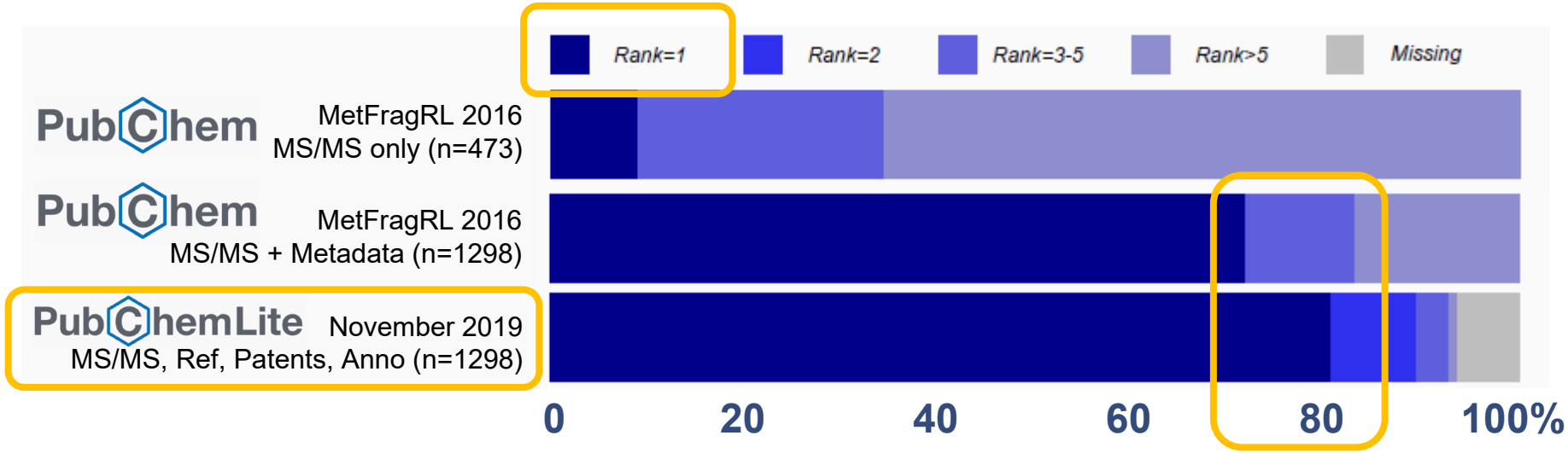




Ruttkies *et al* (2016) DOI: 10.1186/s13321-016-0115-9 & Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

How does PubChemLite perform?

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

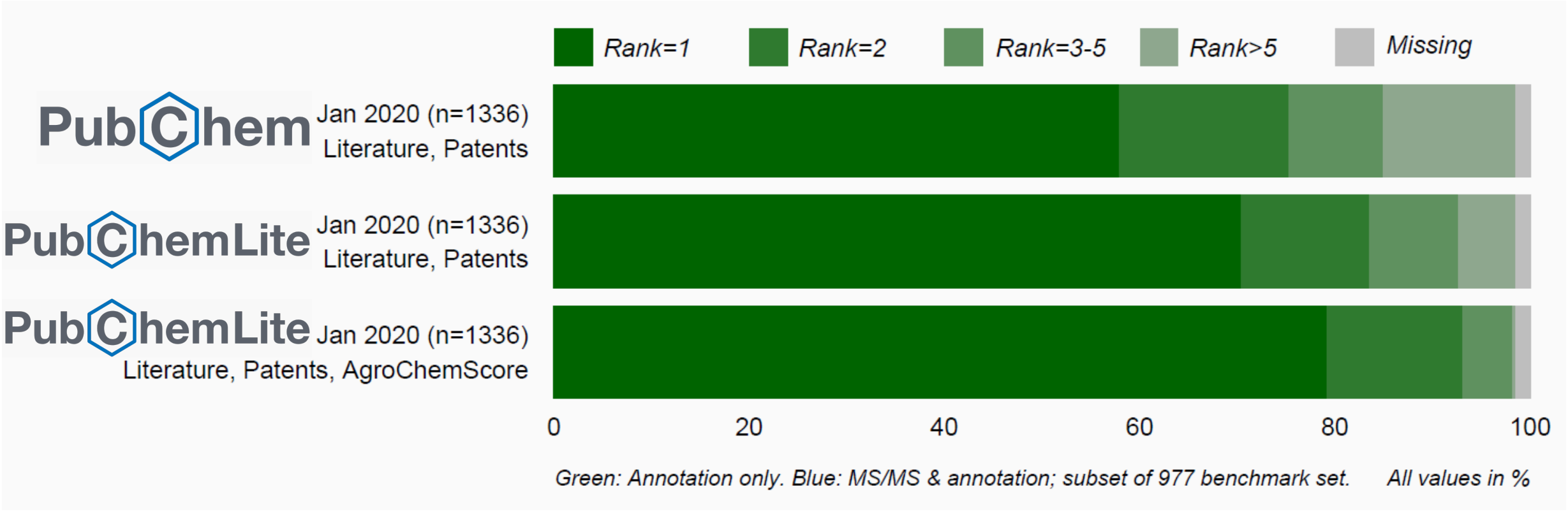


>80 % ranked in first place
~90 % ranked first or second!

} high throughput identification

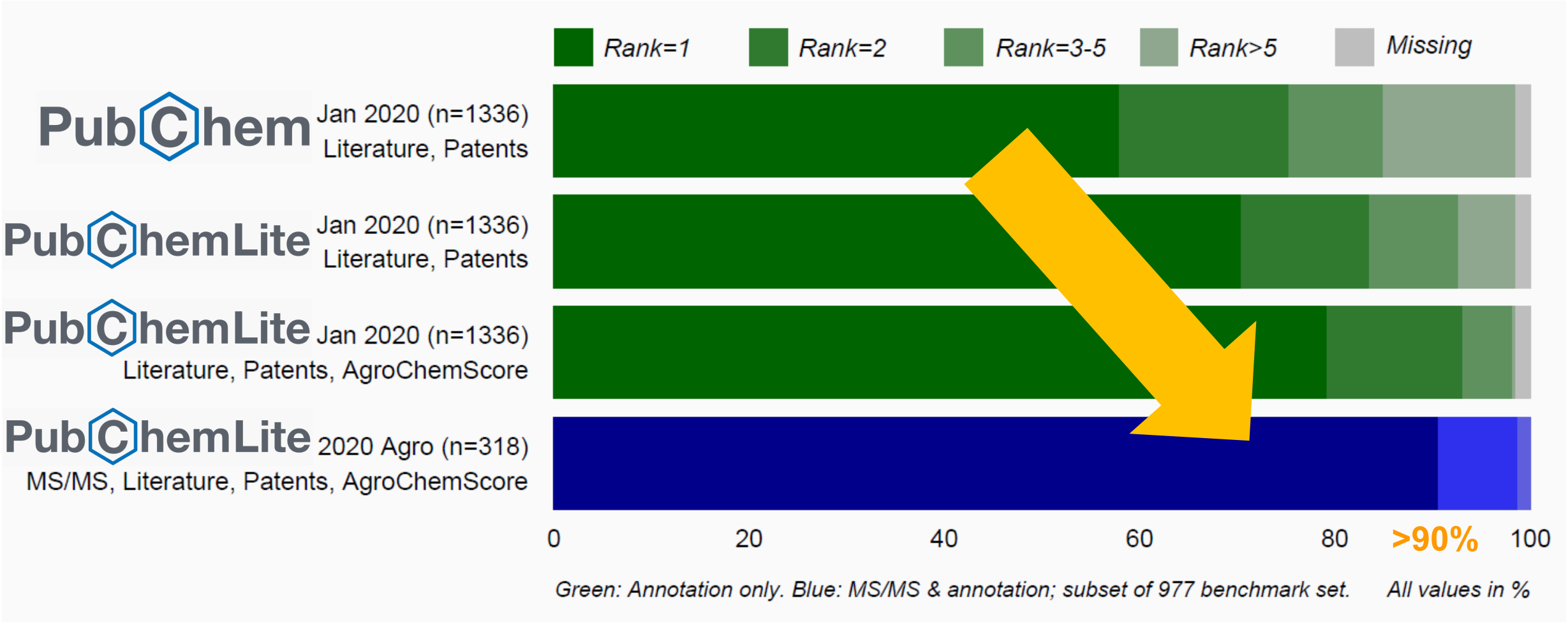
Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Influence of the Annotation Content in PubChemLite



Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Influence of the Annotation Content in PubChemLite

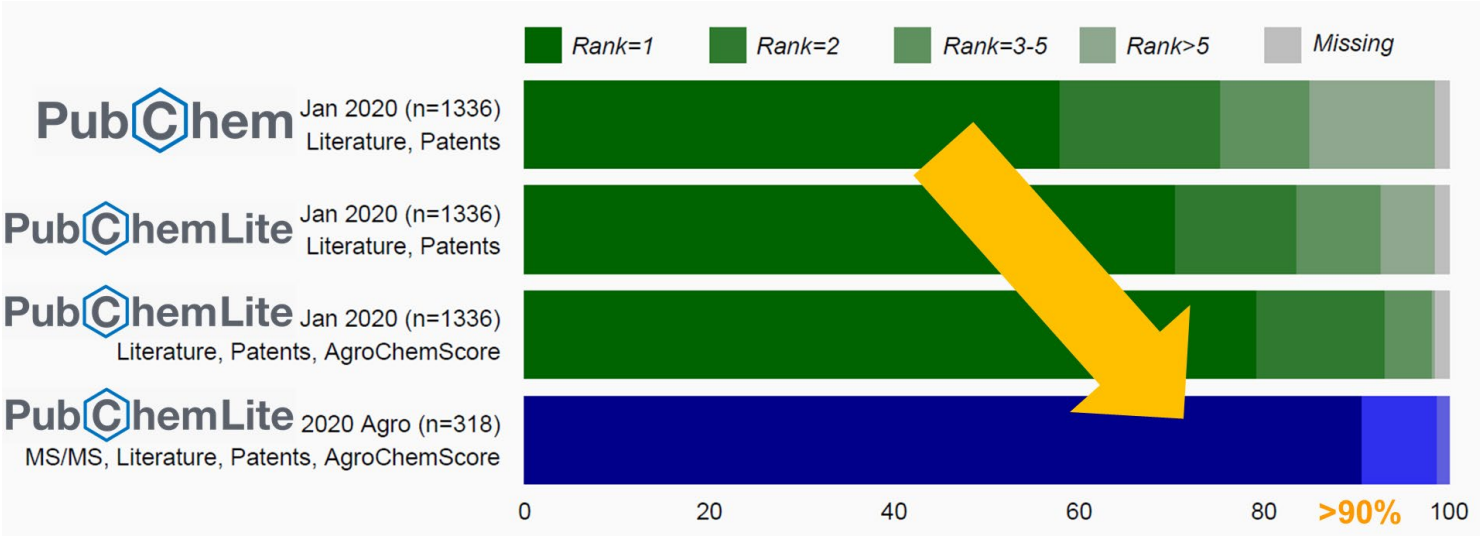


Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

PubChem

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

Influence of the Annotation Content in PubChemLite



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

>97% are Top 2 … 100 % Top 3 …

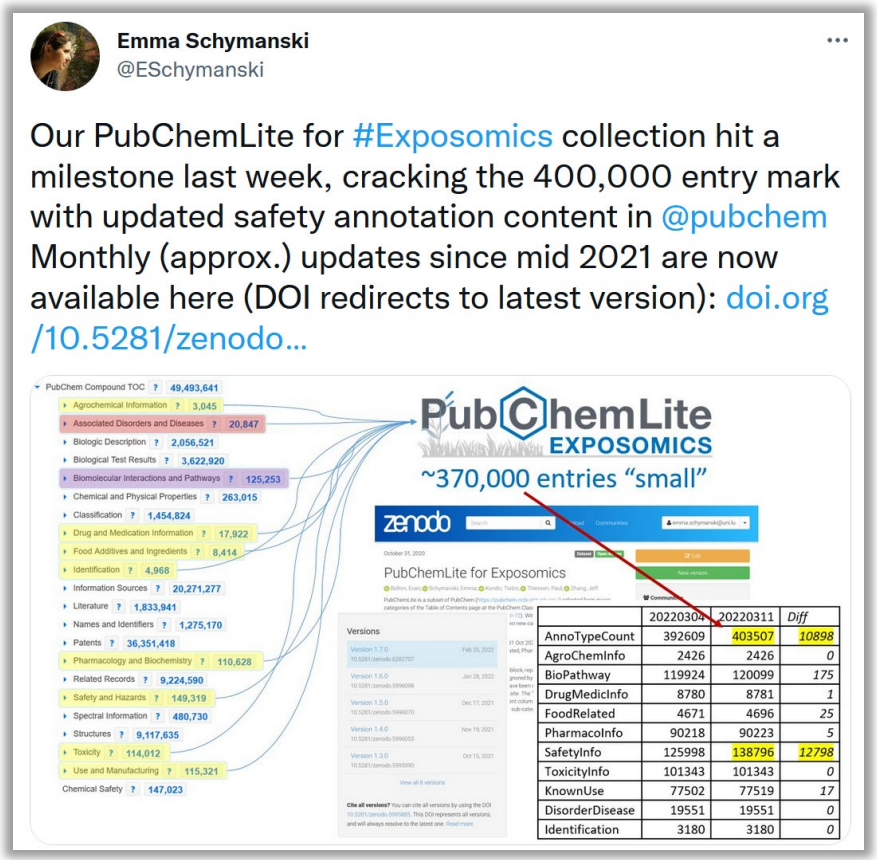
Exposomics: Top 1-3 matches will cover a lot!

(and quickly!)

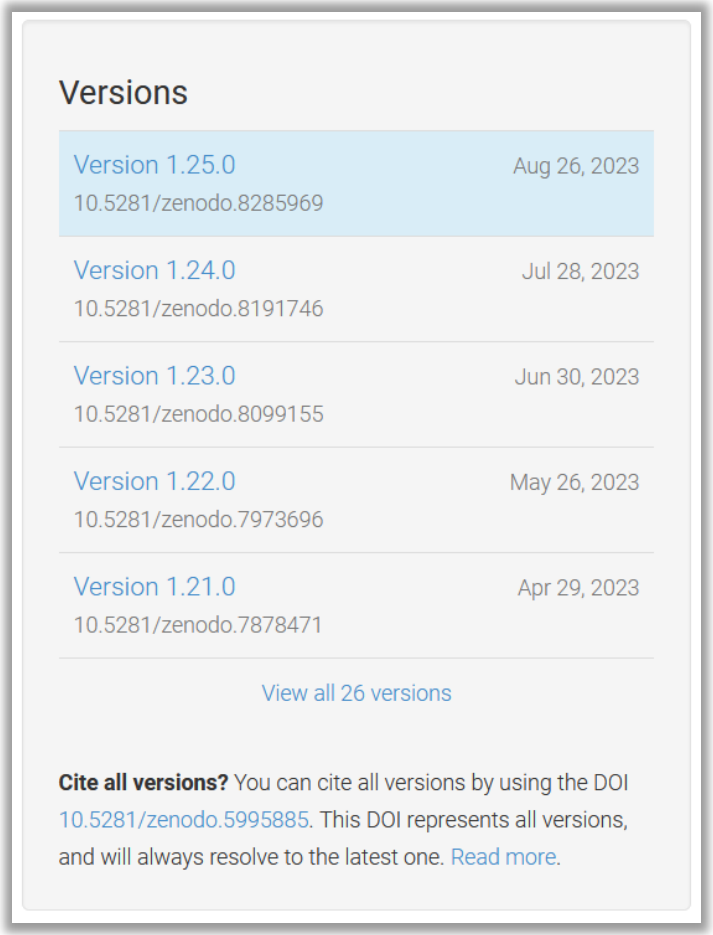
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

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!



