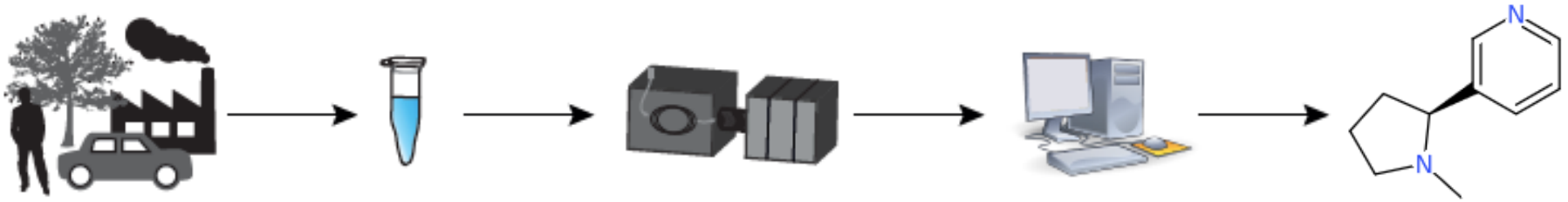
Digital Detective Work:
Connecting Cheminformatics, Mass Spectrometry and our Environment via Open Data



environment | environmental sample | high resolution mass spectrometry | cheminformatics | identify chemicals and toxic effects

Assoc. Prof. Dr. Emma L. Schymanski

FNR ATTRACT Fellow and PI in Environmental Cheminformatics
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
Email: emma.schymanski@uni.lu and @ESchymanski

Plus ECI-LSCB, NORMAN, PubChem, IPB, Eawag, UFZ and many other colleagues who contributed to our science over the years!

Talk available under DOI: 10.5281/zenodo.4266449

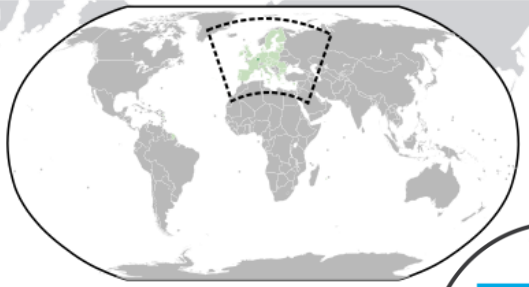
Luxembourg National Research Fund

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

PubChem

760 Data Sources

Explore Data Sources



LCSB

mi.lu

UNIVERSITY OF AMSTERDAM

norman

eawag aquatic research

UFZ

INRA SCIENCE & IMPACT

R3

MassBank.eu

EPA

THE GOVERNMENT OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate and Sustainable Development

Kanton Zürich
Baudirektion
Amt für Abfall, Wasser, Energie und Luft
Gewässerschutz

ENVIRONMENTAL INSTITUTE

UNIVERSITAT JAUME I

UNIVERSITAT ROVIRA i VIRGILI

Research Institute for Pesticides and Water · IUPA

s.lutions

KEMI Kemikalieinspektionen

ETH zürich

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

COLUMBIA MAILMAN SCHOOL OF PUBLIC HEALTH

IPB

WIKIDATA

Stockholm University

elixir LUXEMBOURG

National and Kapodistrian UNIVERSITY OF ATHENS

Maastricht University

UNIVERSITY OF COPENHAGEN

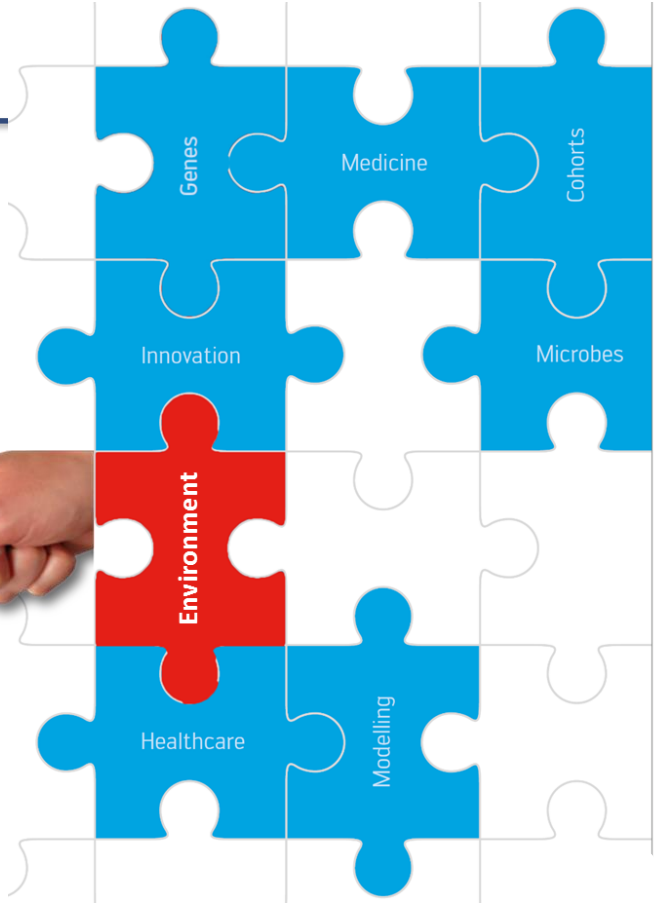
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Source: https://en.wikipedia.org/wiki/File:EU-Luxembourg.svg

LUXEMBOURG

LET'S MAKE IT HAPPEN

Environmental Cheminformatics @ Luxembourg Centre for Systems Biomedicine



Luxembourg National Research Fund

UNIVERSITÉ DU LUXEMBOURG

LUXEMBOURG

LET'S MAKE IT HAPPEN

Members with access to **Environmental Cheminformatics**

Adelene Lai @adelene.lai
Given access 2 months ago

Anjana Elapavalore @anjana.elapavalore
Given access 4 weeks ago

Corey Griffith @corey.griffith
Given access 2 months ago

Emma Schymanski @emma.schymanski It's you
Given access 2 months ago

German Andres Preciat Gonzales @german.preciat
Given access 2 months ago

Hiba Hiba @hiba.hiba
Given access 4 weeks ago

Jessy Krier @jessy.krier
Given access 1 week ago

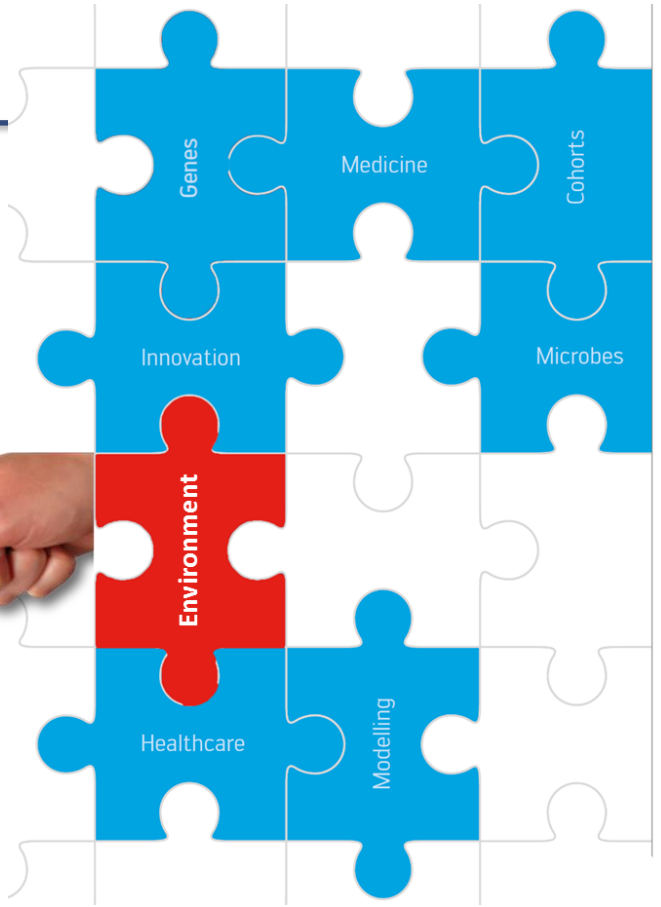
Lorenzo Favilli @lorenzo.favilli
Given access 2 months ago

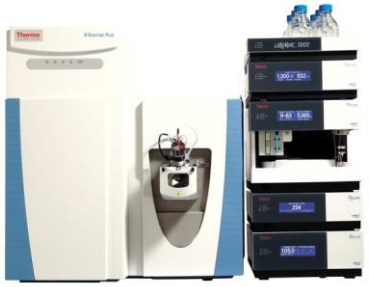
Mira Narayanan @mira.narayanan
Given access 4 weeks ago

Randolph Singh @randolph.singh
Given access 2 months ago

Todor Kondić @todo
Given access 2

R3 LCSB



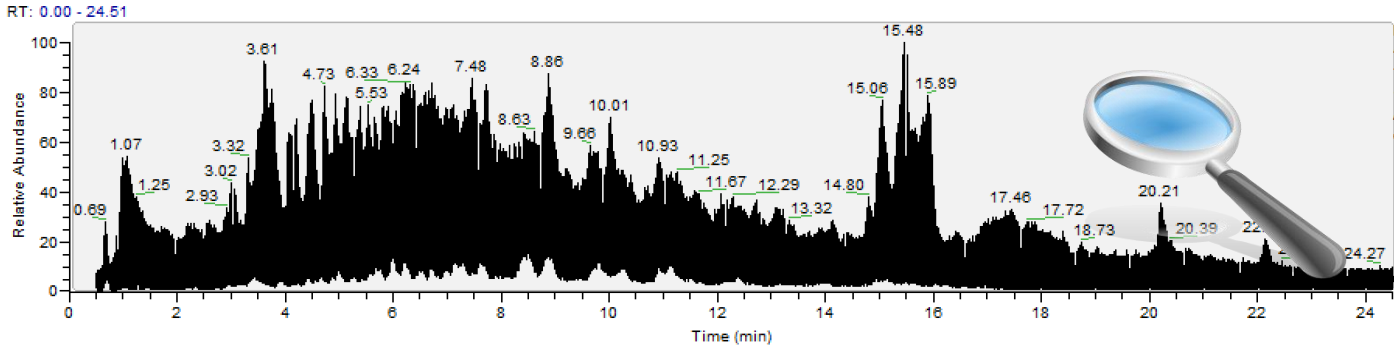


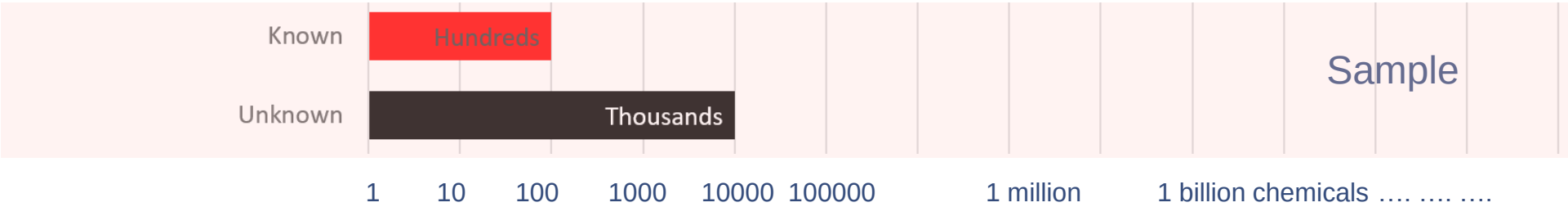
Luxembourg National Research Fund

UNIVERSITÉ DU LUXEMBOURG

Our (Community) Challenge: Identifying Chemicals

High resolution mass spectrometry



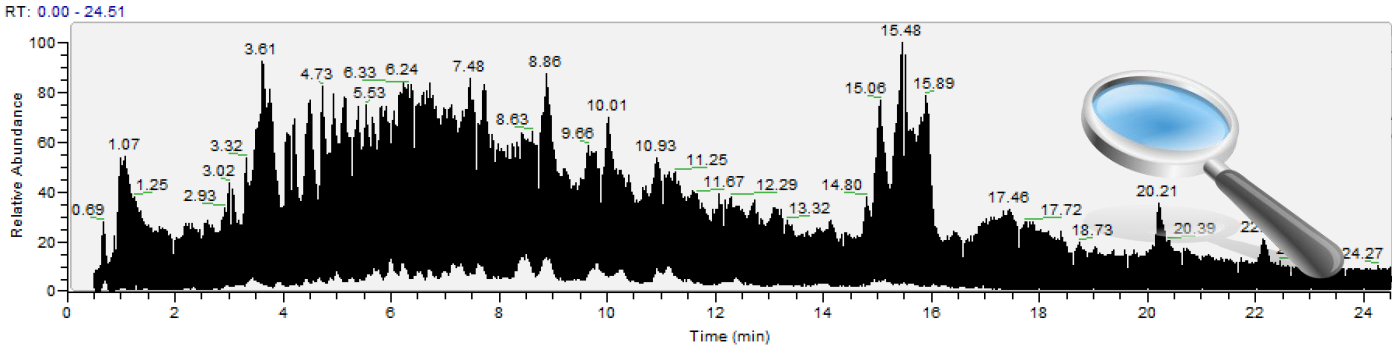


Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Vermeulen *et al.* (2020) DOI: 10.1126/science.aay3164

Our (Community) Challenge: Identifying Chemicals

High resolution mass spectrometry

AND connecting chemical knowledge



Known — Hundreds

Unknown — Thousands

Sample

1 10 100 1000 10000 100000 1 million 1 billion chemicals …. …. ….

