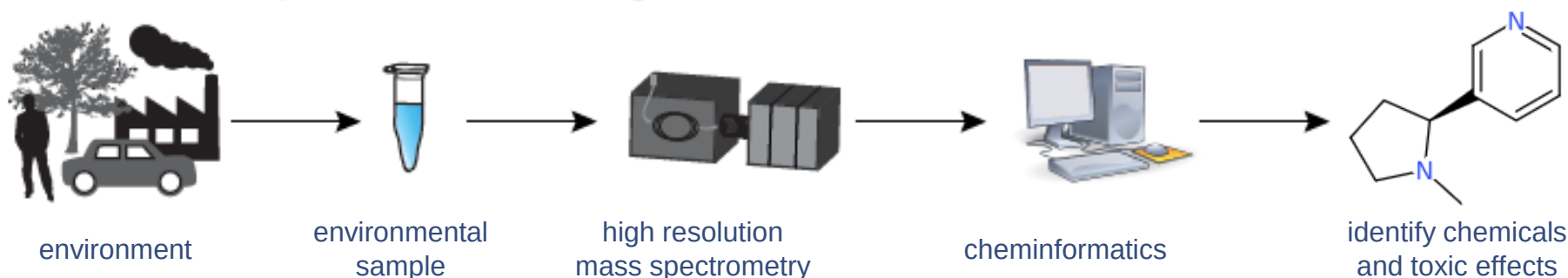


Digital Detective Work: Connecting Cheminformatics, Mass Spectrometry and our Environment



Assoc. Prof. Dr. Emma L. Schymanski¹ & Evan E. Bolton², PhD

¹FNR ATTRACT Fellow and PI in Environmental Cheminformatics
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
Email: emma.schymanski@uni.lu and [@ESchymanski](https://twitter.com/ESchymanski)

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health,
Bethesda, MD 20894, USA. Email: bolton@ncbi.nlm.nih.gov

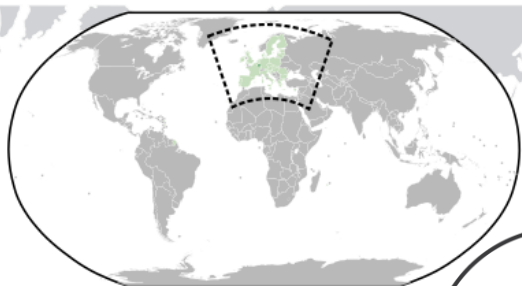


Luxembourg National
Research Fund



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

PubChem



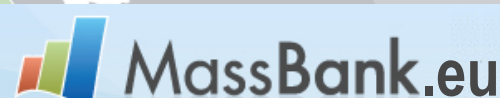
UNIVERSITY OF AMSTERDAM



eawag
aquatic research

756 Data Sources

[Explore Data Sources >](#)



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



UNIVERSITAT
ROVIRA I VIRGILI



solutions



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

COLUMBIA
MAILMAN SCHOOL
OF PUBLIC HEALTH



Maastricht University



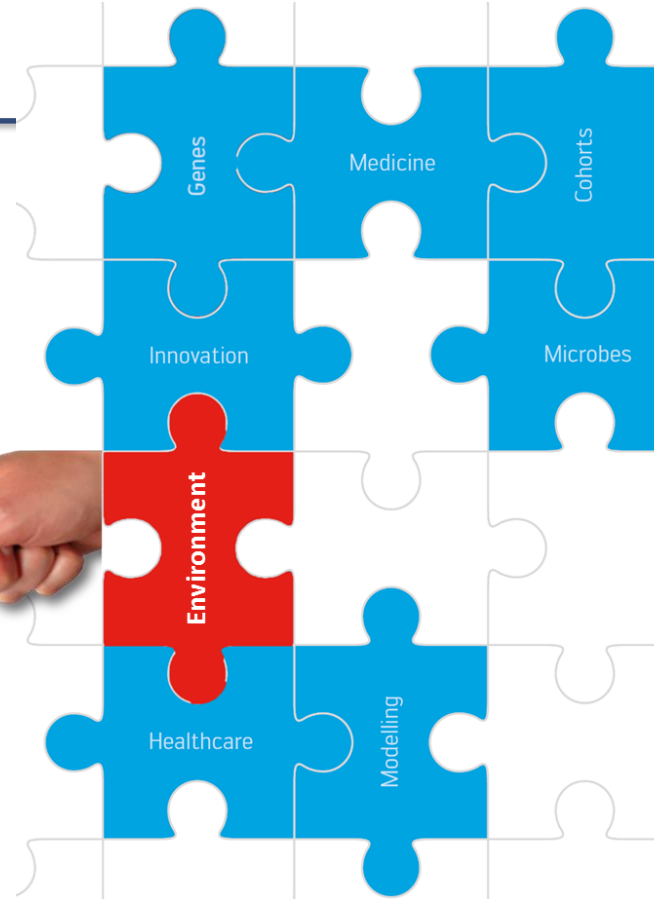
National and Kapodistrian
UNIVERSITY OF ATHENS

UNIVERSITY OF
COPENHAGEN



MEDIZINISCHE
UNIVERSITÄT
INNSBRUCK

Environmental Cheminformatics @ Luxembourg Centre for Systems Biomedicine



Luxembourg National
Research Fund

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



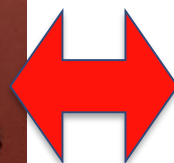
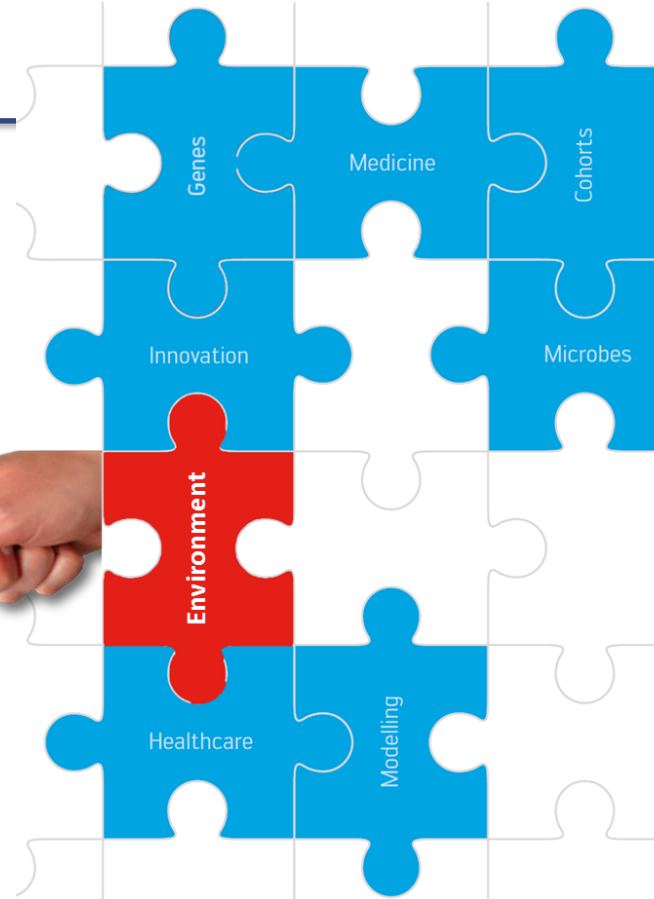
Mira Narayanan @mira.narayanan
Given access 4 weeks ago



Randolph Singh @randolph.singh
Given access 2 months ago



Todor Kondić @todor.kondic
Given access 2 months ago

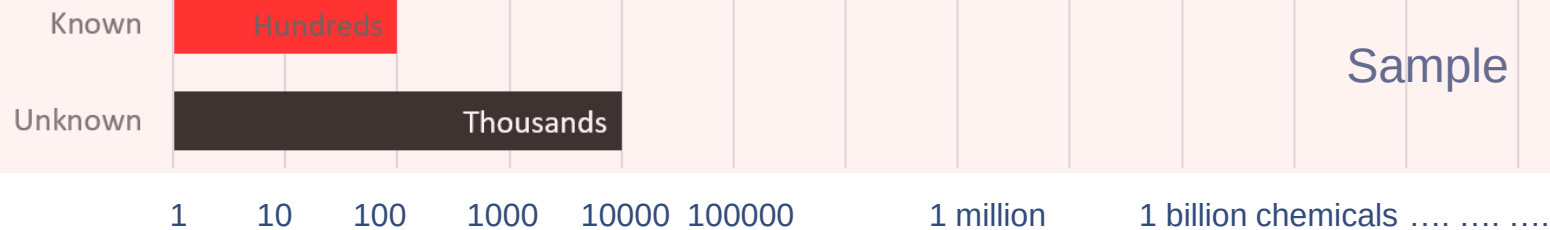
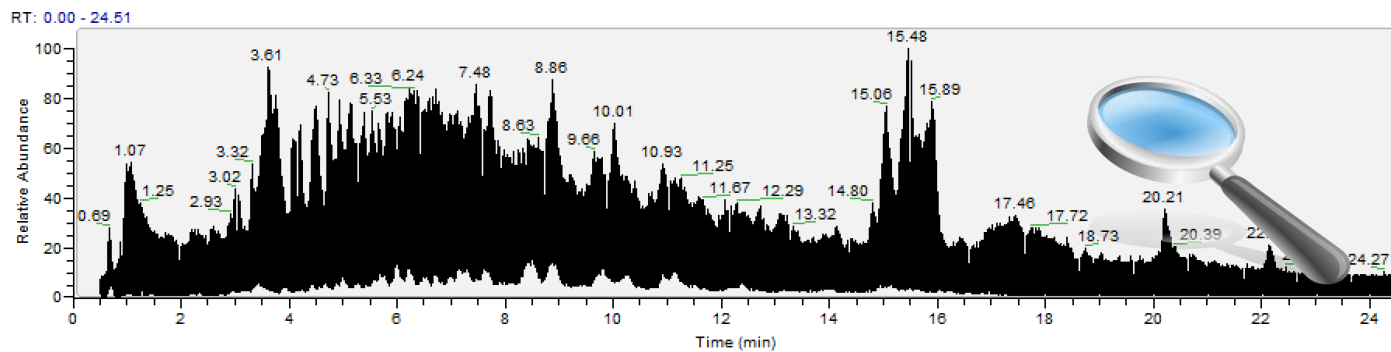