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

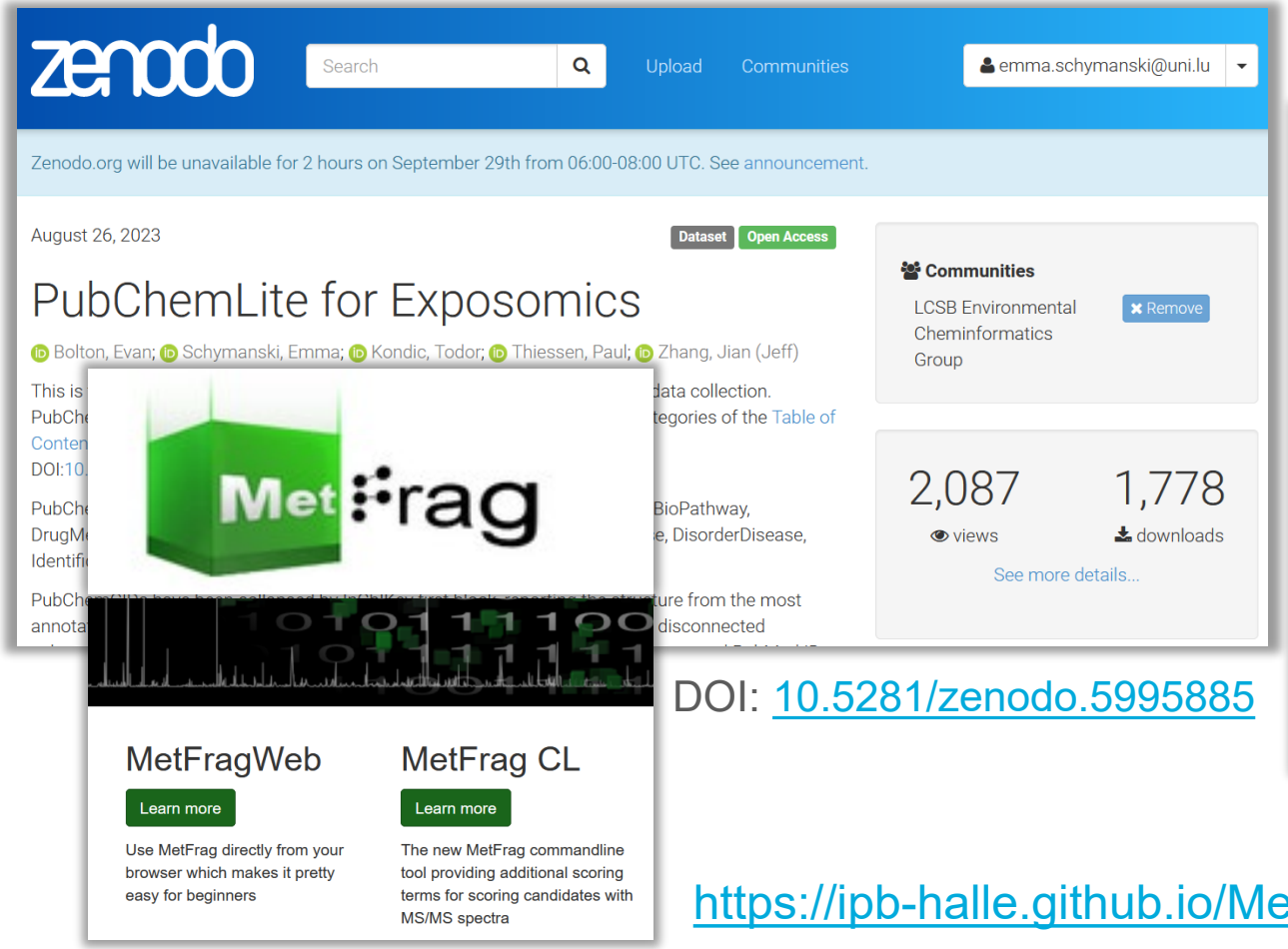


https://twitter.com/ESchymanski/status/1503260666613932033

DOI: 10.5281/zenodo.5995885

Putting PubChemLite into Practice?

- Downloadable & fully integrated in patRoon & MetFrag



DOI: 10.5281/zenodo.5995885

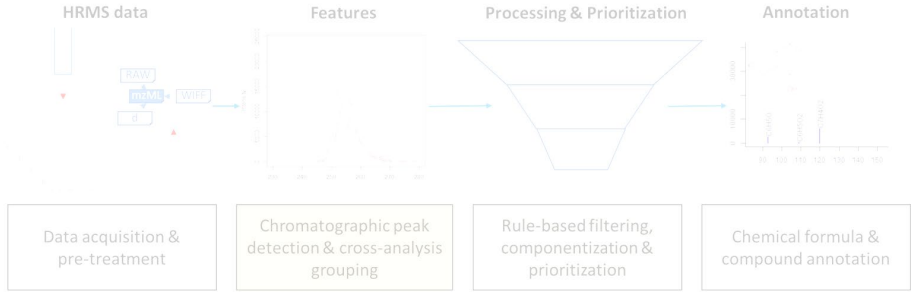
https://rickhelmus.github.io/patRoon/

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

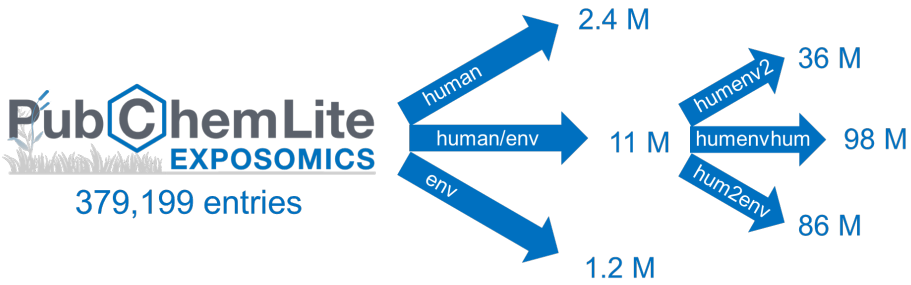
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0 & DOI: 10.5281/zenodo.5995885
Helmus *et al*. (2021) patRoon *J Cheminform* **13,** 1. DOI: 10.1186/s13321-020-00477-w

Outline

- Setting the Scene
  - NT-HRMS & Databases
  - The Great Data Challenge …
- Defining Chemical Space in Exposomics
- How to Tackle Transformations?
- Metabolomics / Exposomics-specific Workflows



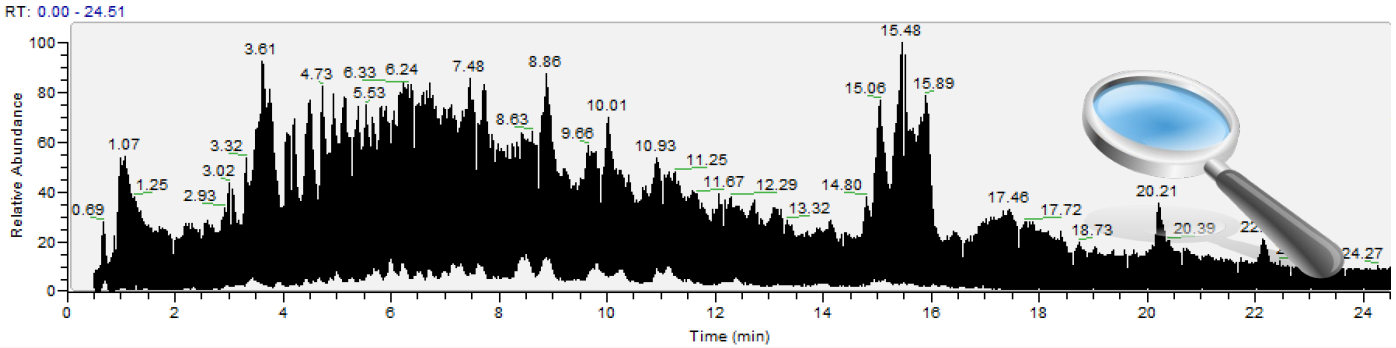
PubChemLite EXPOSOMICS

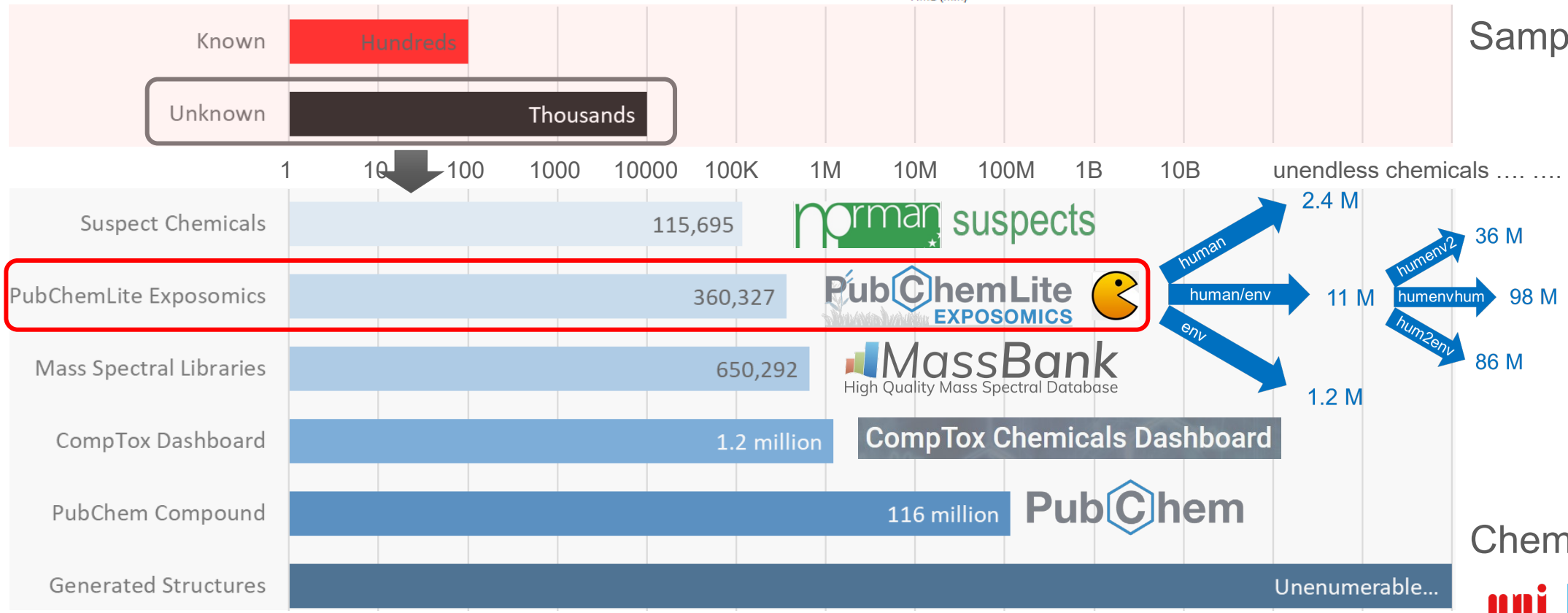


Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution mass spectrometry

AND connecting chemical knowledge





Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 6 Sept. 2023.

Transforming PubChemLite with BioTransformer 3.0

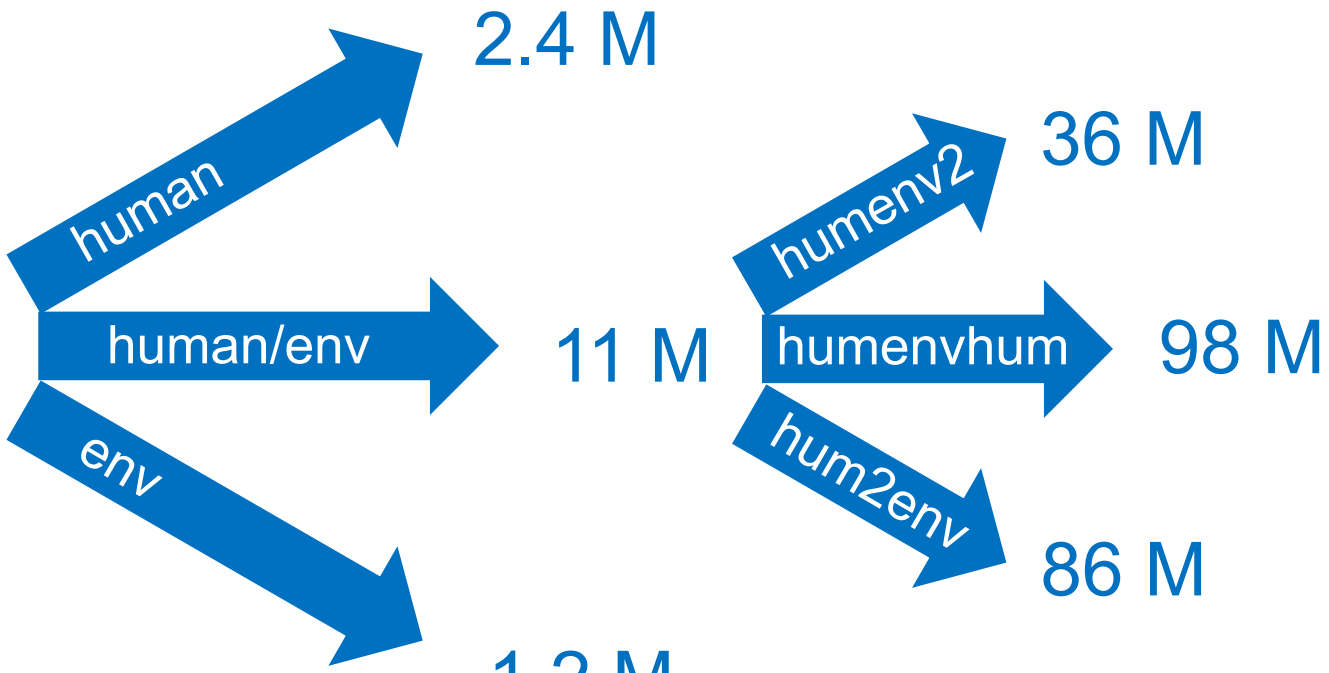
http://biotransformer.ca/

Wishart Lab
UNIVERSITY OF ALBERTA

PubChemLite
EXPOSOMICS

379,199 entries

version 1.0.0; DOI: 10.5281/zenodo.5995886



Combinatorial explosion – more data needed to predict reaction likelihoods

Djoumbou Feunang *et al* (2019). BioTransformer, JCheminf. DOI: 10.1186/s13321-018-0324-5.
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Bolton *et al* (2021) *preliminary calculations;*
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