Suspect Chemicals — 100s-1000s — norman suspects

Mass Spectral Libraries — ~80,000 — MassBank.eu

CompTox Dashboard — 882,000 — EPA Chemicals

PubChem Compound — >111 million — PubChem

1st Gen. PubChem Metabolites — >2 billion — PubChem

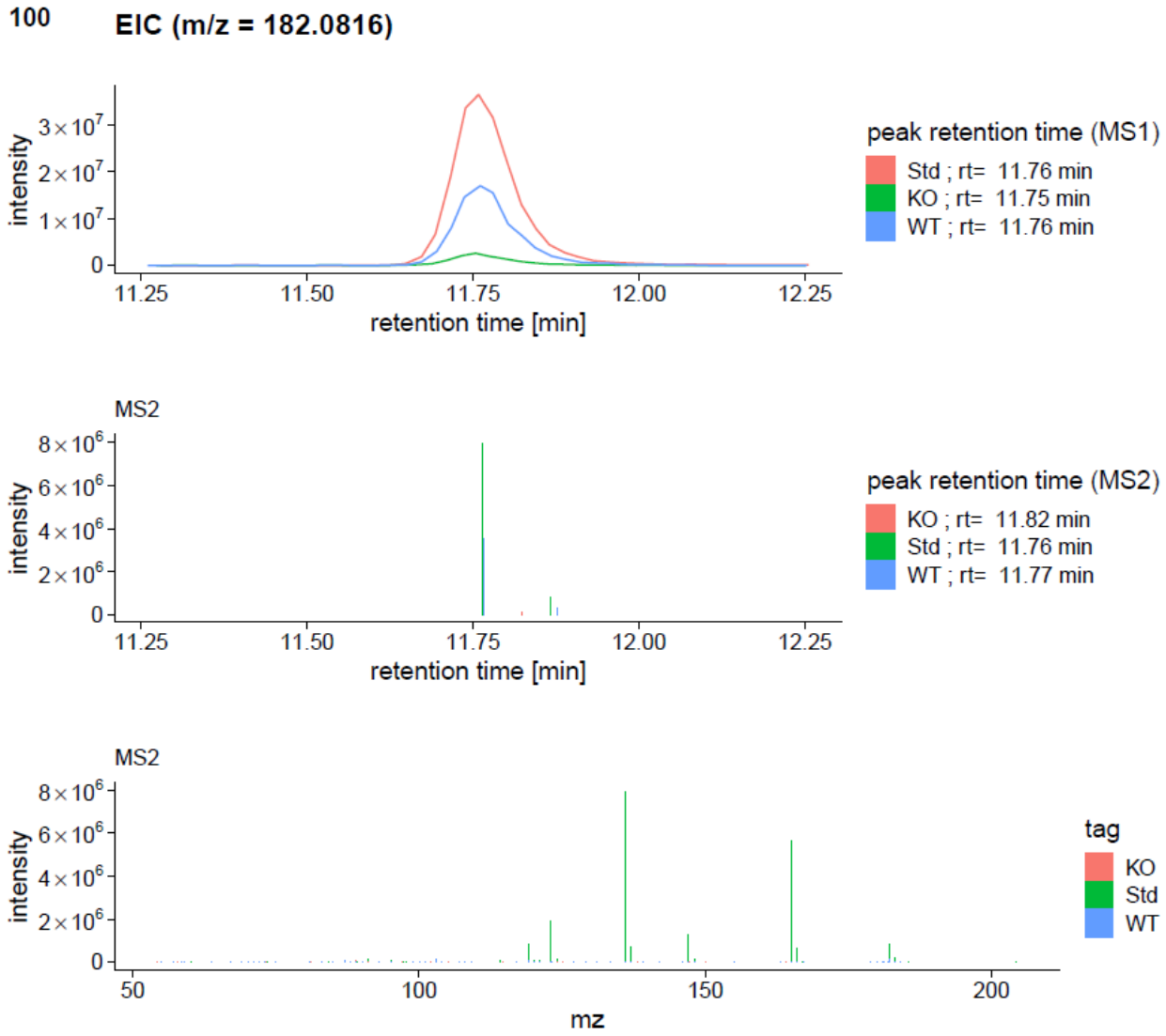
Generated Structures — Millions of billions

Chemicals

Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Vermeulen *et al.* (2020) DOI: 10.1126/science.aay3164

Our Toolkit: Shinyscreen – Extracting Good MS Data

https://git-r3lab.uni.lu/eci/shinyscreen



Anjana Elapavalore, Mira Narayanan, Todor Kondic, Jessy Krier, Hiba Mohammed Taha.

Our Toolkit: NORMAN-SLE: Capturing Expert Knowledge

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

NORMAN WEBSITE | NORMAN DATABASE SYSTEM | HOME | LOGIN

NORMAN SUBSTANCE DATABASE

NORMAN Suspect List Exchange – NORMAN SLE

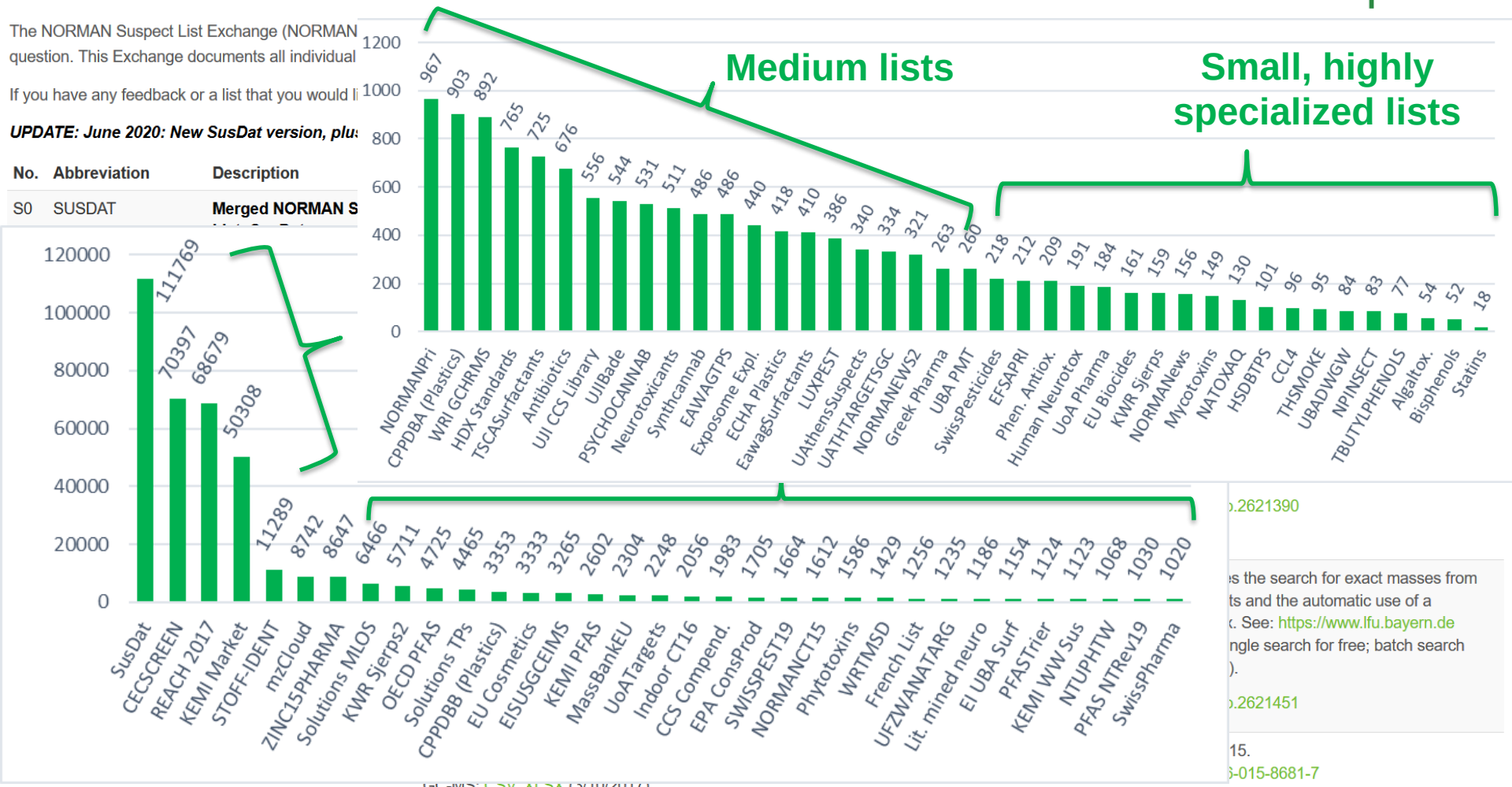
> 73 lists
> 144,980 substances

The NORMAN Suspect List Exchange (NORMAN
question. This Exchange documents all individual

If you have any feedback or a list that you would l

UPDATE: June 2020: New SusDat version, plus

No.	Abbreviation	Description
S0	SUSDAT	**Merged NORMAN S**



.2621390

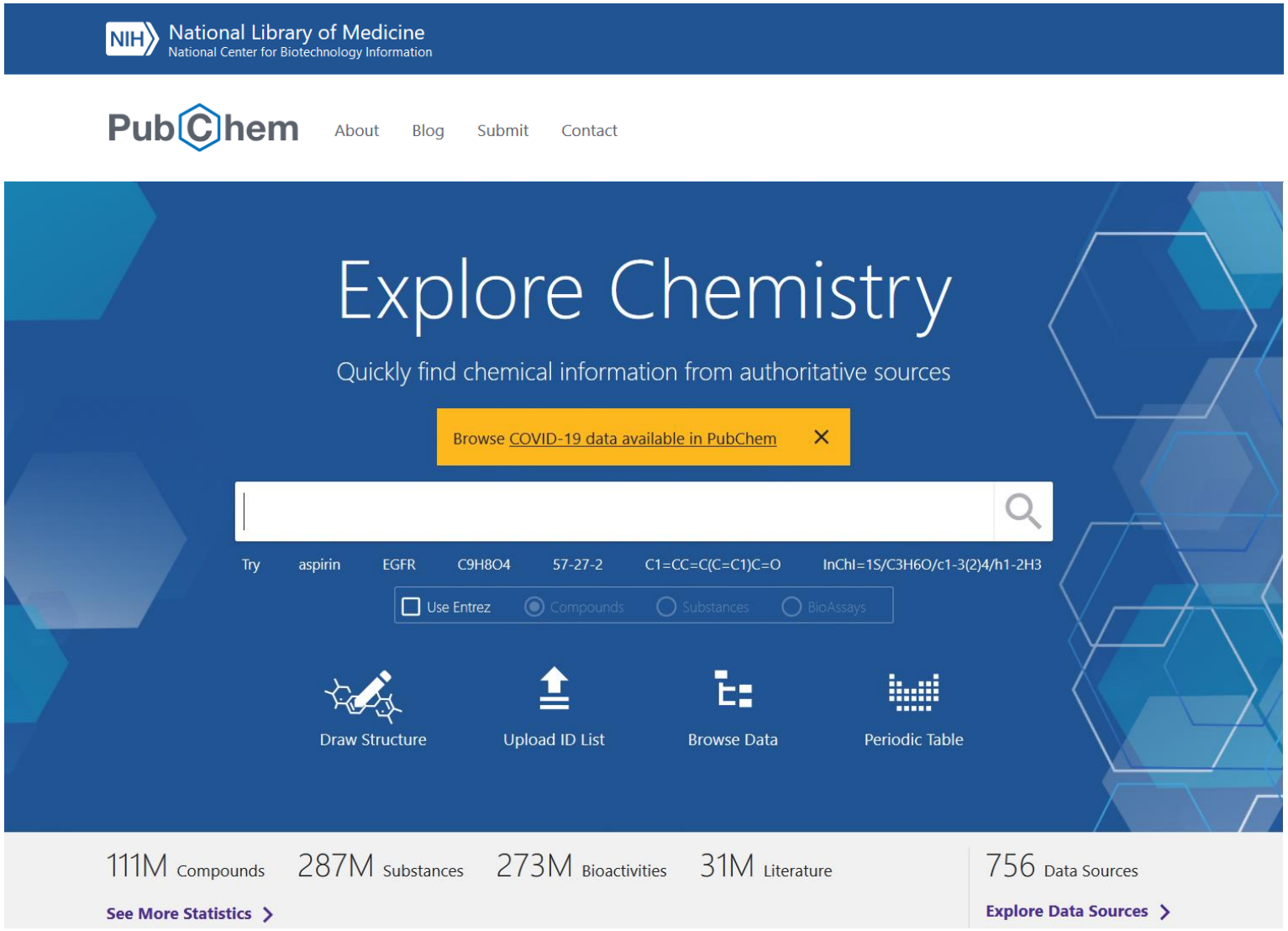
s the search for exact masses from
ts and the automatic use of a
. See: https://www.lfu.bayern.de
ngle search for free; batch search
).

.2621451

15.
3-015-8681-7

Our Toolkit: PubChem – Finding the “known unknowns”

https://pubchem.ncbi.nlm.nih.gov/



Kim S, Chen J, Cheng T, et al. PubChem 2019 update: …. *Nucleic Acids Res*. 2019;47(D1):D1102–D1109. doi:10.1093/nar/gky1033

Our Toolkit: MetFrag – Annotating/Identifying Masses

MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings

Database: PubChem Include references: ☑ Parent Ion: [M+H]+ Calculate

Neutral Mass: 253.966126 Search ppm: 5

Formula:

Identifiers:

Retrieve Candidates Download Candidates

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm: 5

Mzabs: 0.001

Mode: [M+H]+

Tree depth: 2

Group candidates ☑

MS/MS Peak list

```
90.97445 681
106.94476 274
110.02750 110
115.98965 95
117.98540 384
124.93547 613
124.99015 146
125.99793 207
133.95592 777
143.98846 478
144.99625 352
```

Show Spectrum

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, 2016, DOI: 10.1186/s13321-016-0115-9

MetFrag + PubChem + MS/MS only

Status: 2010

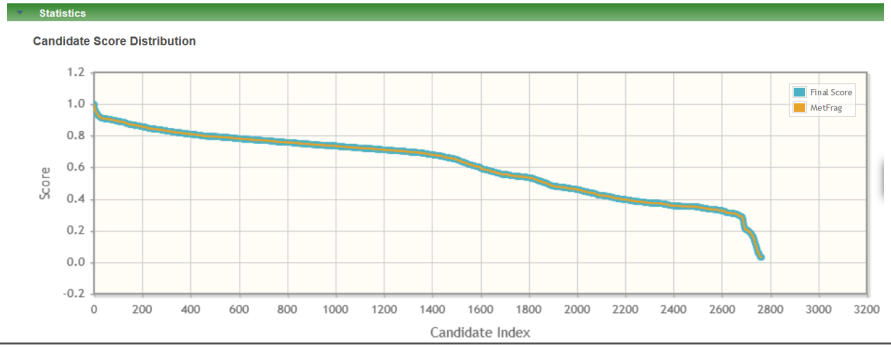
mz [M−H]⁻
213.9637
± 5 ppm

5 ppm
0.001 Da

PubChem

MetFrag

Ranked Candidates



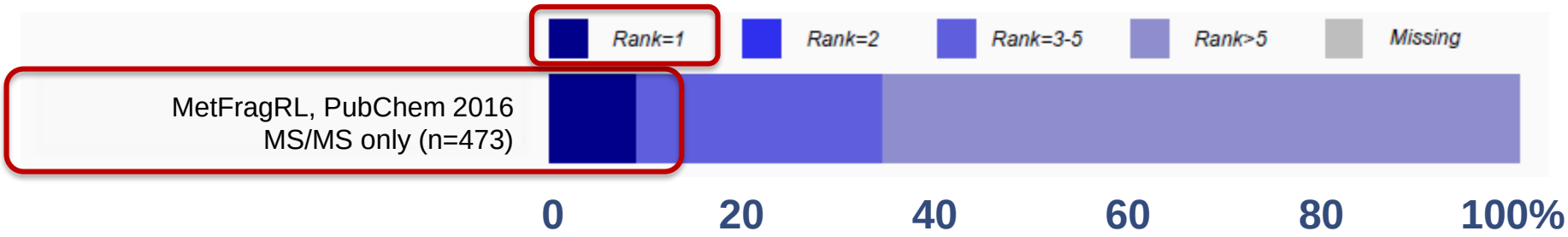
MS/MS

134.0054	339689
150.0001	77271
213.9607	632466

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, 2016, DOI: 10.1186/s13321-016-0115-9