Luxembourg National
Research Fund



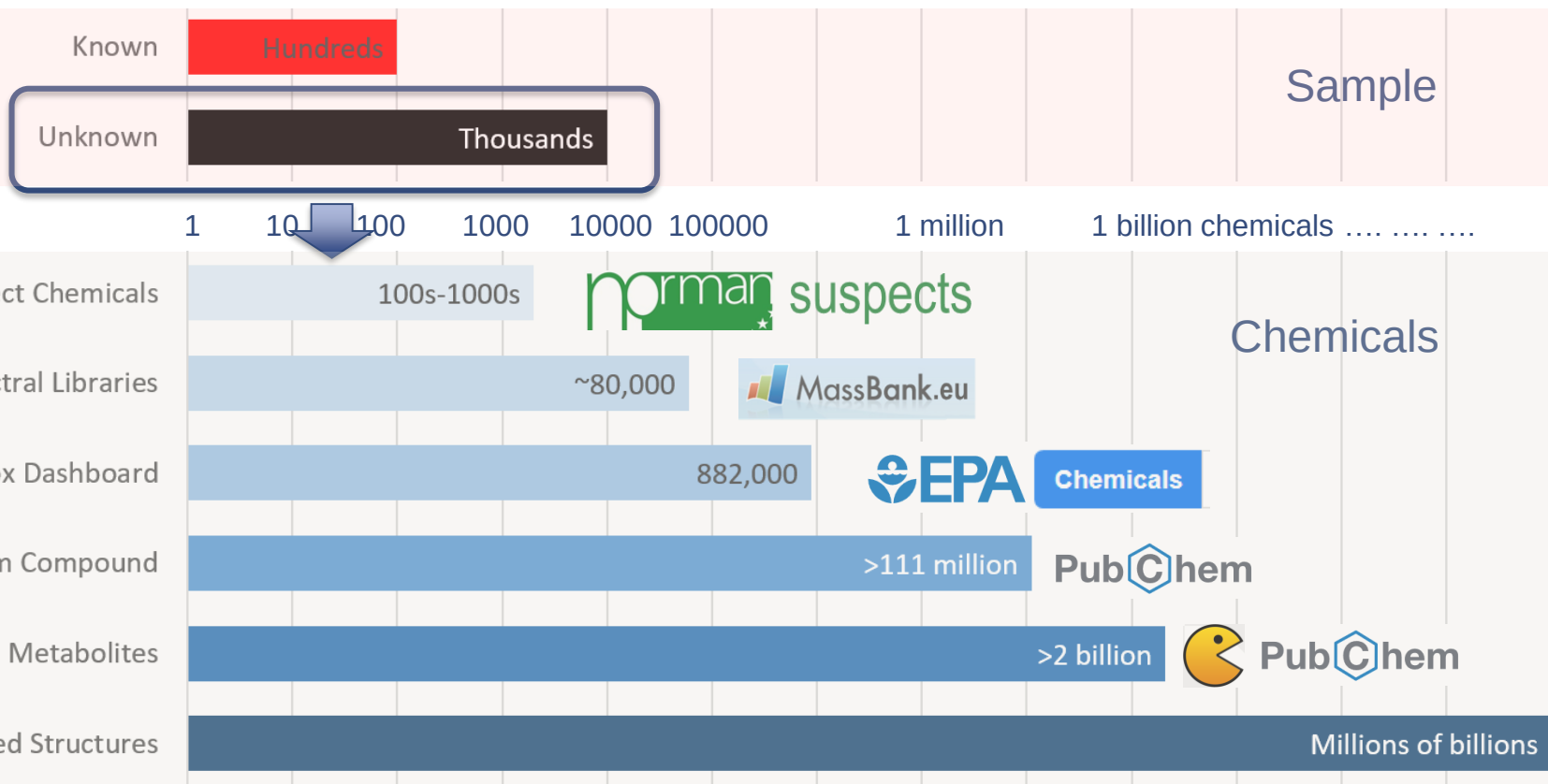
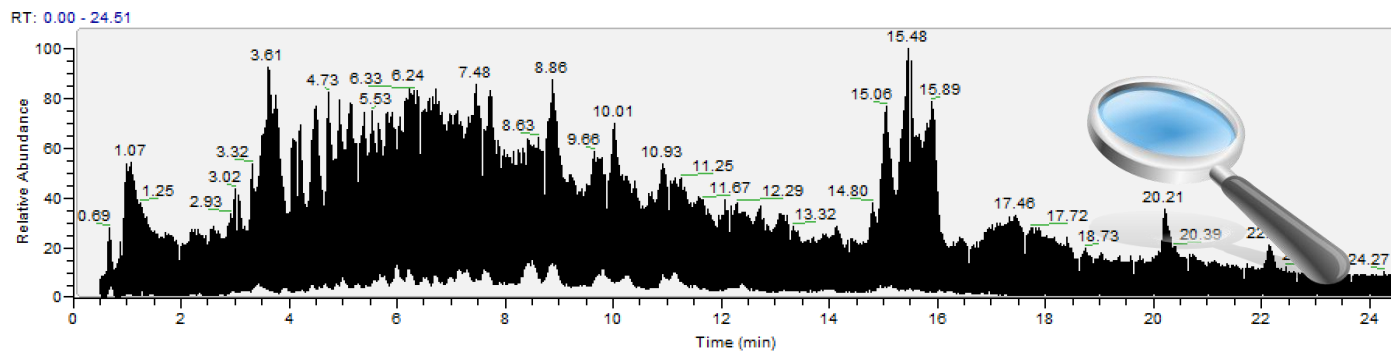
Our (Community) Challenge: Identifying Chemicals

High resolution
mass spectrometry



Our (Community) Challenge: Identifying Chemicals

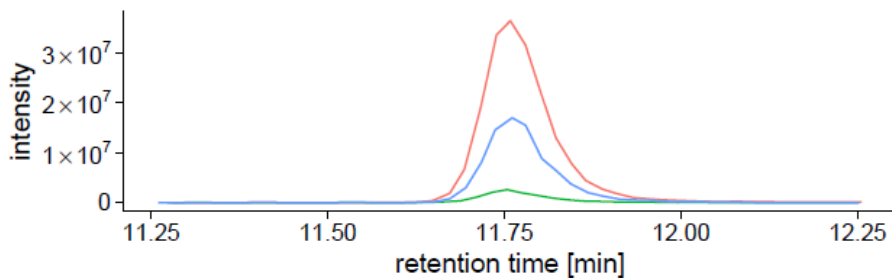
High resolution
mass spectrometry
AND connecting
chemical knowledge



Our Toolkit: ShinyScreen – Extracting Good MS Data

<https://git-r3lab.uni.lu/eci/shinyScreen>

100 **EIC (m/z = 182.0816)**

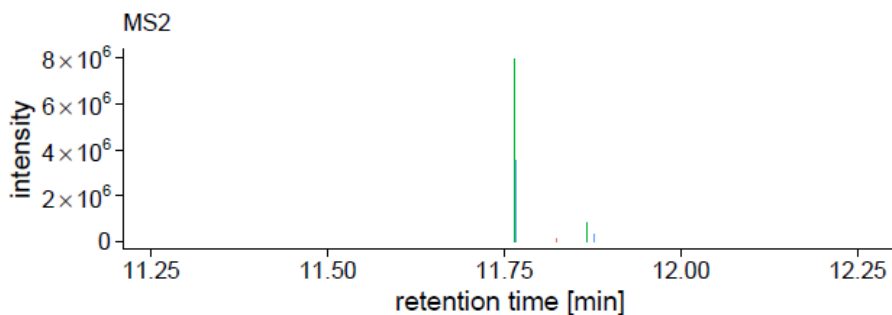


peak retention time (MS1)

Std ; rt= 11.76 min
KO ; rt= 11.75 min
WT ; rt= 11.76 min

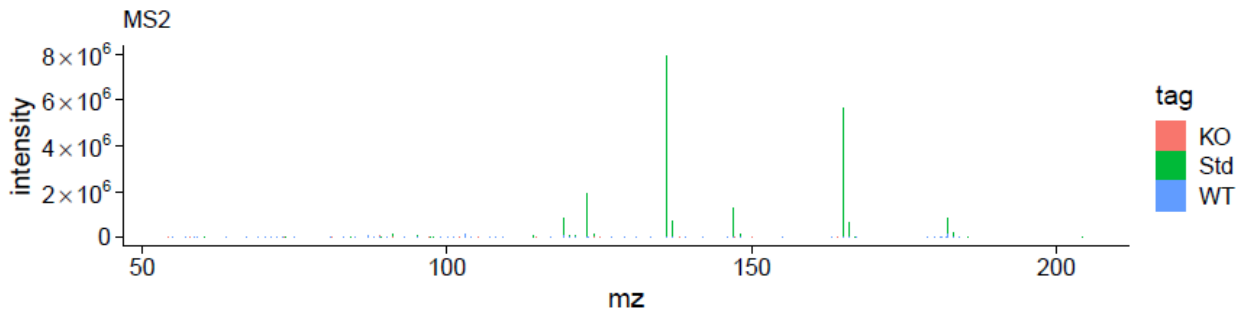


Anjana Elapavalore, Mira Narayanan,
Todor Kondic, Jessy Krier,
Hiba Mohammed Taha.



peak retention time (MS2)

KO ; rt= 11.82 min
Std ; rt= 11.76 min
WT ; rt= 11.77 min



tag

KO
Std
WT



Our Toolkit: NORMAN-SLE: Capturing Expert Knowledge

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

NORMAN SUBSTANCE DATABASE

NORMAN Suspect List Exchange – NORMAN SLE

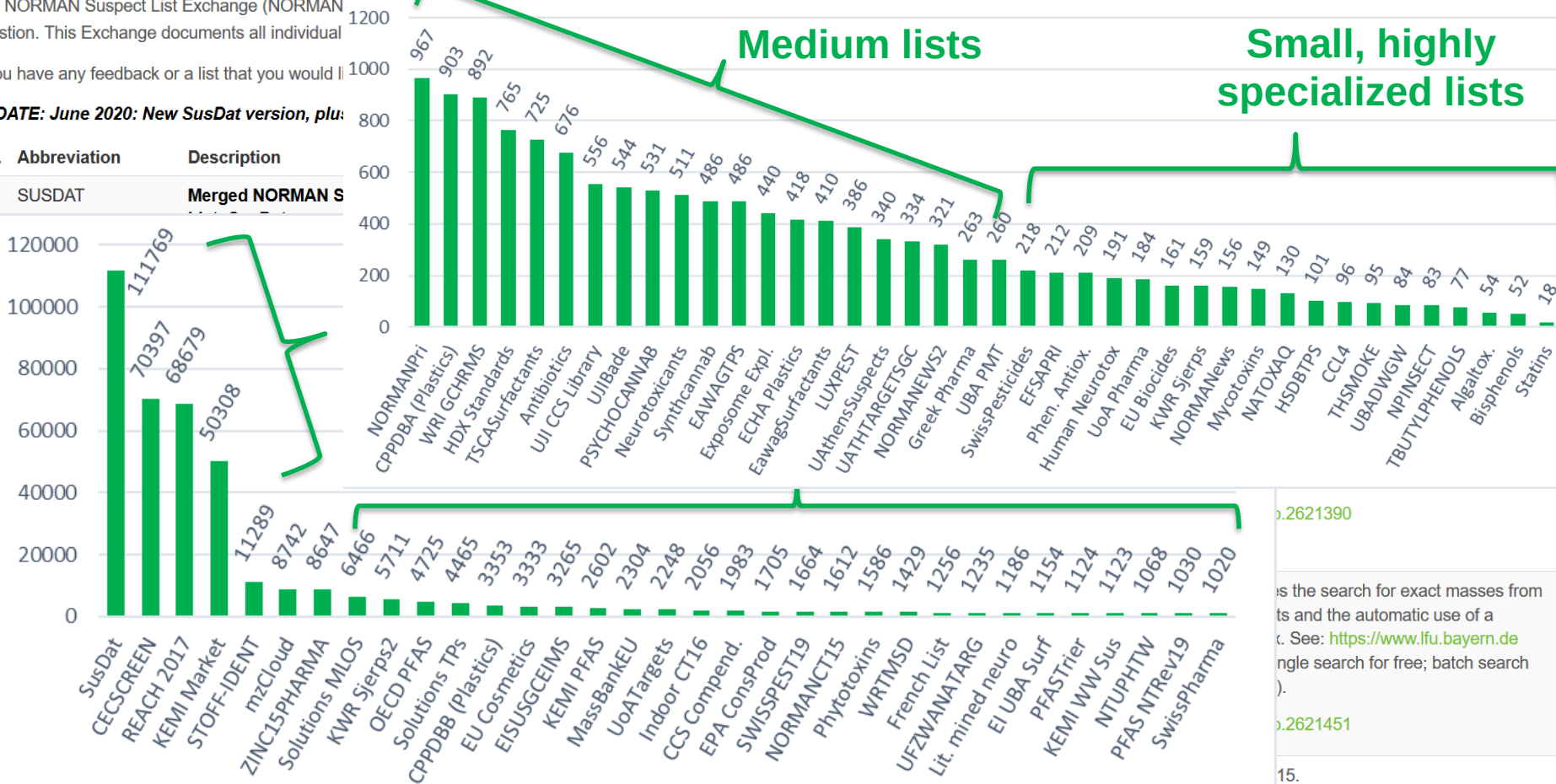
> 73 lists
> 144,980 substances



The NORMAN Suspect List Exchange (NORMAN question. This Exchange documents all individual If you have any feedback or a list that you would li

UPDATE: June 2020: New SusDat version, plu

No.	Abbreviation	Description
S0	SUSDAT	Merged NORMAN S



0.2621390

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

0.2621451

15.
0-015-8681-7

Our Toolkit: PubChem – Finding the “known unknowns”

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

The screenshot shows the PubChem website homepage. At the top, the NIH logo and text "National Library of Medicine National Center for Biotechnology Information" are displayed. Below this is the PubChem logo and navigation links: "About", "Blog", "Submit", and "Contact". The main heading is "Explore Chemistry" with the subtitle "Quickly find chemical information from authoritative sources". A yellow banner highlights "Browse COVID-19 data available in PubChem". A search bar is present with a magnifying glass icon. Below the search bar, a row of suggestions includes "Try aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3". Below this is a filter bar with "Use Entrez" (checked), "Compounds", "Substances", and "BioAssays". Four icons represent "Draw Structure", "Upload ID List", "Browse Data", and "Periodic Table". At the bottom, statistics are listed: "111M Compounds", "287M Substances", "273M Bioactivities", "31M Literature", and "756 Data Sources". Links for "See More Statistics" and "Explore Data Sources" are provided.