FAIR Transformations in PubChem and NORMAN-SLE

norman suspects

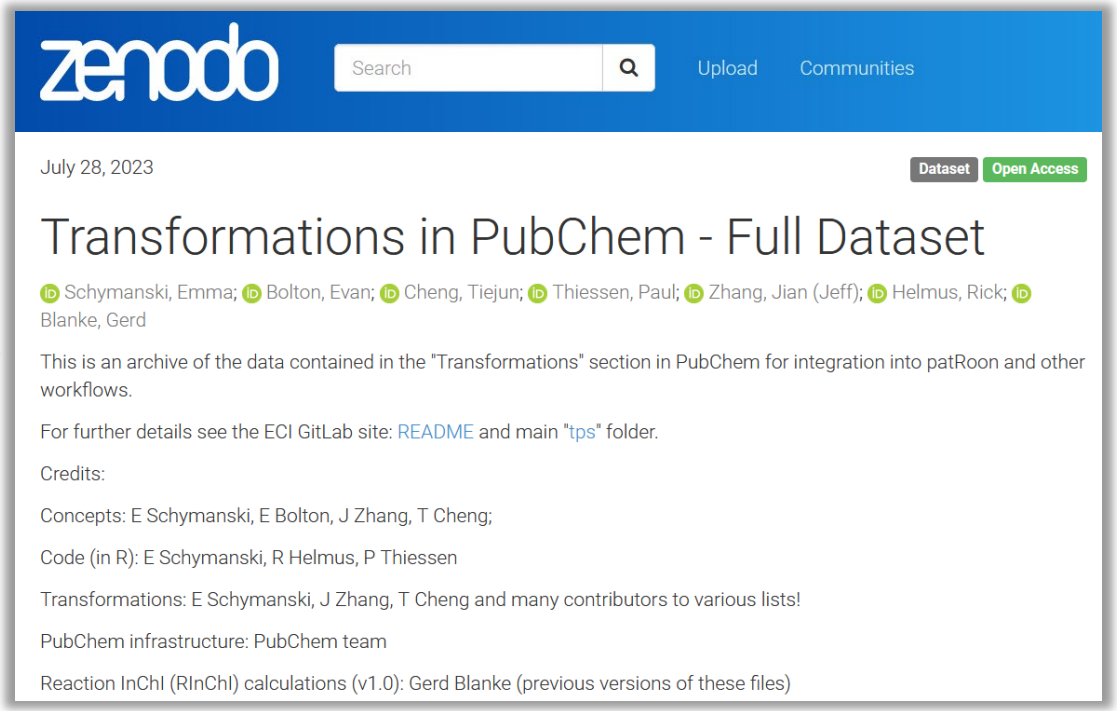
- EAWAGTPS
- HSDB
- MetXBioDB
- REFTPS
- SLUPESTTPS
- SWISSPEST19
- THSTPS
- UACCSCEC

ChEMBL

PubChem Transformations

6,643 chemicals

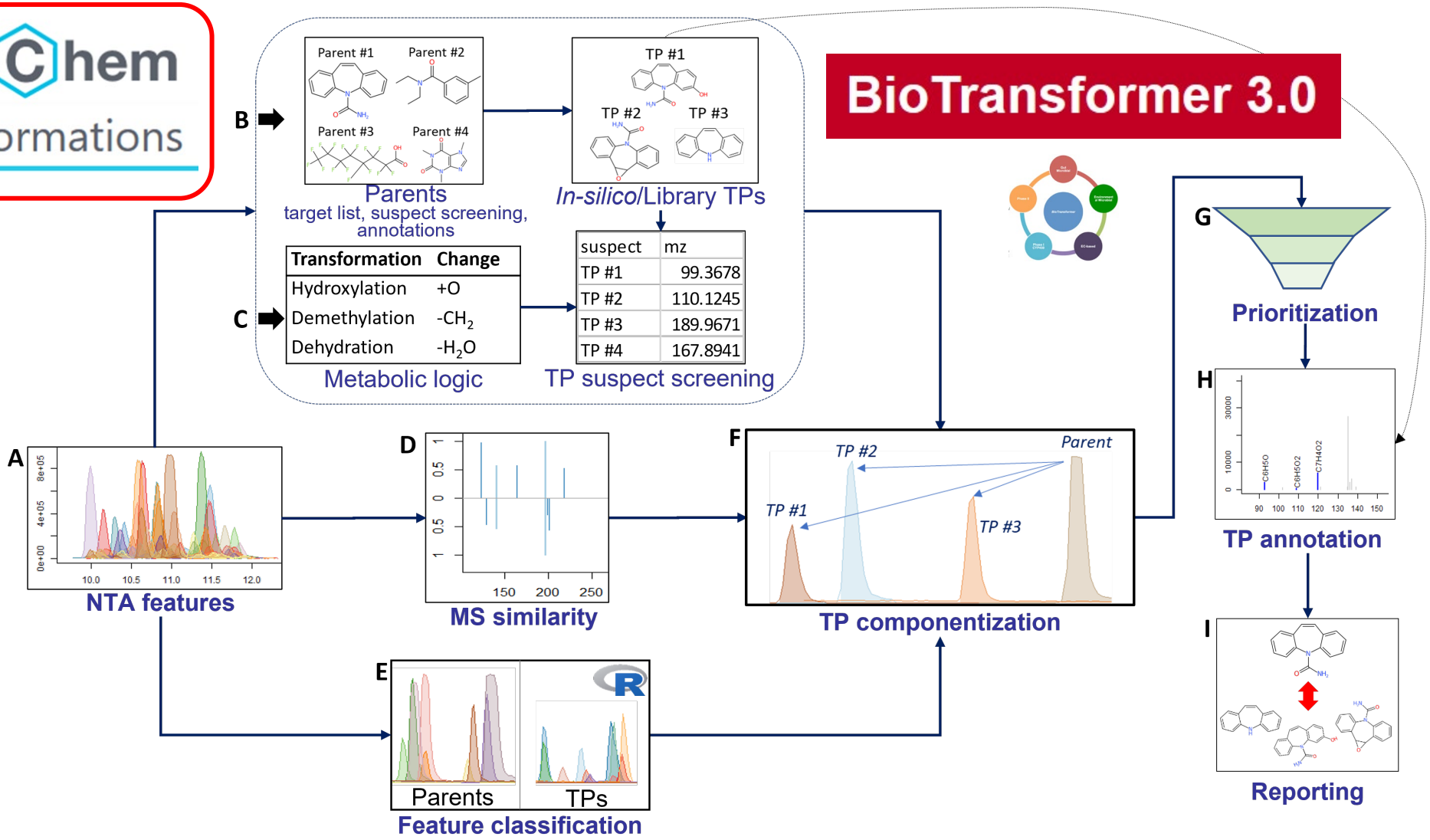
>double the BioTransformer library



Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

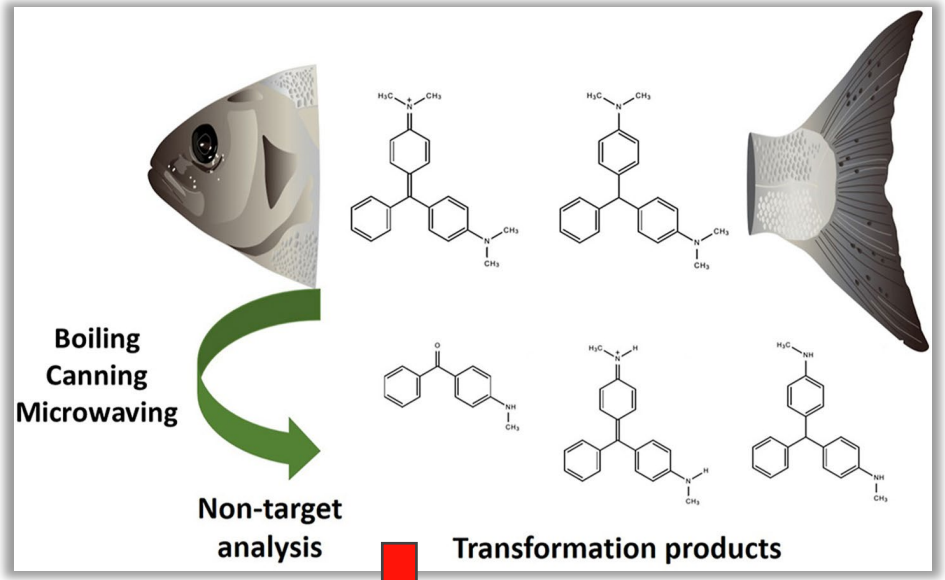
Open Transformation Products Workflows in patRoon 2.0

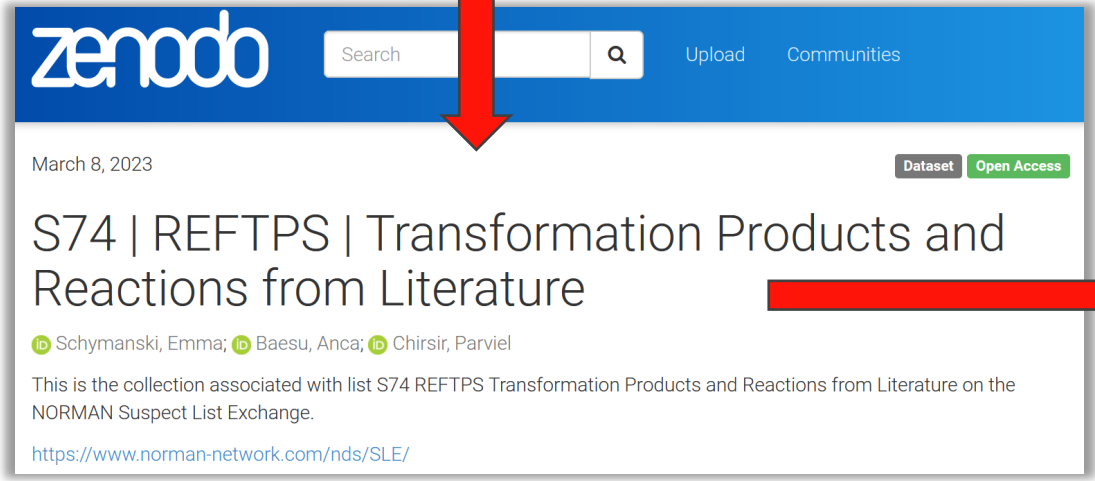
PubChem Transformations

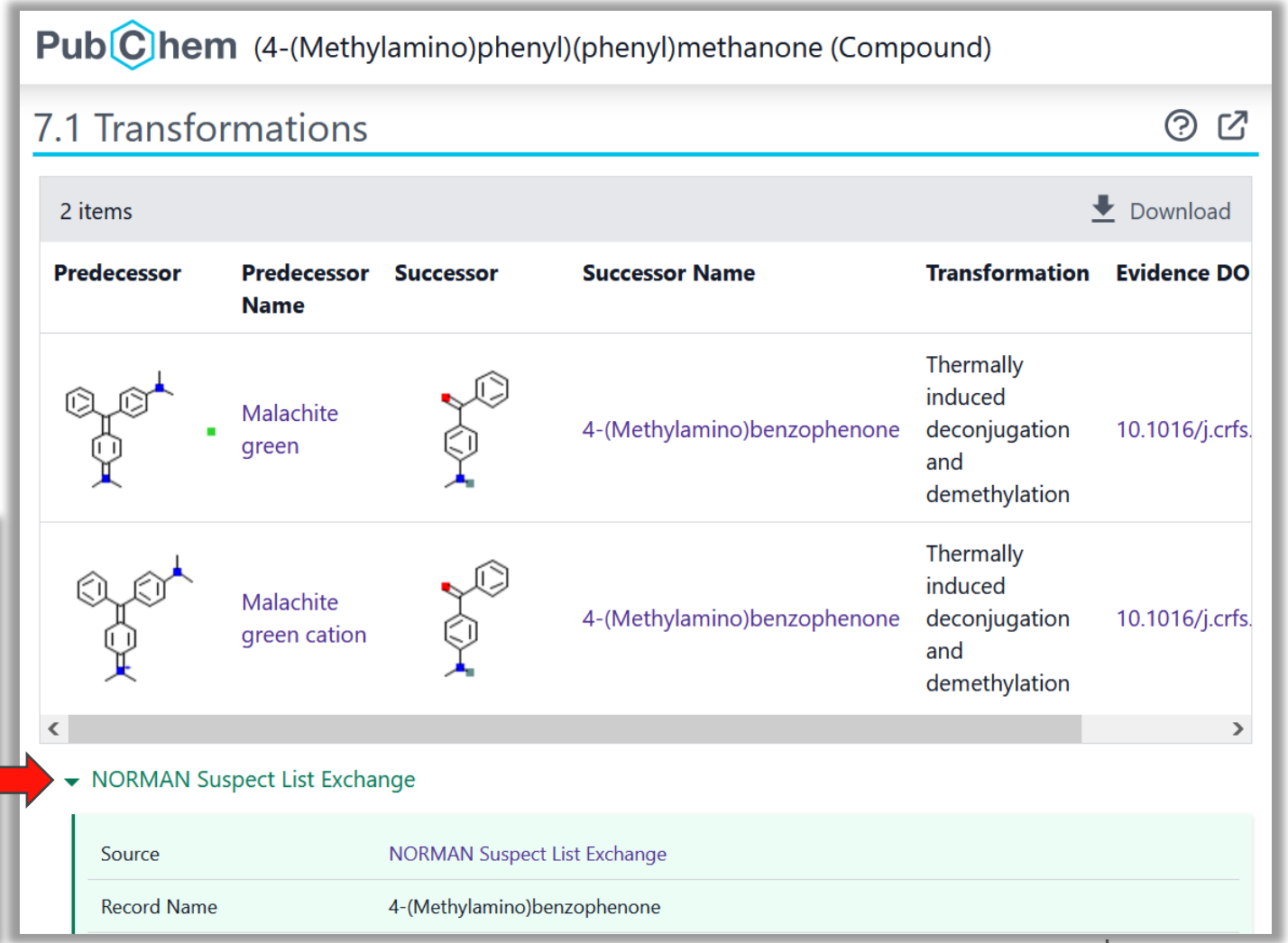


Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

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

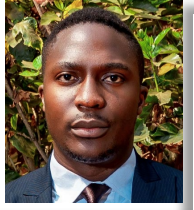






Baesu *et al*. (2021) DOI: 10.1016/j.crfs.2021.09.010. S74 | REFTPS (2021) DOI: 10.5281/zenodo.4318838

FAIRifying PFAS TPs



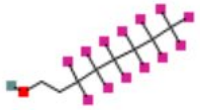
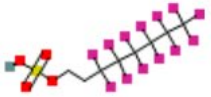
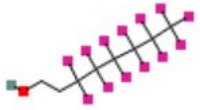
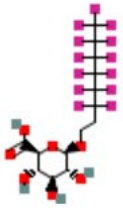
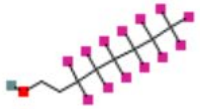
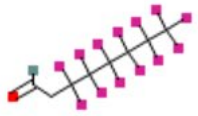
3,3,4,4,5,5,6,6,7,7,8,8,8-Tridecafluorooctan-1-ol (Compound)

7.1 Transformations

29 items

Search

SORT BY Predecessor Name -

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence D
	6:2 FTOH		6:2 FTOH-Sulfate	Conjugation		10.1016/j.cbi.2005.06.007
	6:2 FTOH		6:2 FTOH-Gluc	Conjugation		10.1016/j.cbi.2005.06.007
	6:2 FTOH		6:2 FTAL	Oxidation	24296	10.1016/j.cbi.2005.06.007

https://pubchem.ncbi.nlm.nih.gov/compound/69537#section=Pharmacology-and-Biochemistry

ELSEVIER

Chemico-Biological Interactions

Volume 155, Issue 3, 15 August 2005, Pages 165-180

Metabolic products and pathways of fluorotelomer alcohols in isolated rat hepatocytes

https://doi.org/10.1016/j.cbi.2005.06.007

Jonathan W. Martin [a], Scott A. Mabury [b], Peter J. O'Brien [a]

Show more

+ Add to Me

https://doi.org/1

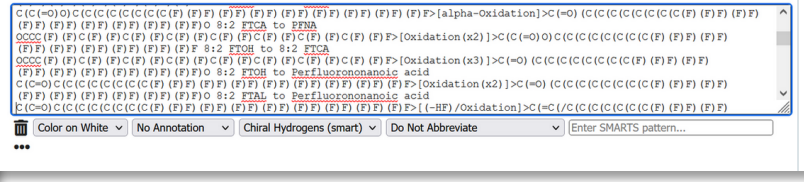
Jon Martin
@AcademicTox

Wow, this was a surprise. My postdoctoral work on metabolism of #PFAS has been FAIRified on @pubchem thanks to @ESchymanski and colleagues. Very exciting to think about how these huge efforts will improve future science for everyone.