Challenge: MetFrag + PubChem + MS/MS only

- MS/MS doesn't provide sufficient information alone …



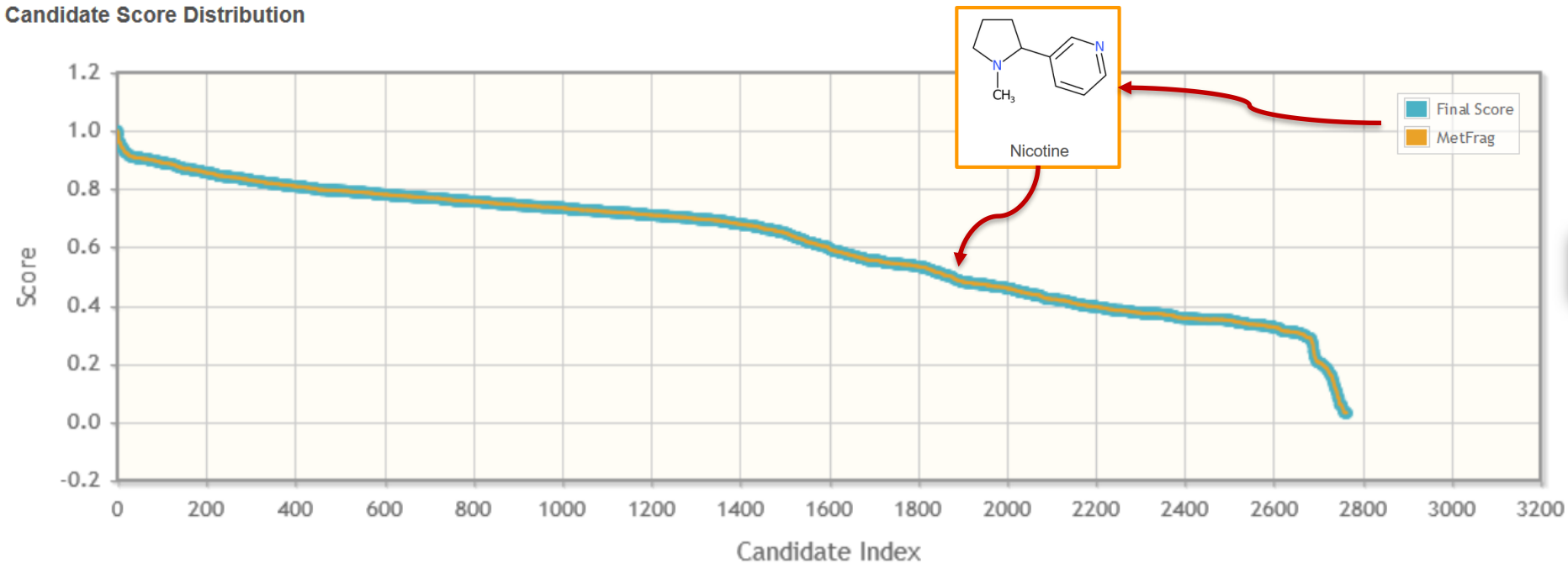
Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

MS/MS does not provide sufficient information alone…

MetFrag

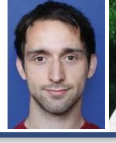
PubChem

Statistics



MetFrag + PubChem + Formula Search + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

MetFrag – MS/MS and MORE!

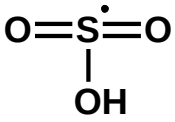


Status: 2016

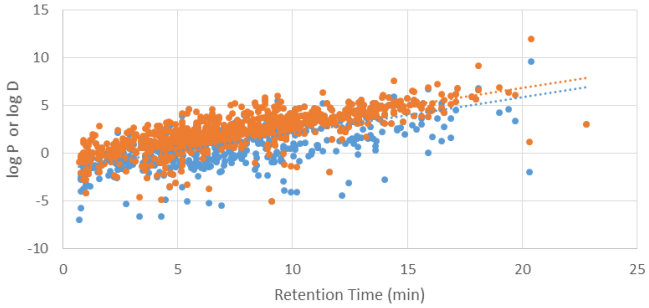
mz [M-H]⁻
213.9637
± 5 ppm

Elements: C,N,S

5 ppm
0.001 Da



RT: 4.54 min
355 InChI/RTs



EPA Chemicals
or
PubChem

MetFrag

MoNA
MassBank of North America
MassBank.eu

References
Tox. Data
Data Sources
Exposure Info
MS-ready links

Literature Patents

Suspect Lists

?TOFF !DENT

norman suspects

EPA Chemicals

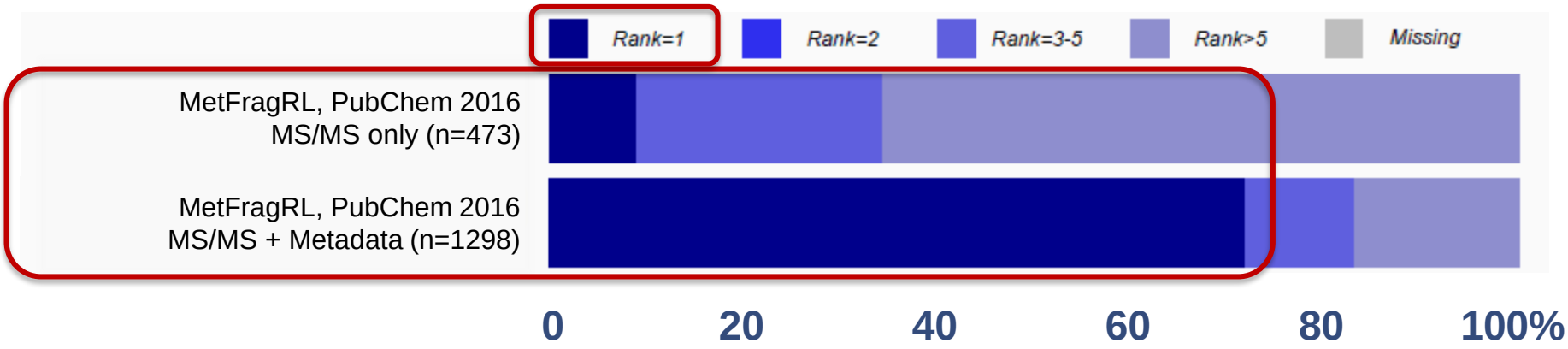
MS/MS

134.0054	339689
150.0001	77271
213.9607	632466

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, 2016, DOI: 10.1186/s13321-016-0115-9

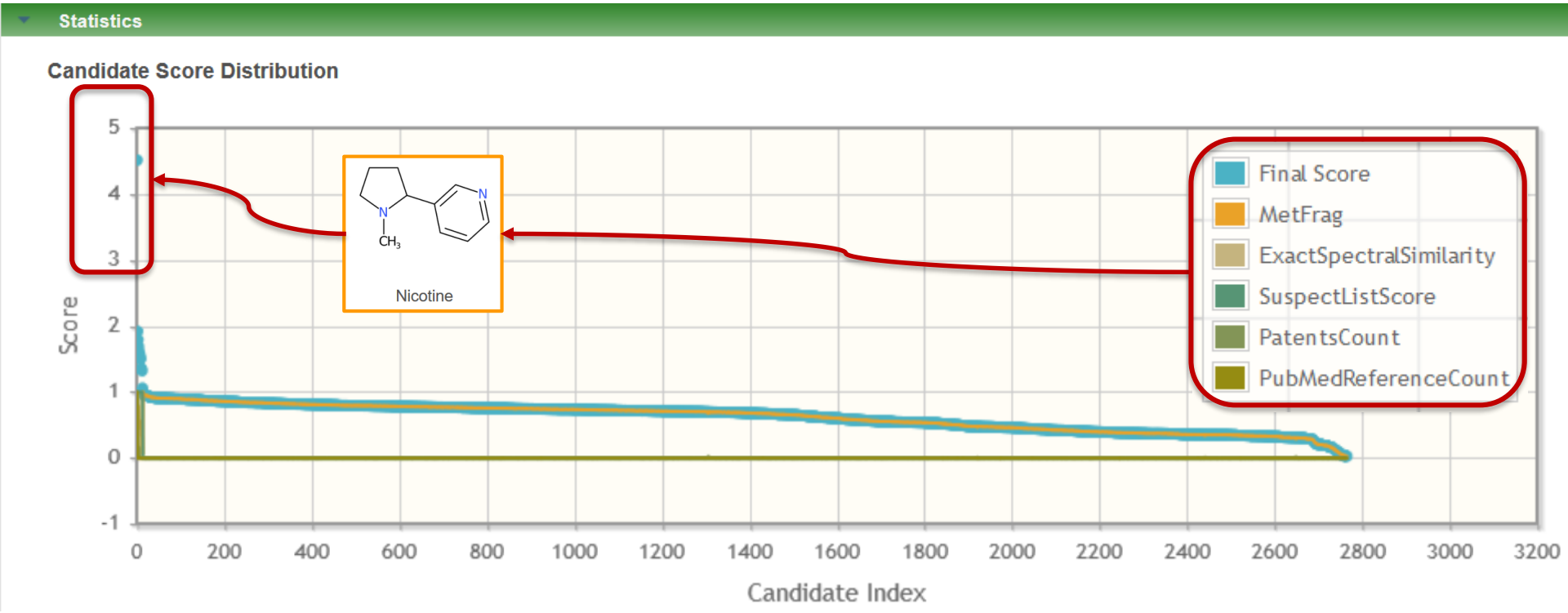
MetFrag + PubChem + MS/MS + Metadata

- Adding literature, references & RT boosts to ~71 % rank 1!



Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

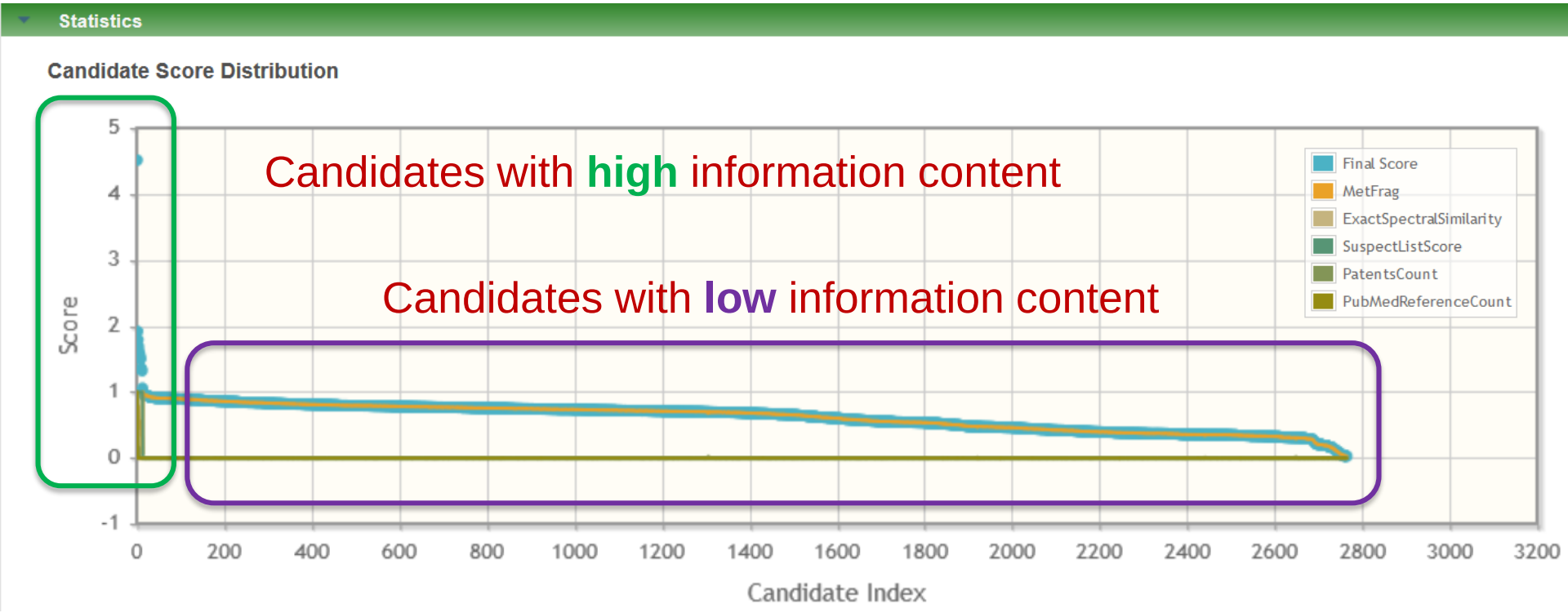
Connecting multiple lines of evidence for identification



MetFrag + PubChem + Formula + MoNA + SusDat + Pat + Refs + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

Connecting multiple lines of evidence for identification

Challenge: the growing number of candidates …
Need: wide coverage and high efficiency!