NIH National Library of Medicine
National Center for Biotechnology Information

PubChem About Blog Submit Contact

Explore Chemistry

Quickly find chemical information from authoritative sources

Browse COVID-19 data available in PubChem X

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

☐ Use Entrez ☒ Compounds ☐ Substances ☐ BioAssays

Draw Structure Upload ID List Browse Data Periodic Table

111M Compounds 287M Substances 273M Bioactivities 31M Literature 756 Data Sources

[See More Statistics >](#) [Explore Data Sources >](#)

Our Toolkit: MetFrag – Annotating/Identifying Masses



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra



Database Settings

Database: Include references: ☒

Parent Ion:

Neutral Mass: Search ppm:

Formula:

Identifiers:

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates ☒

MS/MS Peak list

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

Status: 2010

m/z $[M-H]^-$

213.9637

± 5 ppm

5 ppm

0.001 Da

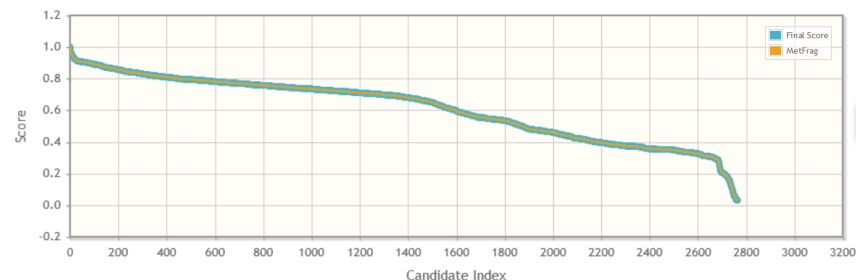
PubChem



Ranked Candidates

Statistics

Candidate Score Distribution



MS/MS

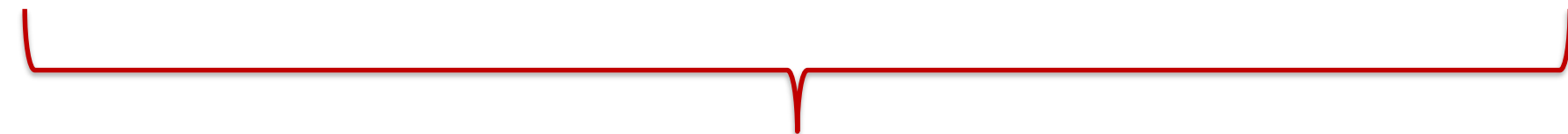
134.0054	339689
150.0001	77271
213.9607	632466

Challenge: MetFrag + PubChem + MS/MS only

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

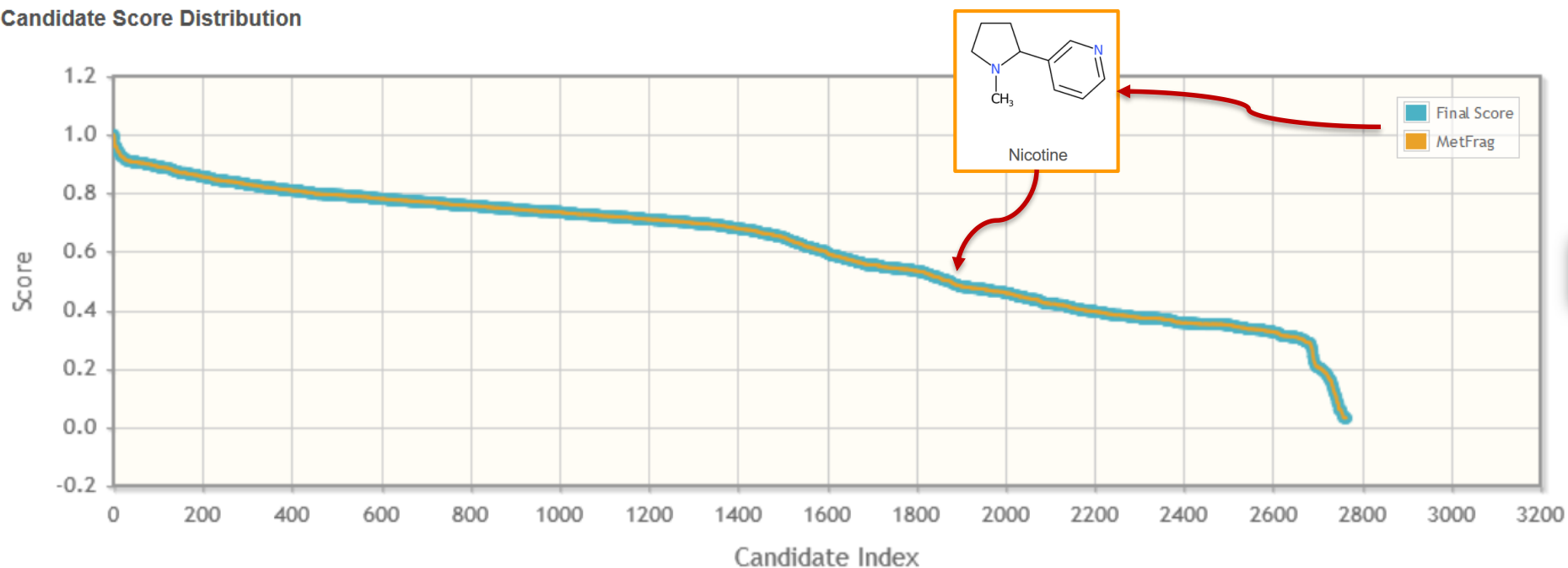


MS/MS does not provide sufficient information alone...



Statistics

Candidate Score Distribution



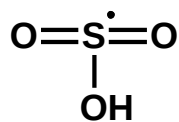
MetFrag – MS/MS and MORE!

Status: 2016

m/z $[M-H]^-$
 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

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



Suspect Lists

?TOFF IDENT

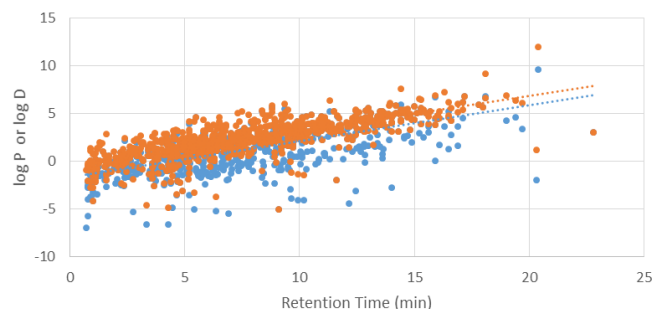
norman suspects

EPA Chemicals

MS/MS

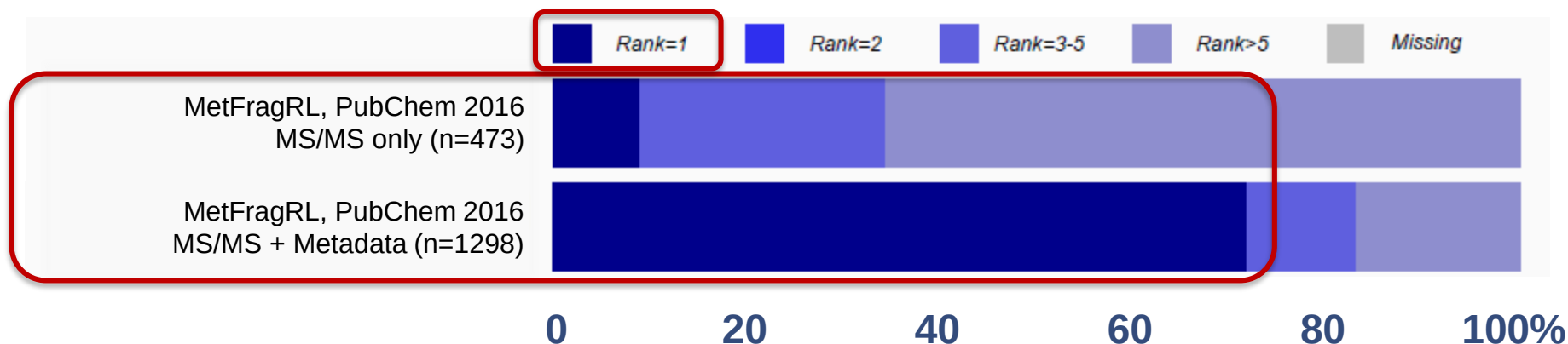
134.0054	339689
150.0001	77271
213.9607	632466

