Finding Small Molecules In Big Data

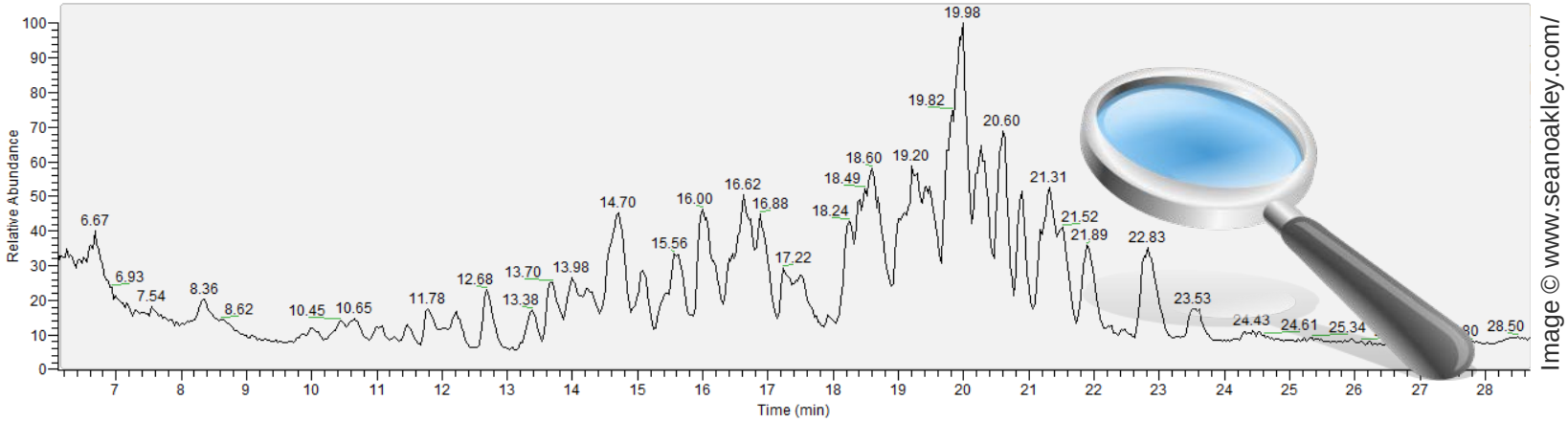


Image © www.seanoakley.com/

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 many colleagues who contributed to my science over the years!

Luxembourg

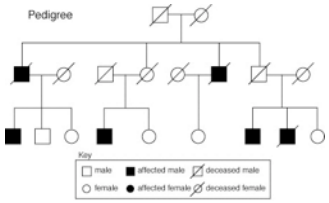
Source: LCSB Communications Team and https://en.wikipedia.org/wiki/File:EU-Luxembourg.svg

University of Luxembourg & LCSB

- Uni Lu was founded in 2003
  - Teenage years!
- LCSB was founded in 2009
  - Young and very dynamic working environment!

10 YEARS
2009 - 2019

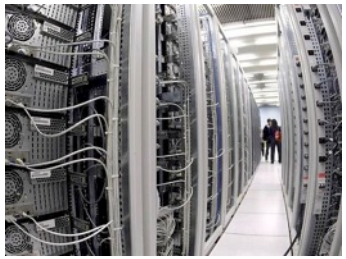
LCSB: Luxembourg Centre for Systems Biomedicine



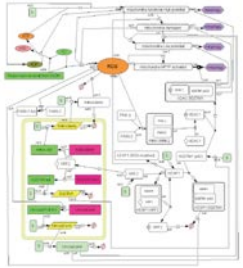
Family Studies

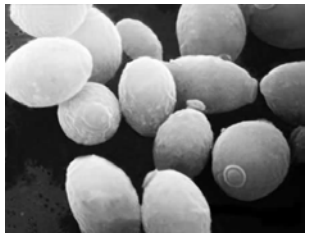
Patients

Longitudinal Cohorts

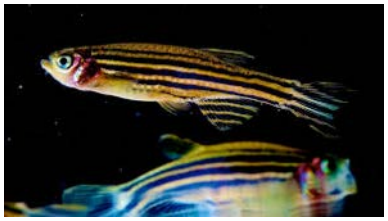


Pathway and Network Analysis
Computational Models
Machine Learning





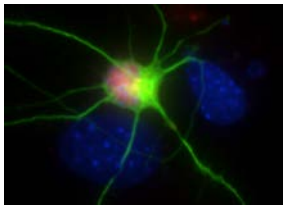
Yeast



Zebrafish

Microbiome

Mouse



Human

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

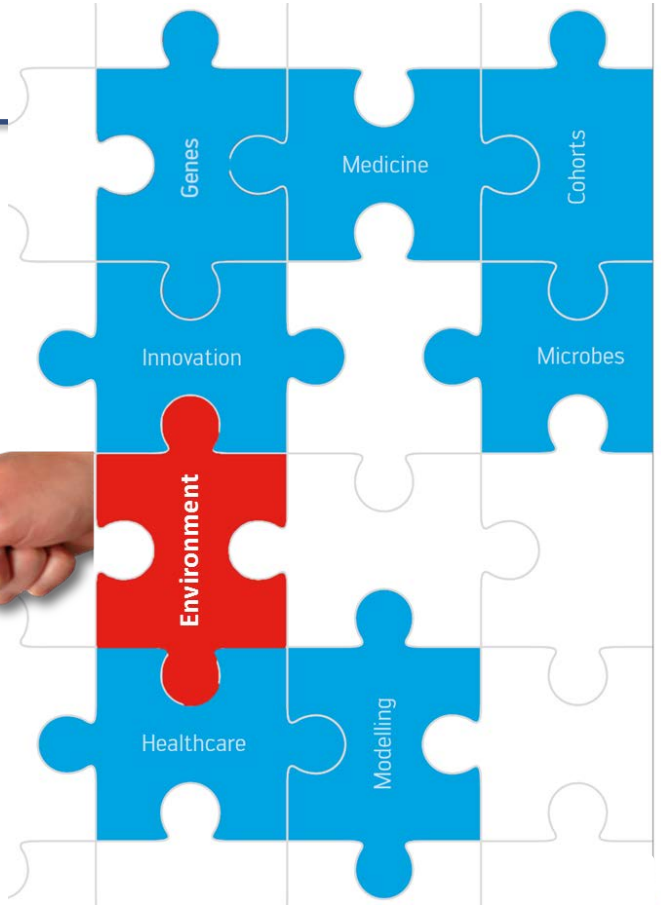
Lorenzo Favilli @lorenzo.favilli
Given access 2 months ago

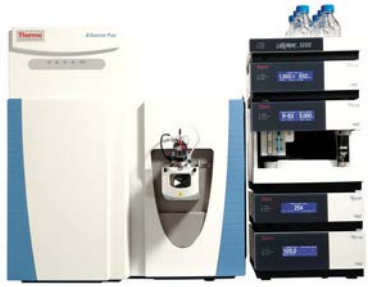
Mira Narayanan @mira.narayanan
Given access 4 weeks ago

Randolph Singh @randolph.singh
Given access 2 months ago

Todor Kondić @todo
Given access 2

R3 LCSB

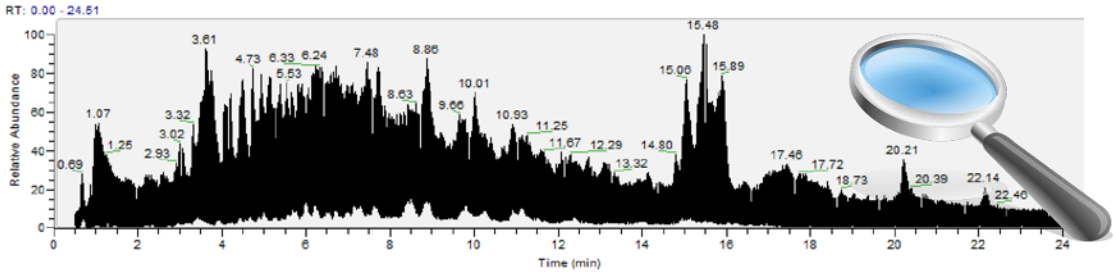




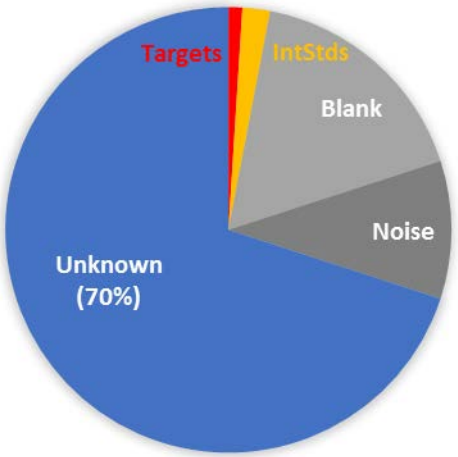
Luxembourg National Research Fund

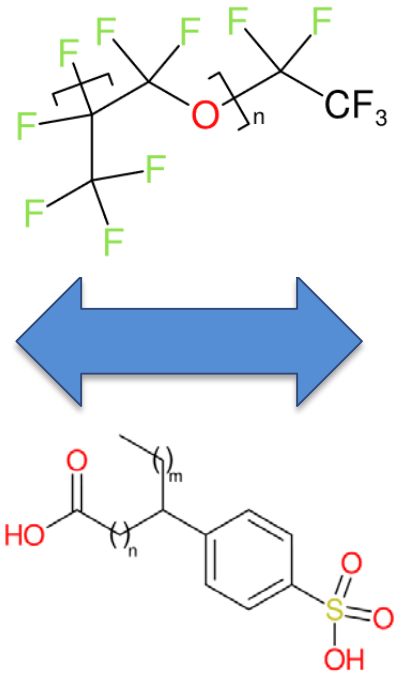
UNIVERSITÉ DU LUXEMBOURG

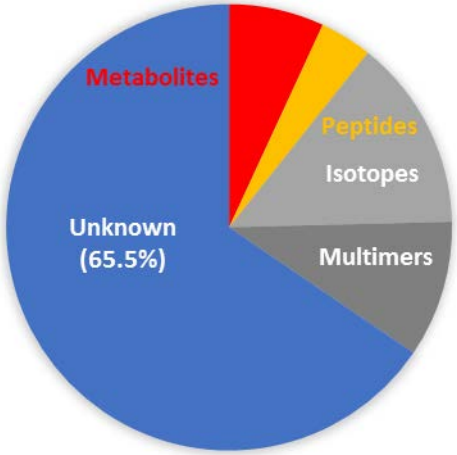
Our challenge? We still have many unknowns …



Wastewater







Cells

(l) Data from Schymanski *et al* 2014, ES&T DOI: 10.1021/es4044374. (r) *E. coli* data provided by N. Zamboni, IMSB, ETH Zürich.

Target, Suspect and Non-Target Screening

KNOWNS	SUSPECTS	SUSPECTS	No Prior Knowledge
TARGET ANALYSIS	SPECTRUM SEARCH	SUSPECT SCREENING	NON-TARGET SCREENING
Targets found	Spectral match	Suspects found	Masses of interest
			(Molecular formula)
			DATABASE SEARCH / STRUCTURE GENERATION

Candidate selection (retention time, MS/MS, calculated properties)

Time, Effort & Number of Compounds….