@PChirsir

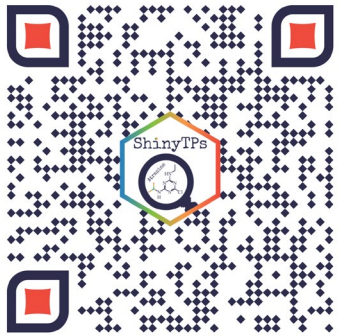
Emma Schymanski @ESchymanski · Jan 8

Spent a rainy afternoon FAIRifying #PFAS data from past @AcademicTox work, thanks to @PChirsir efforts we've committed the first set to @pubchem incl. new structures - hope to bring the 2005 PDF images to #OpenScience life soon! Again @the_cdk Depict makes it so much fun @jwmay !



https://twitter.com/AcademicTox/status/1479841982679765001

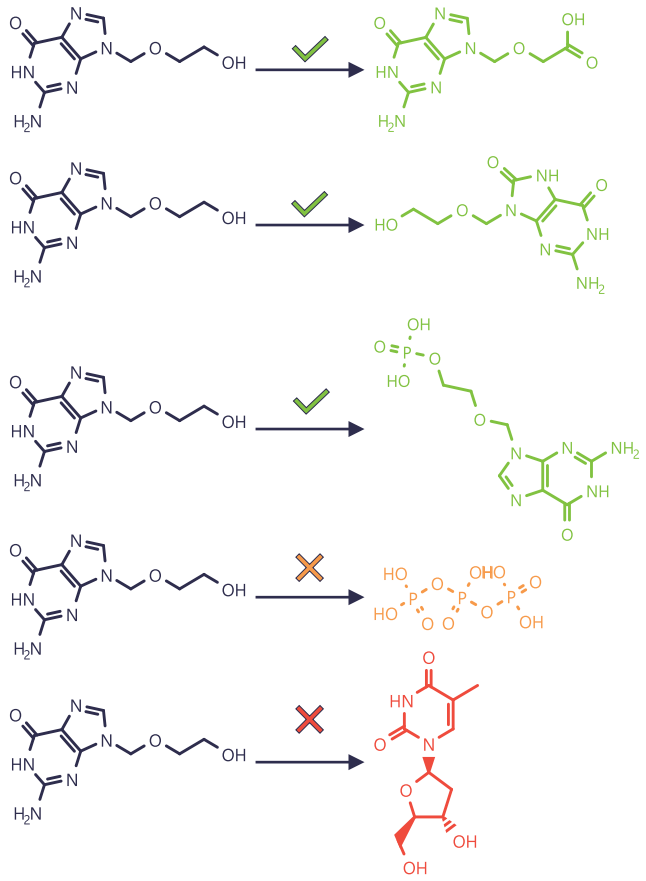
Curating Text-mined Content in PubChem



PubChem Acyclovir (Compound)

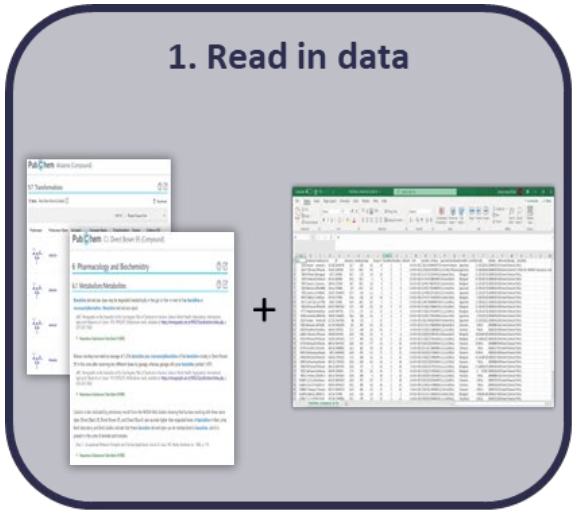
8.6 Metabolism/Metabolites

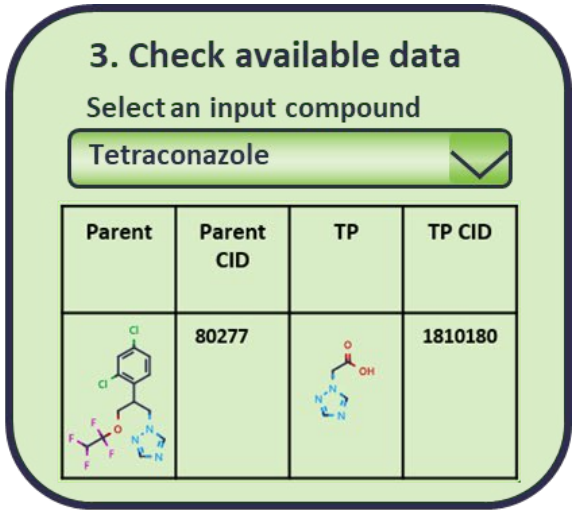
Acyclovir is metabolized partially to **9-carboxymethoxymethylguanine** and minimally to **8-hydroxy-9-(2-hydroxyethoxymethyl)guanine**. In vitro, **acyclovir** also is metabolized to **acyclovir monophosphate**, diphosphate, and **triphosphate** in cells infected with herpes viruses, principally by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

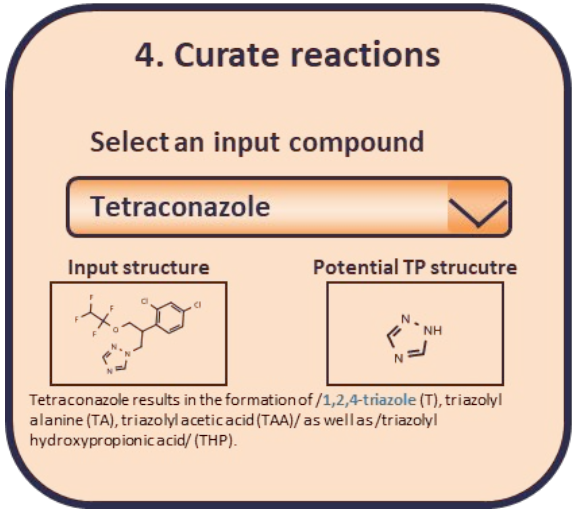


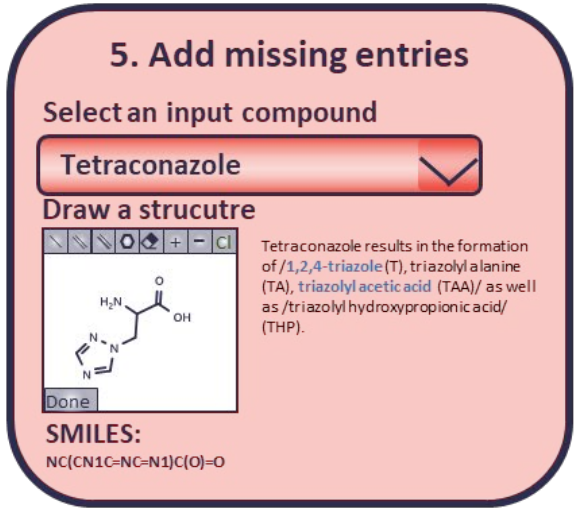
Palm *et al* (2023) ShinyTPs, DOI: 10.26434/chemrxiv-2023-xm41h (under review). Image by Emma Palm.

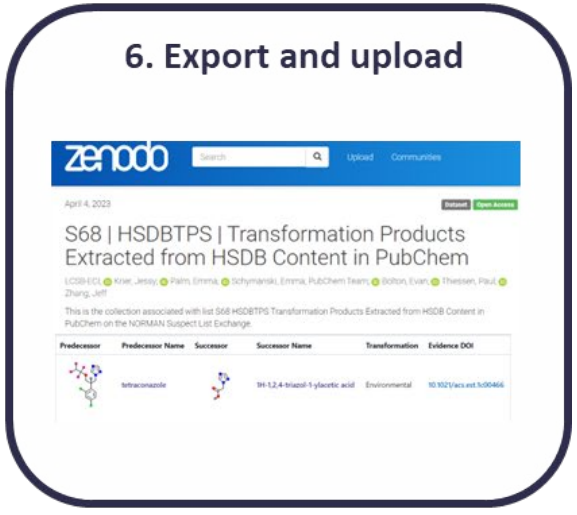
ShinyTPs: Curating TPs from Text Mining Results





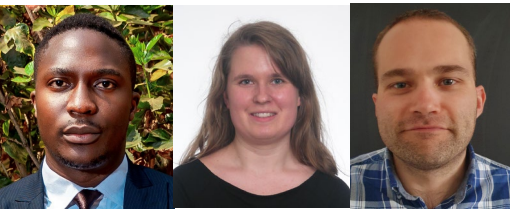






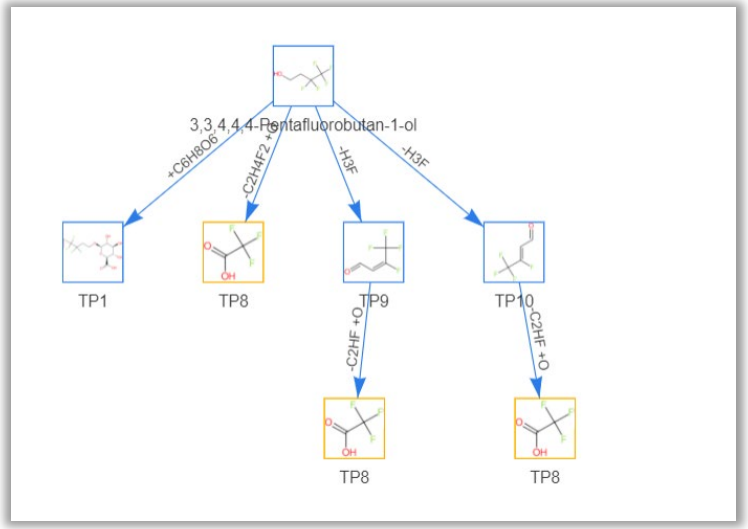
Palm *et al* (2023) ShinyTPs, DOI: 10.26434/chemrxiv-2023-xm41h (under review). Image created by Emma Palm.

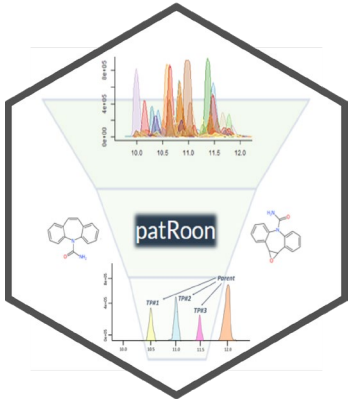
Applying TP NT-HRMS Workflows with patRoon

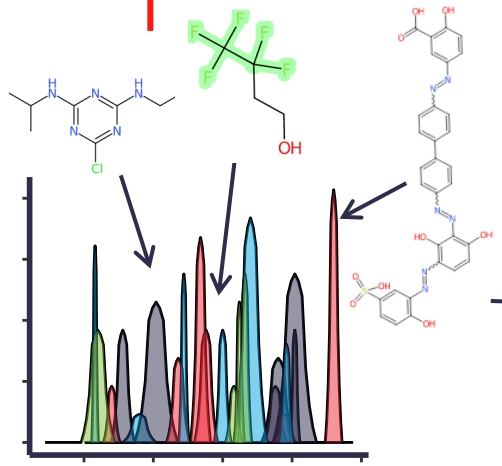


ShinyTPs

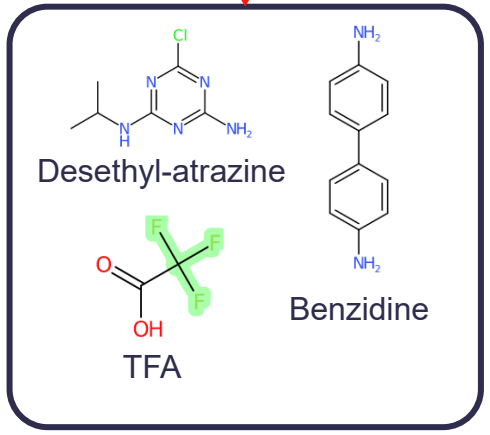
PubChem Transformations



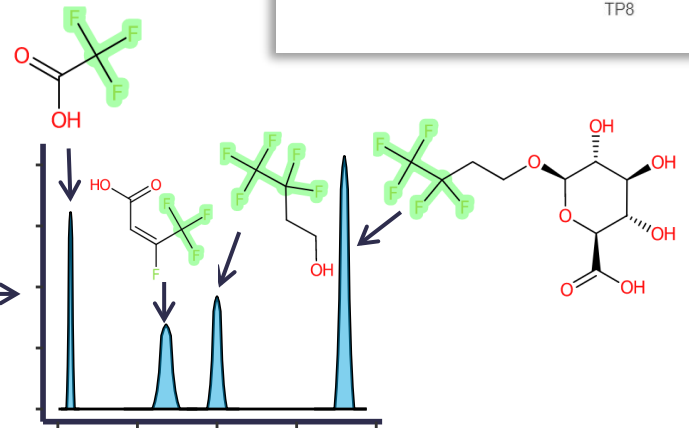




Suspect/Non-target screening



TP library

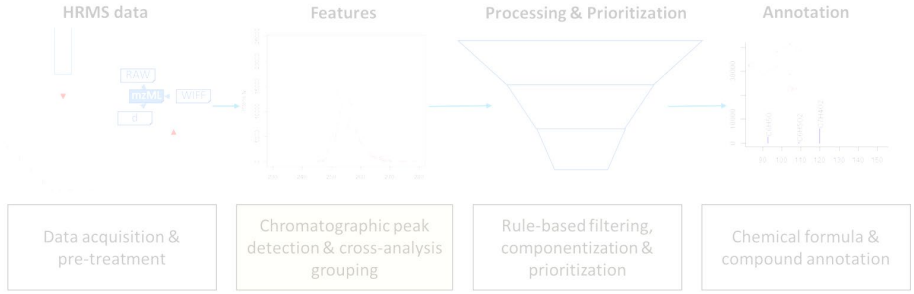


TP suspect screening

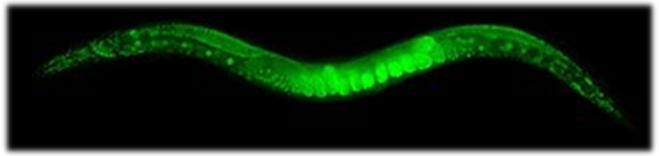
Mod. from Palm *et al* (2023) ShinyTPs, DOI: 10.26434/chemrxiv-2023-xm41h, Palm *et al* (2023). SETAC Presentation, DOI: 10.5281/zenodo.7885829

Outline

- Setting the Scene
  - NT-HRMS & Databases
  - The Great Data Challenge …
- Defining Chemical Space in Exposomics
- How to Tackle Transformations?
- Metabolomics / Exposomics-specific Workflows



PubChemLite EXPOSOMICS



Finally … Do we always need the entire chemical space?