MoNA
 MassBank of North America
 MassBank.eu

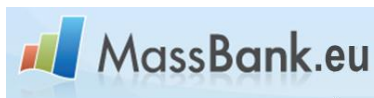


MetFrag + PubChem + MS/MS + Metadata

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

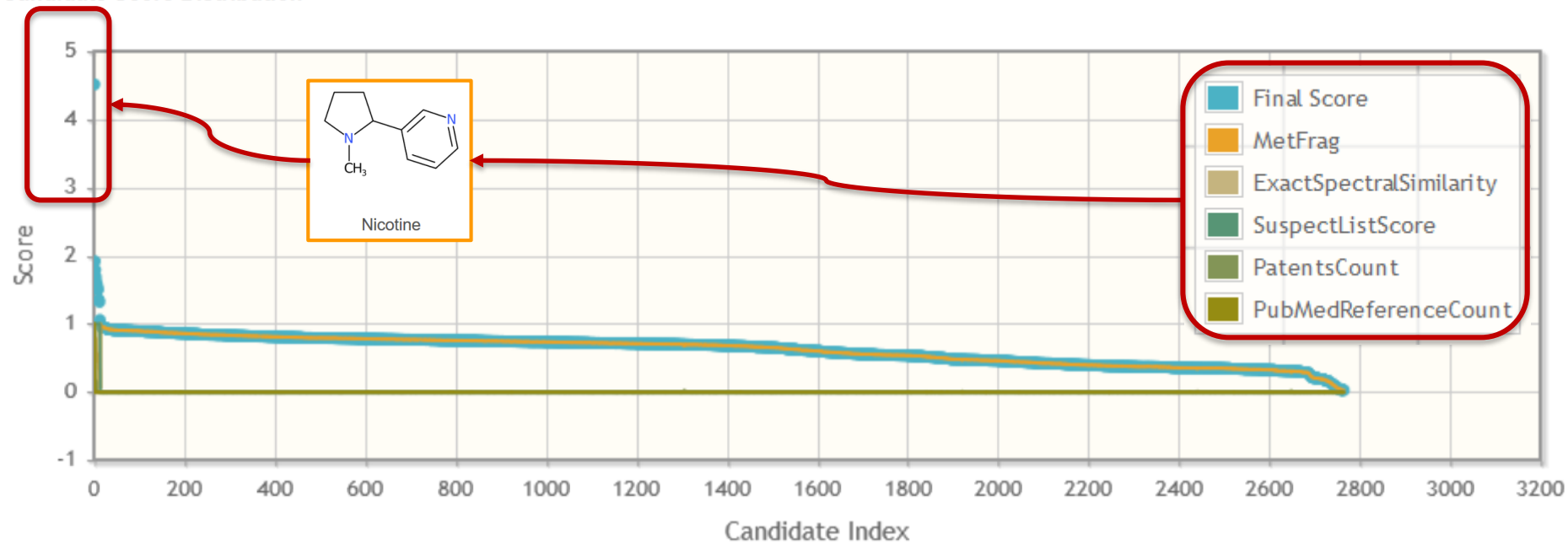


Connecting multiple lines of evidence for identification

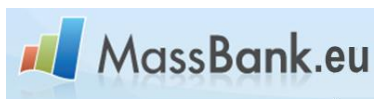


Statistics

Candidate Score Distribution



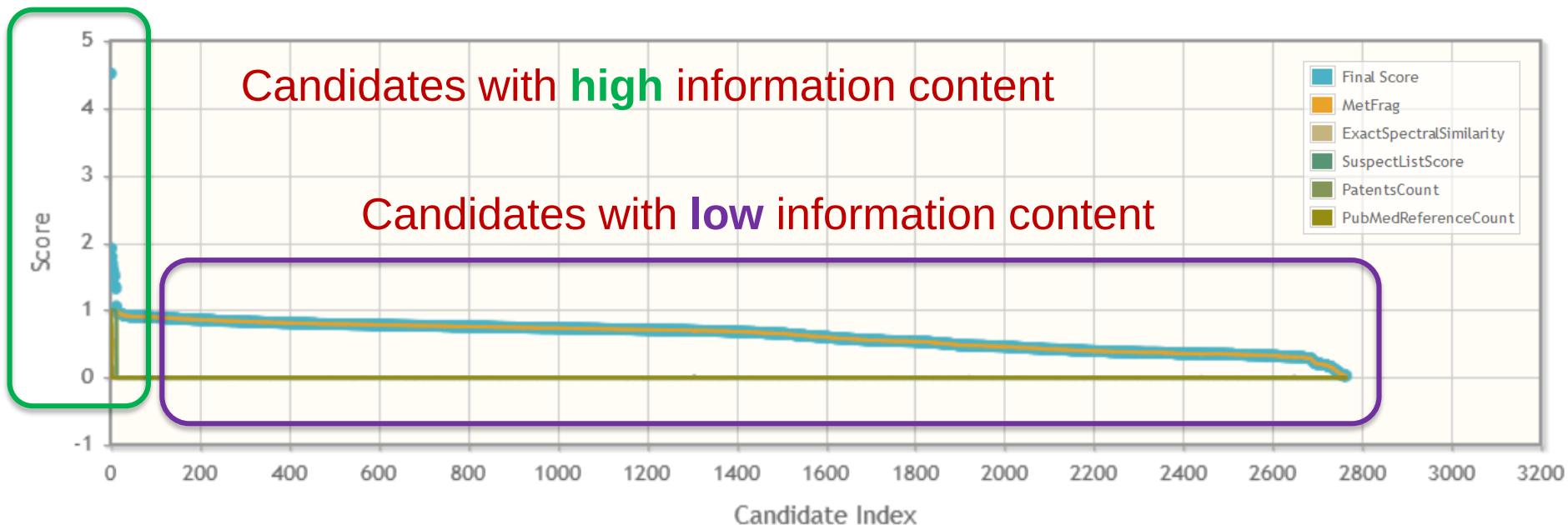
Connecting multiple lines of evidence for identification



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

Statistics

Candidate Score Distribution



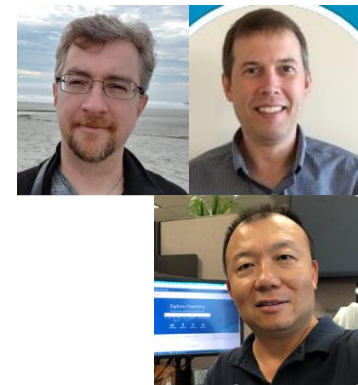
PubChem Compound TOC

Agrochemical Information	33,765,953
Biologic Description	2,002
Biological Test Results	1,539,532
Biomolecular Interactions and Pathways	3,467,416
Chemical and Physical Properties	109,610
Classification	237,729
Diseases	18,569,627
Drug and Medication Information	5,746
Food Additives and Ingredients	7,447
Identification	15,955
Information Sources	20,654,780
Literature	1,668,437
Names and Identifiers	1,310,169
Patents	22,144,888
Pharmacology and Biochemistry	130,367
Related Records	5,297,096
Safety and Hazards	125,607
Spectral Information	761,478
Structures	5,926,225
Toxicity	114,554
Use and Manufacturing	108,745
Chemical Safety	122,739

111 million ... OR ...
the most relevant / annotated?

Environmental Use Case:
PubChemLite tier0: 316 K

Exposomics Use Case:
PubChemLite tier1: 360 K



zenodo Search Upload Communities emma.schymanski@uni.lu

January 14, 2020 Dataset Open Access

PubChemLite tier0 and tier1

Bolton, Evan; Schymanski, Emma

Database Settings

Database: PubChemLite_14Jan2020

Neutral Mass: 162.11576 Search ppm: 5

Formula: C10H14N2

Identifiers:

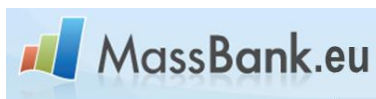
Retrieve Candidates 37 Candidates

Communities
LCSB Environmental Cheminformatics Group [Remove]

503 views 482 downloads
[See more details...](#)

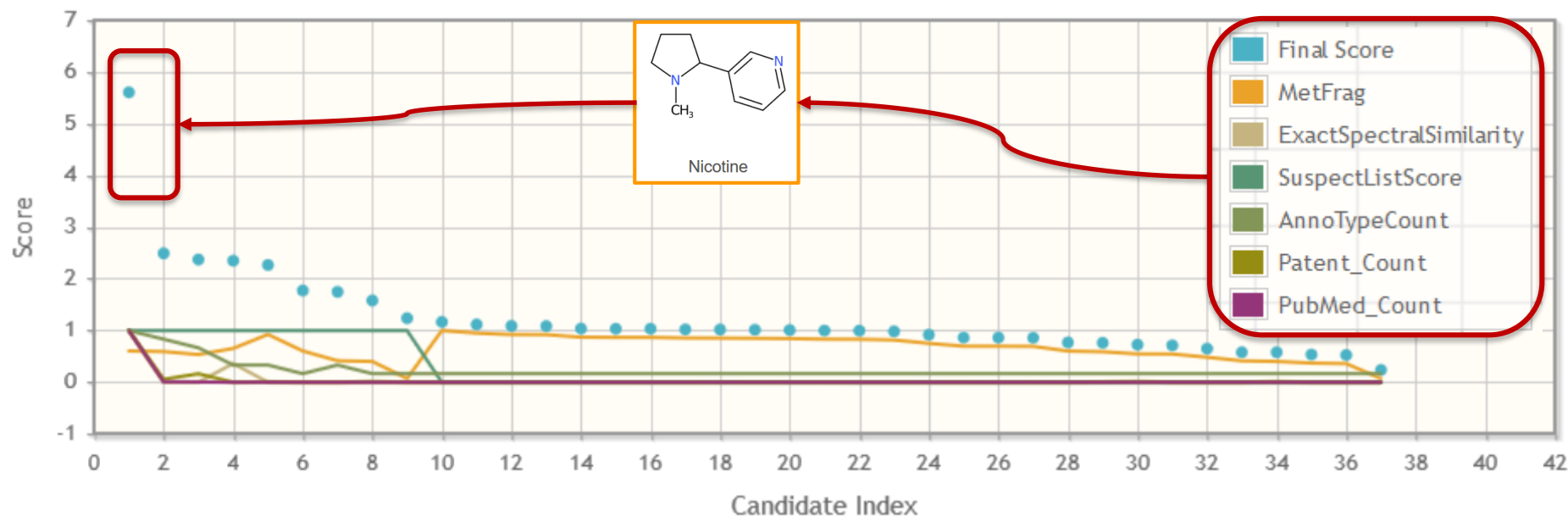
Metrag

PubChemLite: tailor-made database + metadata



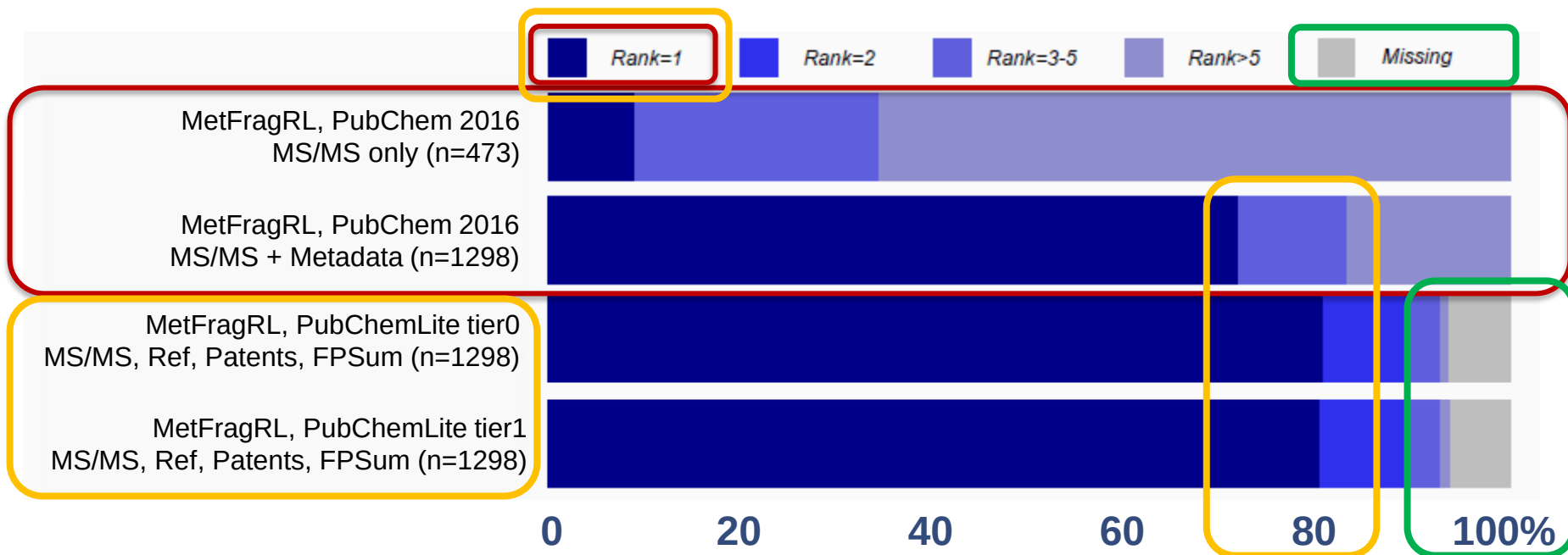
Statistics

Candidate Score Distribution



How does PubChemLite perform?