MetFrag + PubChem + Formula + MoNA + SusDat + Pat + Refs + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

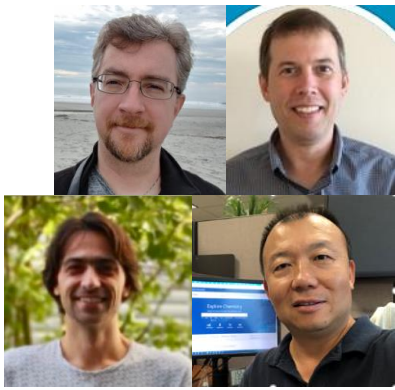
The 111 million PubChem Challenge …

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

- PubChem Compound TOC ? 49,493,641
  - Agrochemical Information ? 3,045
  - Associated Disorders and Diseases ? 20,847
  - Biologic Description ? 2,056,521
  - Biological Test Results ? 3,622,920
  - Biomolecular Interactions and Pathways ? 125,253
  - Chemical and Physical Properties ? 263,015
  - Classification ? 1,454,824
  - Drug and Medication Information ? 17,922
  - Food Additives and Ingredients ? 8,414
  - Identification ? 4,968
  - Information Sources ? 20,271,277
  - Literature ? 1,833,941
  - Names and Identifiers ? 1,275,170
  - Patents ? 36,351,418
  - Pharmacology and Biochemistry ? 110,628
  - Related Records ? 9,224,590
  - Safety and Hazards ? 149,319
  - Spectral Information ? 480,730
  - Structures ? 9,117,635
  - Toxicity ? 114,012
  - Use and Manufacturing ? 115,321
- Chemical Safety ? 147,023

111 million … OR … the most relevant / annotated?

PubChemLite for Exposomics
371,663 entries



zenodo

Search | Upload | Communities | emma.schymanski@uni.lu

October 31, 2020 — Dataset — Open Access

PubChemLite for Exposomics

Bolton, Evan; Schymanski, Emma; Kondic, Todor; Thiessen, Paul; Zhang, Jeff

PubChemLite is a subset of PubChem (https://pubchem.ncbi.nlm.nih.gov/) selected from major

Edit | New version

Communities

LCSB Environmental Cheminformatics Group — Remove

690 views | 763 downloads

See more details...

Database Settings

Database: PubChemLite_31Oct2020

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16ClN5

Identifiers:

Retrieve Candidates | 4 Candidates

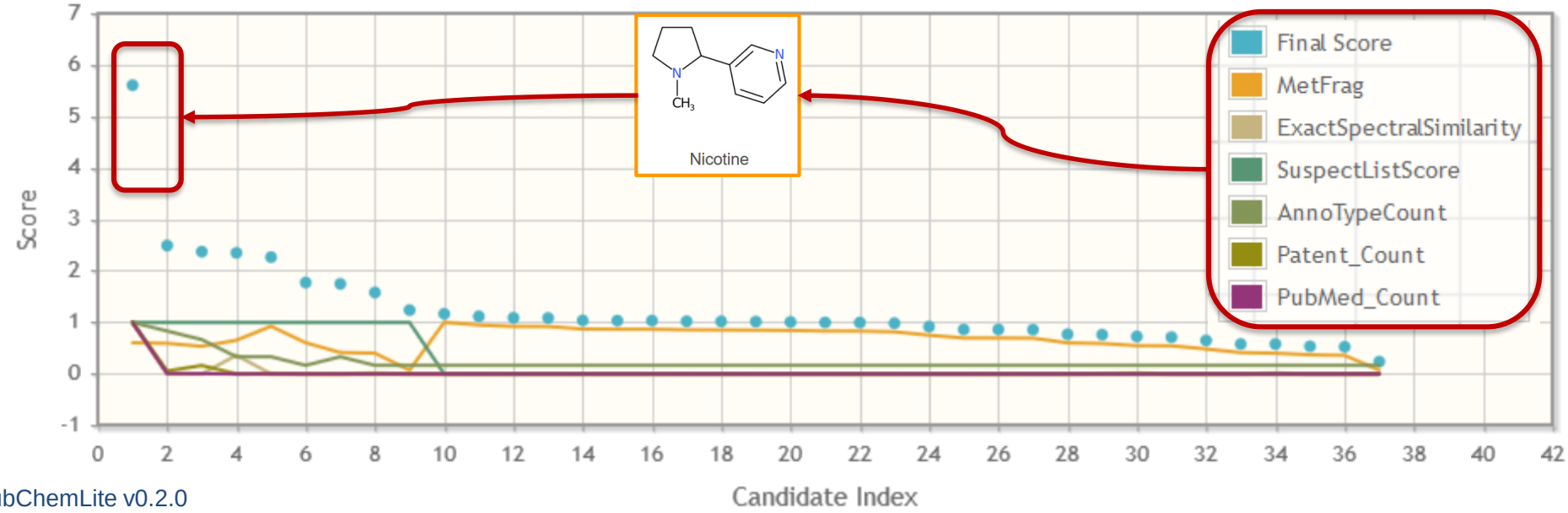
MetFrag

Bolton, Schymanski, Kondic, Thiessen, Zhang (2020). PubChemLite for Exposomics (v 0.3.0) DOI: 10.5281/zenodo.4183801

PubChemLite: tailor-made database + metadata

Statistics

Candidate Score Distribution



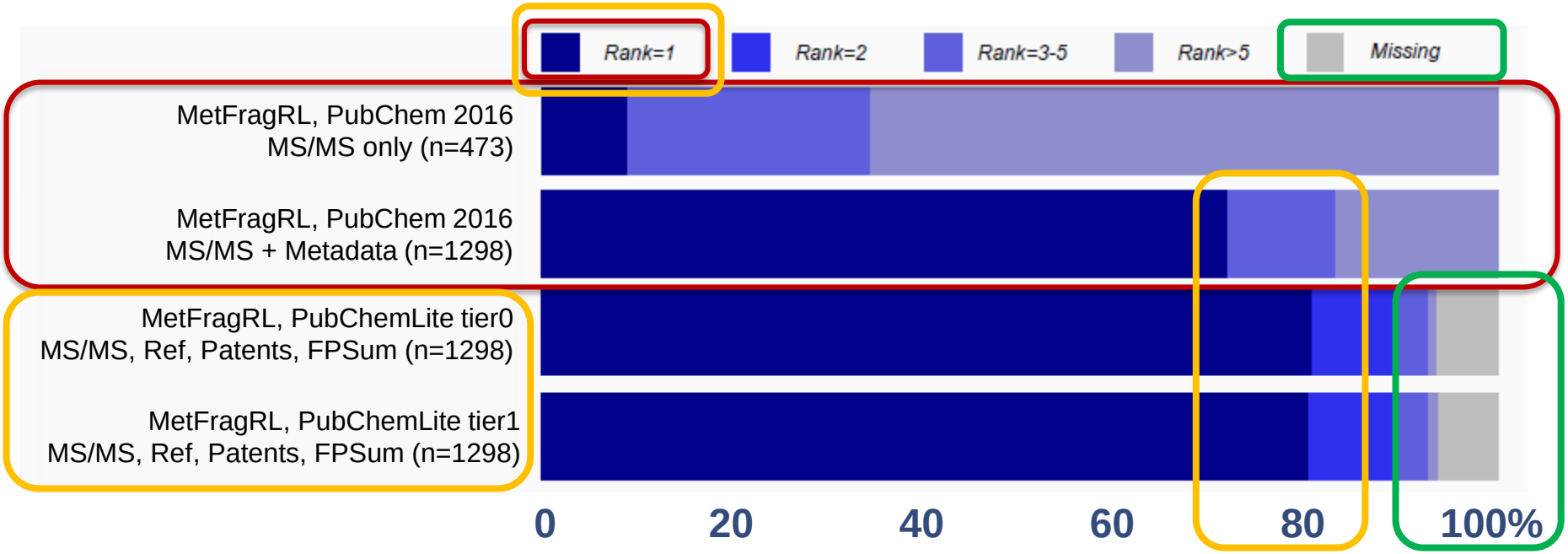
PubChemLite v0.2.0

MetFrag+PubChemLite+Formula+MoNA+SusDat+Pat+Refs+Anno + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

How does PubChemLite perform?

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

- 111 M => 300 K … how does this influence performance?

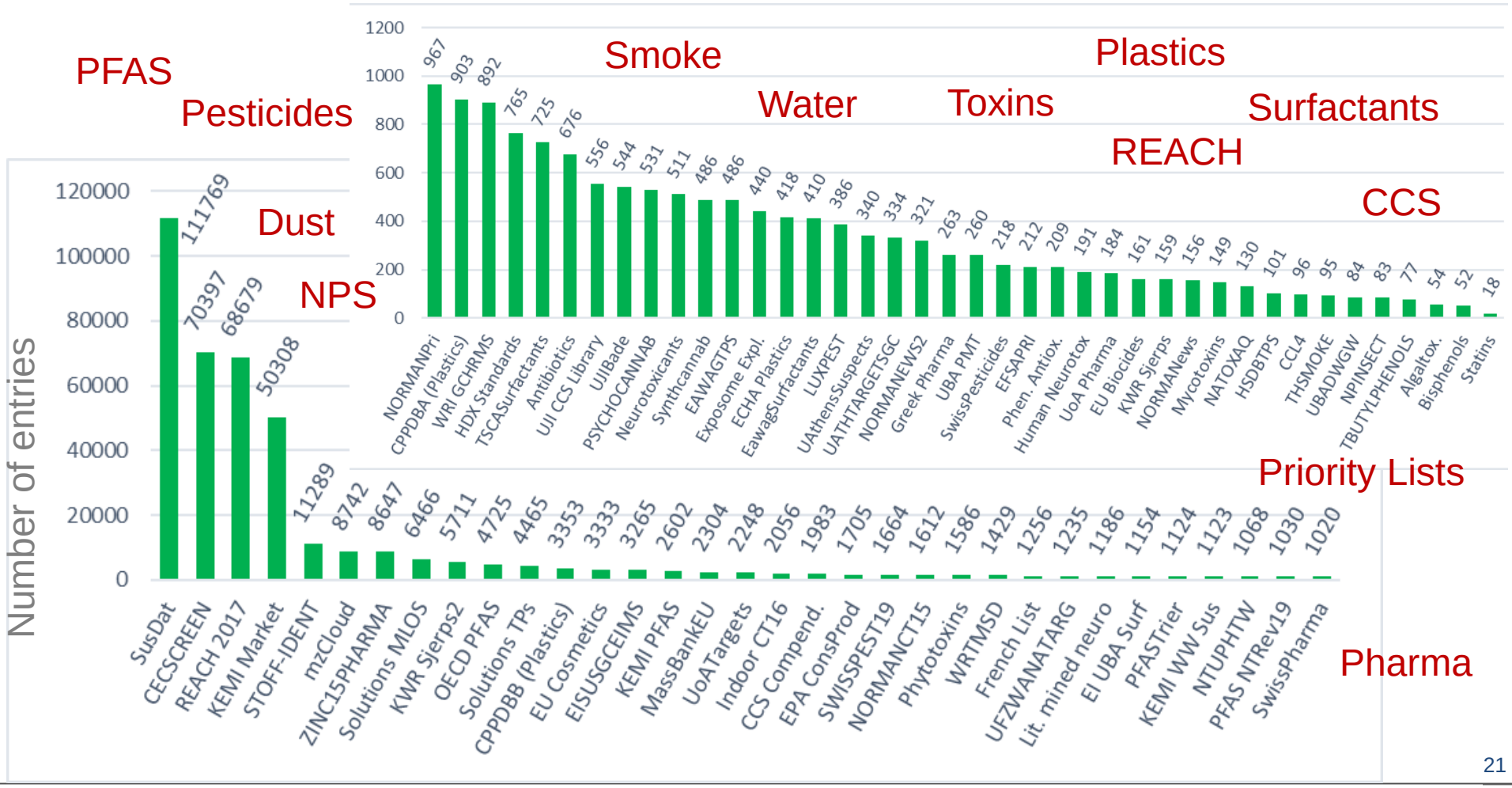


Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

NORMAN Suspect List Exchange: Capturing Knowledge

- https://www.norman-network.com/nds/SLE/
- https://zenodo.org/communities/norman-sle
- https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

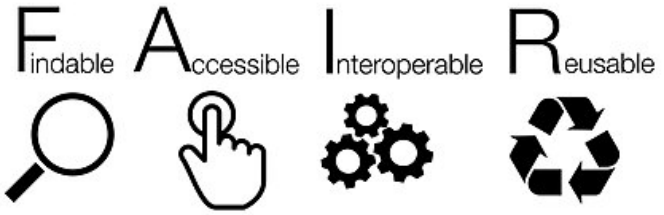
>73 lists!



Schymanski *et al*. (in prep.) https://www.norman-network.com/nds/SLE/

NORMAN-SLE on Zenodo

https://zenodo.org/communities/norman-sle/



zenodo

Search

emma.schymanski@uni.lu

Communities

NORMAN Suspect List Exchange

Remove

NORMAN Suspect Li

Recent uploads

Search NORMAN Suspect List Exchange

November 21, 2019 (NORMAN-SLE S61.0.1.2) | Dataset | Open A

S61 | UJICCSLIB | Collision Cross Sectior

Celma, Alberto; Fabregat-Safont, David; Ibàñe Félix; Sancho, Juan Vicente;

This is the collection associated with list S61 UJICCSl https://www.norman-network.com/nds/SLE/ A list of compounds (both positive and negative ionization mo instrument) pr

Uploaded on July 30, 2020

2 more version(s) exist for this record

1,049 views

1,166 downloads

See more details...

Publication date:
November 21, 2019

DOI:
DOI 10.5281/zenodo.3966751

Keyword(s):
Collision Cross Section | Ion Mobility | Suspect Screening | IMS-MS | CCS

Related identifiers:
Supplement to https://www.norman-network.com/nds/SLE/

New upload

NORMAN Suspect List Exchange

This is a public repository (under development) for suspect lists currently available on the NORMAN Suspect List Exchange: https://www.norman-network.com/?q=suspect-list-exchange

Read more

Curated by:
schymane

Curation policy:
This community will collect data that is

NORMAN-SLE on PubChem

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

https://pubchem.ncbi.nlm.nih.gov/source/NORMAN Suspect List Exchange

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

PubChem NORMAN Suspect List Exchange

Organization:	NORMAN Network (c/o Emma Schymanski)
Category:	Research and Development
URL:	https://www.norman-network.com/nds/SLE/
Contact Name:	Emma Schymanski
Address:	6 avenue du Swing, Belvaux, Luxembourg, 4367
Data Source ID:	23819
Data in PubChem:	144,980 Live Substances 11,208 Annotations 1 Classification
Last Updated:	2020/08/01

Huge thanks to Evan Bolton, Jian Zhang (Jeff), Paul Thiessen, Ben Shoemaker and the PubChem team for this!