PubChem

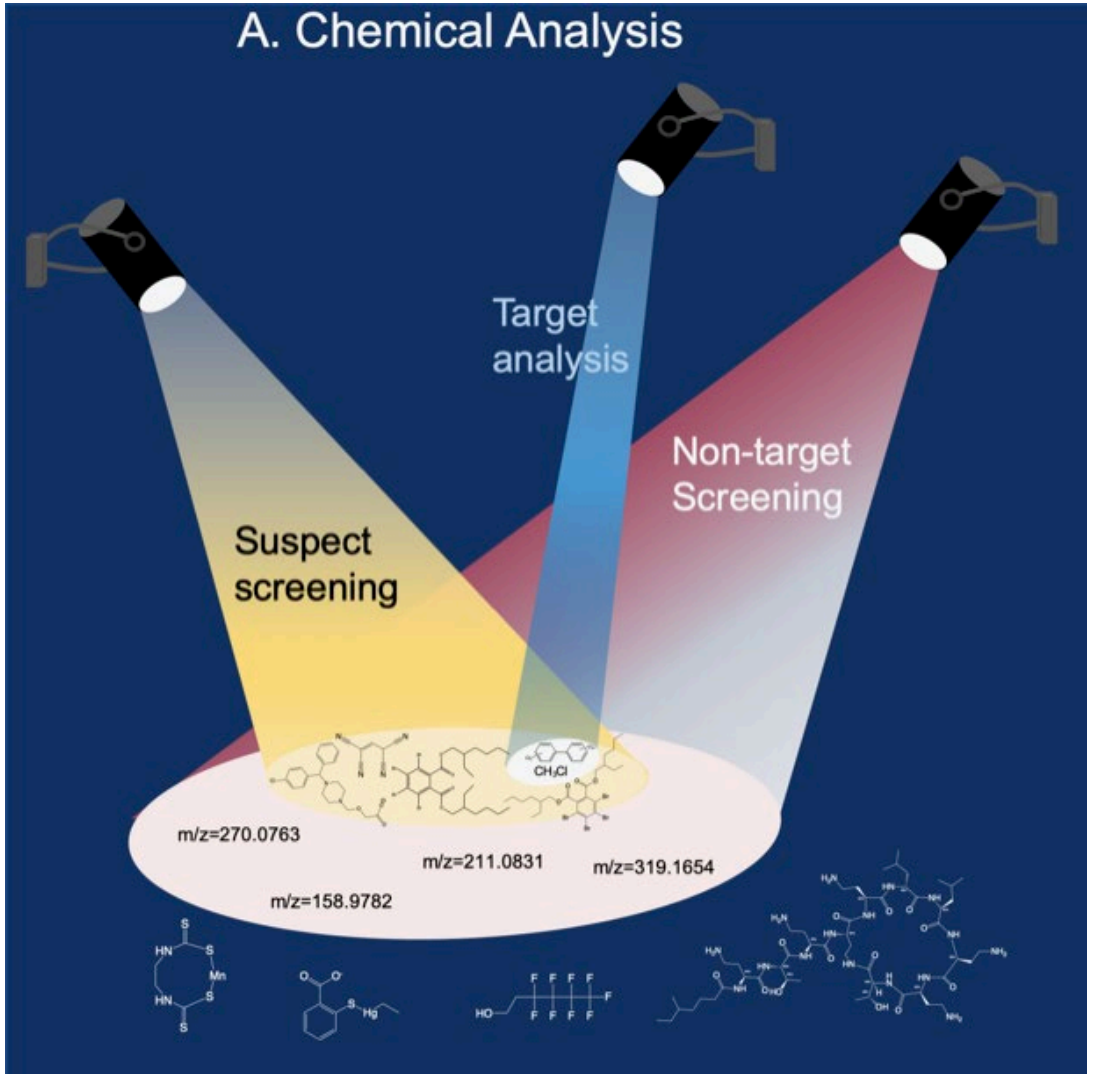
116 million

PubChemLite EXPOSOMICS

360,327

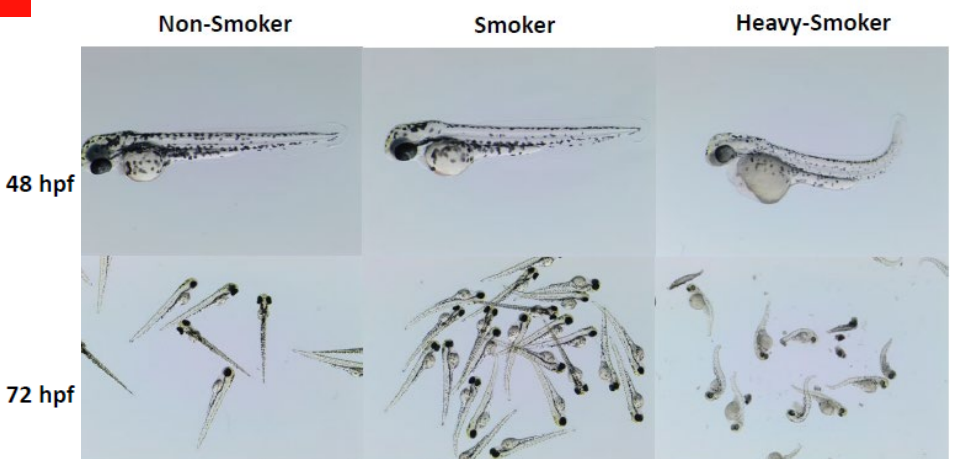
~1400 metabolites

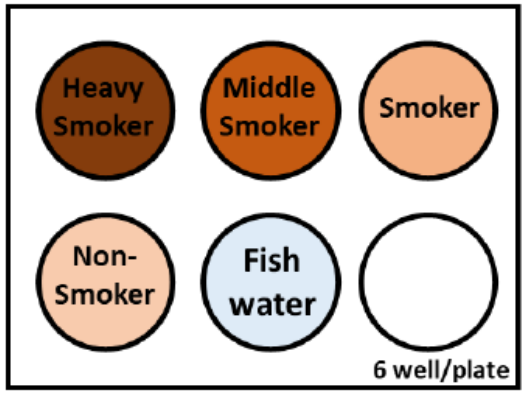
~3600 metabolites



Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

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

First Previous Page 1 of 247 Next

PubChem

3.3 Metabolites

423 items

Search SORT BY Compound CID - Increasing

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

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 (als headers for MetFrag import.

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

Menu

C_elegans_metabolites.Rmd

9.79 KB

Edit

```
title: "Finding _C. elegans_ Metabolites on PubChem"
author: "Emma SCHYMANSKI"
date: "19/01/2022"
output: pdf_document
```

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

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.

Topic	Documentation
Taxonomy	*C. elegans* as PDF or RMarkdown

PubChem ECI GitLab Docs

PubChem Caenorhabditis elegans (Taxonomy)

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

First Previous Page 1 of 233 Next Last

ECI Group, LCSB, University of Luxembourg

3.4 Natural Products

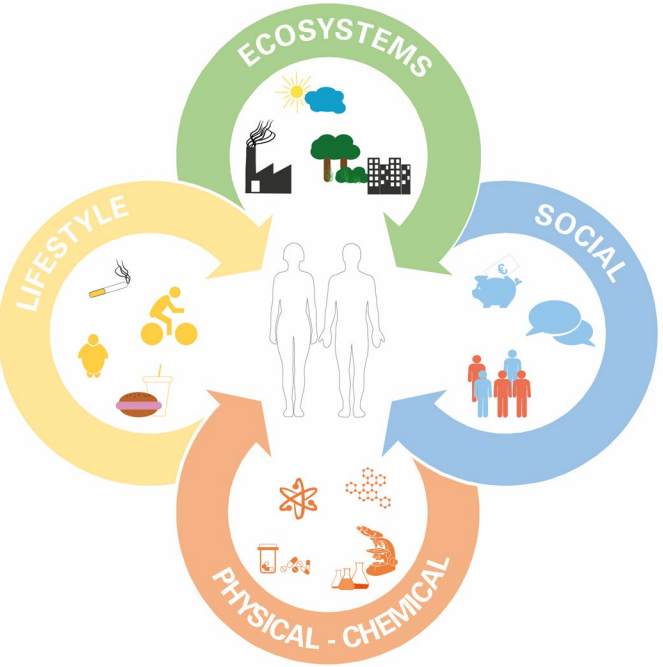
712 items Download

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

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:

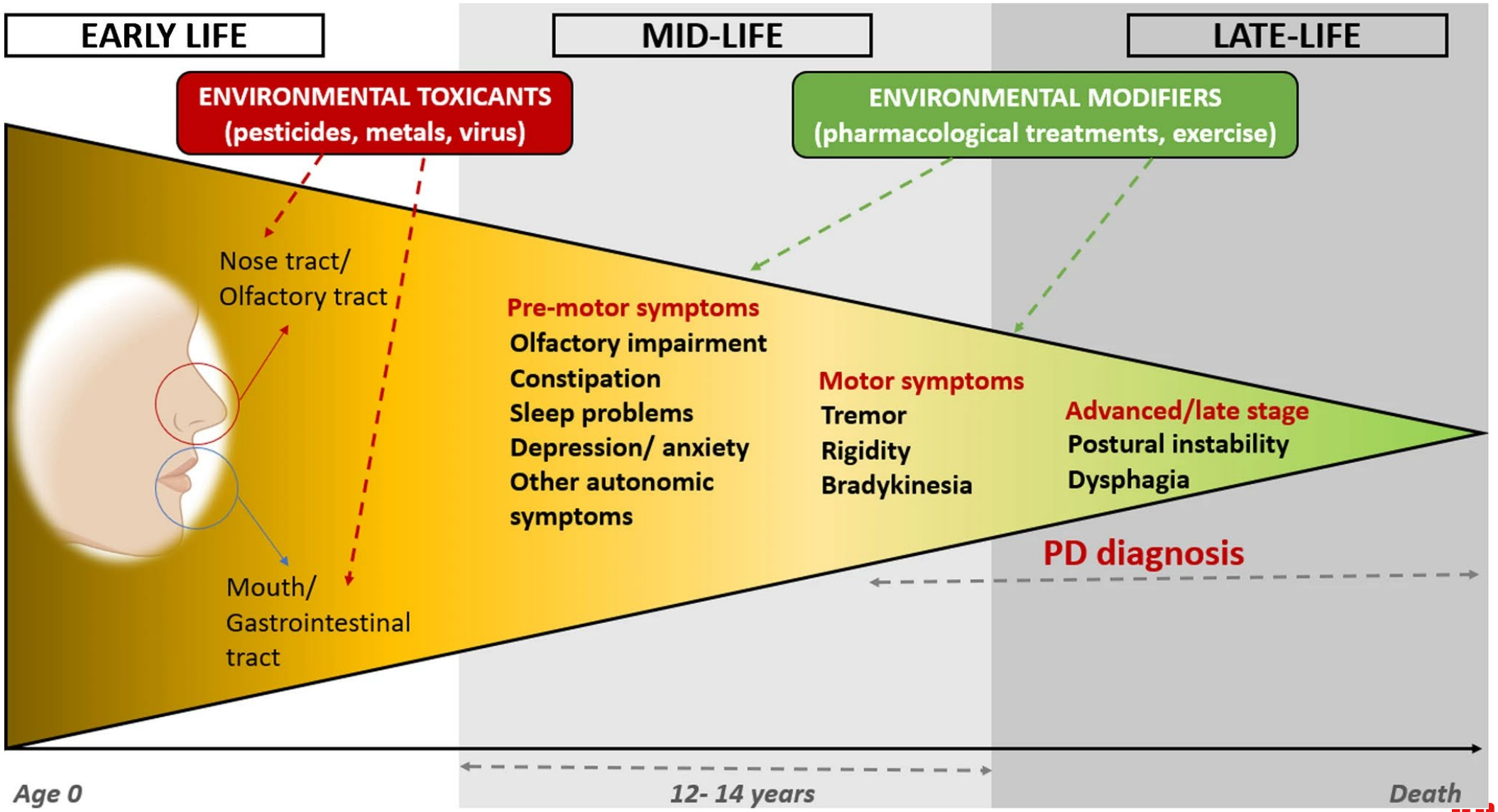
  "The cumulative measure of environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, diet, behaviour and endogenous processes"

- (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)

Image c/o Lucie Debroux LCSB, modified from images by Utrecht Uni & Vermeulen *et al.* (2020). Science. DOI: 10.1126/science.aay3164. Wild (2005) DOI: 10.1158/1055-9965.EPI-05-0456; Miller & Jones (2014) DOI: 10.1093/toxsci/kft251; Landrigan *et al* (2018) DOI: 10.1016/S0140-6736(17)32345-0

Exposomics and Neurodegenerative Disease: Timeline!