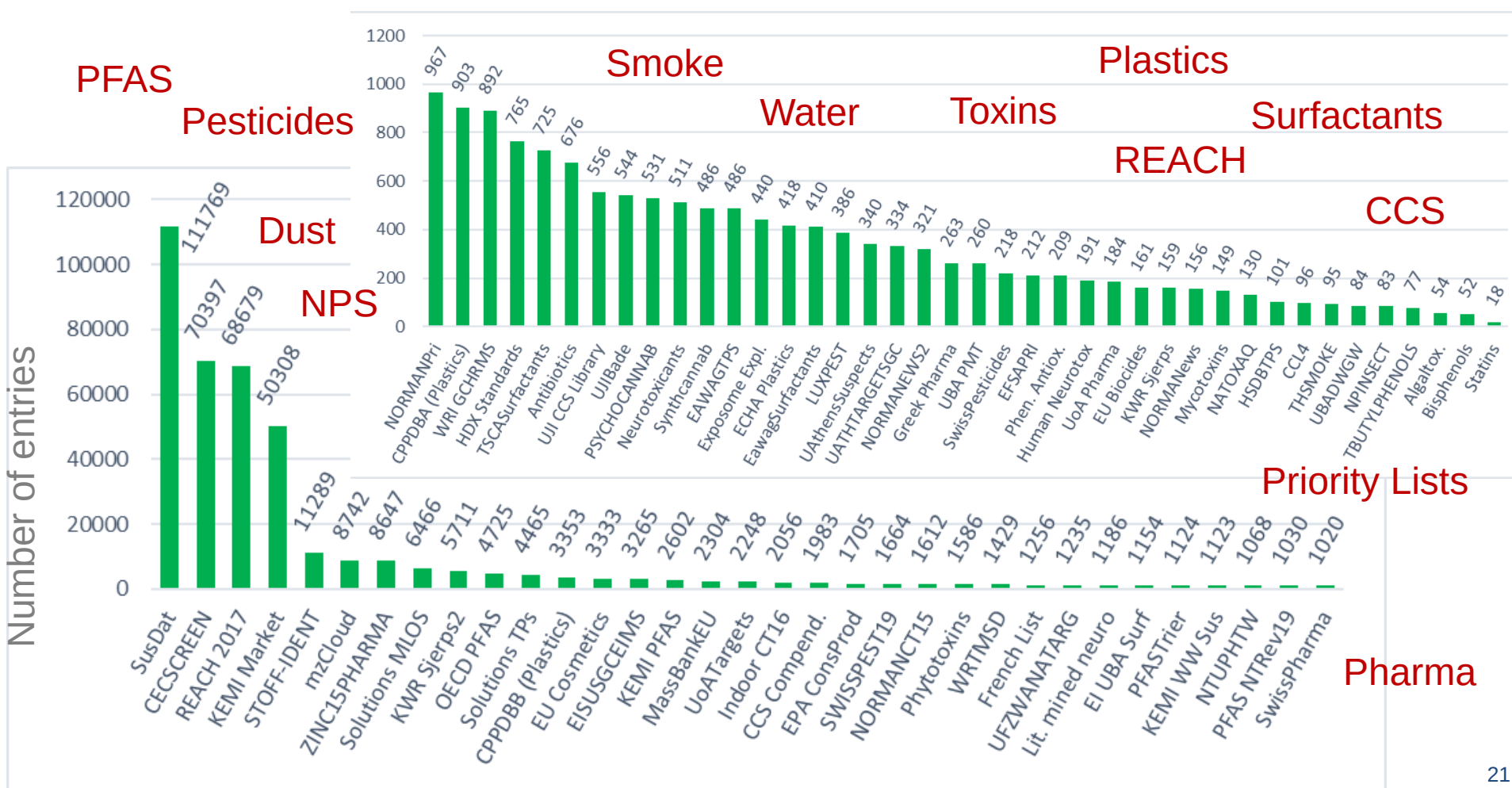
- 111 M => 300 K ... how does this influence performance?



NORMAN Suspect List Exchange: Capturing Knowledge

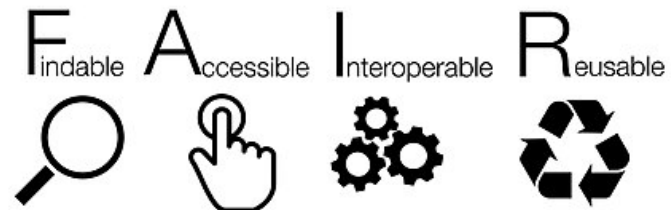
- <https://www.norman-network.com/nds/SLE/>
- <https://zenodo.org/communities/norman-sle>
- <https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101>

>73 lists!



NORMAN-SLE on Zenodo

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



zenodo

Search

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 Section

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

This is the collection associated with list S61 UJICCSLIB. A list of compounds (both positive and negative ionization mode instrument) pr

Uploaded on July 30, 2020

2 more version(s) exist for this record

Communities

NORMAN Suspect List Exchange

Remove

1,049

views

1,166

downloads

[See more details...](#)

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

Publication date:

November 21, 2019

DOI:

DOI [10.5281/zenodo.3966751](https://doi.org/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/>



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



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



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



○ <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 Display zero count nodes?

None

Compound

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**

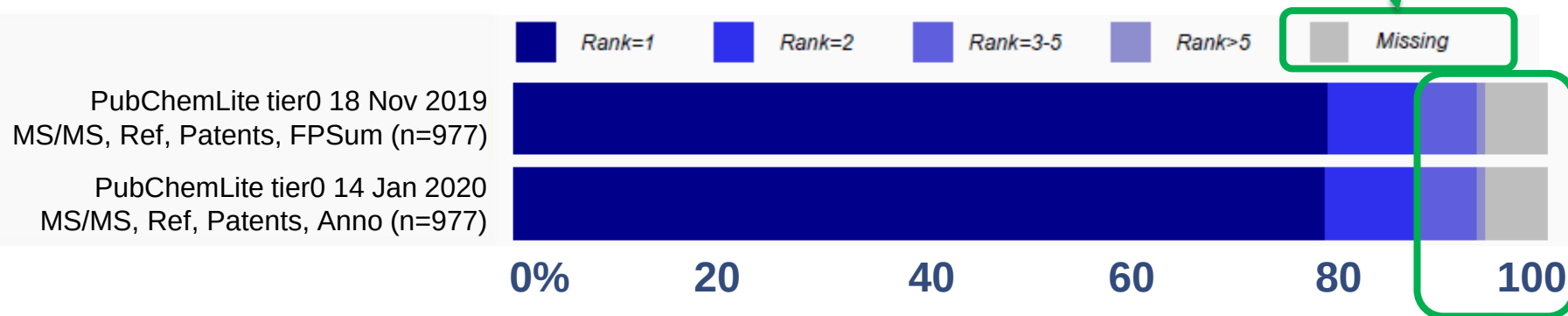
▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? **647**

▶ 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**

○ Assessing the missing entries ...



▼ NORMAN Suspect List Exchange Classification ? 117,037

▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and

▶ 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

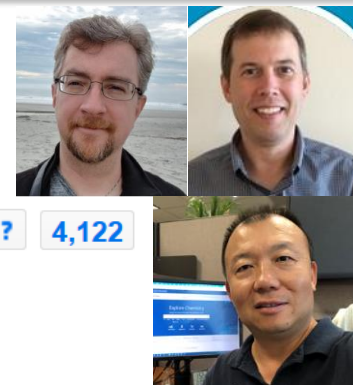
▶ S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574

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

▶ S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 97

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 PubChem Terbutylazine (Compound)

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 ? 1

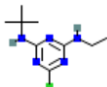
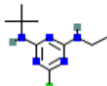
S08 | ATHENSSUS | University of Athens Surfactants and Susp

S09 | PFASTRIER | PFAS Suspect List of fluorinated substance

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
	Terbutylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine	10.5281/zenodo.3827
	Terbutylazine	Deethylation		Terbutylazine-desethyl	10.1007/s13361-017-

PubChem Terbutylazine (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.

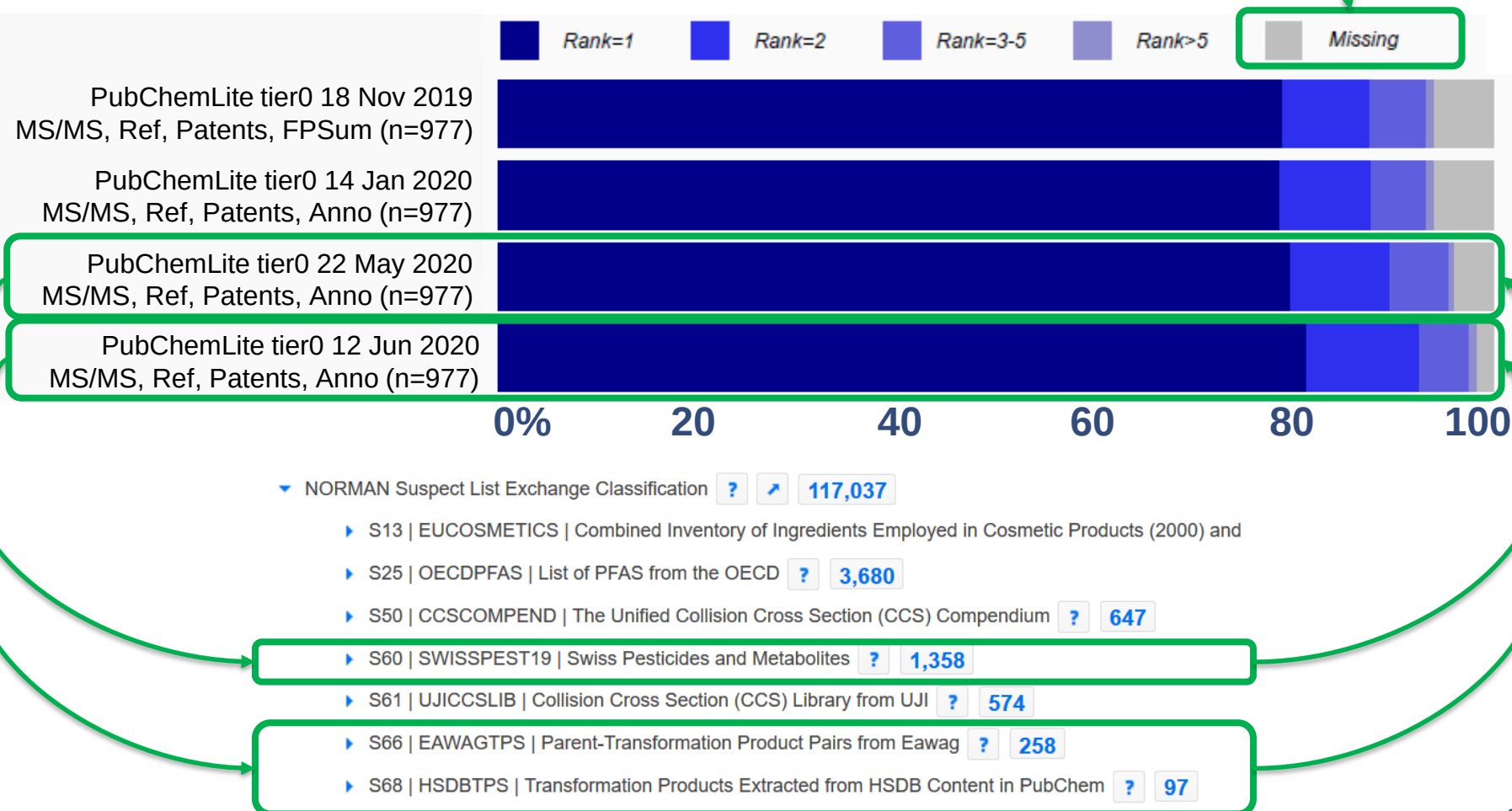
S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