Finding Small Molecules in (Swiss) Wastewater

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

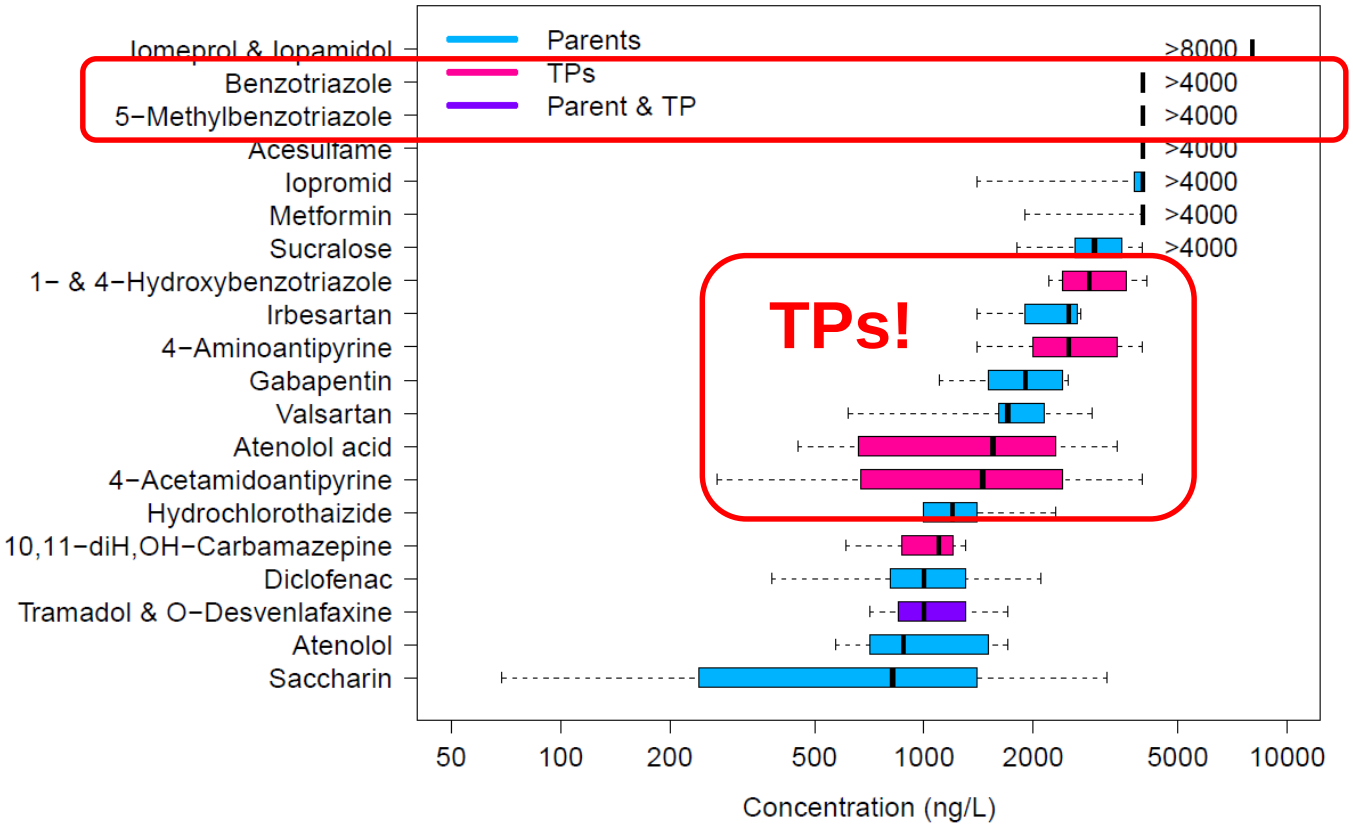
Target Analysis: Status Quo (>364 targets)

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

TARGET ANALYSIS

Targets found



Confirmation and quantification of compounds present

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

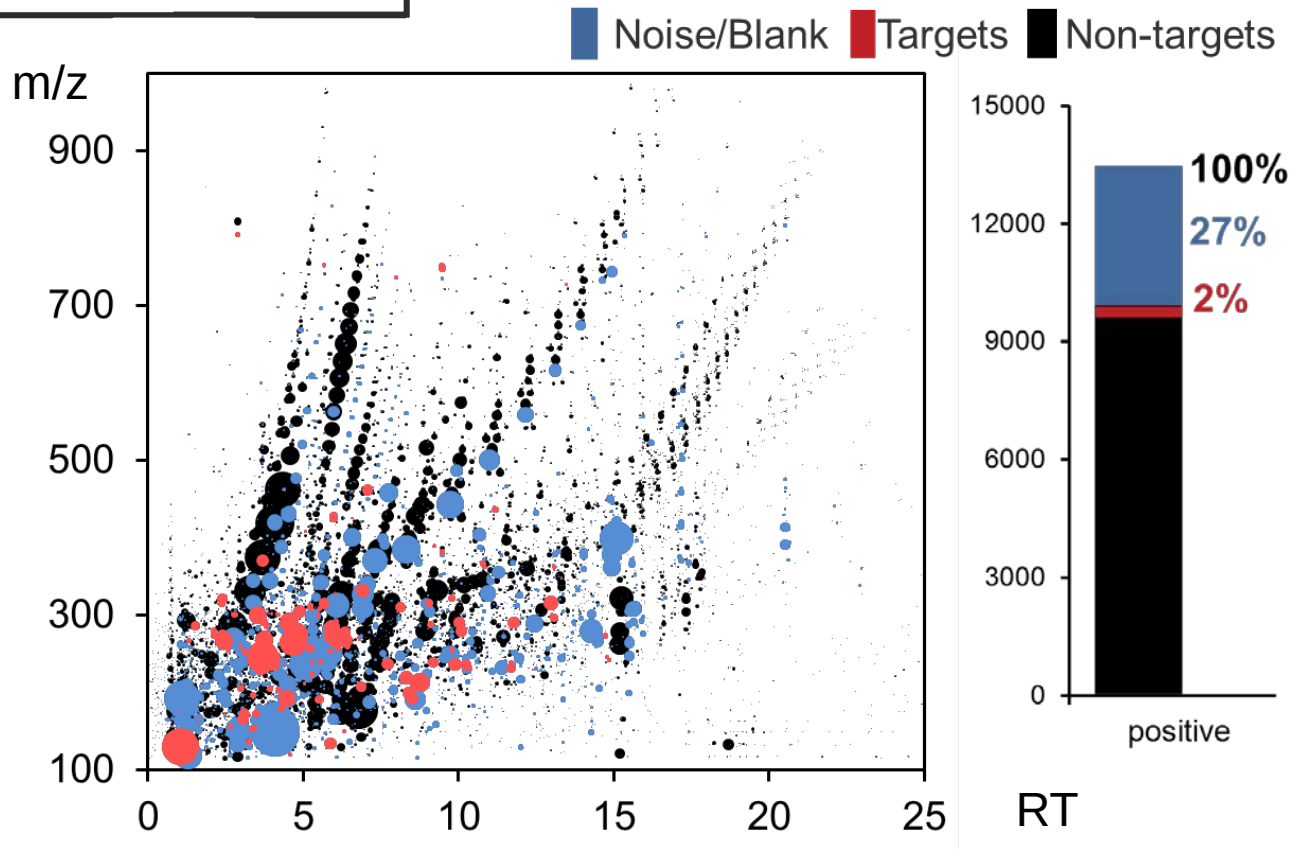
Target Analysis: Status Quo (>364 targets)

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

TARGET ANALYSIS

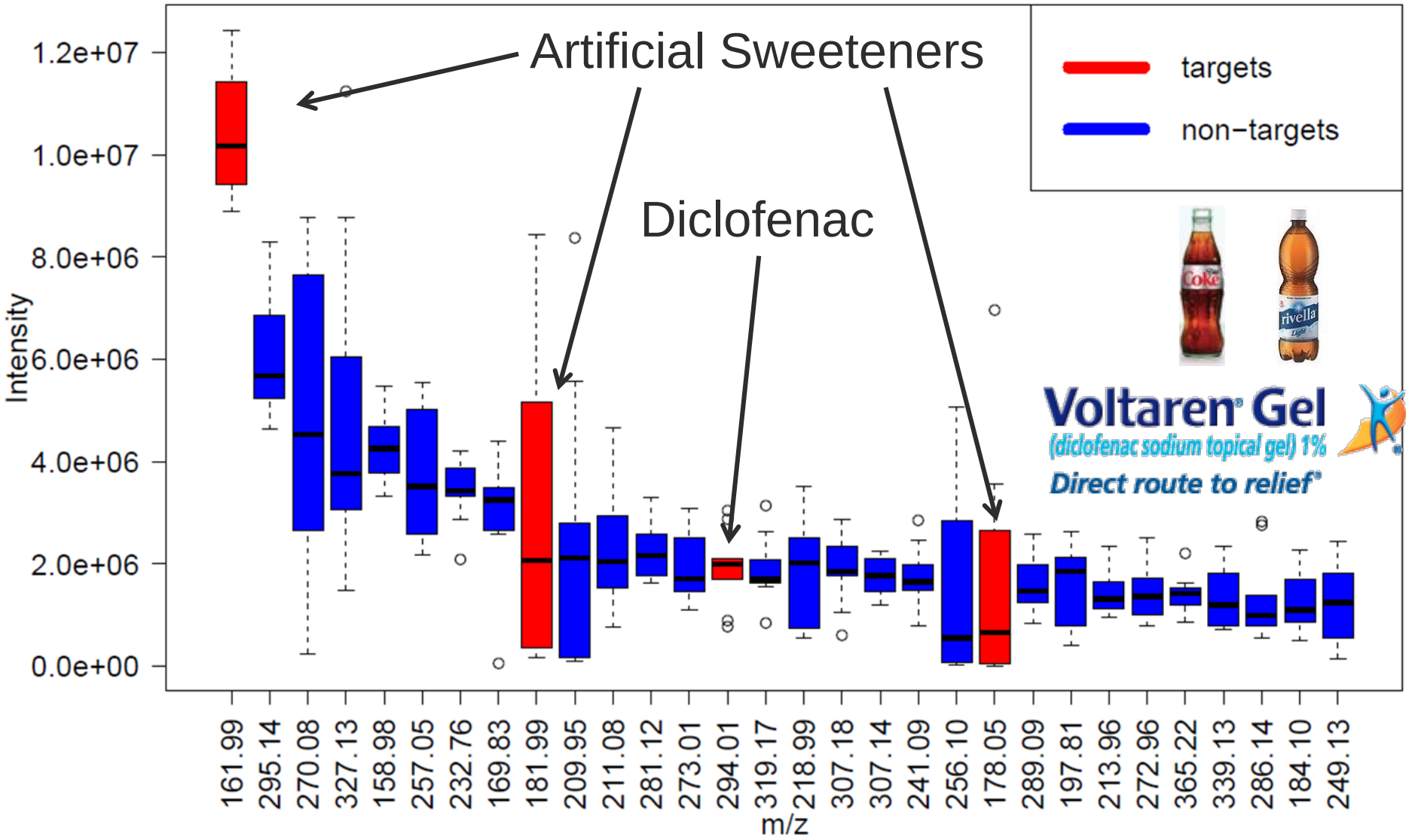
Targets found



Confirmation and quantification of compounds present

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

Swiss Wastewater: Top 30 Peaks (ESI-)