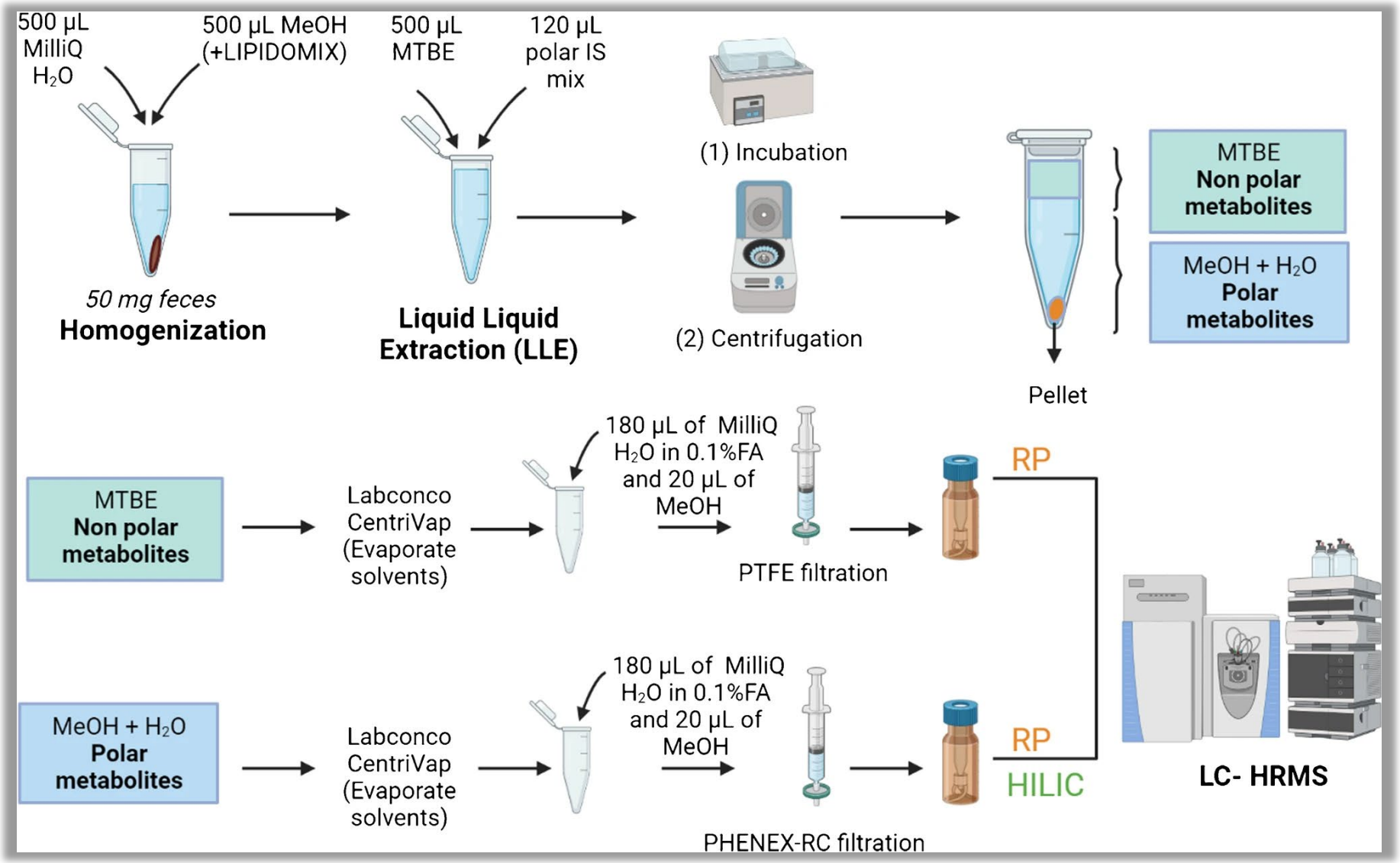
MICROH



Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

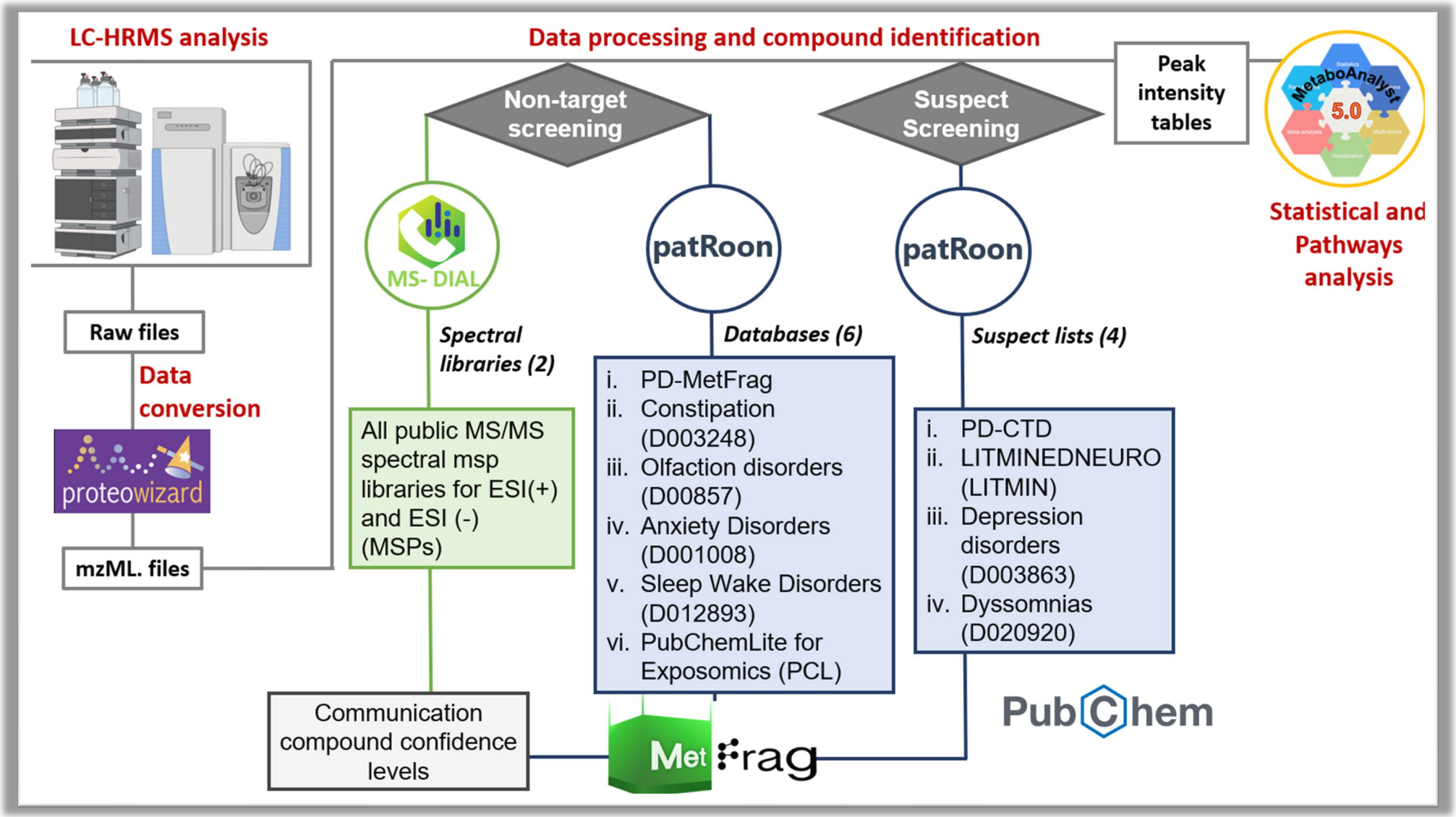
Exposomics: Sample Prep. for Broad Range of Analysis

MICROH



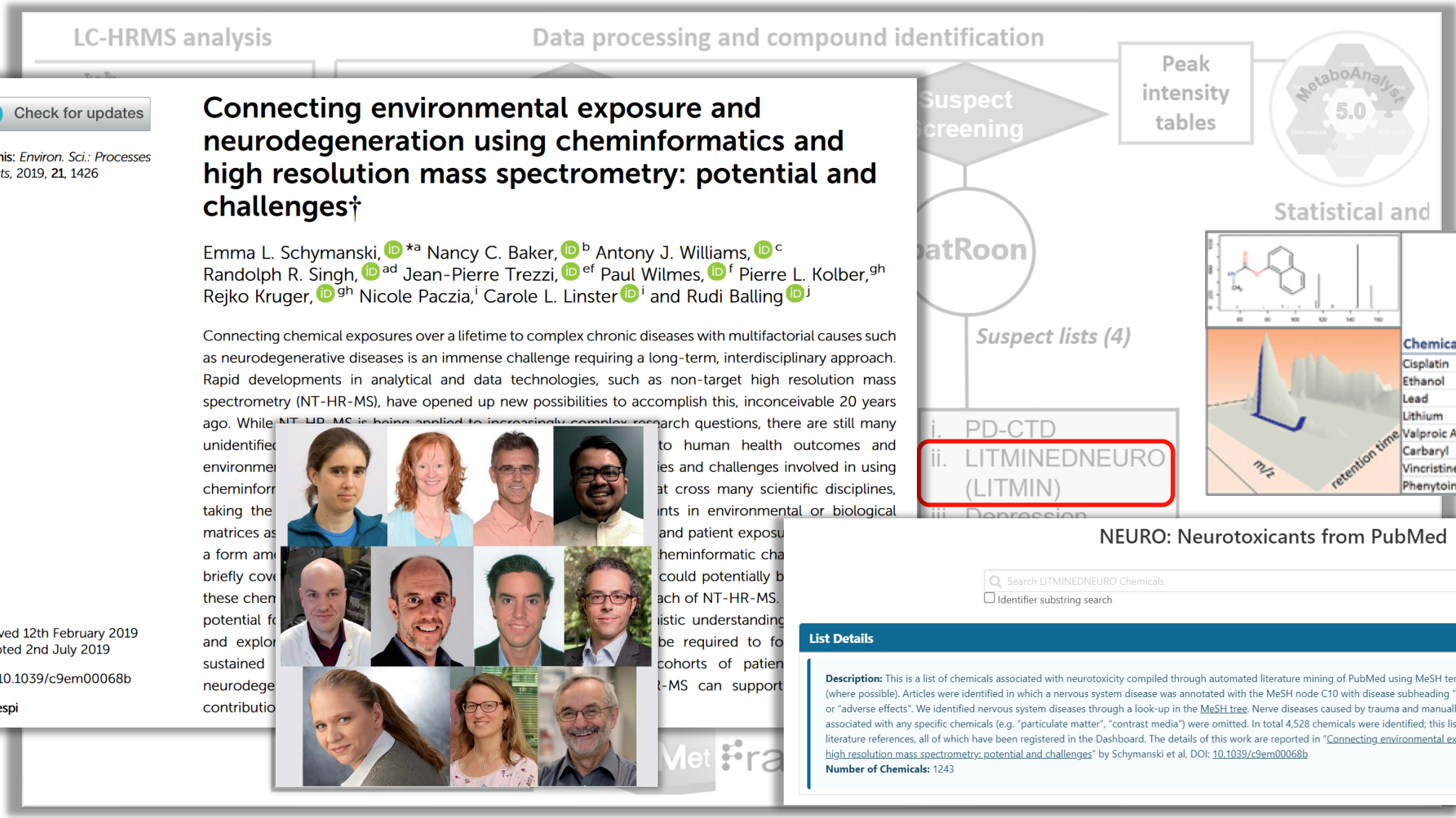
Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

Exposomics: Disease-specific Databases and Interpretation



Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

Exposomics: Disease-specific Databases: LITMINEDNEURO



Check for updates

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

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

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

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 to human health outcomes and environmen ies and challenges involved in using cheminform at cross many scientific disciplines, taking the nts in environmental or biological matrices as and patient exposu a form am heminformatic cha briefly cov could potentially b these chem ach of NT-HR-MS. potential f istic understanding and explor be required to fo sustained cohorts of patien neurodege -MS can support contributio

Received 12th February 2019
Accepted 2nd July 2019

DOI: 10.1039/c9em00068b

rsc.li/espi

Schymanski *et al*. (2019). Environ Sci: Proc Impact. DOI: 10.1039/C9EM00068B
Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

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

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

Database

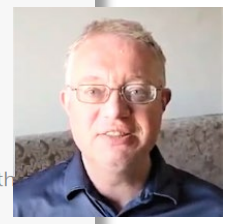
PubChem Protein, Gene, Pathway, and Taxonomy Data Collections: Bridging Biology and Chemistry through Target-Centric Views of PubChem Data

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

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

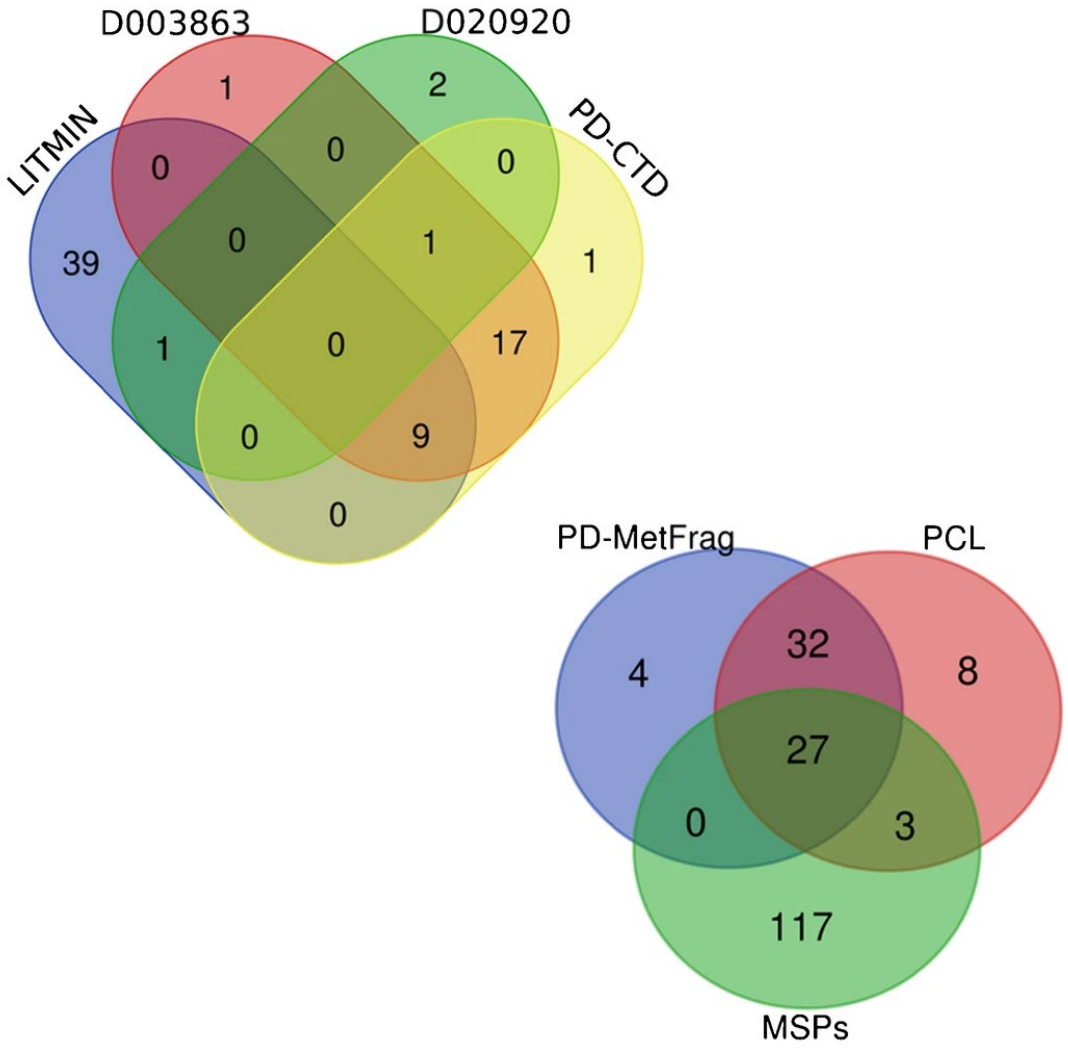
Leonid Zaslavsky*†, Tiejun Cheng†, Asta Gindulyte†, Siqian He†, Sunghwan Kim†, Qingliang Li†, Paul Thiessen†, Bo Yu† and Evan E. Bolton†