► NORMAN Suspect List Exchange

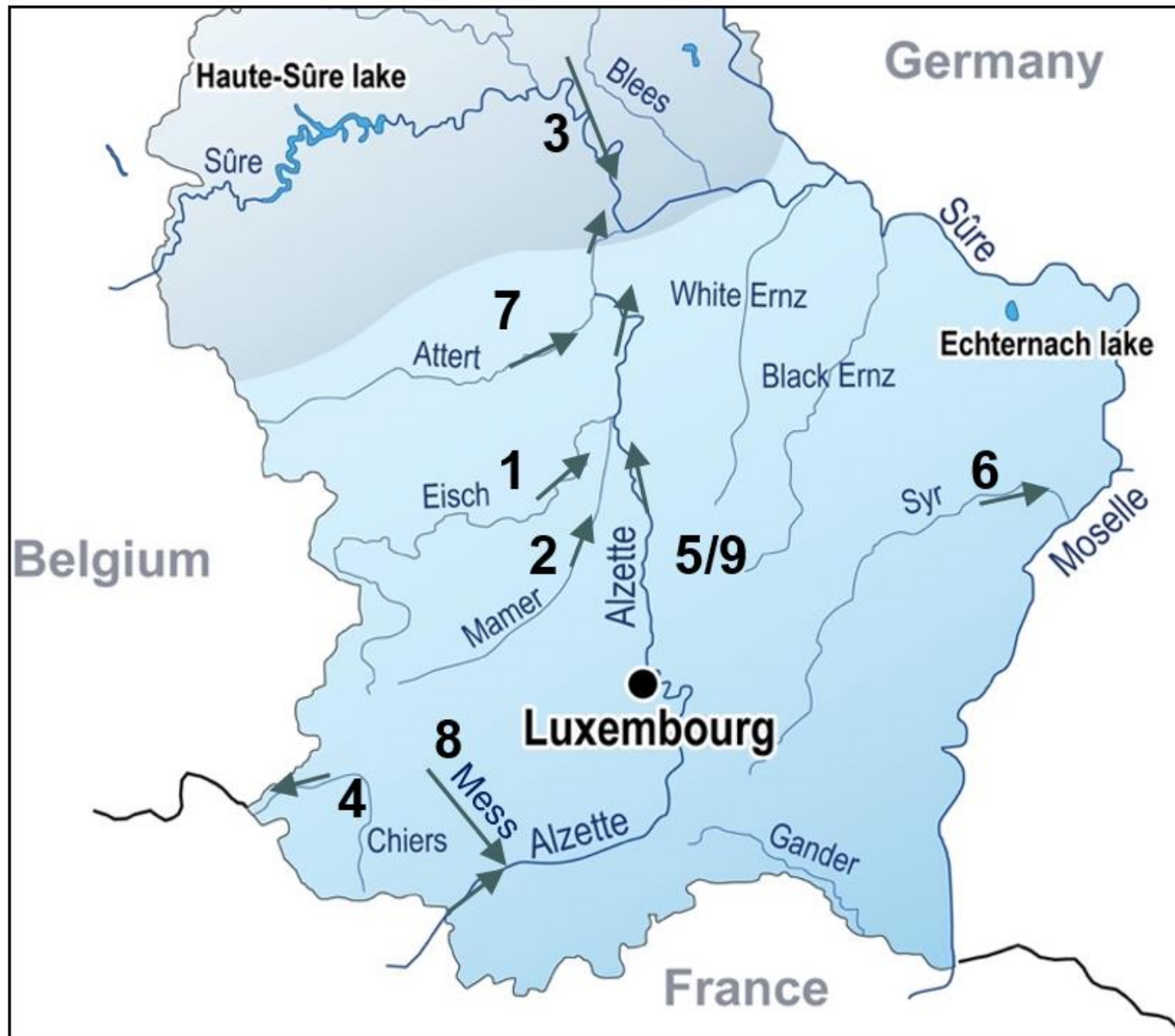
Terbutylazine has known environmental transformation products that include CSAA036479, CSAA04949, CSCD648241, CSCD692760, GS31398, MT1, GS 26379, MT13, GS 23158, Terbutylazine metabolite MT14, Terbutylazine metabolite MT23, and Terbutylazine metabolite MT24.

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

○ Assessing the missing entries ...



“Real Life” Application



Suspect Screening for Lux-Relevant Pesticides



zenodo

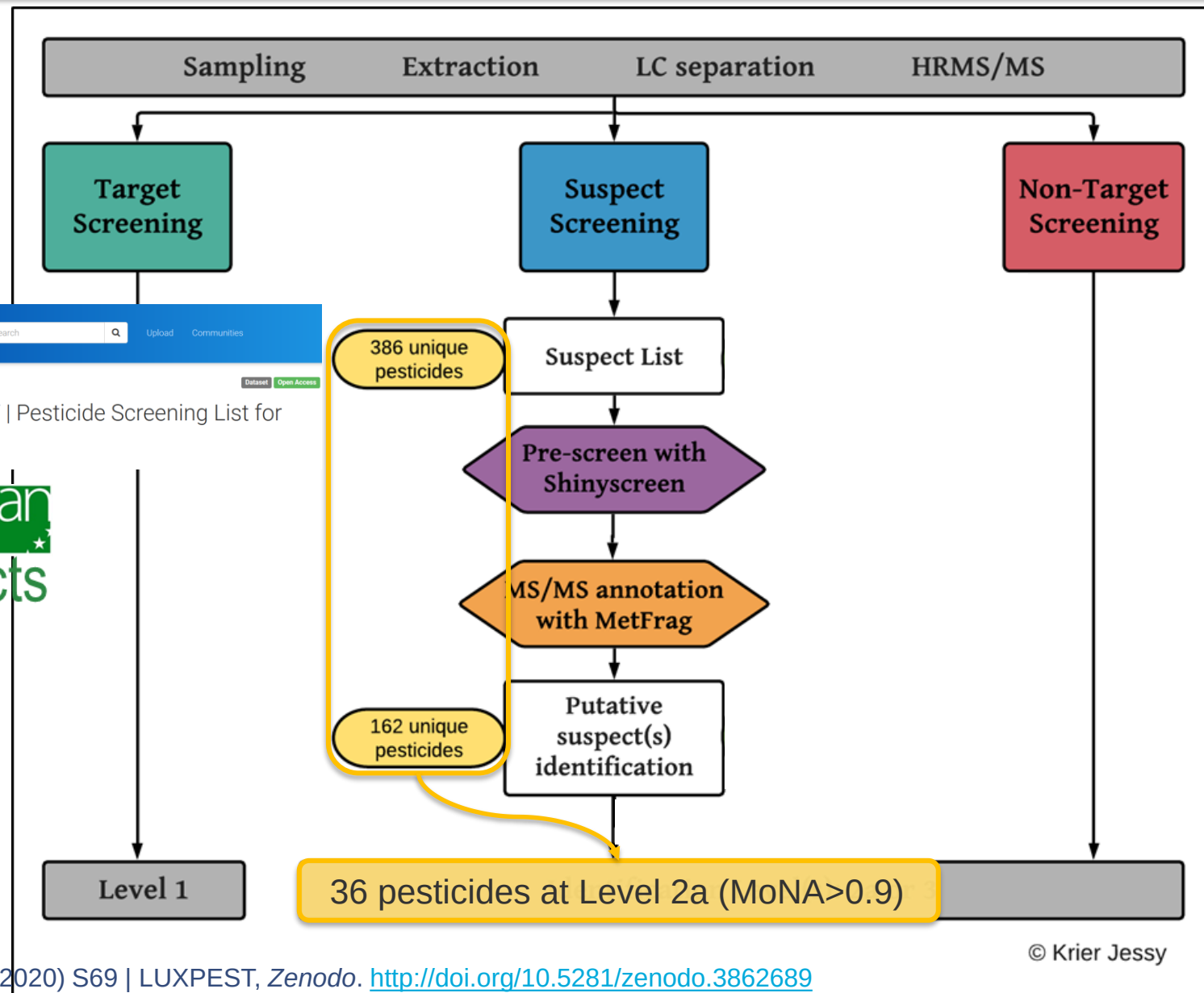
Search [] Upload Communities

May 28, 2020

S69 | LUXPEST | Pesticide Screening List for Luxembourg

Krier, Jessy

norman suspects

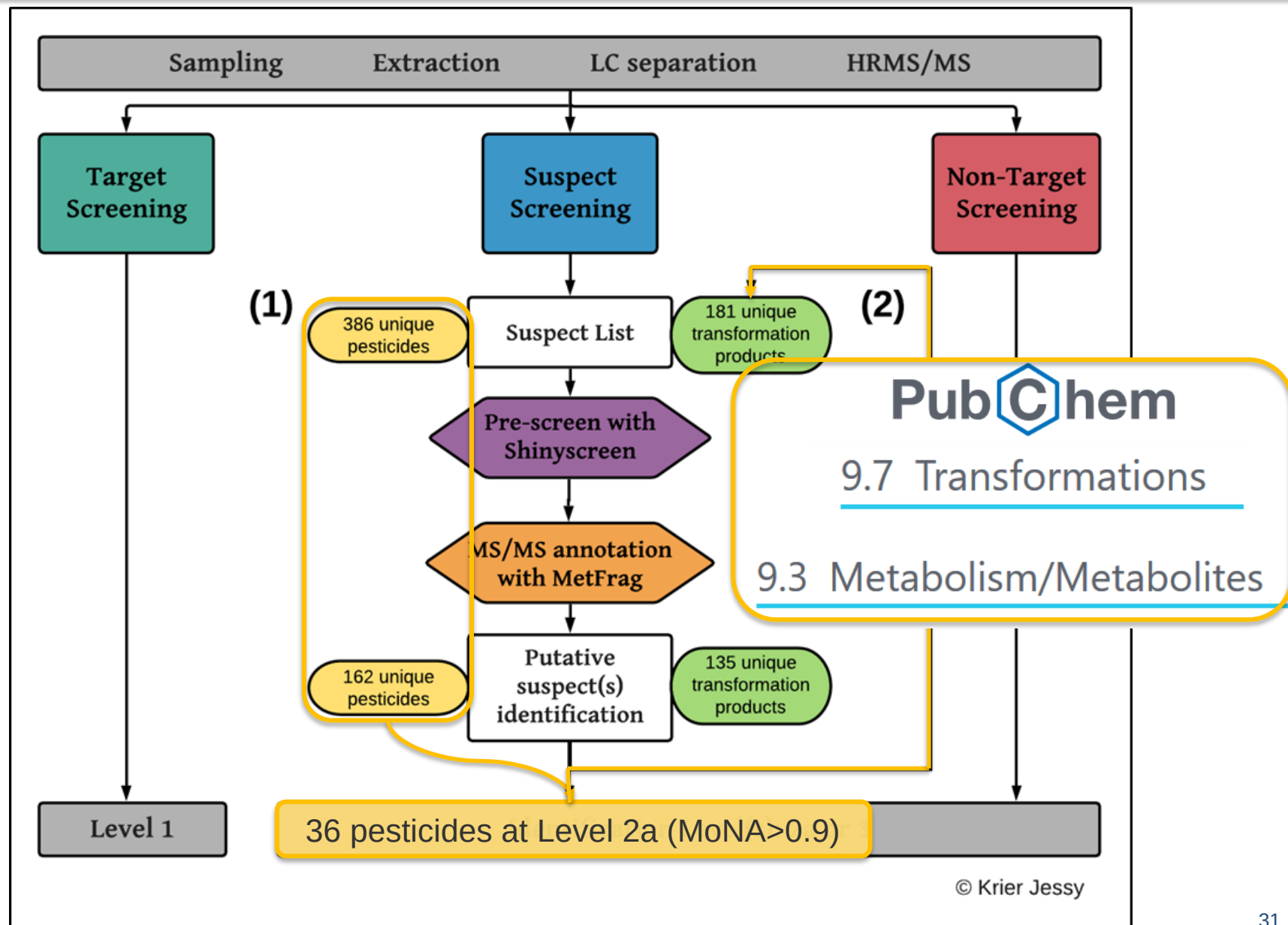


© Krier Jessy

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



Literature Mining for Metabolites / TPs – and Curation

PubChem Terbutylazine (Compound)

8.3 Metabolism/Metabolites



Metabolism of terbutylazine 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 Terbutylazine (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, respectively. Degradation of the **triazine ring** did not occur. **Ammeline** and **ammeline** dealkylated/hydroxylated metabolites common to all triazines, were

USEPA; Reregistration Eligibility Decision (RED) Database for Terbutylazine (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)

In mammals, following oral administration, ...a de-ethyl metabolite for one of the products formed by oxidation of one **ethyl** group of the tert-butyl

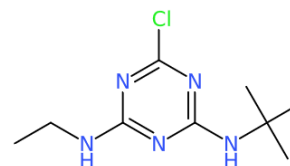
Tomlin CDS, ed. Terbutylazine (5915-41-3). In: The e-Pesticide Manual, Version 10.0. The British Crop Protection Council.