NORMAN-SLE on PubChem

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

- https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

PubChem Classification Browser

Help

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). More...

Select classification NORMAN Suspect List Exchange

Search selected classification by Keyword | Enter desired search term | Search

Classification description (from NORMAN Suspect List Exchange)

The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring questions. More...

Data type counts to display None | Compound

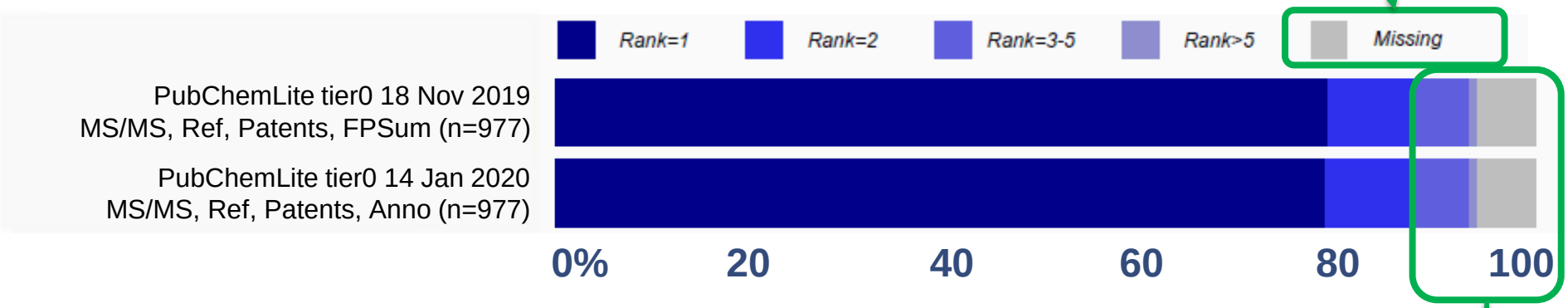
Display zero count nodes? Yes | No

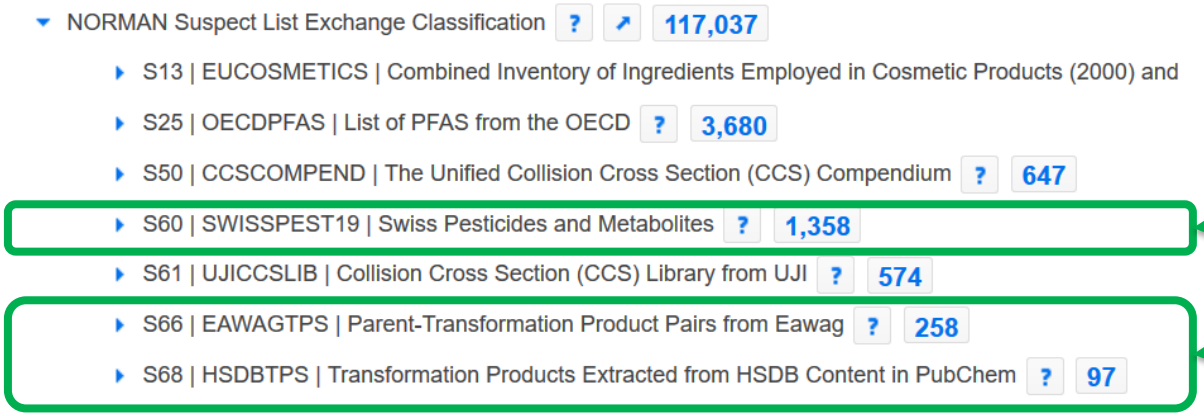
Browse NORMAN Suspect List Exchange Tree

- NORMAN Suspect List Exchange Classification ? 130,864
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 4,131
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 647
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,354
  - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
  - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258

Missing Entries in PubChemLite

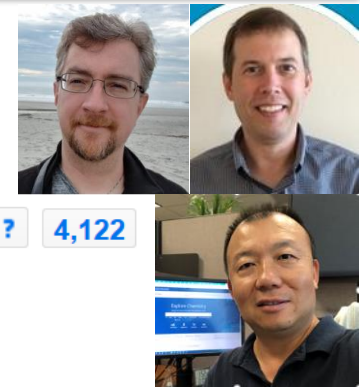
- Assessing the missing entries …





Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

Transformation Products: Filling the Data Gaps!



PubChem NORMAN Suspect List Exchange

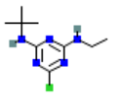
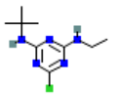
- NORMAN Suspect List Exchange Classification ? 117,037
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 4,122
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 647
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites ? 1,358
  - 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 ? 97
  - S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
  - S72 | NTUPHTW | Pharmaceutically Active Substances from
- S00 | SUSDAT | Merged NORMAN Suspect List: SusDat ?
- S01 | MASSBANK | NORMAN Compounds in MassBank EU
- S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of W
- S03 | NORMANCT15 | NORMAN Collaborative Trial Targets an
- S04 | UJIBADE | Target List from UJI used in Bade et al 2015
- S05 | KWRSJERPS | KWR Drinking Water Suspect List ?
- S06 | ITNANTIBIOTIC | Antibiotic List from the ITN MSCA ANS
- S07 | EAWAGSURF | Eawag Surfactants Suspect List ? 17
- S08 | ATHENSSUS | University of Athens Surfactants and Susp
- S09 | PFASTRIER | PFAS Suspect List of fluorinated substance

PubChem Terbuthylazine (Compound)

8.5 Transformations

Page 3 of 25 items View More Rows & Details Download

SORT BY Please Choose One

Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbuthylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine	10.5281/zenodo.382
	Terbutylazine	Deethylation		Terbutylazine-desethyl	10.1007/s13361-017-

PubChem Terbuthylazine (Compound)

Microbiocides, Algicides, Herbicides

S69 | LUXPEST | Pesticide Screening List for Luxembourg | DOI:10.5281/zenodo.3862688

▸ NORMAN Suspect List Exchange

Pesticides -> Herbicides -> Triazine herbicides -> Chlorotriazine herbicides

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

▸ NORMAN Suspect List Exchange

7.2 Agrochemical Transformations

Terbutylazine has known environmental transformation products that include Terbutylazine-2-hydroxy, Terbutylazine-desethyl, and Terbutylazine-desethyl-2-hydroxy.

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

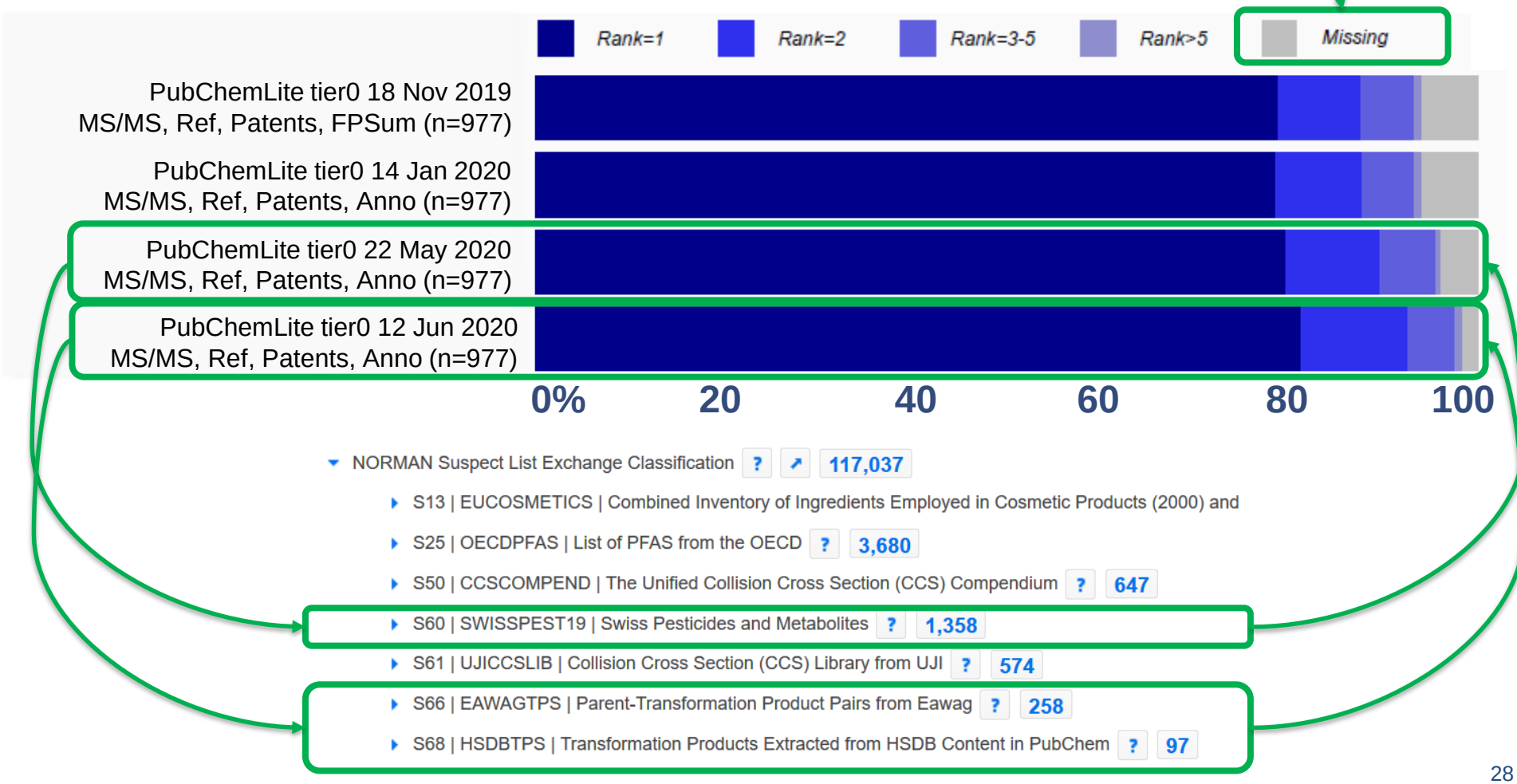
▸ NORMAN Suspect List Exchange

Terbutylazine has known environmental transformation products that include CSAA036479, CSAA04949, CSCD648241, CSCD692760, GS31398, MT1, GS 26379, MT13, GS 23158, Terbutylazine metabolite MT14, Terbutylazine metabolite MT23, and Terbutylazine metabolite MT24.

S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 | DOI:10.5281/zenodo.3544759