Pictures: www.coca-cola-com; www.rivella.ch; www.voltargengel.com

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

Suspect Screening: Different Approaches

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

Suspect List

TARGET ANALYSIS

SUSPECT SCREENING

Targets found

Suspects found

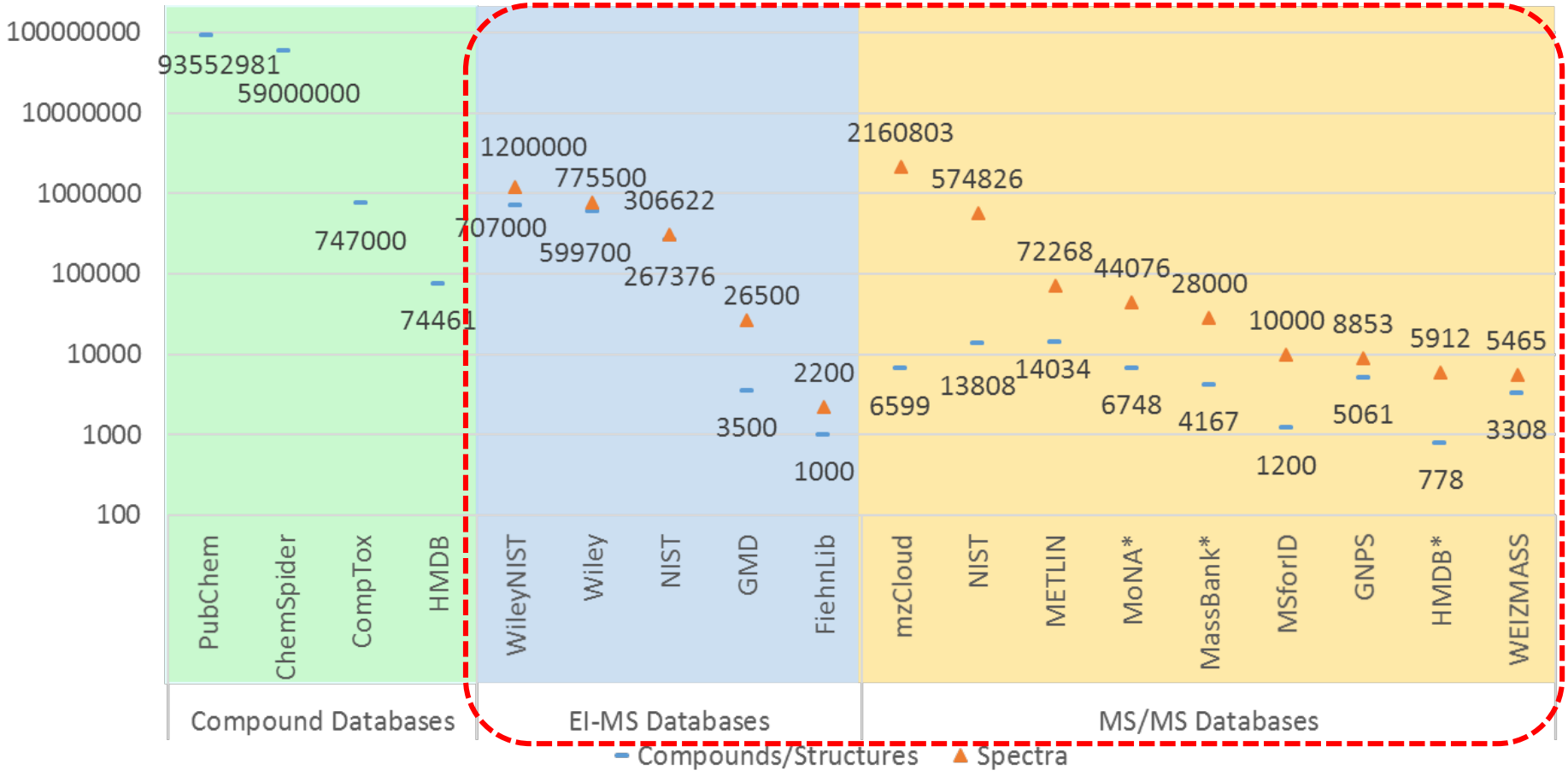
- Search in ***mass spectral*** libraries
- Screen for ***predicted*** transformation products of known parent compounds
- Look for ***"well known"*** substances without reference standards
- Screen for ***known*** homologue series

Candidate selection (retention time, MS/MS, calculated properties)

Confirmation and quantification of compounds present

Searching Mass Spectral Libraries

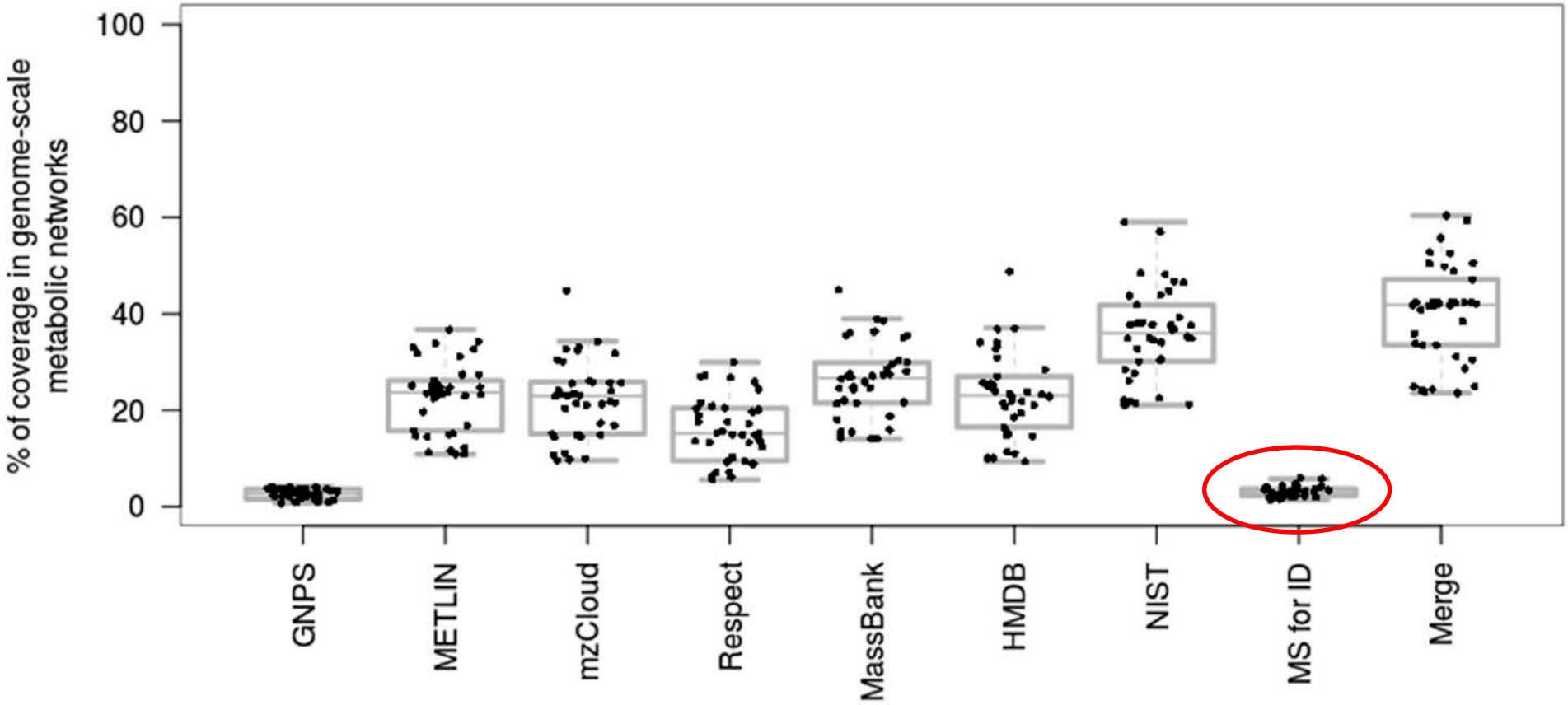
- … which one?



Peisl, Schymanski & Wilmes, 2018 *Anal. Chim. Acta*, DOI: 10.1016/j.aca.2017.12.034

Mind the Gap!

- Best library to choose depends highly on your dataset
  - Example: MSforID (https://msforid.com/) is poor for metabolic networks – but great for forensic toxicology!

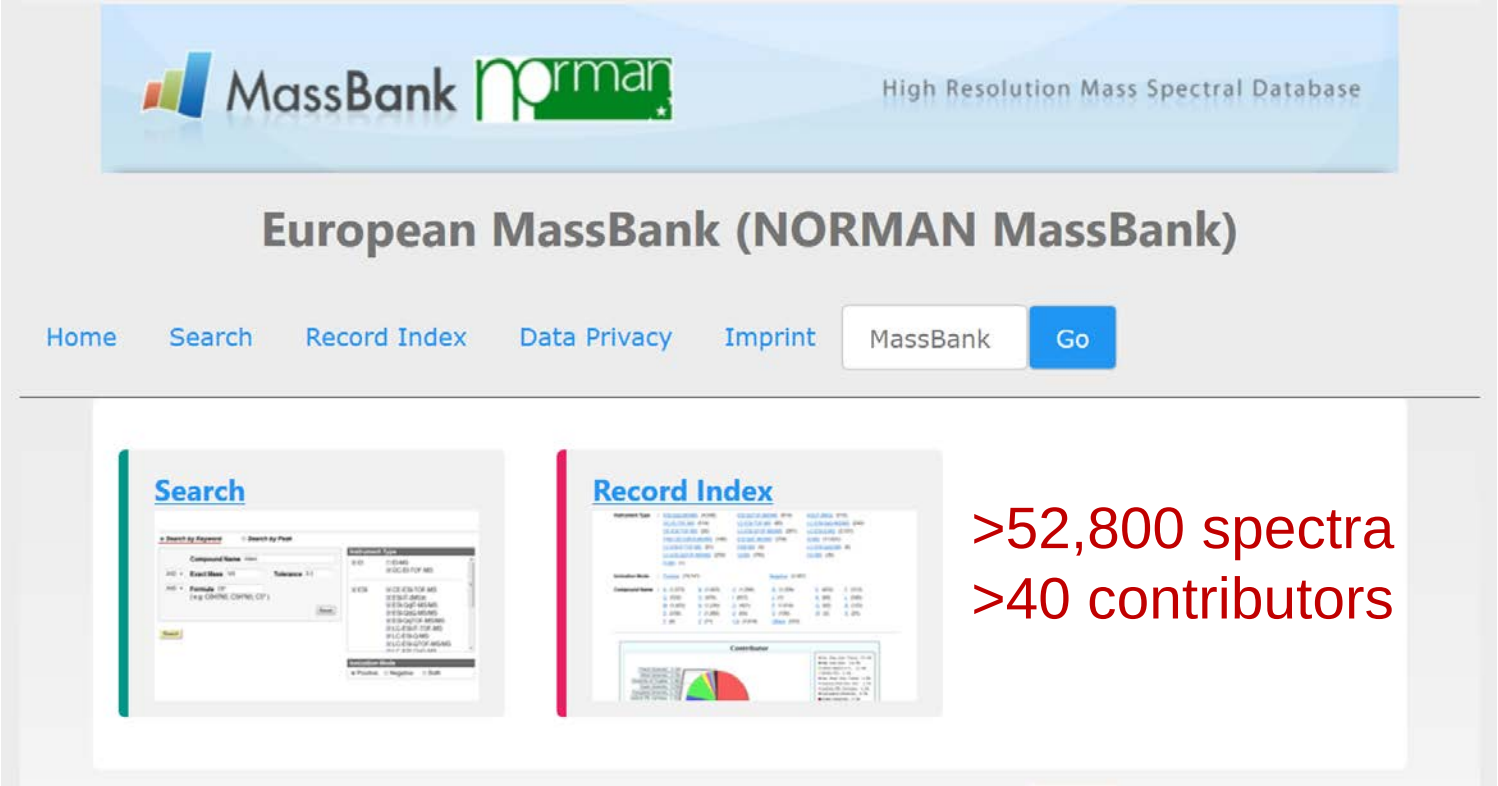


Frainay, C. *et al.* (2018) "Mind the Gap: …" Metabolites: http://www.mdpi.com/2218-1989/8/3/51