► Hazardous Substances Data Bank (HSDB)

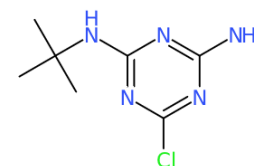


```
CCNC1=NC(=NC(=N1)C1)NC(C)(C)C Terbutylazine CID:22206
NC1=NC(=NC(=N1)C1)NC(C)(C)C desethyl-terbutylazine CID:108201
CCNC1=NC(=NC(=N1)C1)N des-t-butyl-terbutylazine CID:13878
CCNC1=NC(=NC(=N1)O)NC(C)(C)C 2-hydroxy-terbutylazine CID:135495928
CCNC1=NC(=NC(=N1)C1)NC(C)(C)C(O) (hydroxy-t-butyl)-Terbutylazine CID:779516
NC1=NC(=NC(=N1)C1)N didealkyl-terbutylazine CID:18831
OC1=NC(=NC(=N1)C1)N hydroxy-didealkyl-terbutylazine CID:135438601
NC1=NC(=NC(=N1)C1)N didealkyl-terbutylazine CID:18831
```

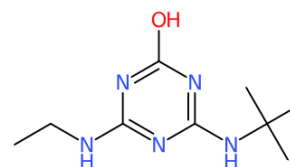
Color on White No Annotation Chiral Hydrogens (smart) Do Not Abbreviate



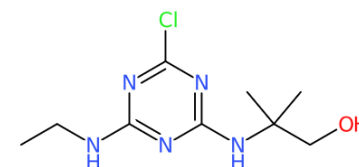
Terbutylazine CID:22206



desethyl-terbutylazine CID:108201



2-hydroxy-terbutylazine CID:135495928



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



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 of
PubChem on the NOR

<https://www.norman-suspects.org/>

HSDBTPS is a list of r

HSDB (Hazardous Substances)

[10.5281/zenodo.38274](https://www.norman-suspects.org/10.5281/zenodo.38274)

Entries automatically

Preview

Predecessor_CID

13450

13450

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

LCSB-ECI & PubChem Team. DOI [10.5281/zenodo.3890392](https://doi.org/10.5281/zenodo.3890392)

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

added new CIDs to HSDBTPS

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

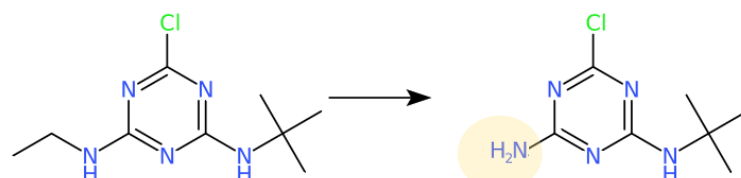
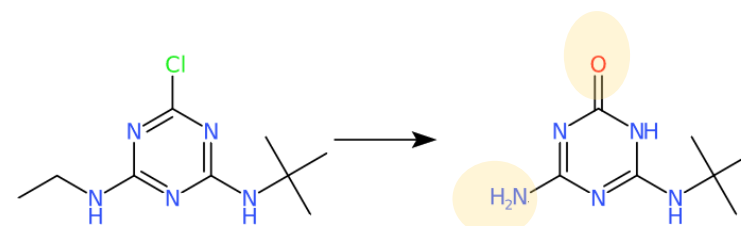
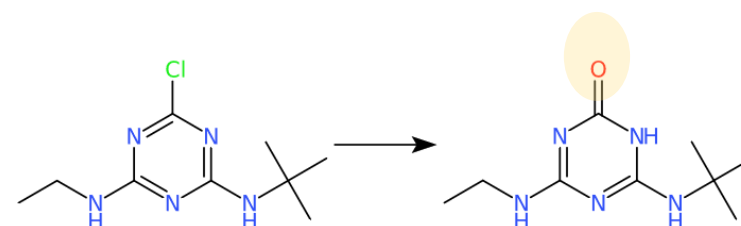
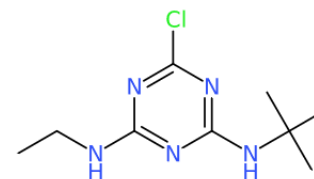
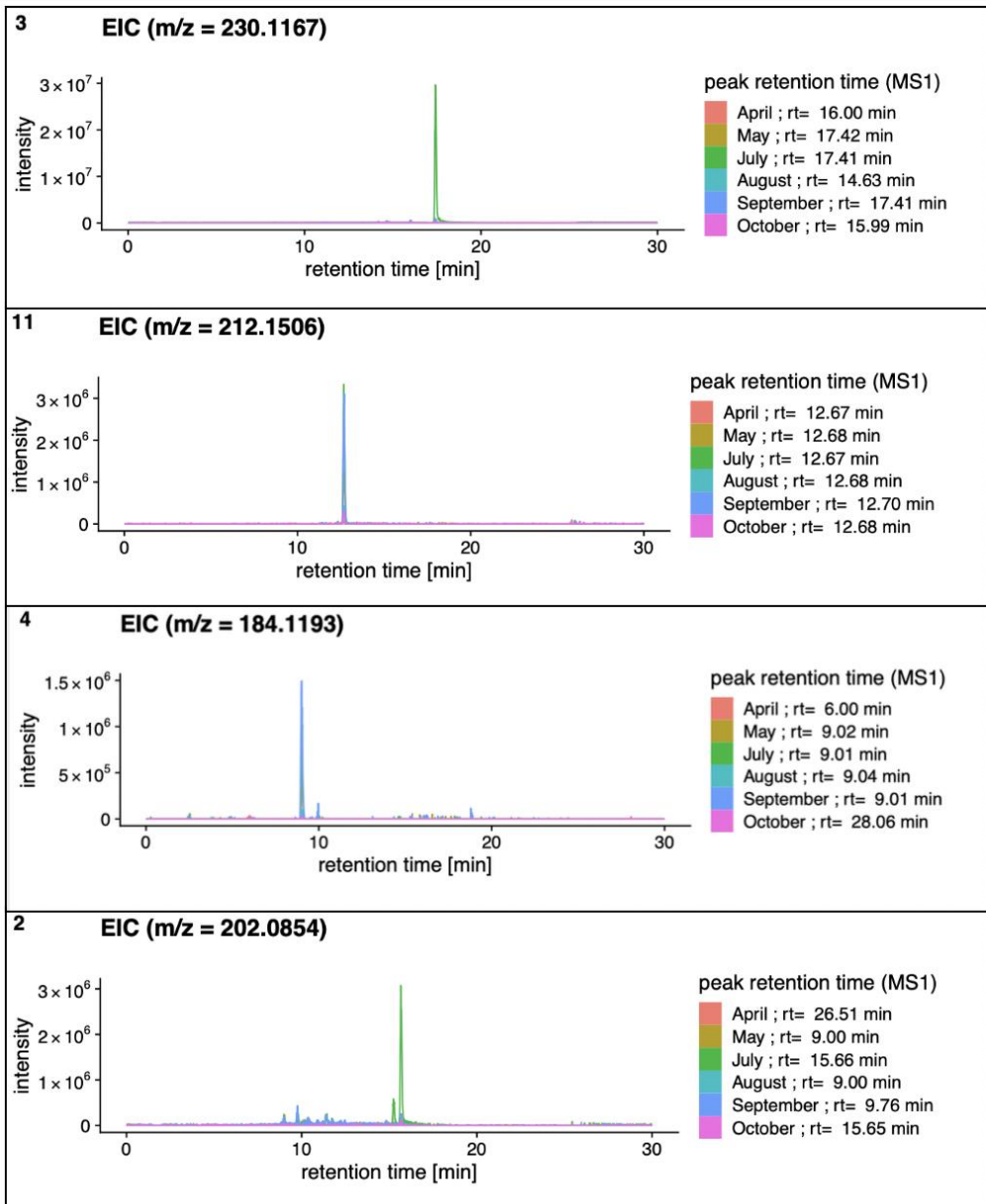
8.5 Transformations

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.12000506
	Terbutryn	Mammalian metabolism		2-[[[4-(Ethylamino)-6-methylsulfanyl]-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.5281/zenodo.38274

Pesticides & TPs over time (tent. IDs)



More Examples? Thirdhand Smoke (THS) in Dust



Metrag

Database Settings

Database:

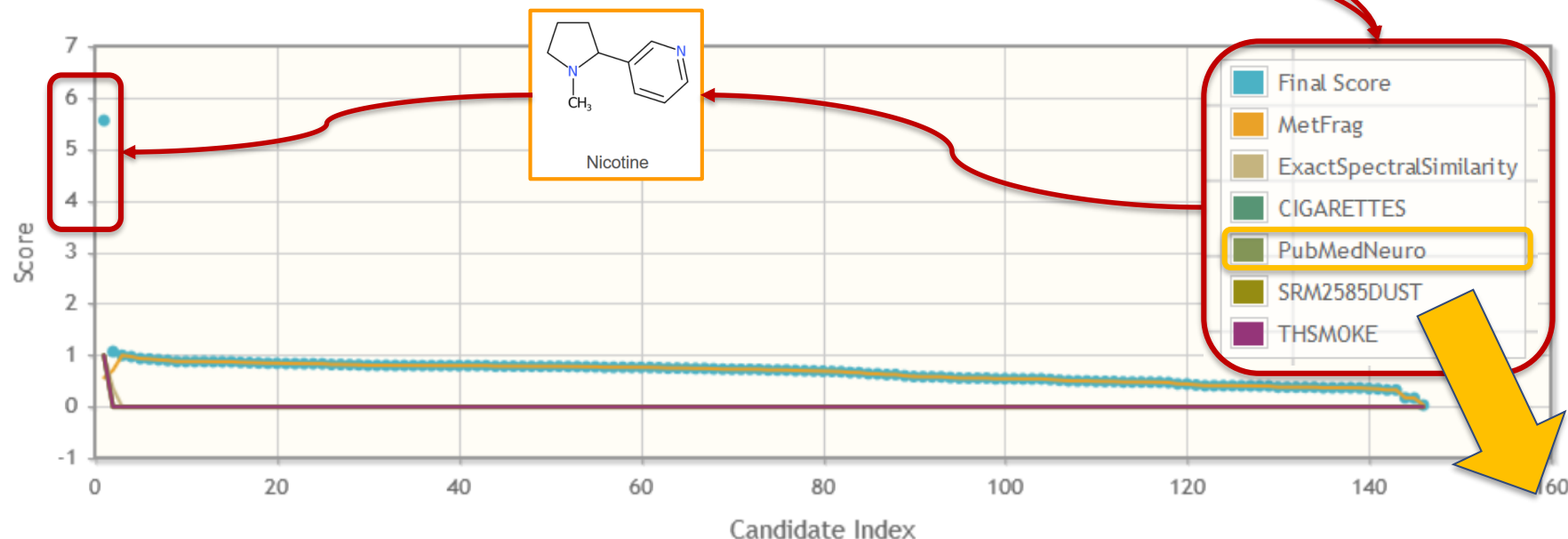
Neutral Mass:

Formula:

Identifiers:

Candidate Filter & Score Settings

Candidate Score Distribution



New MetaData: Disease-Specific Reference Counts

S37 LitMinedNeuro: DOI: [10.5281/zenodo.3242298](https://doi.org/10.5281/zenodo.3242298)