Missing Entries in PubChemLite

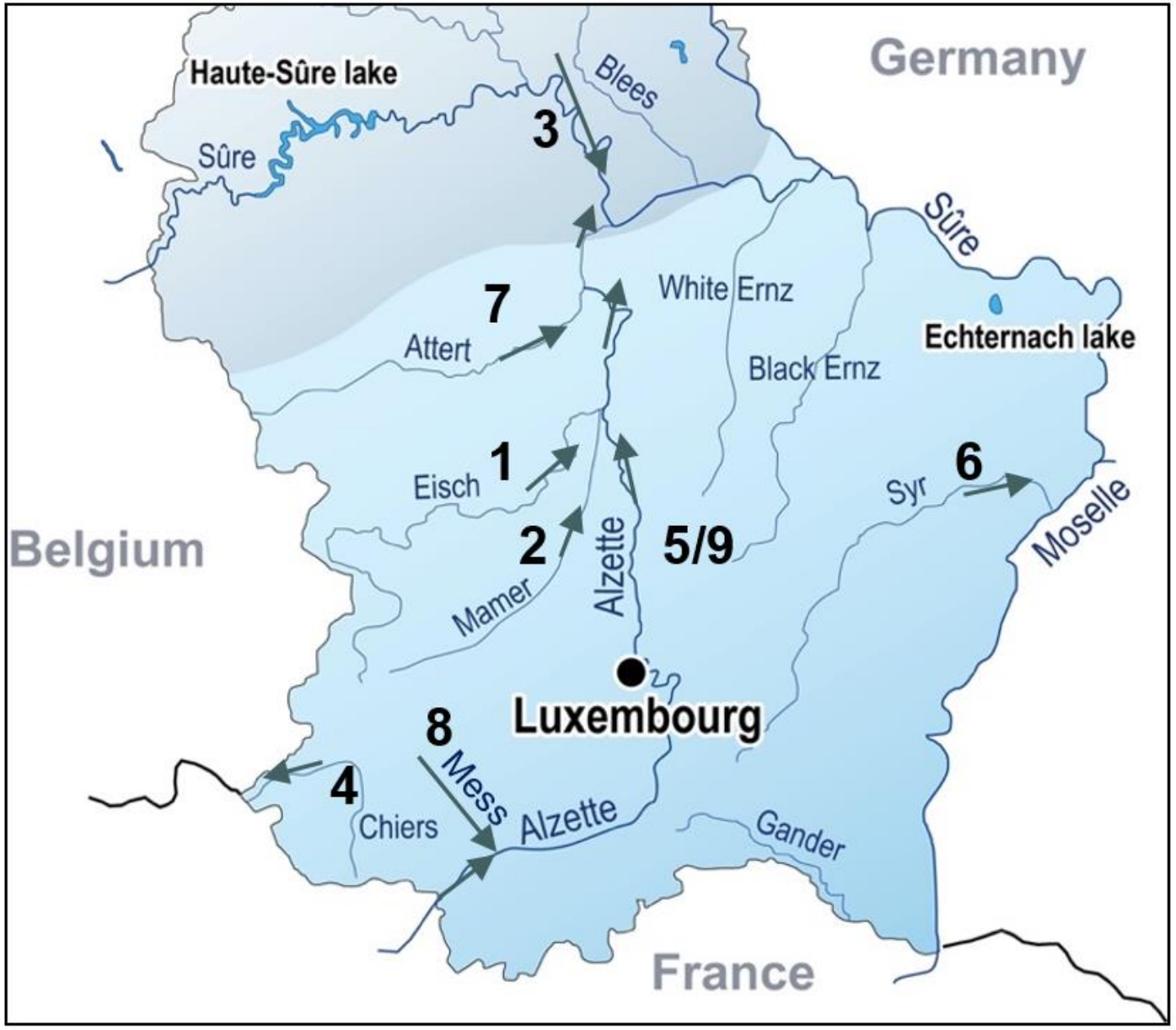
- Assessing the missing entries …



Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

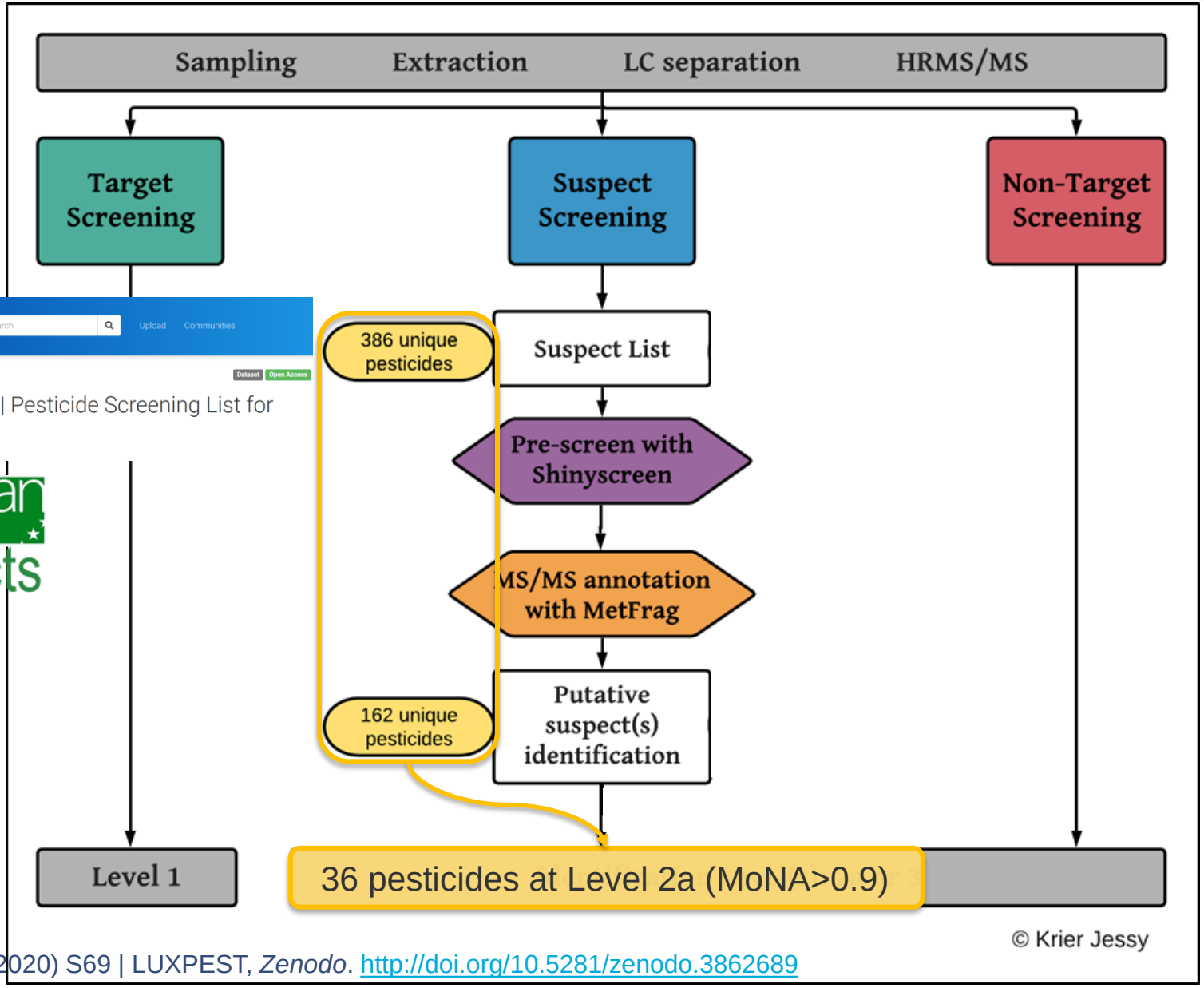
“Real Life” Application

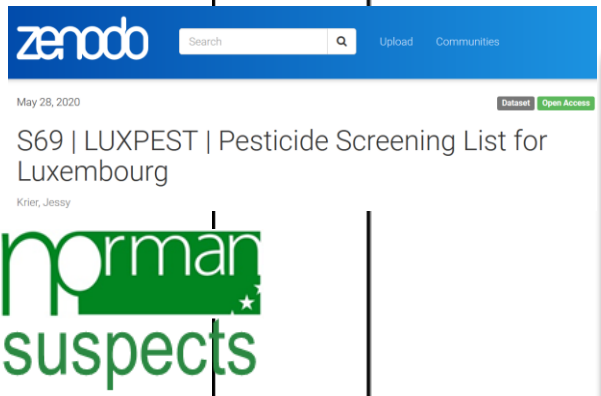
THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development



Jessy Krier (2020) Master’s Thesis, defended July 2020. Figure 34.

Suspect Screening for Lux-Relevant Pesticides



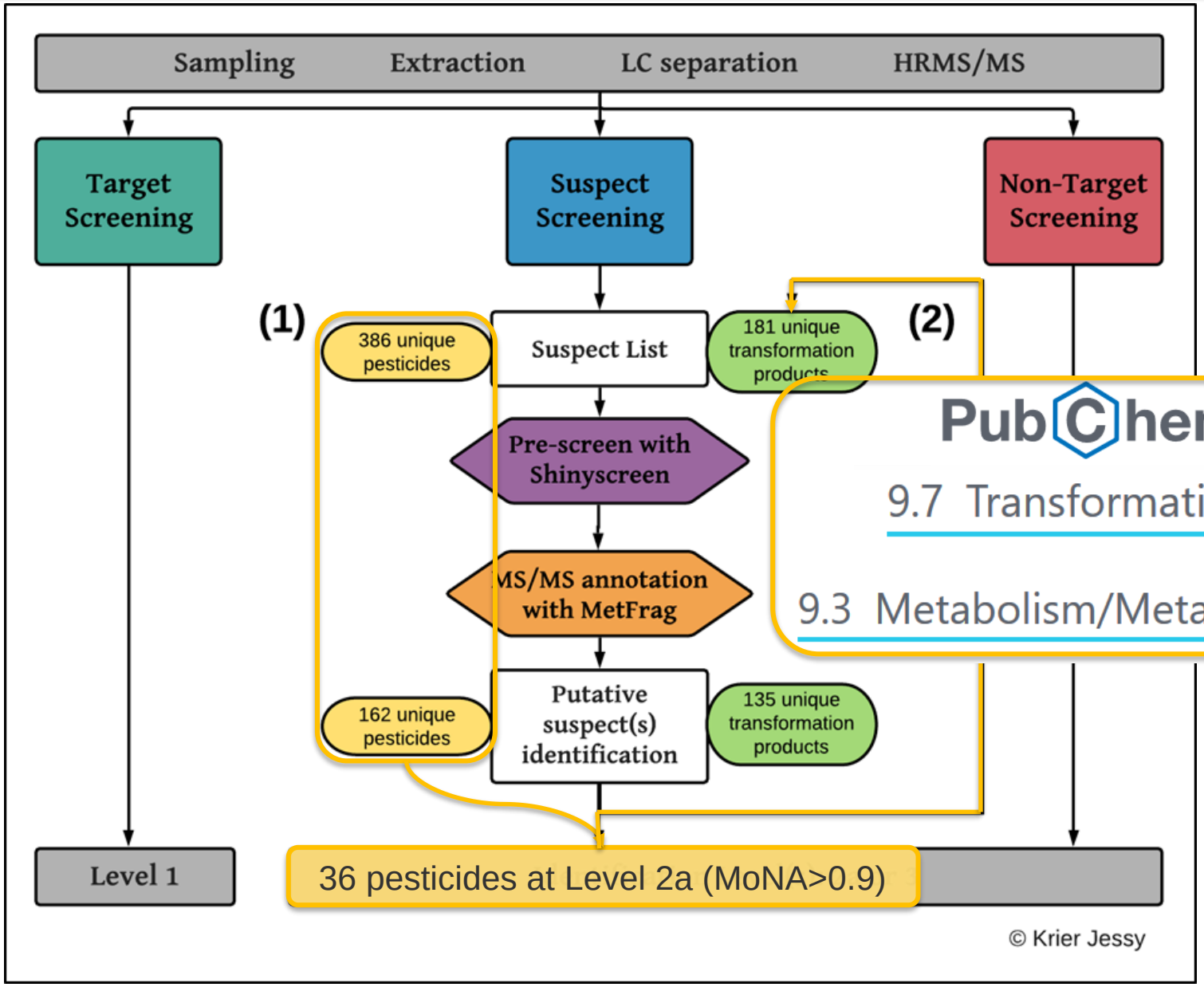


norman suspects

Jessy Krier (2020) S69 | LUXPEST, *Zenodo*. http://doi.org/10.5281/zenodo.3862689

Jessy Krier (2020) Master's Thesis, defended July 2020. Figure 8

Data-Driven TP/Metabolite Search



Jessy Krier (2020) Master's Thesis, defended July 2020. Figure 8.

Literature Mining for Metabolites / TPs – and Curation

PubChem Terbuthylazine (Compound)

8.3 Metabolism/Metabolites

Metabolism of terbuthylazine in rats is similar to other chloro-s-triazine herbicides. The major routes of metabolism are hydrolysis of the chlorine moiety and mono- or didealkylation. Hydroxylation of one or both of the dealkylated amine groups may also occur.

USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.12 (March 1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/reregistration/status.htm

▸ Hazardous Substances Data Bank (HSDB)

Urine and feces contained up to 25 and 15 identified metabolites, re
Degradation of the triazine ring did not occur. Ammeline and amme
dealkylated/hydroxylated metabolites common to all triazines, were

*USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (
1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/*

▸ Hazardous Substances Data Bank (HSDB)

In mammals, following oral administration, ...a de-ethyl metabolite fo
of products formed by oxidation of one methyl group of the tert-but

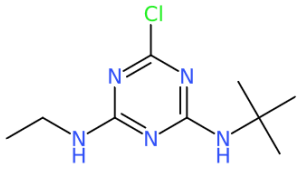
*Tomlin CDS, ed. Terbuthylazine (5915-41-3). In: The e-Pesticide Manual, Versi
Protection Council.*

▸ Hazardous Substances Data Bank (HSDB)

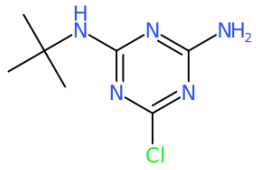
DEPICT D K

```
CCNC1=NC(=NC(=N1)Cl)NC(C)(C)C Terbutylazine CID:22206
NC1=NC(=NC(=N1)Cl)NC(C)(C)C desethyl-terbutylazine CID:108201
CCNC1=NC(=NC(=N1)Cl)N des-t-butyl-terbutylazine CID:13878
CCNC1=NC(=NC(=N1)O)NC(C)(C)C 2-hydroxy-terbutylazine CID:135495928
CCNC1=NC(=NC(=N1)Cl)NC(C)(C)C(O) (hydroxy-t-butyl)-Terbutylazine CID:779516
NC1=NC(=NC(=N1)Cl)N didealkyl-terbutylazine CID:18831
OC1=NC(=NC(=N1)Cl)N hydroxy-didesalkyl-terbutylazine CID:135438601
```

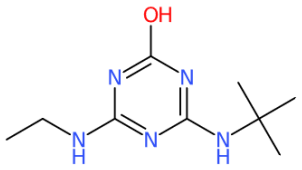
Color on White | No Annotation | Chiral Hydrogens (smart) | Do Not Abbreviate



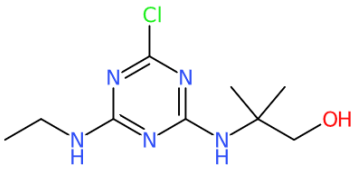
Terbutylazine CID:22206



desethyl-terbutylazine CID:108201



2-hydroxy-terbutylazine CID:135495928



(hydroxy-t-butyl)-Terbutylazine CID:779516

https://www.simolecule.com/cdkdepict/depict.html

Living Data Connections

zenodo Search Upload Communities emma.schymanski@uni.lu

June 11, 2020

Dataset Open Access

Edit

New version

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

LCSB-ECI; Krier, Jessy; Schymanski, Emma; PubChem Team; Bolton, Evan; Thiessen, Paul; Zhang, Jeff

This is the collection a
PubChem on the NOR

https://www.norman-

HSDBTPS is a list of r
HSDB (Hazardous Su
10.5281/zenodo.382

Entries automatically

Communities

LCSB Environmental ... Remove

... oup

... ist Remove

File Edit View Repository Branch Help

Current repository **pubchem** | Current branch **master** | Fetch origin Last fetched 2 minutes ago

Changes 2 | History

No branches to compare

Update extractAnnotations.R
Emma Schymanski • Jun 9, 2020

HSDB Ref Info
Emma Schymanski • Jun 8, 2020

added new CIDs to HSDBTPS
Emma Schymanski • Jun 8, 2020

Update PCLite_eval_support.R
Emma Schymanski • Jun 8, 2020

Added S69 LUXPEST
Emma Schymanski • May 28, 2020

Update PCLite_eval_support.R
Emma Schymanski • May 25, 2020

Update user_PCLite_eval.R
Emma Schymanski • May 20, 2020

Update PCLite_eval_support.R

added new CIDs to HSDBTPS

Emma Schy

Added newly r

annotations\tp

annotations\tp

annotations\tp

...\S68_HSDBTI

annotations\tp

Preview

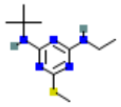
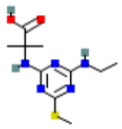
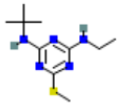
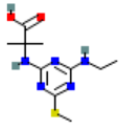
Predecessor_CID
13450
13450