MassBank EU

http://massbank.eu/MassBank

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



MassBank-data validation status

build passing

Creating High-Quality Mass Spectra

RMassBank

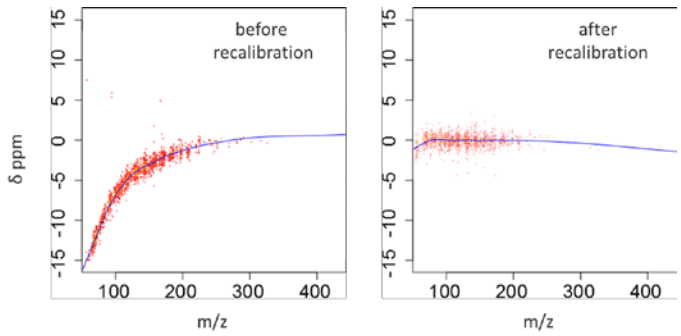
LC-MS/MS raw data

compound list: SMILES, name

online resources: CTS, CACTUS

Automatic MS and MS/MS Recalibration and Clean-up Remove interfering peaks

Spectral Annotation with - Experimental Details - Compound Information



MassBank records

structure files

nprman

MassBank.eu

https://github.com/MassBank/RMassBank/

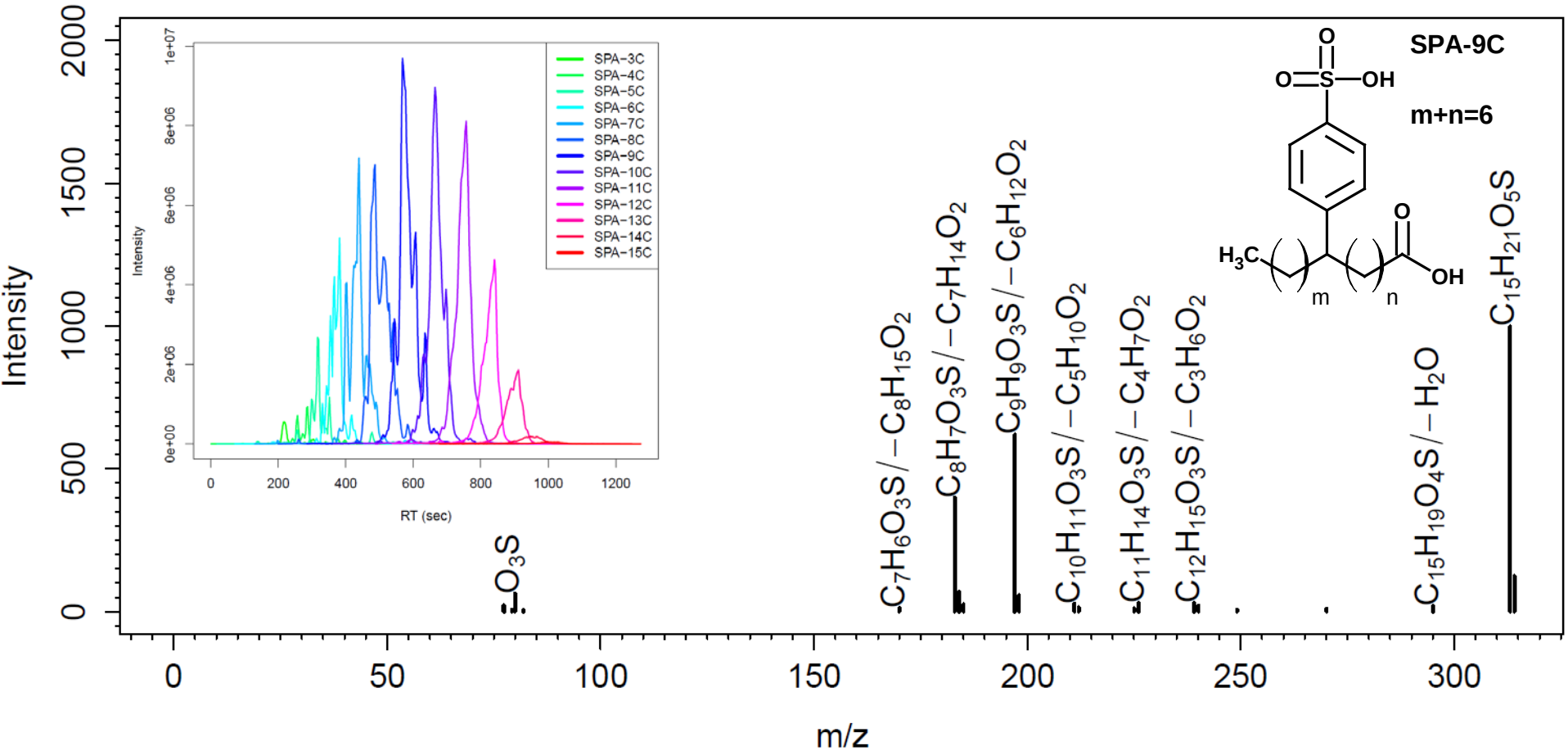
http://bioconductor.org/packages/RMassBank/

Stravs, Schymanski, Singer and Hollender, 2013, Journal of Mass Spectrometry, 48, 89–99. DOI: 10.1002/jms.3131

Communicating Mass Spectra for Mixtures

Chromatography and MS/MS Annotation

https://github.com/MassBank/RMassBank/



Formulas: http://sourceforge.net/projects/genform/
Meringer *et al,* 2011, *MATCH* 65, 259-290
Data: Schymanski *et al*. 2014, ES&T, 48: 1811-1818. DOI: 10.1021/es4044374

Literature: LIT00034,35
Sample: ETS00002
Standard: ETS00016,17,19,20

Stravs *et al*. (2013), *J. Mass Spectrom,* 48(1):89-99. DOI: 10.1002/jms.3131

Suspect Screening: Different Approaches

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

Suspect List

TARGET ANALYSIS

SUSPECT SCREENING

Targets found

Suspects found

- Search in ***mass spectral*** libraries
- Screen for ***predicted*** transformation products of known parent compounds
- Look for ***"well known"*** substances without reference standards
- Screen for ***known*** homologue series

Candidate selection (retention time, MS/MS, calculated properties)

Confirmation and quantification of compounds present

Suspect Screening: Benzotriazole TPs

HPLC separation and HR-MS/MS

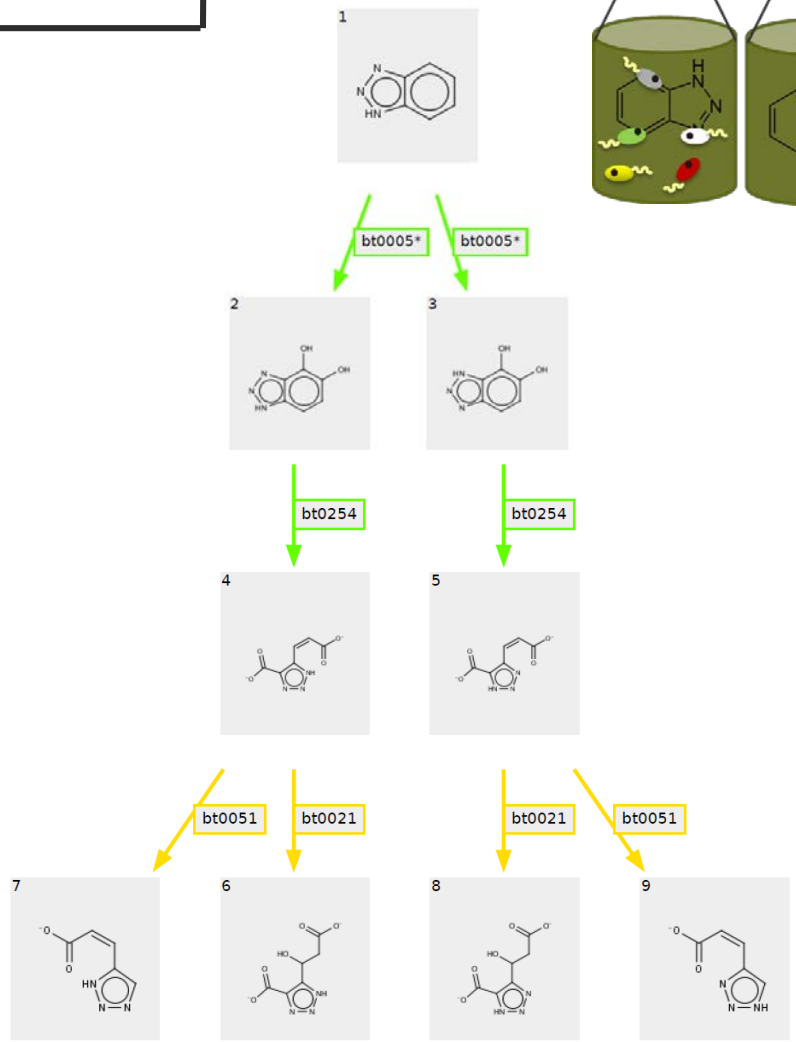
28 Suspects

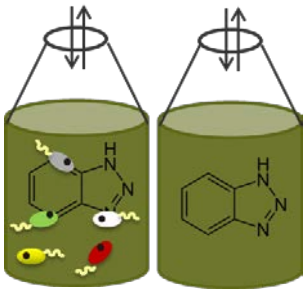
SUSPECT SCREENING

11 masses for
6 suspect formulas

7 with MS/MS
1 reference std.

1 TP confirmed
1 TP “likely”, no std.





[UM-PPS]
↓
Eawag-PPS
↓
enviPath

Huntscha *et al.* 2014, *ES&T,* 48(8), 4435-4443. DOI: 10.1021/es405694z

Suspect Screening: Benzotriazole TPs

HPLC separation and HR-MS/MS

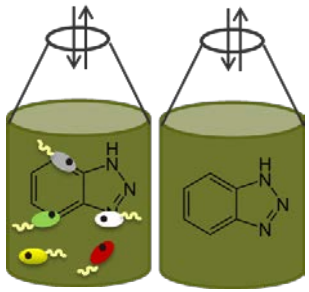
28 Suspects

SUSPECT SCREENING

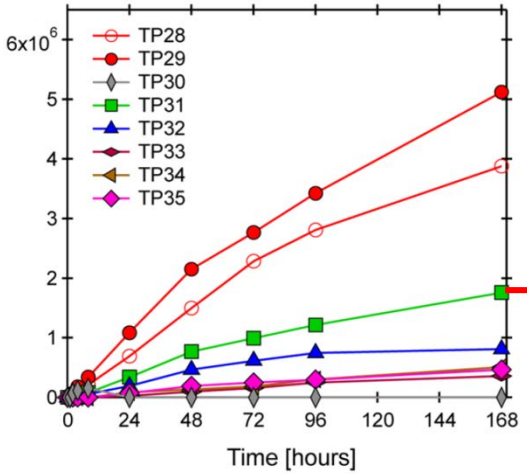
11 masses for
6 suspect formulas

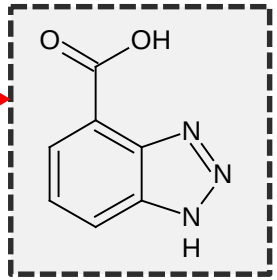
7 with MS/MS
1 reference std.

1 TP confirmed
1 TP "likely", no std.

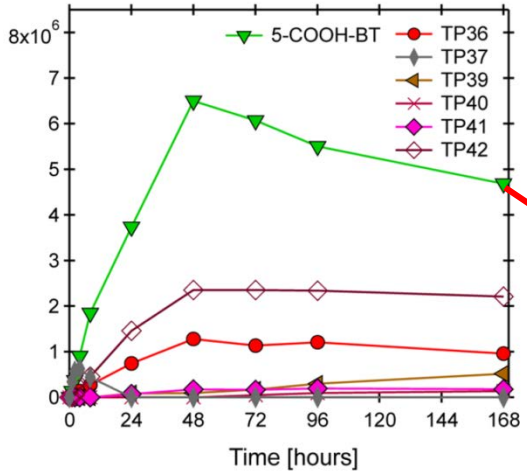


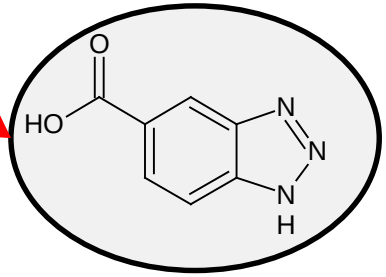
[UM-PPS] → Eawag-PPS → enviPath





- Predicted with Eawag-PPS
- No standard
- Not in databases (at that time)





- Confirmed with reference std.
- Observed in WWTP effluents

Huntscha *et al.* 2014, *ES&T,* 48(8), 4435-4443. DOI: 10.1021/es405694z

Suspect Screening: Different Approaches

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

Suspect List

TARGET ANALYSIS

SUSPECT SCREENING

Targets found

Suspects found

- Search in ***mass spectral*** libraries
- Screen for ***predicted*** transformation products of known parent compounds
- Look for ***"well known"*** substances without reference standards
- Screen for ***known*** homologue series

Candidate selection (retention time, MS/MS, calculated properties)

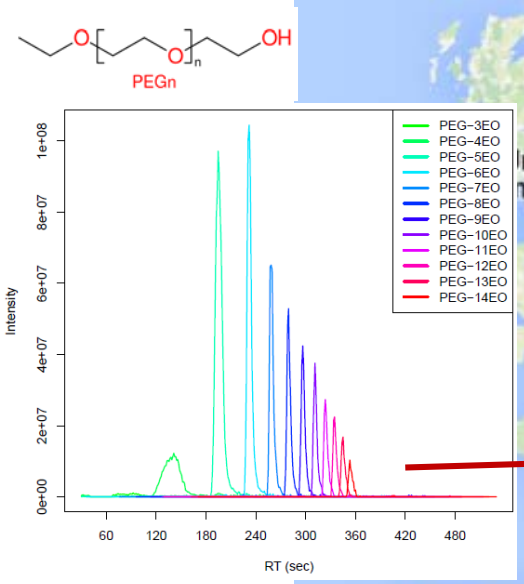
Confirmation and quantification of compounds present

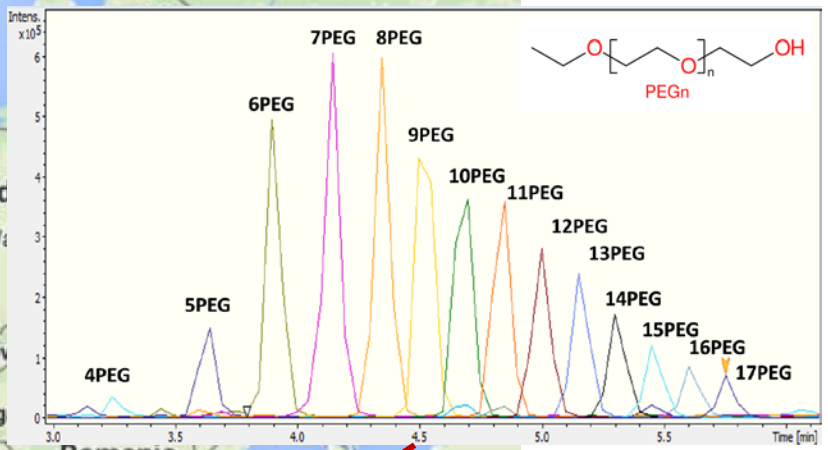
(European) Environmental Community (subset!)