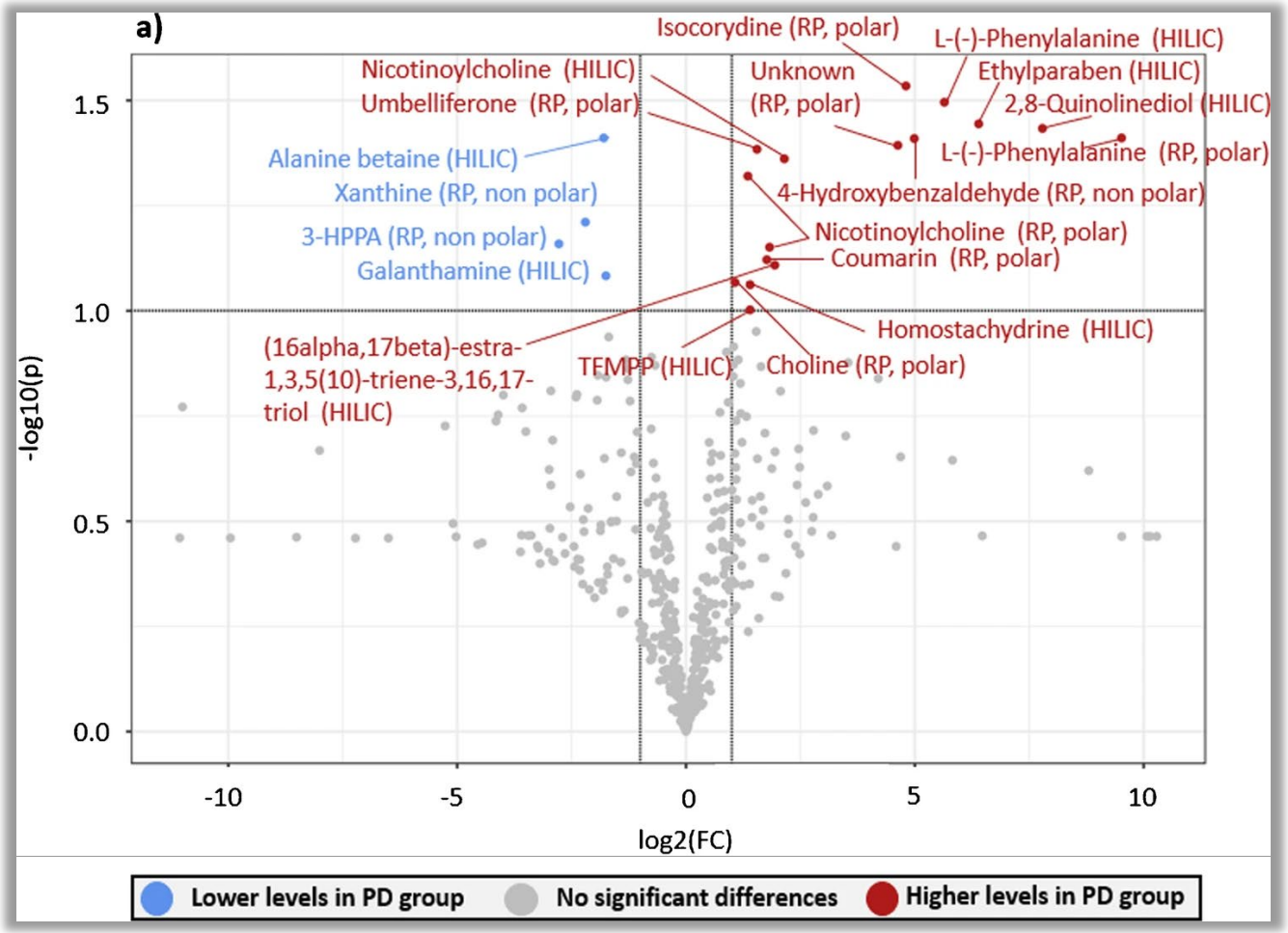
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States



Kim S. et al. (2022). DOI: 10.1016/j.jmb.2022.167514 & Zaslavsky *et al.* (2022) DOI: 10.3389/frma.2021.689059

Exposomics: Disease-specific Data Interpretation



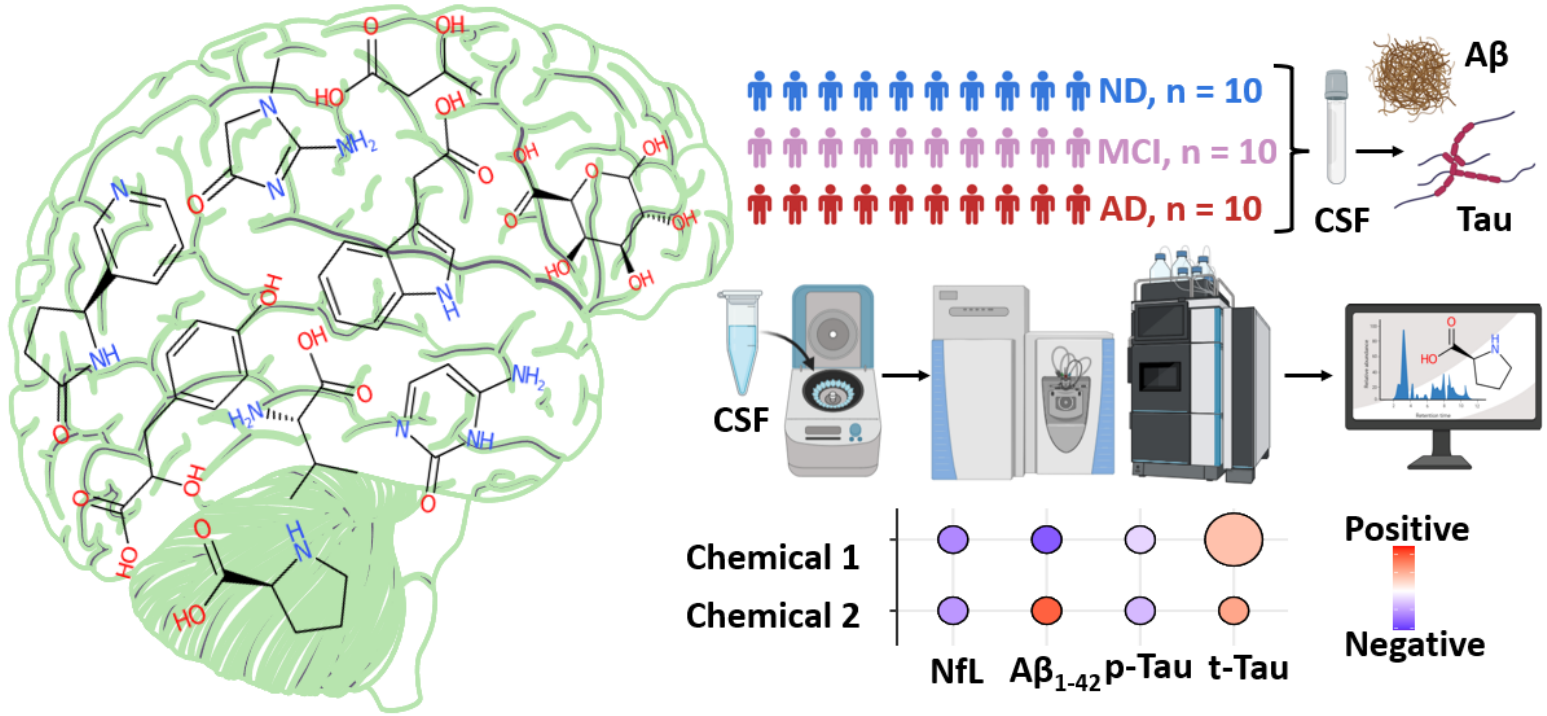


Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

Exposomics: Disease-specific Databases and Interpretation

MICROH

Information available varies widely depending on disease and sample

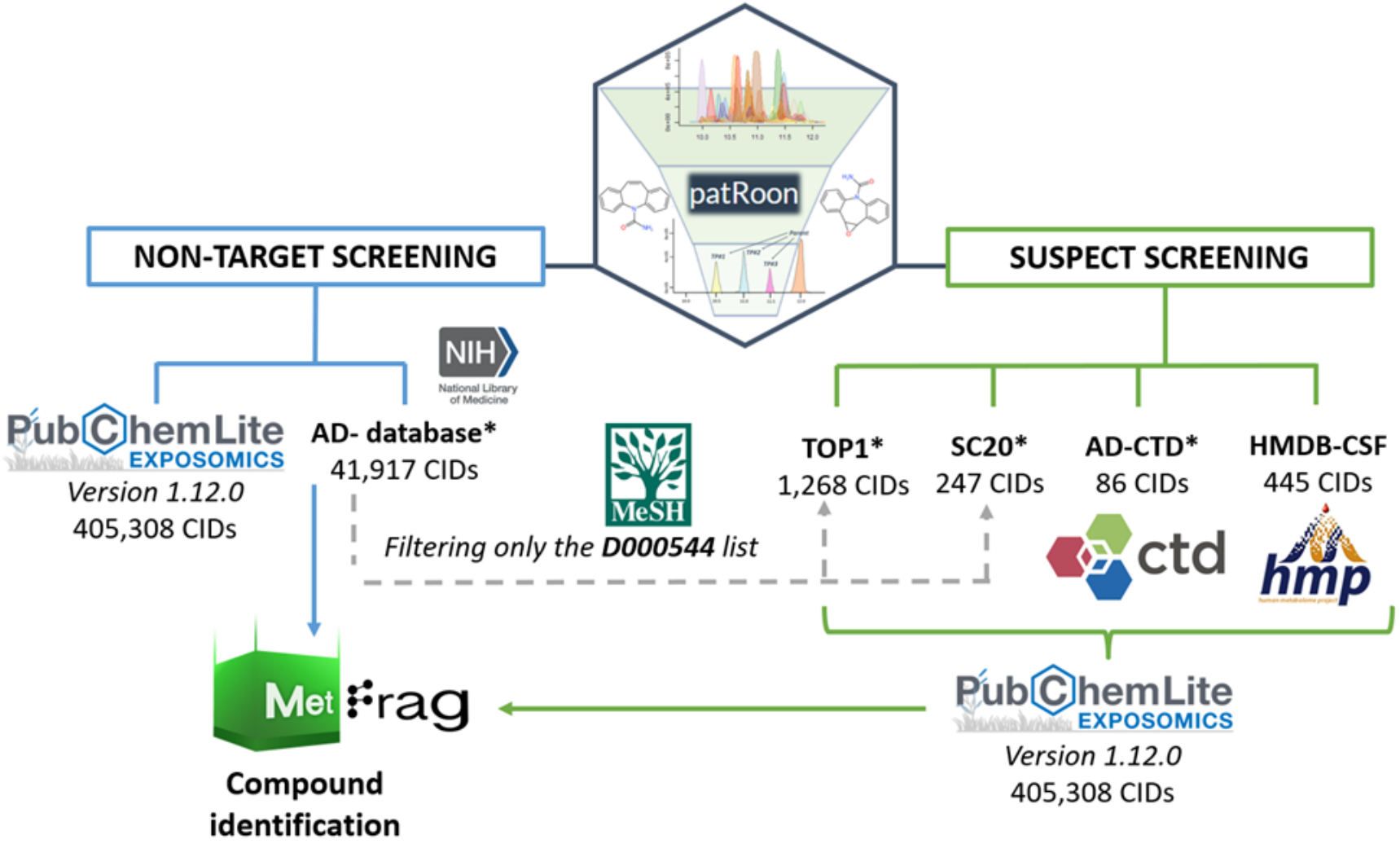


Talavera Andújar B. et al. (2023). ChemRxiv. DOI: 10.26434/chemrxiv-2023-6j2gm

Exposomics: Disease-specific Databases and Interpretation

MICROH

Information available varies widely depending on disease and sample



Talavera Andújar B. et al. (2023). ChemRxiv. DOI: 10.26434/chemrxiv-2023-6j2gm

Identification Evidence & Confidence Levels …

MICROH

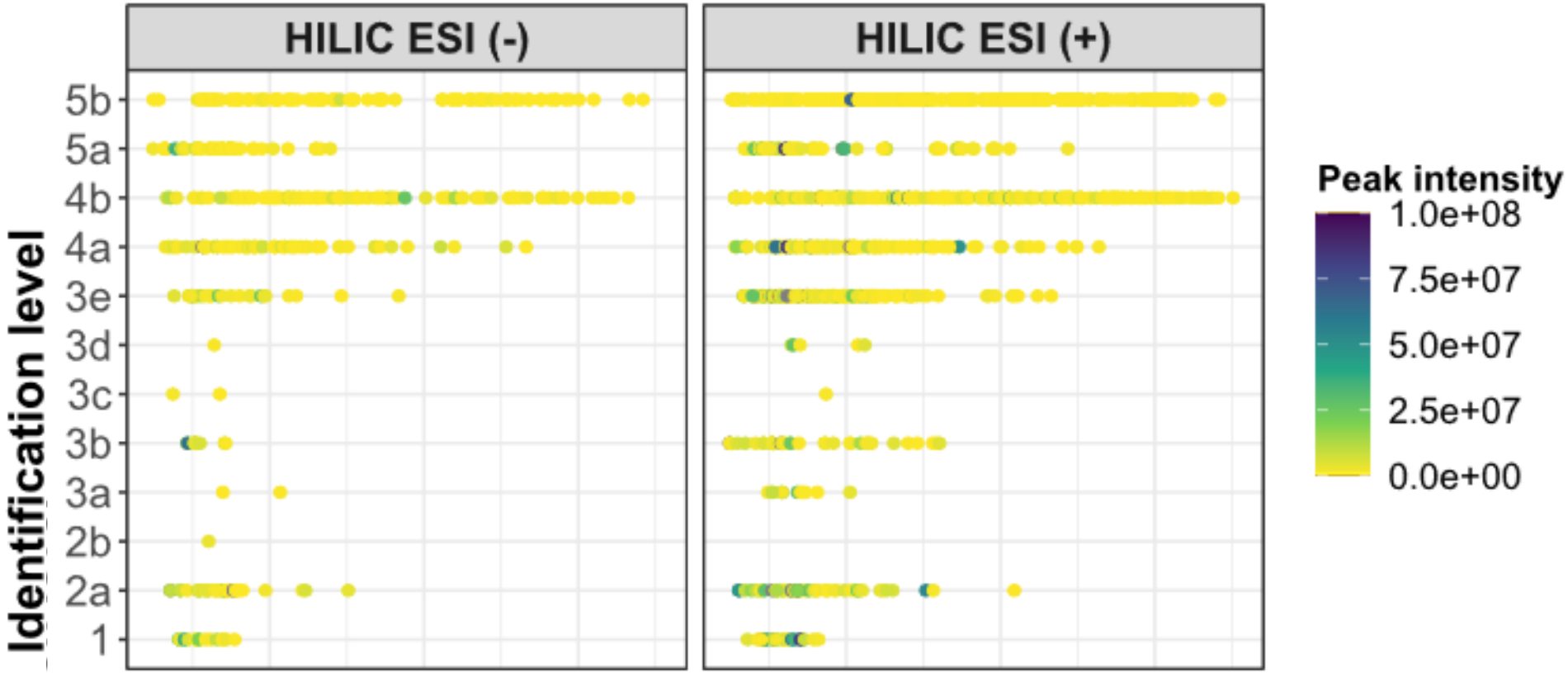
	Minimum requirements	Means used for identification
Level 1	Confirmed structure with reference standard	Internal library
Level 2a	>= 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product 70-100	
	Fragment presence 50-100	
Level 2b	>= 3 ion fragments matching	MS-DIAL MSPs libraries
	Dot product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

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

Talavera Andújar B. et al. (2023). ChemRxiv. DOI: 10.26434/chemrxiv-2023-6j2gm

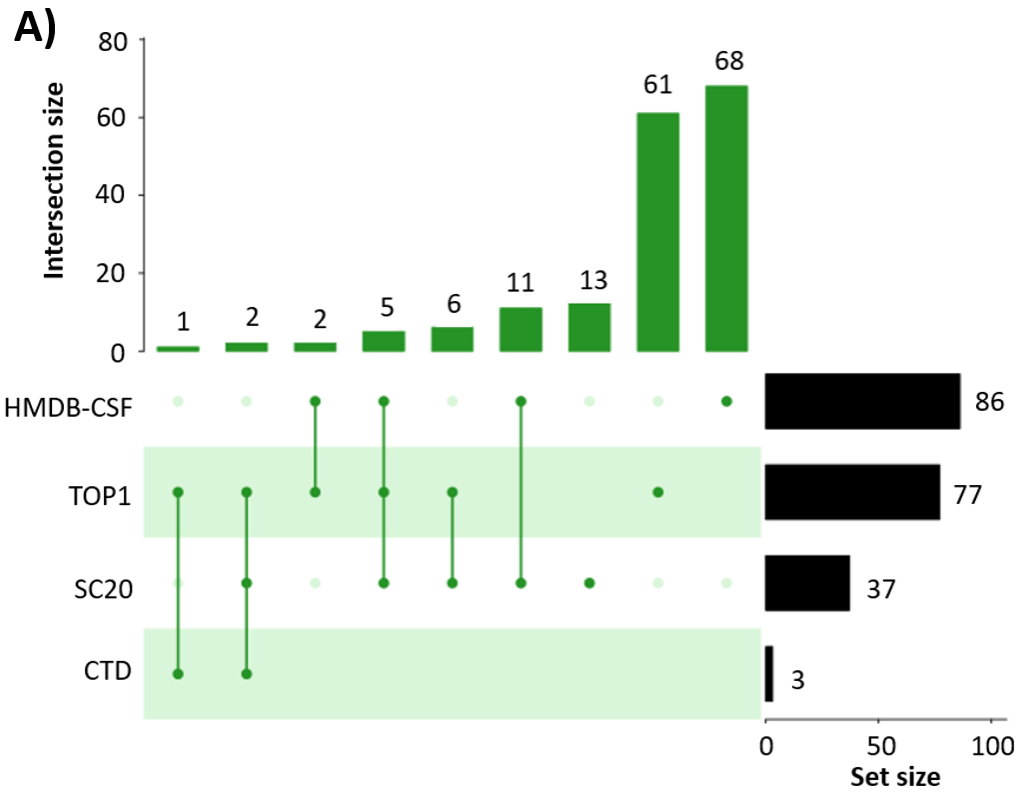
m/z & Intensity by Annotation Confidence Level & Method