8.5 Transformations

19 items View More Rows & Details

Download

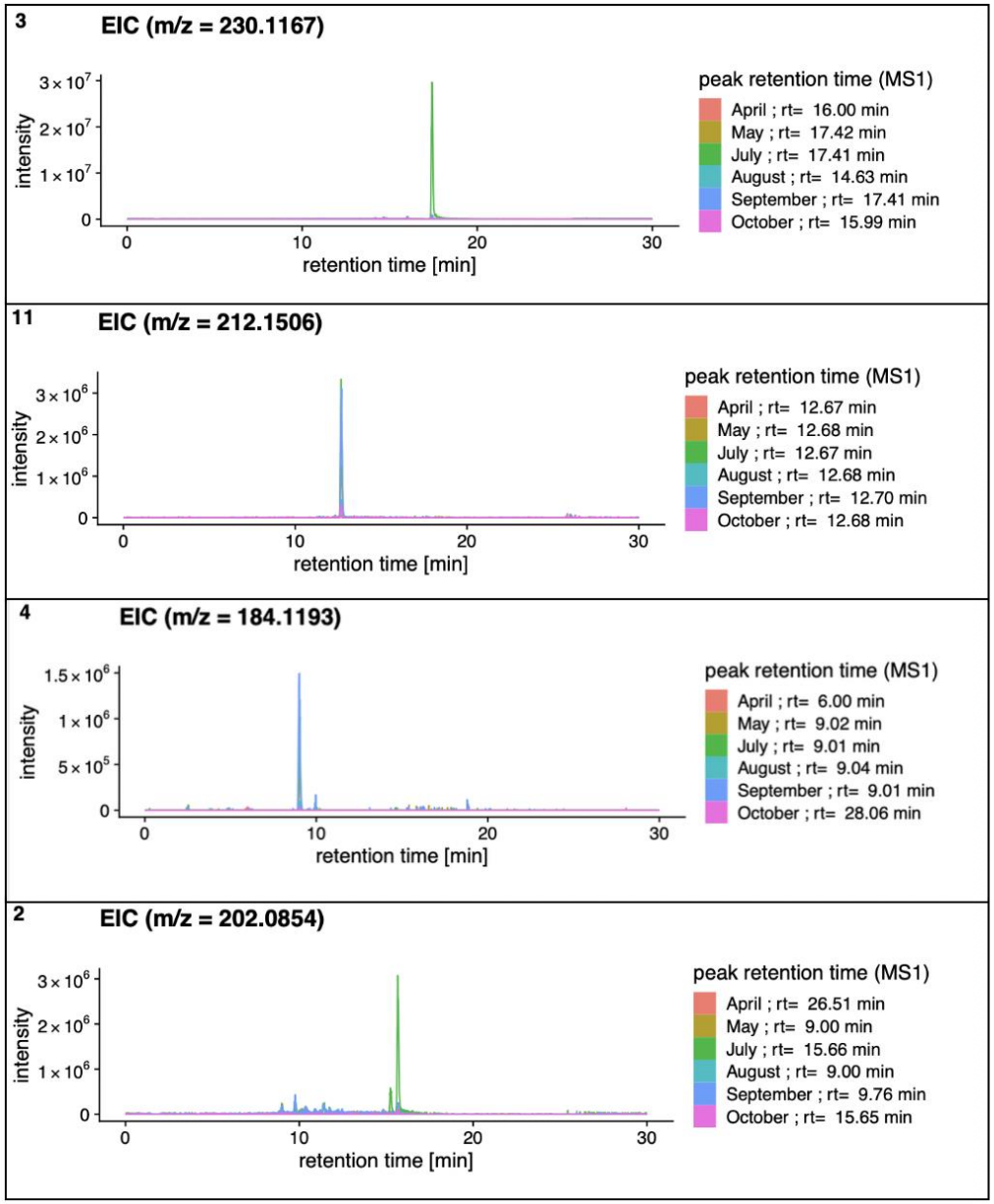
SORT BY Please Choose One

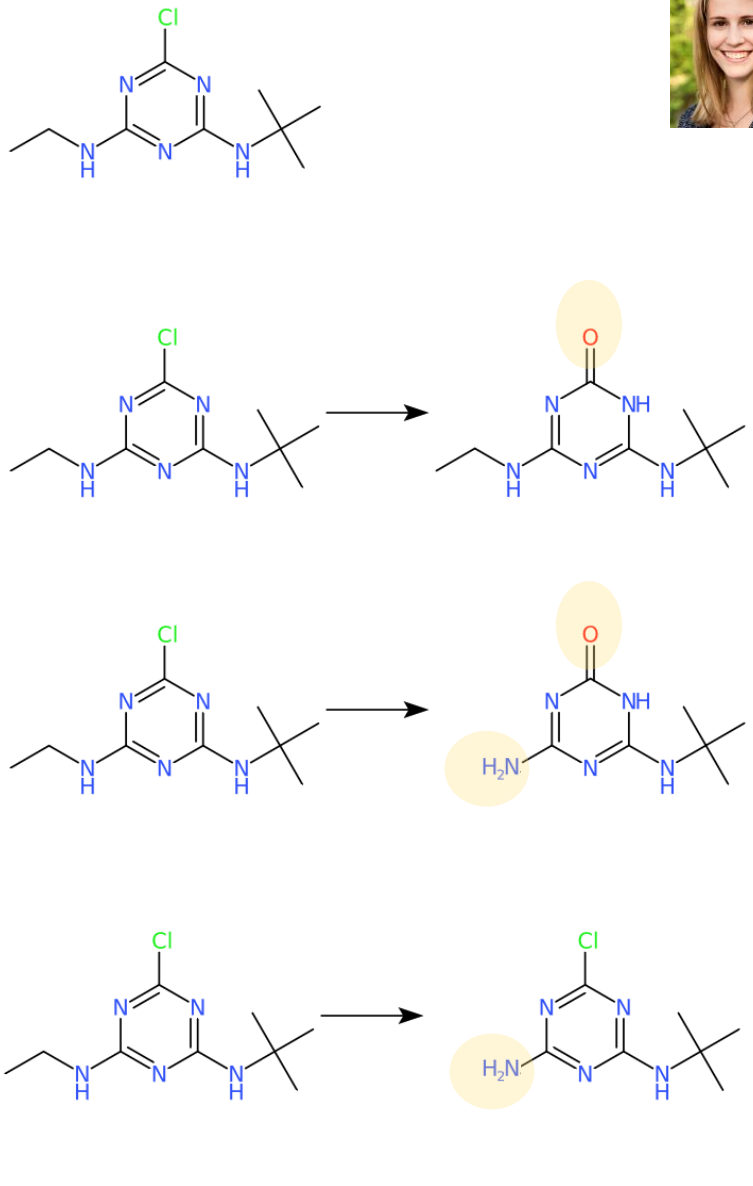
Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.1002/bms.1200506
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.5281/zenodo.38274

https://git-r3lab.uni.lu/eci/pubchem/

LCSB-ECI & PubChem Team. DOI 10.5281/zenodo.3890392

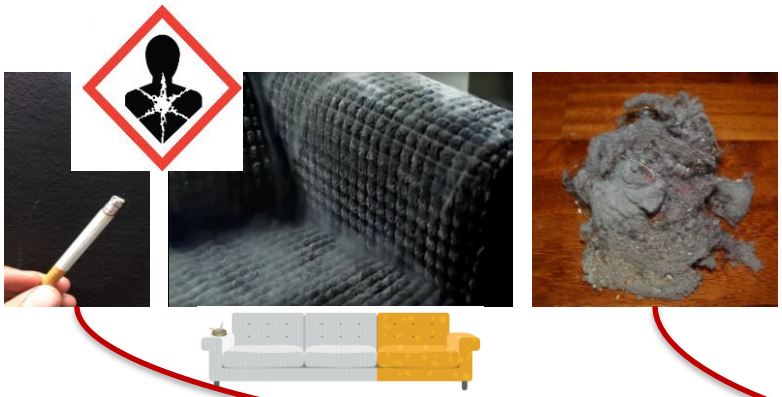
Pesticides & TPs over time (tent. IDs)



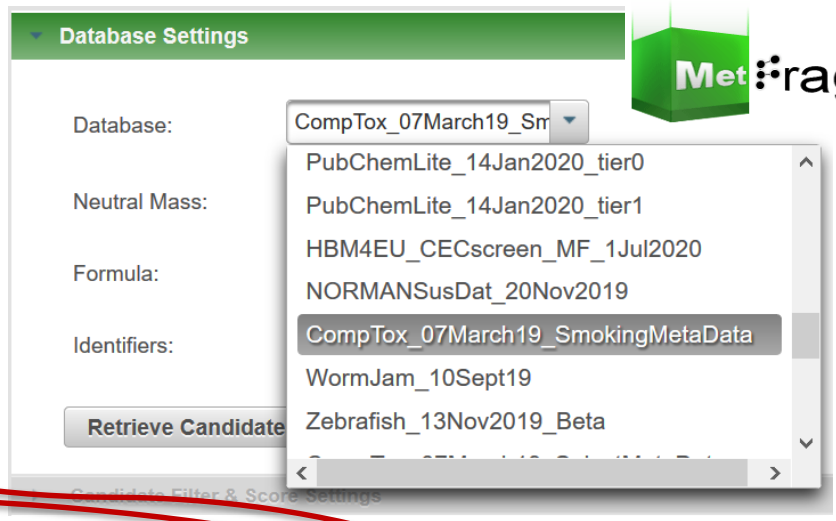


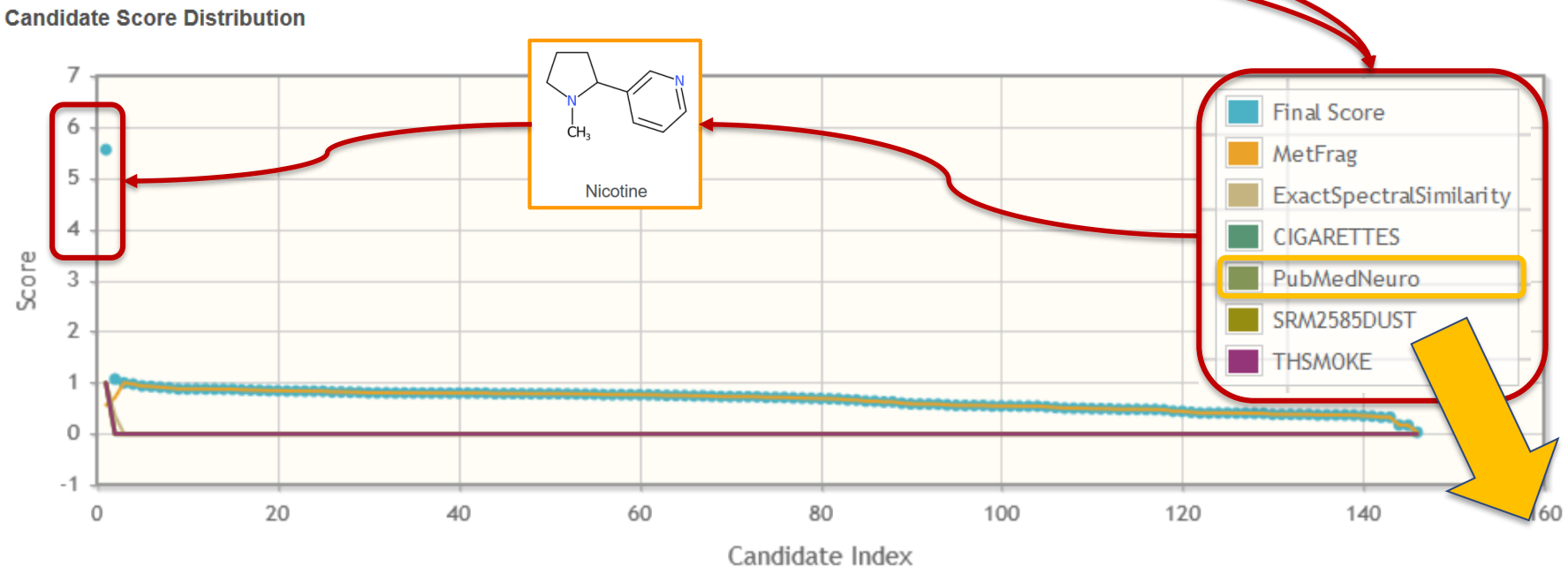
Jessy Krier (2020) Master's Thesis, defended July 2020. Modified from Figure 33.

More Examples? Thirdhand Smoke (THS) in Dust



+



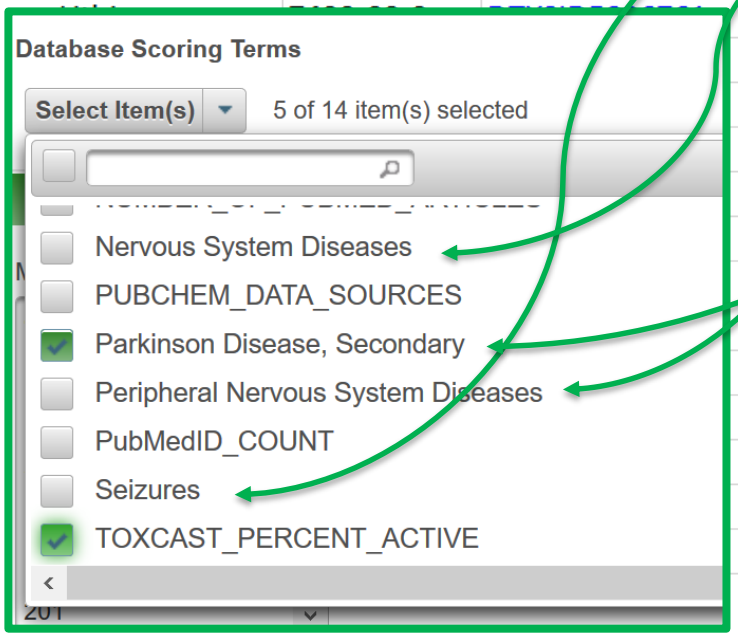


Schymanski, EL, Torres, S, & Ramirez, N. (2020). Thirdhand Smoke in House Dust. Zenodo. http://doi.org/10.5281/zenodo.3613472