Schymanski *et al.* 2015, ABC, DOI: 10.1007/s00216-015-8681-7

European (World-)Wide Exchange of Suspects

NORMAN Suspect List Exchange:
https://www.norman-network.com/nds/SLE/





Schymanski et al. 2015, *ABC*, DOI: 10.1007/s00216-015-8681-7

Suspect Screening – NORMAN SLE

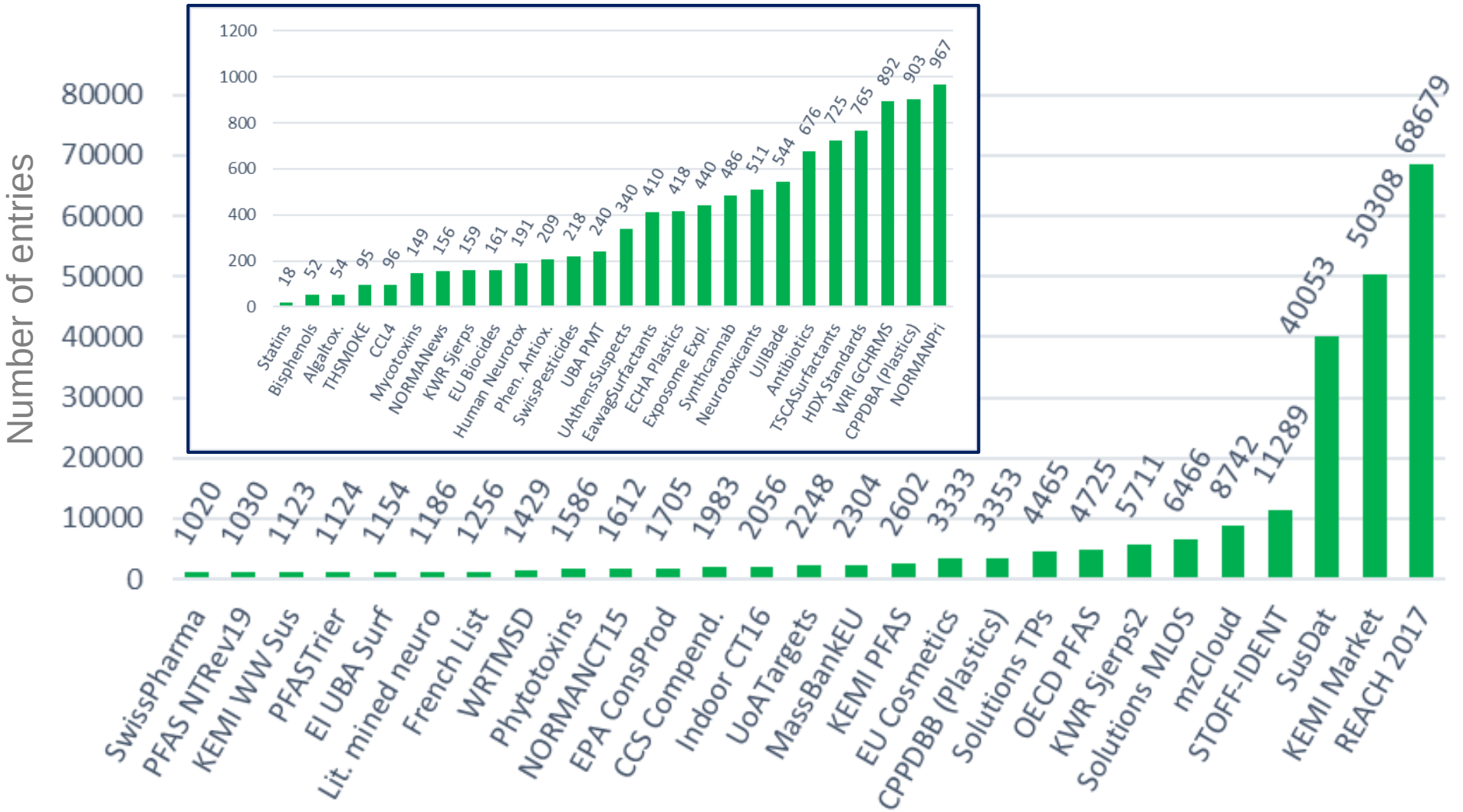
NORMAN Suspect List Exchange – NORMAN SLE

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

No.	Abbreviation	Description	Link to full list	Link to InChIKey list	References
S0	SUSDAT	**Merged NORMAN Suspect List: SusDat**	Interactive Data table (updating…) CompTox SUSDAT List	MS-ready InChIKeys (1/03/2018)	A merged list of >40,000 structures from suspect lists. See interactive version. Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. *Work in progress ... please report any issues!* DOI: 10.5281/zenodo.2664077
S1	MASSBANK	**NORMAN Compounds in MassBank**	CSV, XLSX with Fragments (3/10/2017) CompTox MassBank EU Reference List CompTox MassBank EU Special Cases CompTox Fragment Download	MassBankEUInChIKeys (17/06/2019)	www.massbank.eu Stravs *et al.* 2013. DOI: 10.1002/jms.3131 DOI: 10.5281/zenodo.2621390
S2	STOFFIDENT	**HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances**	STOFF-IDENT Contents (6/09/2017) CompTox STOFF-IDENT List	STOFF-IDENT InChIKeys (6/09/2017)	The database enables the search for exact masses from target or unknown lists and the automatic use of a Retention Time Index. See: https://www.lfu.bayern.de/stoffident/#!home (single search for free; batch search after free registration). DOI: 10.5281/zenodo.2621451
S3	NORMANCT15	**NORMAN Collaborative Trial Targets and Suspects**	LC-MS: CSV, XLSX (3/10/2017) GC-MS: CSV, XLSX (3/10/2017) CompTox NORMANCT15 List	LC-MS InChIKeys (31/10/2016) GC-MS InChIKeys	Schymanski *et al.* 2015. DOI: 10.1007/s00216-015-8681-7 DOI: 10.5281/zenodo.2621478

Suspect Screening

- NORMAN Suspect List Exchange (NORMAN-SLE): 58 lists!

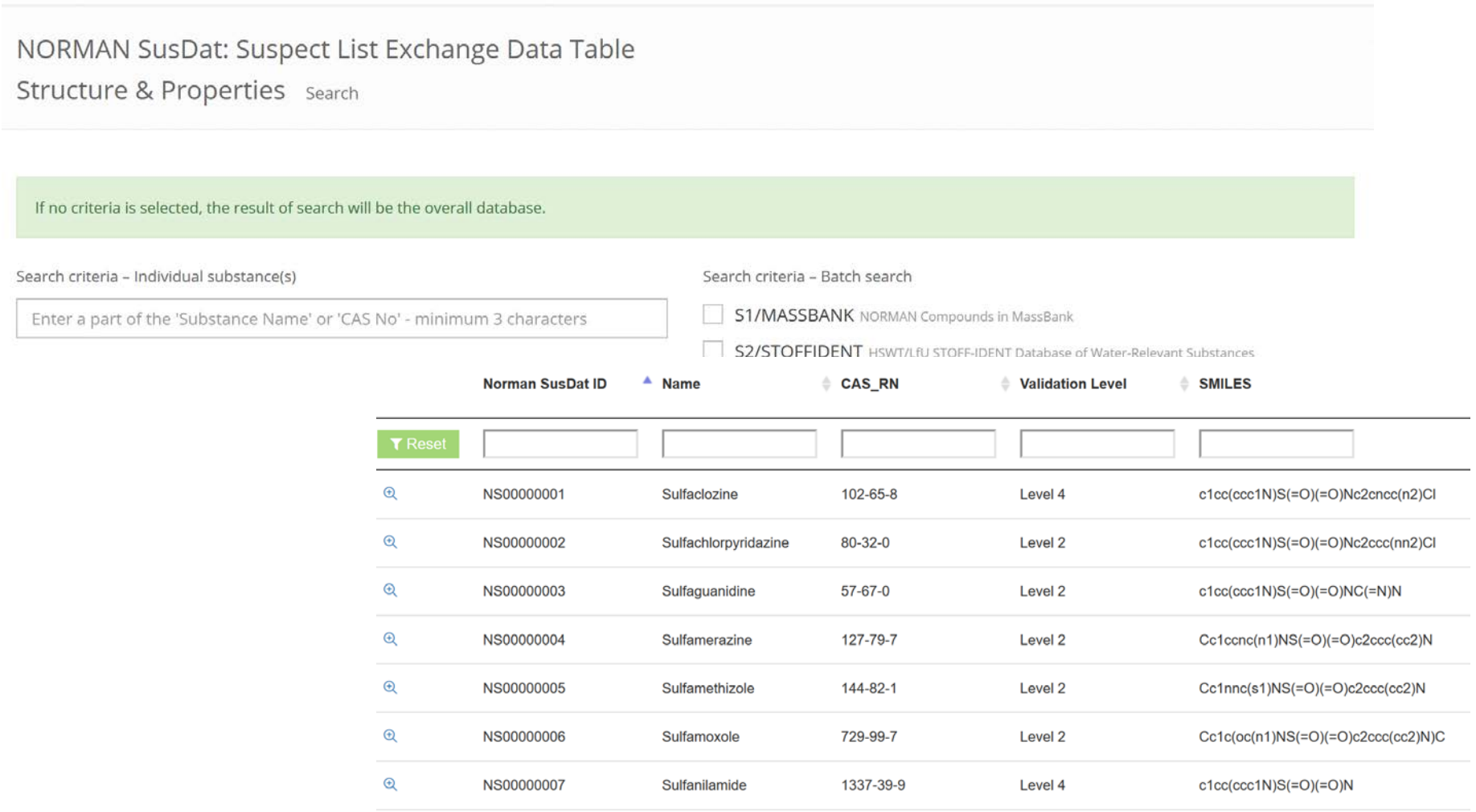


Schymanski *et al*. (in prep.) https://www.norman-network.com/nds/SLE/

NORMAN SusDat (merged list)

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

https://comptox.epa.gov/dashboard/chemical_lists/susdat



Schymanski, Aalizadeh *et al*. in prep; https://www.researchgate.net/project/Supporting-Mass-Spectrometry-Through-Cheminformatics

Suspect Screening: Different Approaches

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

Suspect List

TARGET ANALYSIS

SUSPECT SCREENING

Targets found

Suspects found

- Search in ***mass spectral*** libraries
- Screen for ***predicted*** transformation products of known parent compounds
- Look for ***"well known"*** substances without reference standards
- Screen for ***known*** homologue series

Candidate selection (retention time, MS/MS, calculated properties)