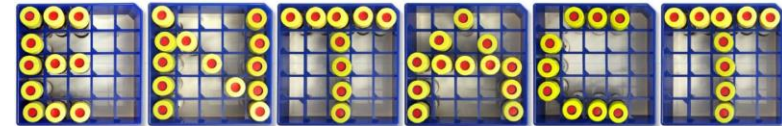
Chemical	CAS RN	DSSToxID	PMID Ct	Seizures	Nervous System Diseases	Peripheral Nervous System Diseases	Brain Diseases	Muscular Diseases	Basal Ganglia Diseases	Parkinson Disease, Secondary	Coma	Hallucinations	Tremor	Memory Disorders	Central Nervous
Cisplatin	15663-27-1	DTXSID4024983	1032	20	47	140	13	0	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

Database Scoring Terms

Select Item(s) 5 of 14 item(s) selected

- ☐ Nervous System Diseases
- ☐ PUBCHEM_DATA_SOURCES
- ☒ Parkinson Disease, Secondary
- ☐ Peripheral Nervous System Diseases
- ☐ PubMedID_COUNT
- ☐ Seizures
- ☒ TOXCAST_PERCENT_ACTIVE

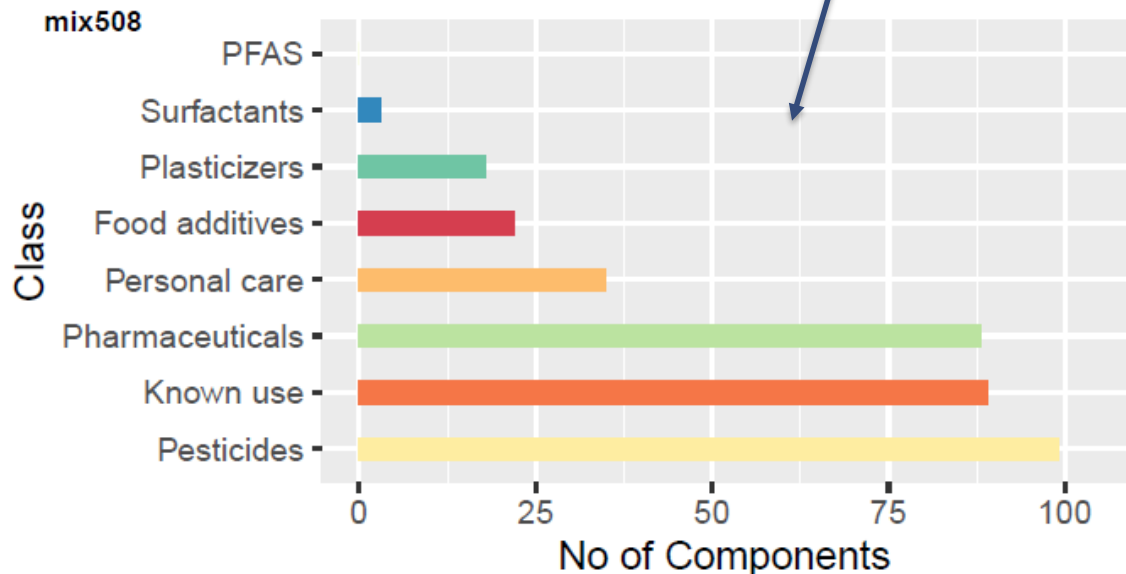
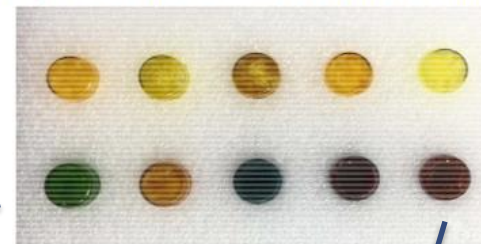
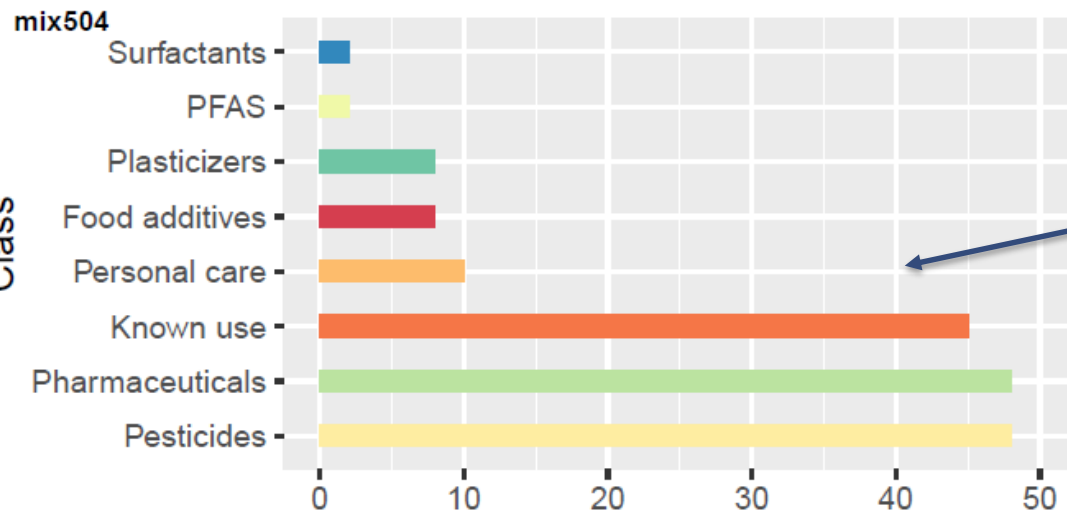
Future Perspectives



○ Rapid classification / interpretation of large collections

No of Components

95, 185 or 365 substances/mixture



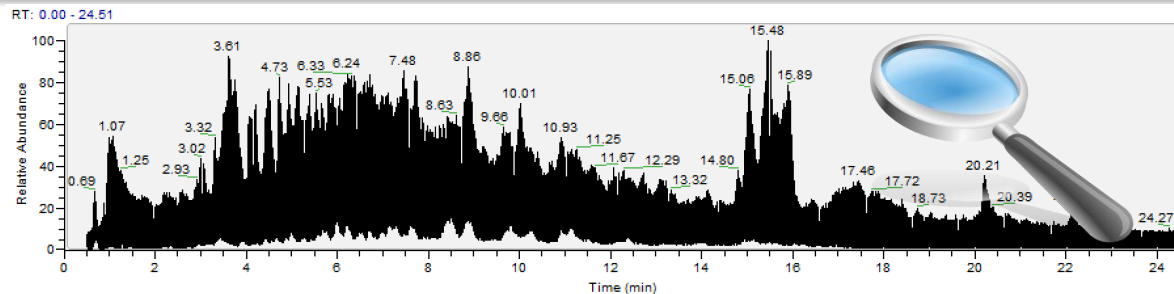
ENTACT: Elin Ulrich *et al.* 2018,
https://www.epa.gov/sites/production/files/2018-06/documents/comptox_cop_6-28-18.pdf and
Ulrich, EM, *et al.* (2019) *Anal Bioanal Chem.*
DOI: [10.1007/s00216-018-1435-6](https://doi.org/10.1007/s00216-018-1435-6)

Classification Figures:

Anjana Elapavalore, ECI (Master's thesis)

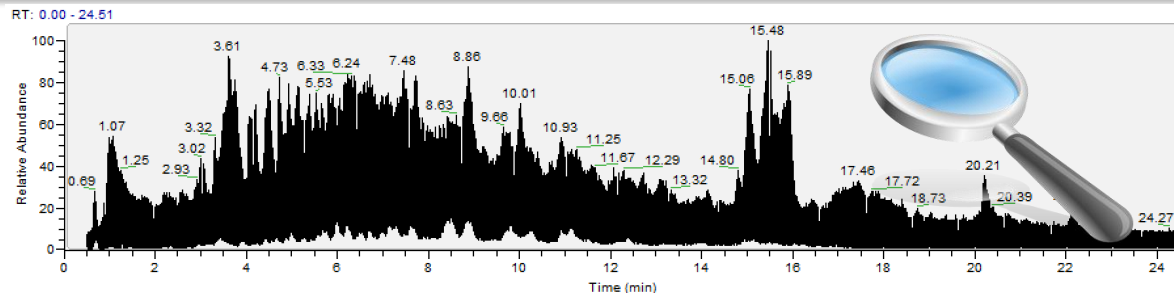
Take Home Messages

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



Take Home Messages

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

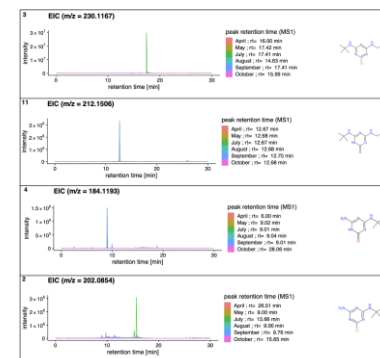


- “**Digital Detective Work**” is ...
 - Capturing expert knowledge
Machine and **human** readable
 - **Connecting** this to environmental observations
 - Identifying and **closing knowledge gaps**
 - Supporting **interpretation** of complex data

PubChem Terbutylazine (Compound)

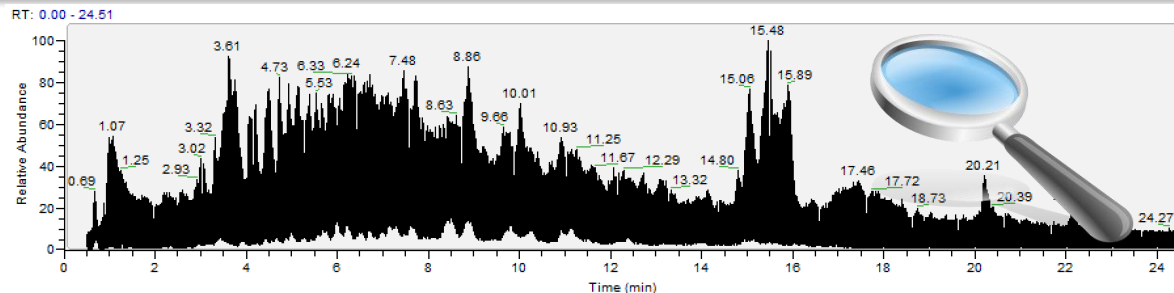
8.5 Transformations

	Terbutylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine
	Terbutylazine	In vitro (rats, pigs, humans) metabolism		2-[[4-Chloro-6-(ethylamino)-1,3,5-triazin-2-yl]amino]-2-methylpropan-1-ol
	Terbutylazine	Deethylation		Terbutylazine-desethyl



Take Home Messages

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



- “**Digital Detective Work**” is ...

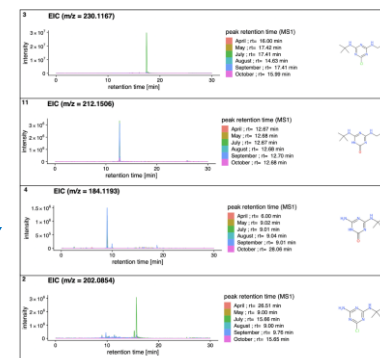
- Capturing expert knowledge
Machine and **human** readable
- **Connecting** this to environmental observations
- Identifying and **closing knowledge gaps**
- Supporting **interpretation** of complex data

PubChem Terbutylazine (Compound)

8.5 Transformations

	Terbutylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine
	Terbutylazine	In vitro (rats, pigs, humans) metabolism		2-[[4-Chloro-6-(ethylamino)-1,3,5-triazin-2-yl]amino]-2-methylpropan-1-ol
	Terbutylazine	Deethylation		Terbutylazine-desethyl

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



Acknowledgements



Luxembourg National
Research Fund



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**

Philippe Diderich (eau.etat.lu)



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/>





Community Efforts!



MassBank consortium

norman

ART2
CURE