New MetaData: Disease-Specific Reference Counts

S37 LitMinedNeuro: DOI: 10.5281/zenodo.3242298

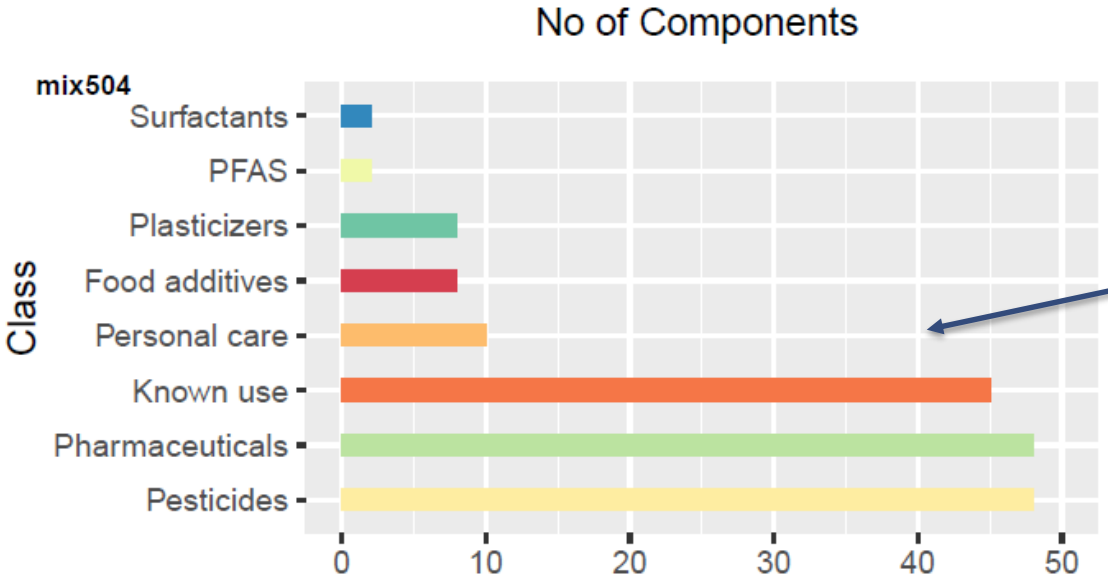
Chemical	CAS RN	DSSToxID	PMID Ct	Seizures	Nervous System Diseases	Peripheral Nervous System Diseases	Brain Diseases	Muscular Diseases	Basal Ganglia Diseases	Parkinson Disease, Secondary	Coma	Hallucinations	Tremor	Memory Disorders	Central Nervous
Cisplatin	15663-27-1	DTXSID4024983	1032	20	47	140	13	0	4	1	1	0	1	2	4
Ethanol	64-17-5	DTXSID9020584	768	100	23	11	18	26	1	3	20	6	17	54	2
Lead	7439-92-1	DTXSID2024161	740	28	107	68	102	4	2	2	1	3	4	19	30
			689	30	50	9	22	5	36	13	25	6	93	12	15
			666	32	10	3	65	6	10	18	45	5	18	4	2
			638	1	24	0	11	0	6	289	0	0	5	0	1
			567	17	59	125	15	5	1	1	5	3	2	1	8
			560	37	24	25	16	9	3	1	9	3	8	4	6
			555	6	6	1	10	6	153	51	4	4	11	1	0
			530	151	16	0	8	0	2	3	3	8	6	12	11
			489	8	3	0	3	2	2	0	9	4	1	0	5
			485	4	43	217	9	14	0	0	0	0	0	1	2
			477	13	41	1	105	4	0	0	1	0	1	13	12
			464	150	26	15	3	2	0	0	8	4	6	2	10
			451	17	25	1	79	4	0	1	5	0	1	9	18
			450	6	79	22	23	2	3	5	2	2	38	7	25

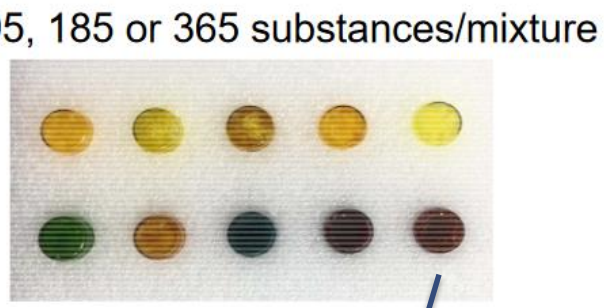


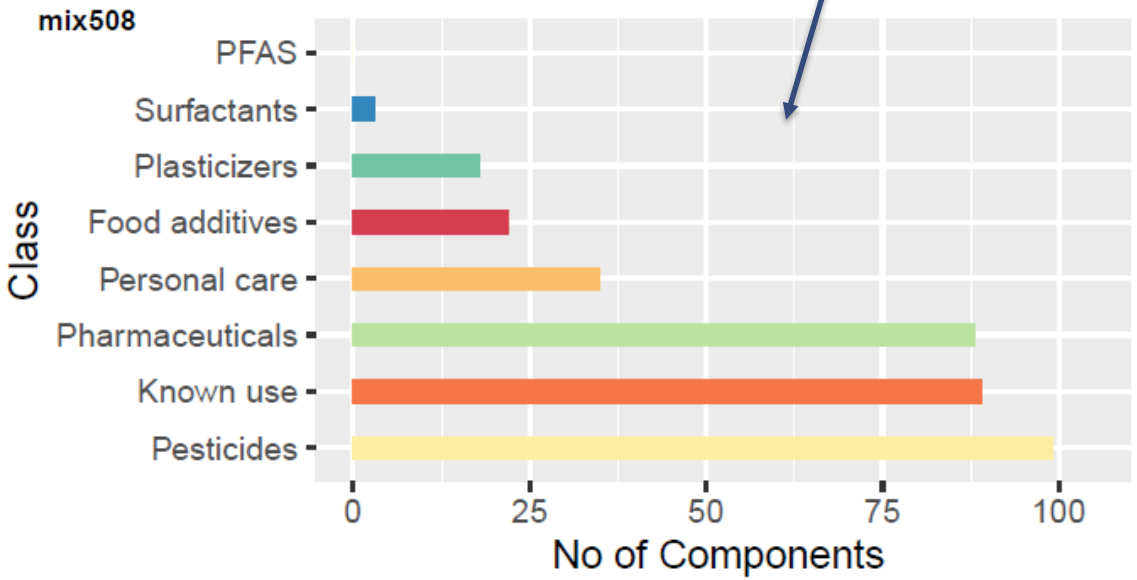
Schymanski *et al.* DOI: 10.1039/C9EM00068B (Perspective) Environ. Sci.: Processes Impacts, 2020.

Future Perspectives

- Rapid classification / interpretation of large collections







#BioHackEU20

BIOHACKATHON-EUROPE

ENTACT: Elin Ulrich *et al.* 2018, https://www.epa.gov/sites/production/files/2018-06/documents/comptox_cop_6-28-18.pdf and

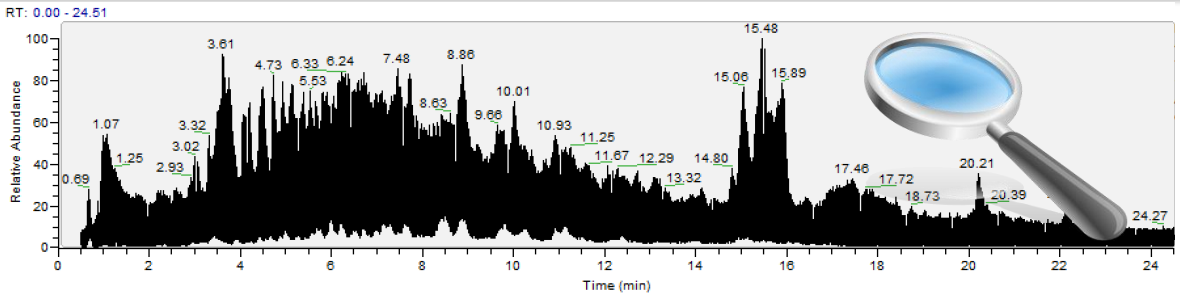
Ulrich, EM, et al. (2019) Anal Bioanal Chem. DOI: 10.1007/s00216-018-1435-6

Classification Figures:

Anjana Elapavalore, ECI (Master's thesis)

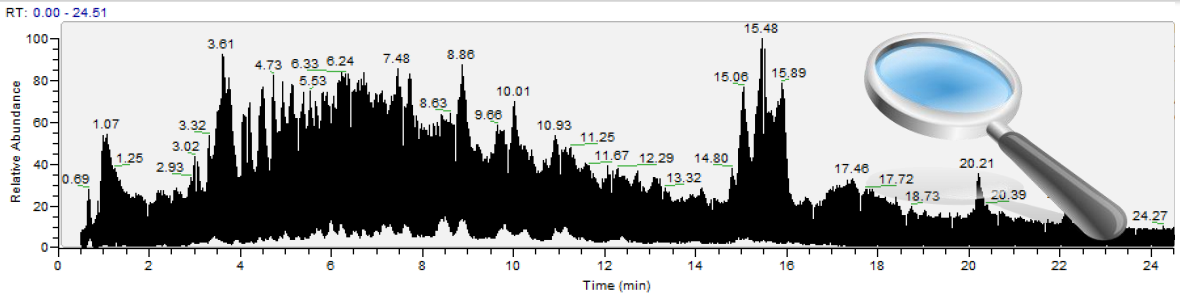
Take Home Messages

- Challenge:
  Increase ***% identified***
  Improve ***interpretation***

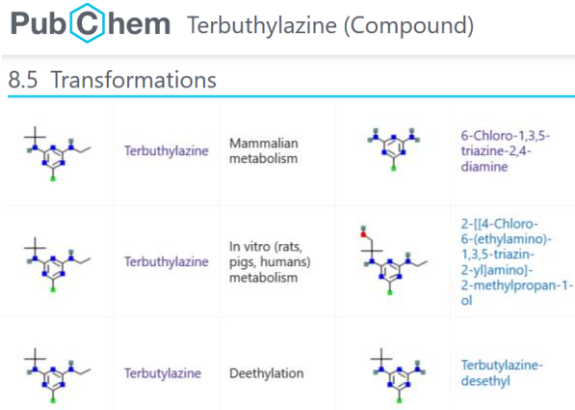


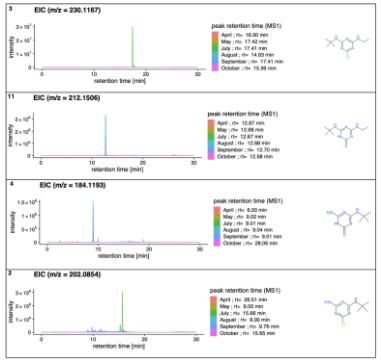
Take Home Messages

- Challenge:
  Increase ***% identified***
  Improve ***interpretation***



- ***"Digital Detective Work" is …***
  - Capturing expert knowledge
    Machine and ***human*** readable
  - ***Connecting*** this to environmental observations
  - Identifying and ***closing knowledge gaps***
  - Supporting ***interpretation*** of complex data

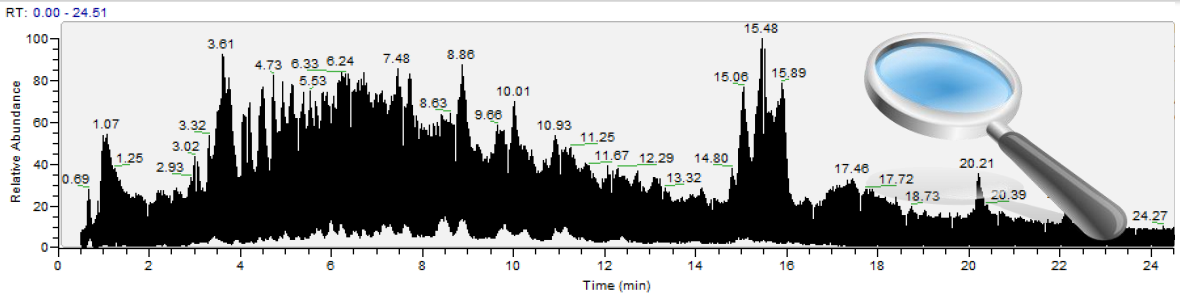




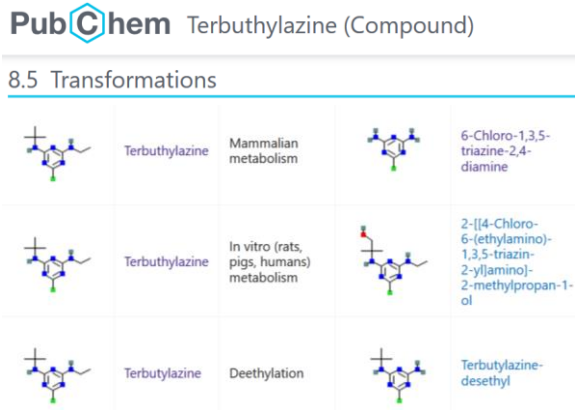
More? Check out our presentations at: https://zenodo.org/communities/lcsb-eci/

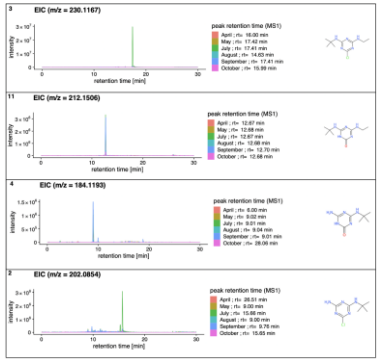
Take Home Messages

- Challenge:
  Increase ***% identified***
  Improve ***interpretation***



- ***"Digital Detective Work" is …***
  - Capturing expert knowledge
    Machine and ***human*** readable
  - ***Connecting*** this to environmental observations
  - Identifying and ***closing knowledge gaps***
  - Supporting ***interpretation*** of complex data





- Finally … information in the public domain helps ***everybody***
  - You never know when it will help you ☺

More? Check out our presentations at: https://zenodo.org/communities/lcsb-eci/

Acknowledgements

Open Data for Chemistry

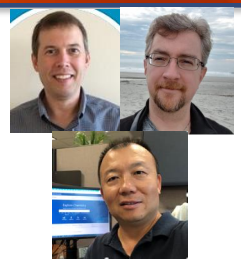
ROYAL SOCIETY OF CHEMISTRY

Luxembourg National Research Fund

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

PubChem

THE GOVERNMENT OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate and Sustainable Development



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

Christoph Ruttkies, Steffen Neumann, MetFrag Team

Simone Witzmann, Vero Briche, Rick Helmus, Jessy Krier Todor Kondic, Emma Schymanski, Anjana Elapavalore, Hiba Mohammed Taha; Mira Narayanan, Corey Griffith, Lorenzo Favilli, German Preciat, Adelene Lai, Randolph Singh Begoña Talavera

Philippe Diderich (eau.etat.lu)

R3

elixir LUXEMBOURG

#BioHackEU20

{BH} BIOHACKATHON-EUROPE

emma.schymanski@uni.lu and @ESchymanski

Further Information:

https://pubchem.ncbi.nlm.nih.gov/

https://git-r3lab.uni.lu/eci/shinyscreen

https://git-r3lab.uni.lu/eci/pubchem/

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

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

norman

IPB

LCSB

uni.lu UNIVERSITÉ DU LUXEMBOURG

Slides available under DOI: 10.5281/zenodo.4266449


SEVENTH FRAMEWORK PROGRAMME
eawag
aquatic research
Community Efforts!
modelkey
solutions
MassBank consortium
ECHIDNA
ART2 CURE