Confirmation and quantification of compounds present

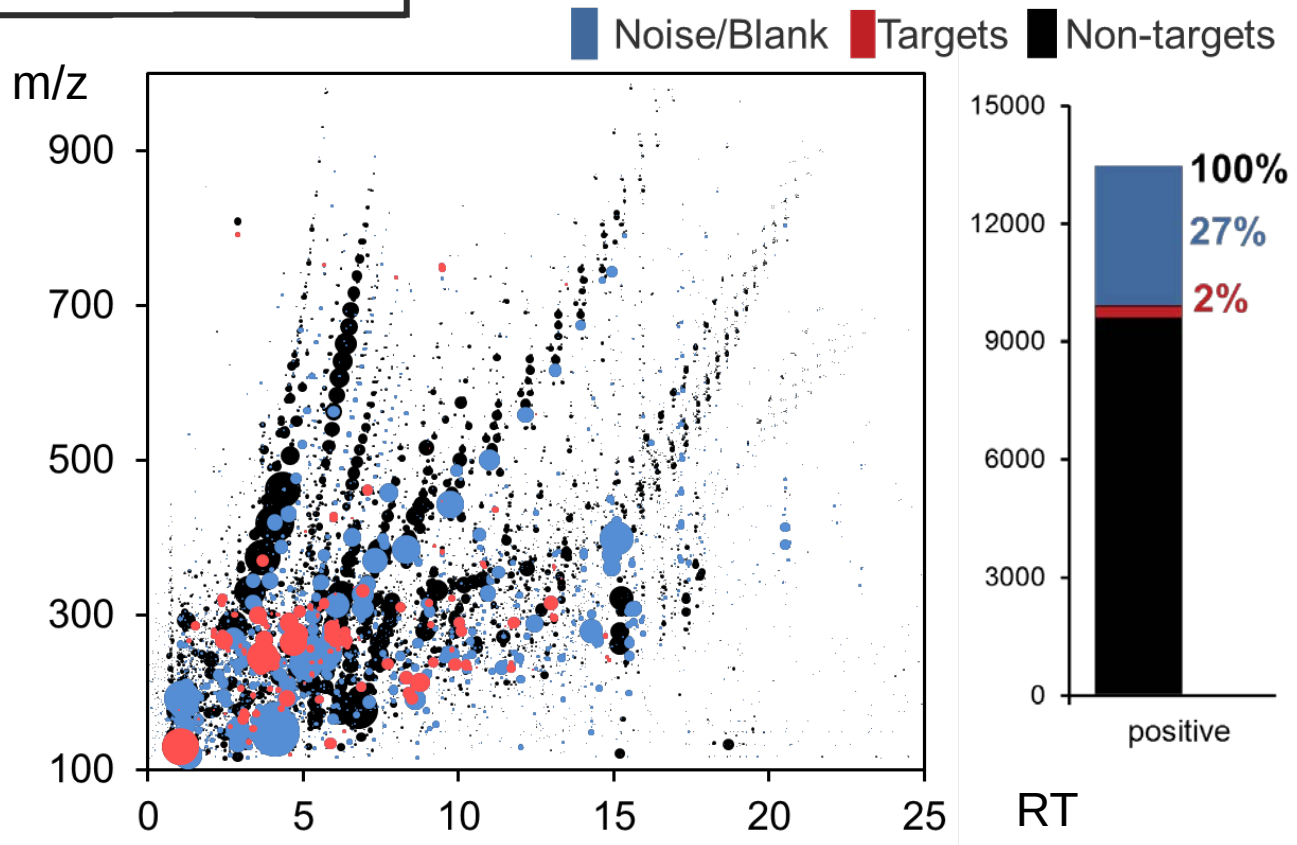
RECAP: Target Analysis: Status Quo (>364 targets)

Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Target List

TARGET ANALYSIS

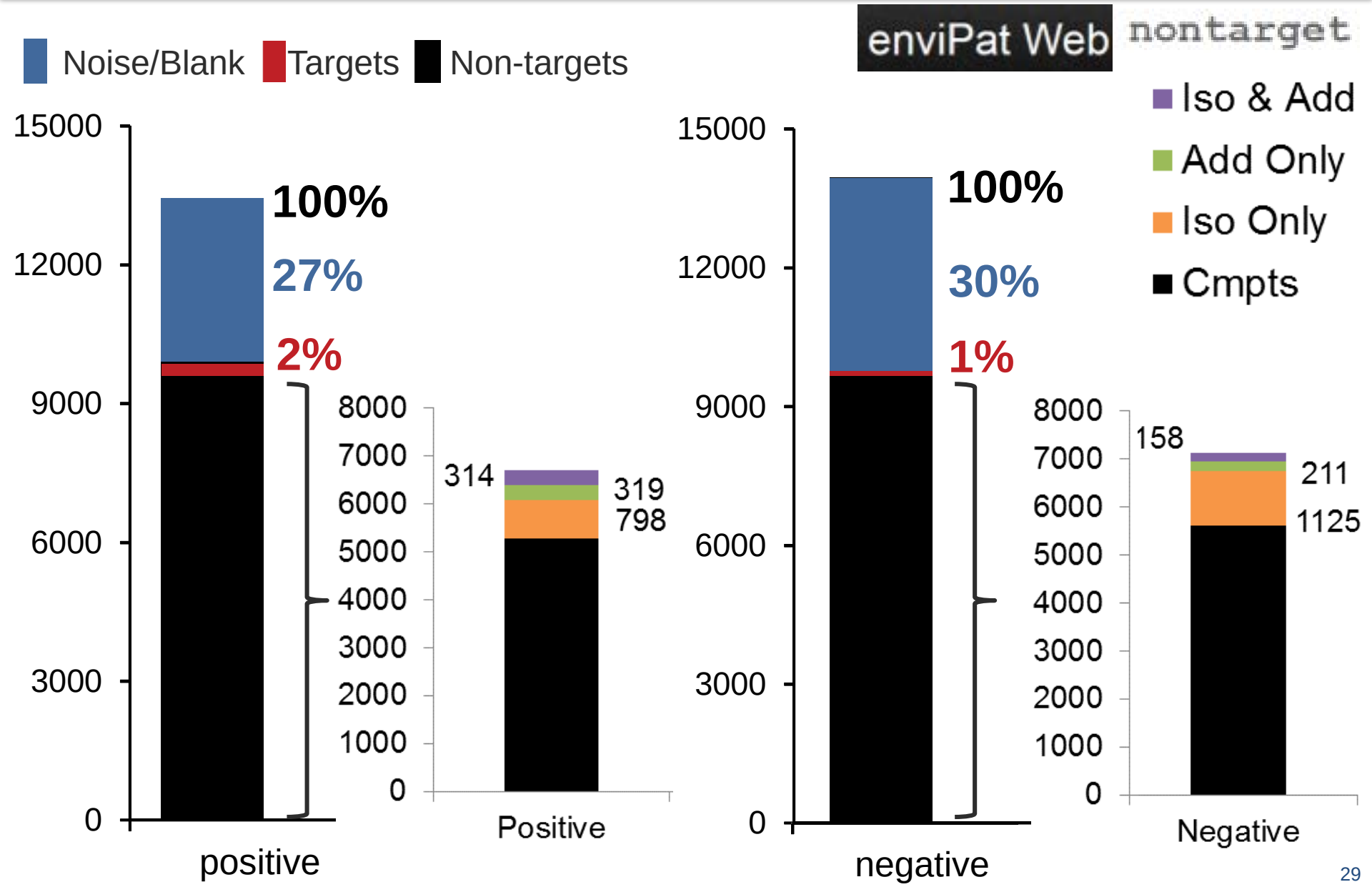
Targets found



Confirmation and quantification of compounds present

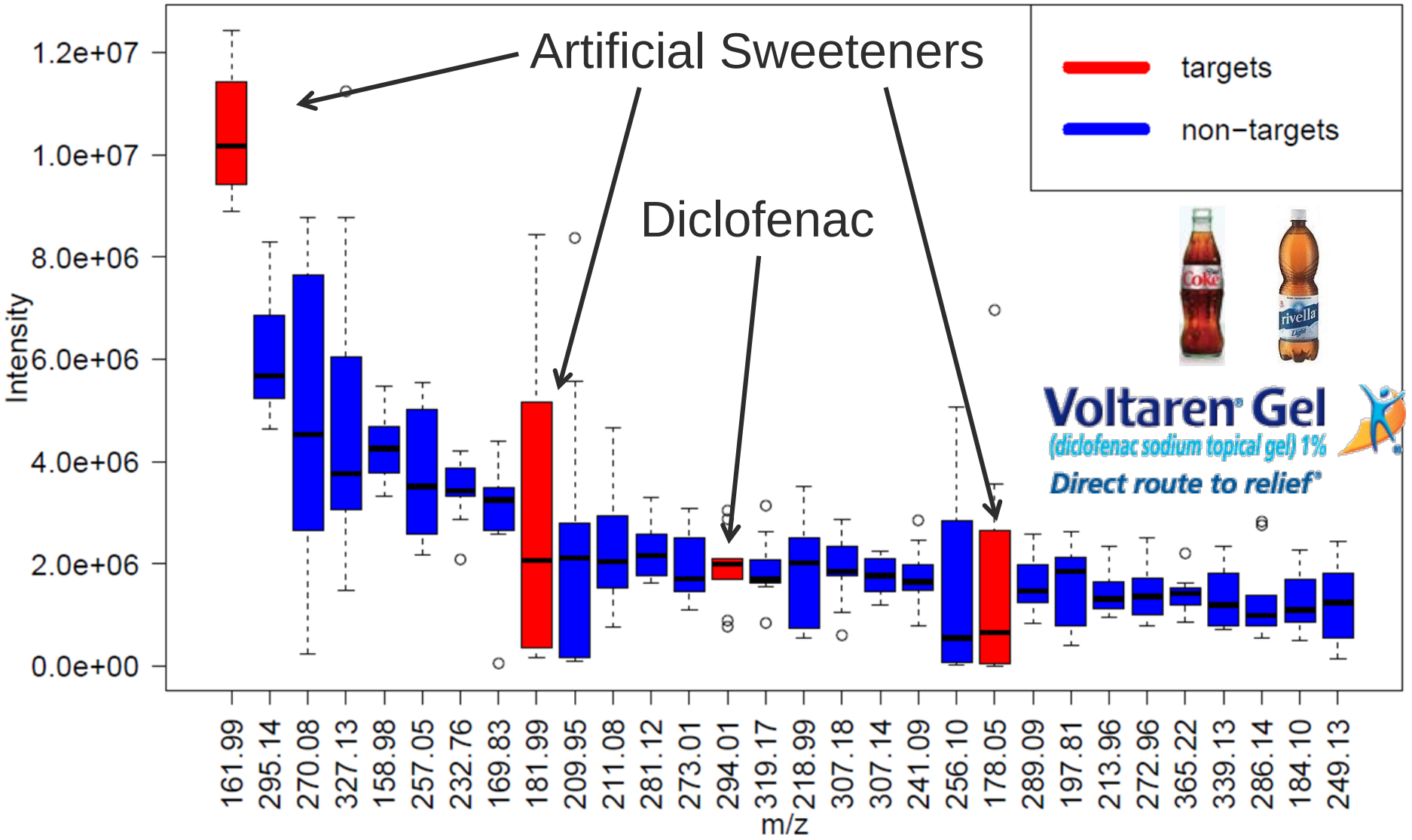
Schymanski *et al*. (2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