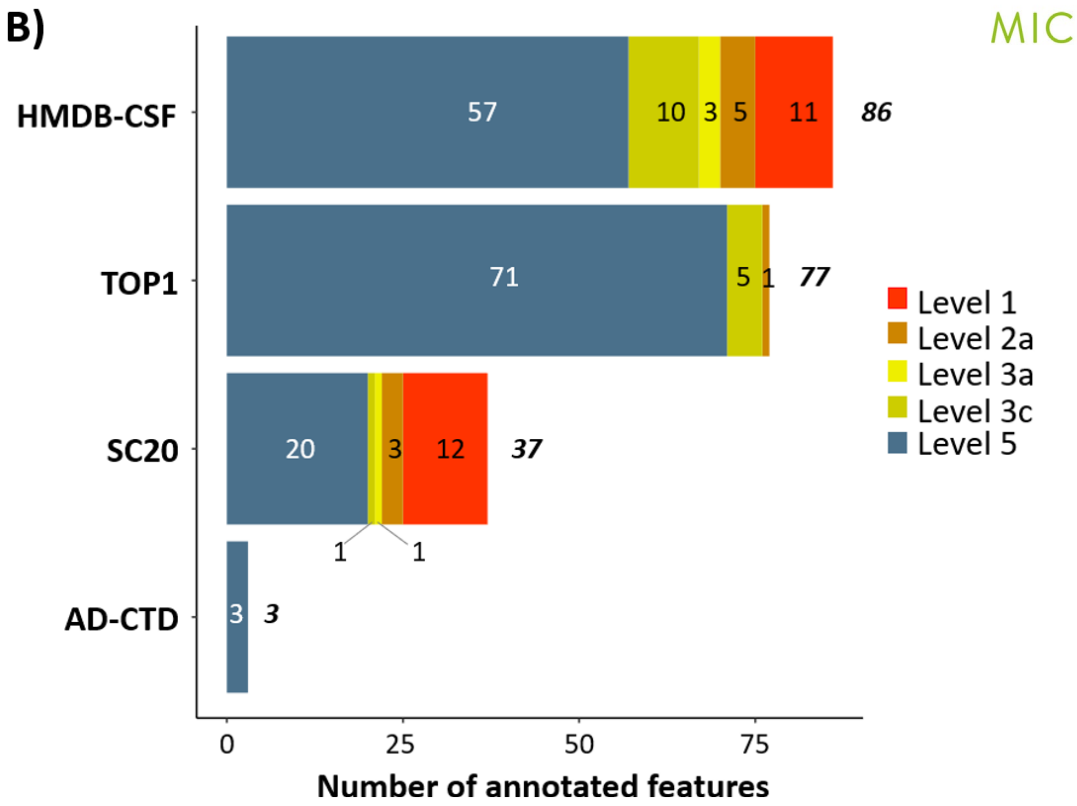
MICROH



Talavera Andújar B. et al. (2023). ChemRxiv.
DOI: 10.26434/chemrxiv-2023-6j2gm

Annotation Level and Overlap between Databases

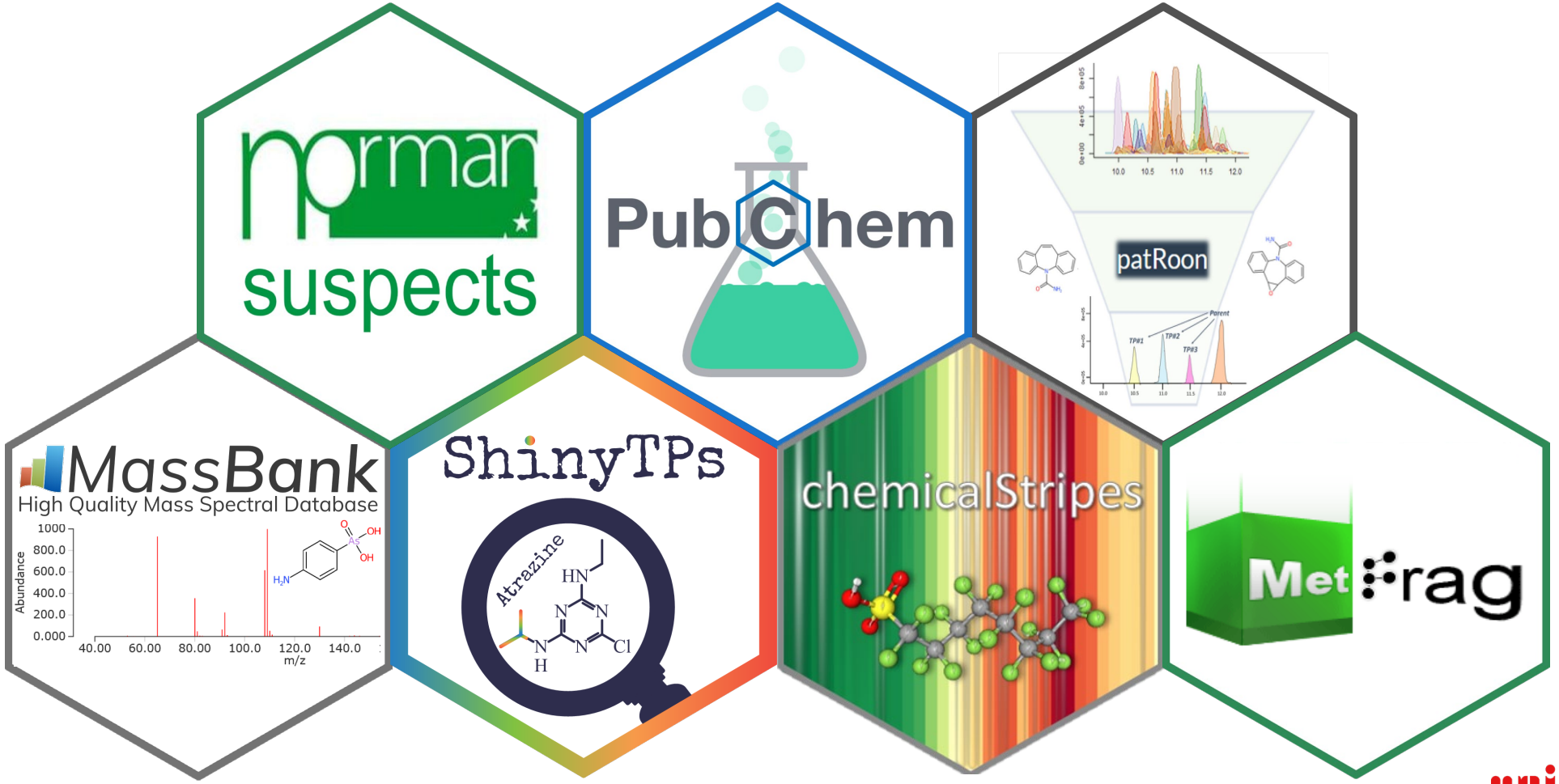




Talavera Andújar B. et al. (2023). ChemRxiv. DOI: 10.26434/chemrxiv-2023-6j2gm

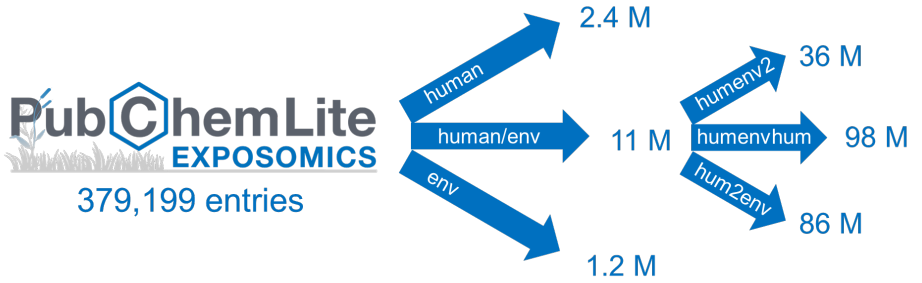
“Take home” Messages

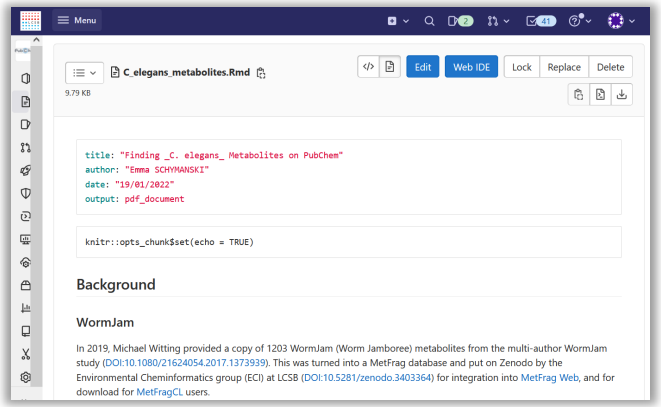
- Open, FAIR data supports NT-HRMS workflows



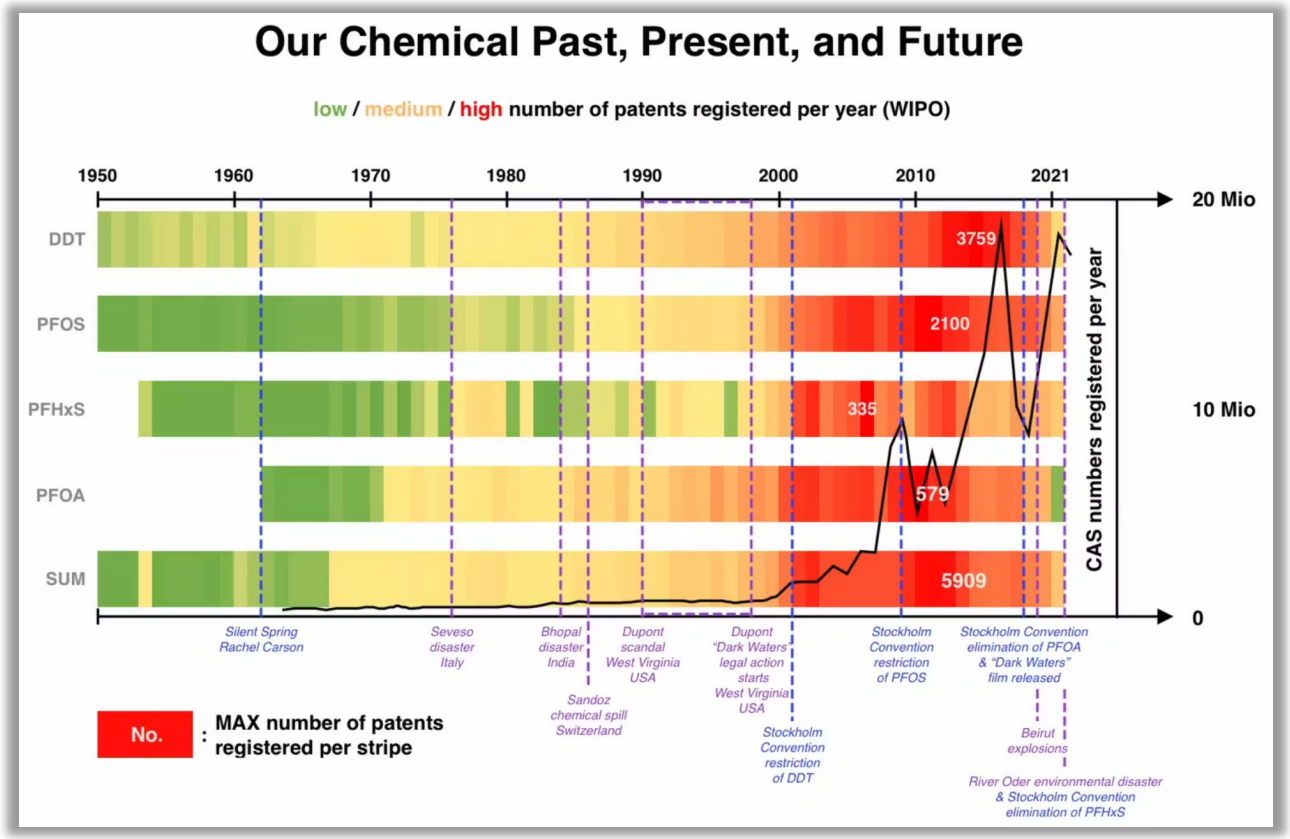
“Take home” Messages

- Dynamic, Open Collection: PubChemLite EXPOSOMICS
  - Expert knowledge to fill the gaps
  - Annotation content is extremely powerful
- Transformations and Metabolism
  - More data needed to improve predictions
- Workflows available in patRoon
- 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

Arp, Aurich *et al* (2023) ES&T, DOI: 10.1021/acs.est.3c01735

Acknowledgements!

Today's slides:

DOI: 10.5281/zenodo.8343924

Email: emma.schymanski@uni.lu

Twitter/X: @ESchymanski

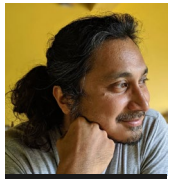
@schymane@mstdn.social

ORCID: 0000-0001-6868-8145

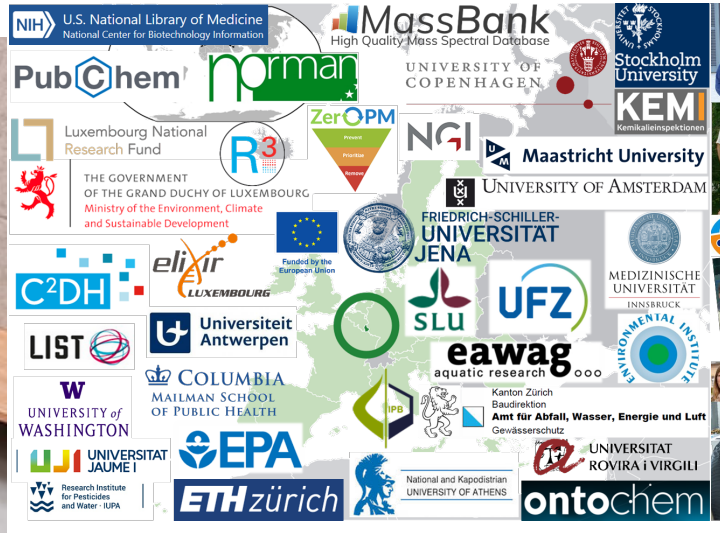
Today's slides

zenodo

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



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