Grouping Isotopes and Adducts



Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

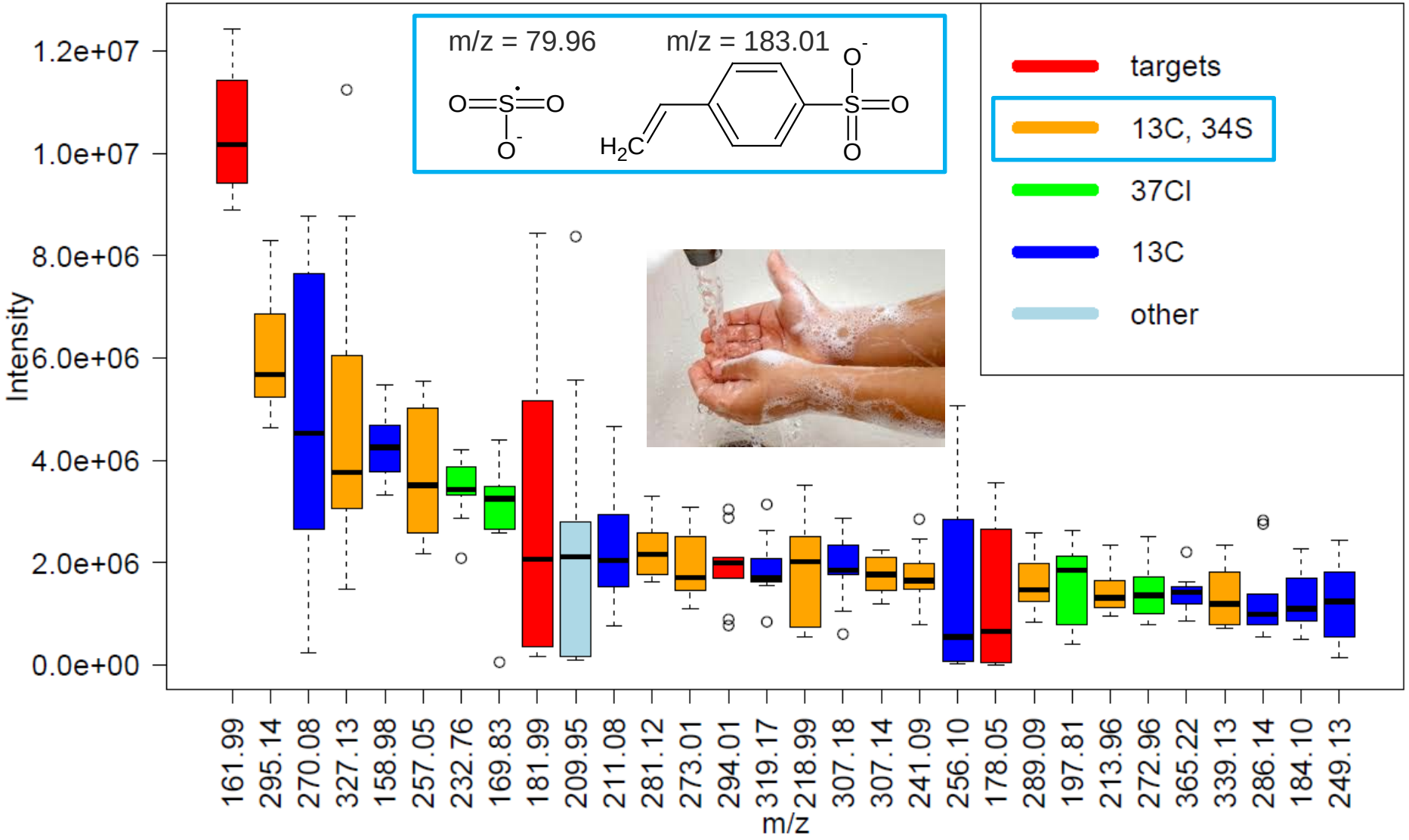
Recap: Swiss Wastewater: Top 30 Peaks (ESI-)



Pictures: www.coca-cola-com; www.rivella.ch; www.voltargengel.com

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

Swiss Wastewater: Top 30 Peaks (ESI-)



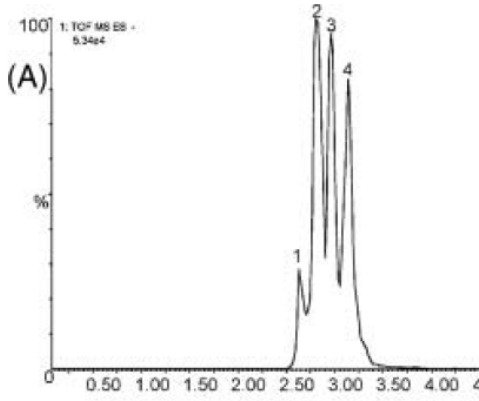
Picture: www.momsteam.com

Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

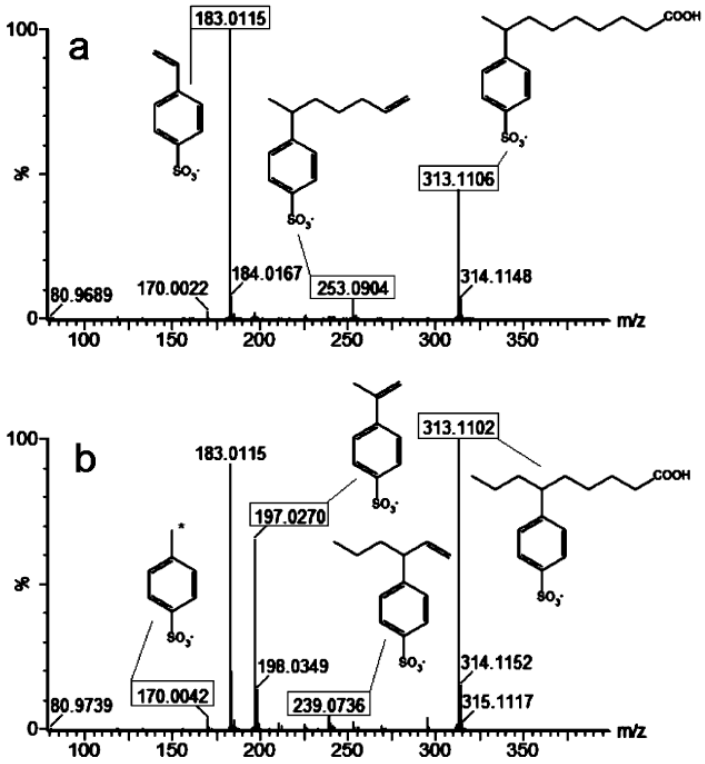
Surfactant Screening From Literature

Literature sources

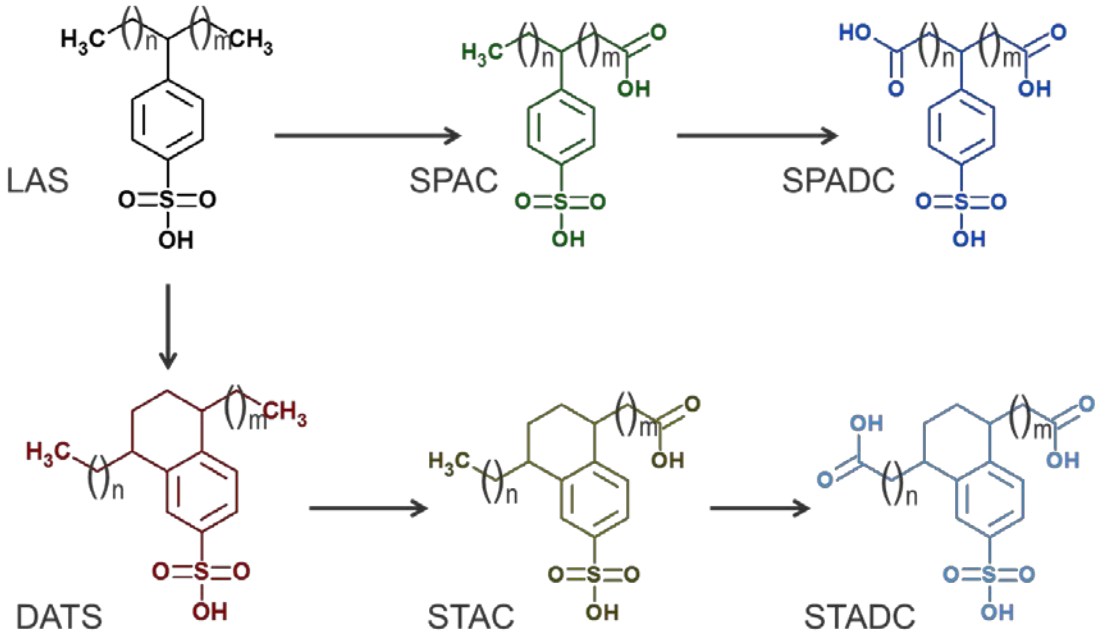
- Formulas, masses (ions), retention times and intensities
- Spectra of selected compounds (different instruments)



Gonzalez et al. Rapid Comm. Mass Spec. 2008, 22: 1445-54



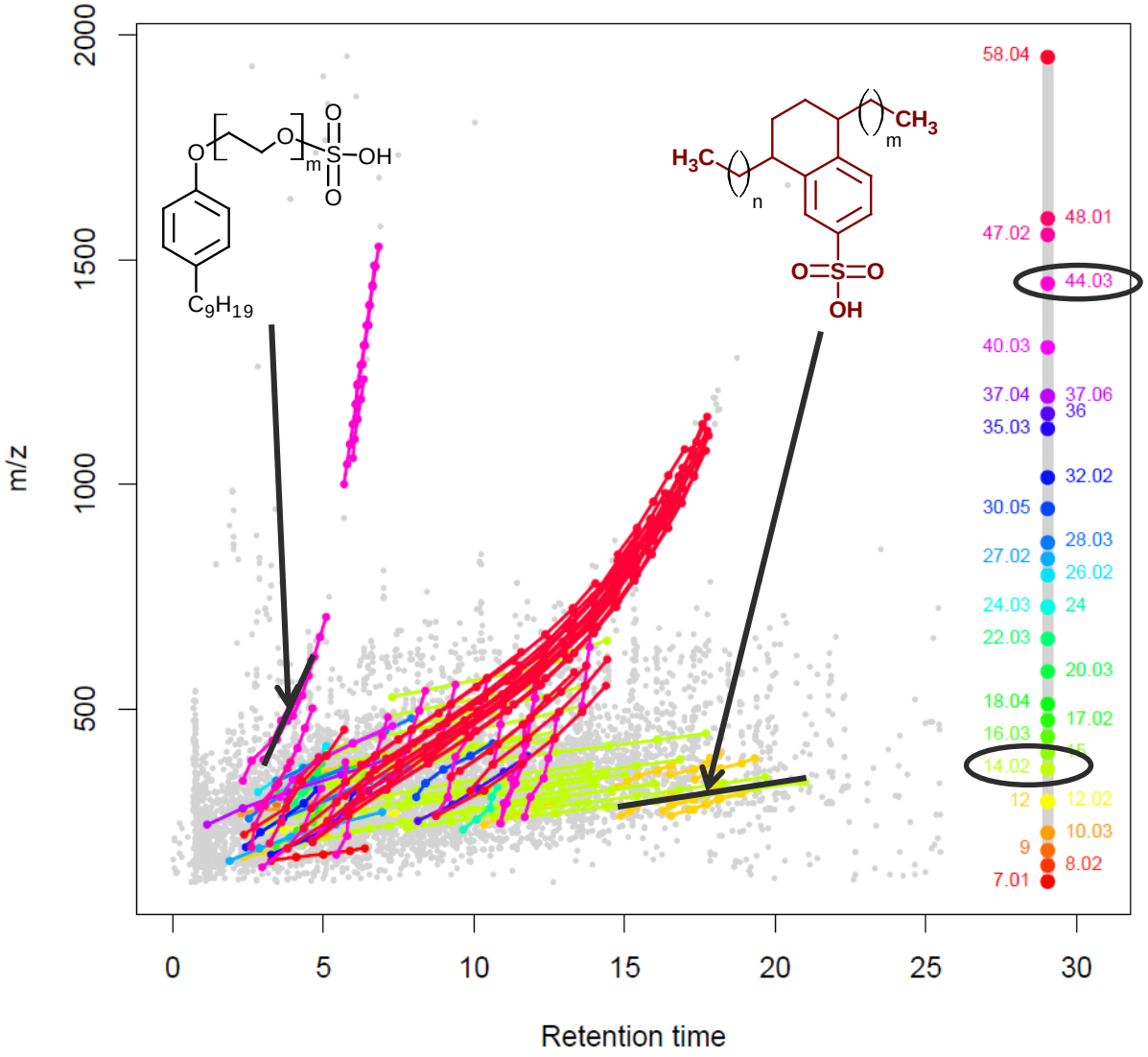
Lara-Martin et al. EST. 2010, 44: 1670-1676



Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

Homologous Series Detection

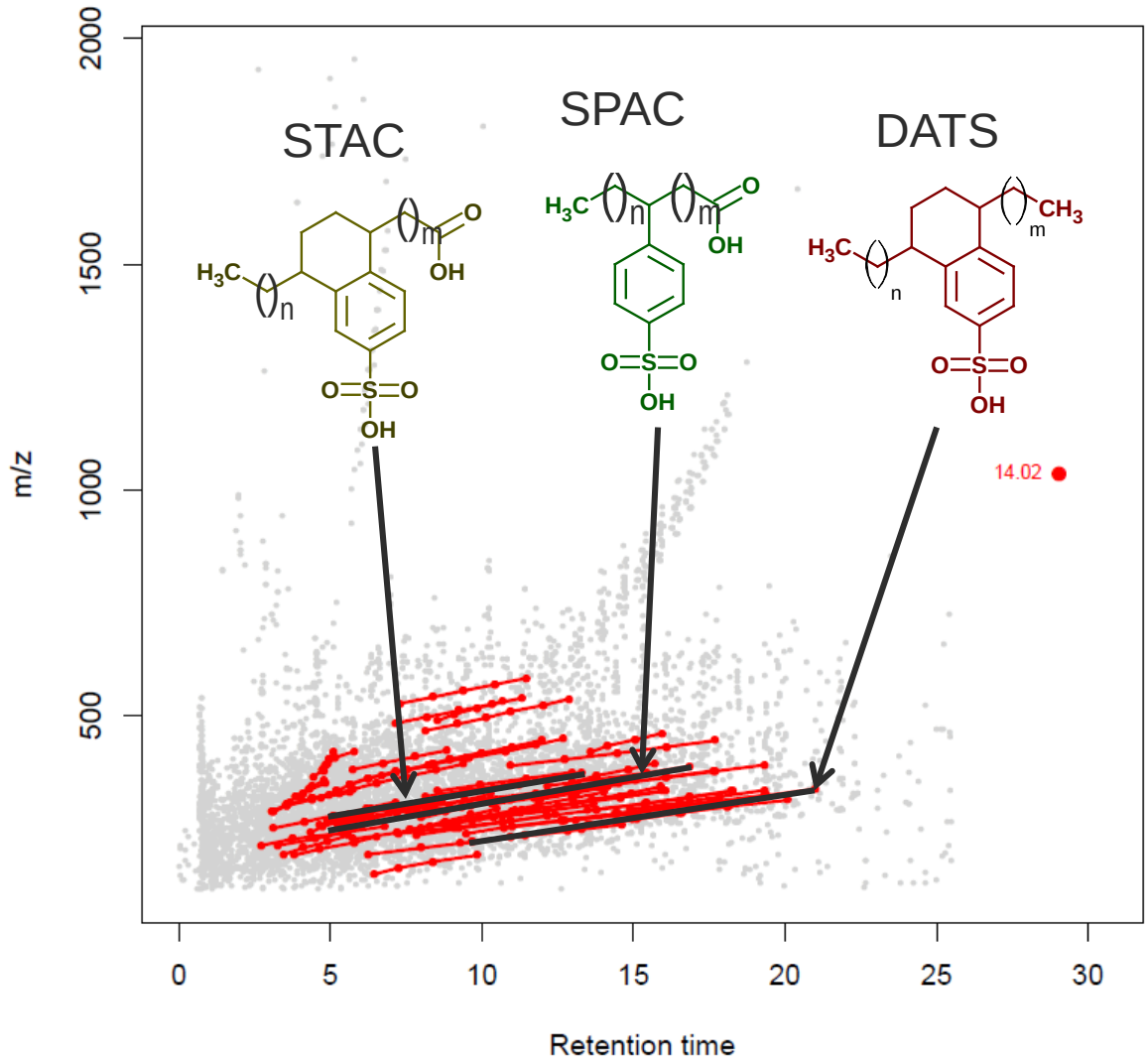
Search for discrete mass differences



http://www.envihomolog.eawag.ch/

M. Loos & H Singer, 2017. J. Cheminf. DOI: 10.1186/s13321-017-0197-z & Schymanski et al. 2014, ES&T DOI: 10.1021/es4044374

Homologous Series Detection



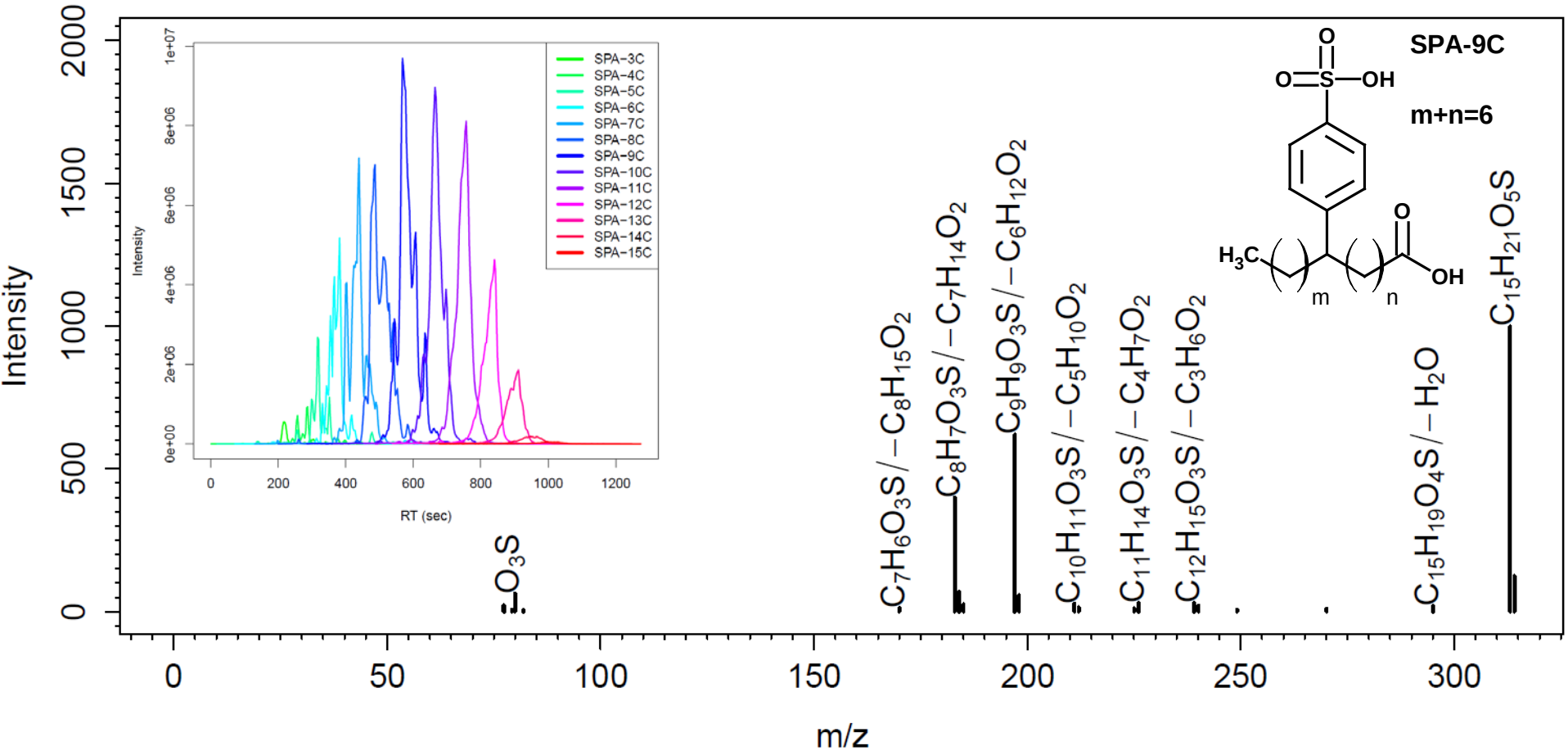
http://www.envihomolog.eawag.ch/

M. Loos & H Singer, 2017. J. Cheminf. DOI: 10.1186/s13321-017-0197-z & Schymanski et al. 2014, ES&T DOI: 10.1021/es4044374

Supporting Evidence for Homologues

Chromatography and MS/MS Annotation

https://github.com/MassBank/RMassBank/

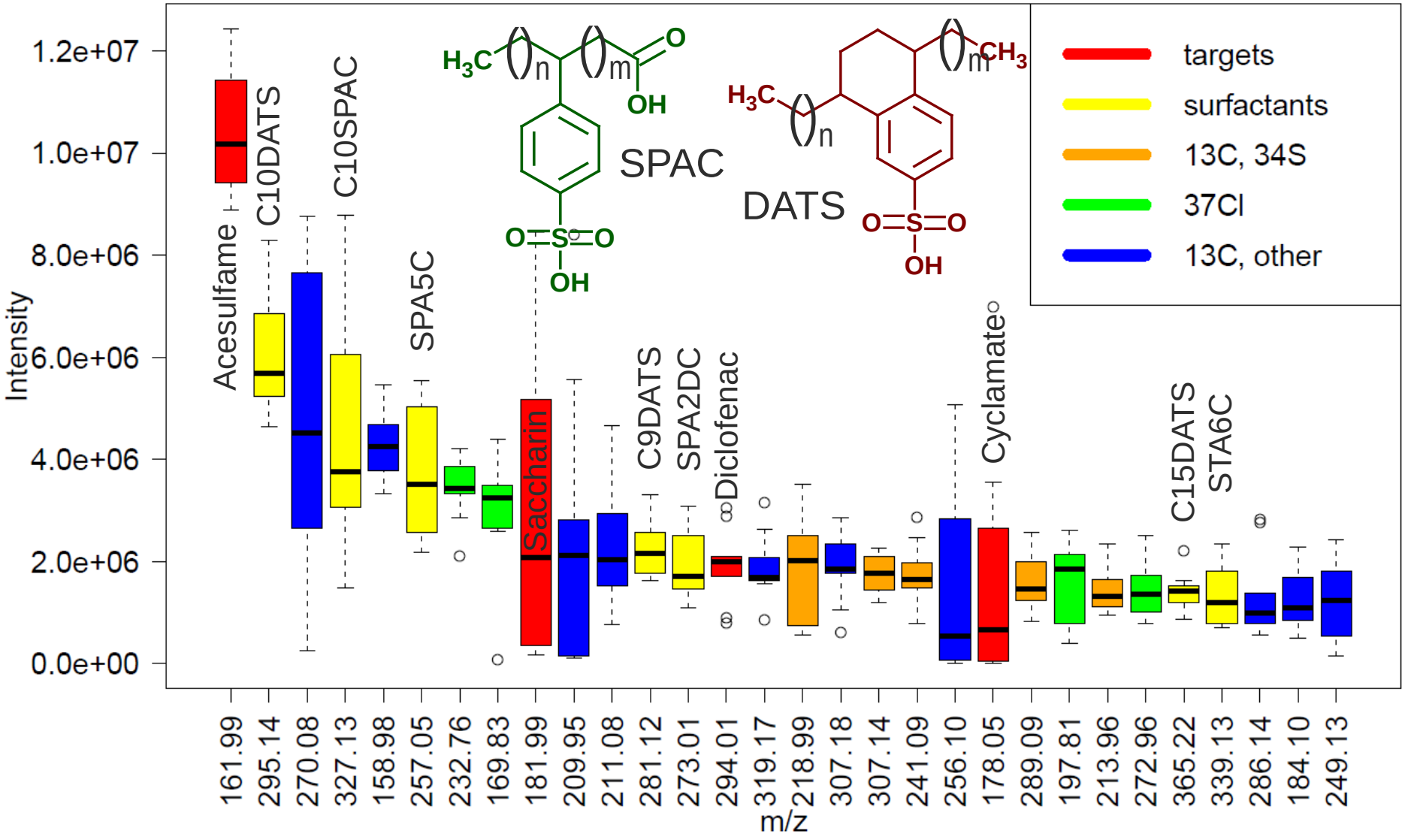


Formulas: http://sourceforge.net/projects/genform/
Meringer *et al,* 2011, *MATCH* 65, 259-290
Data: Schymanski *et al*. 2014, ES&T, 48: 1811-1818. DOI: 10.1021/es4044374

Literature: LIT00034,35
Sample: ETS00002
Standard: ETS00016,17,19,20

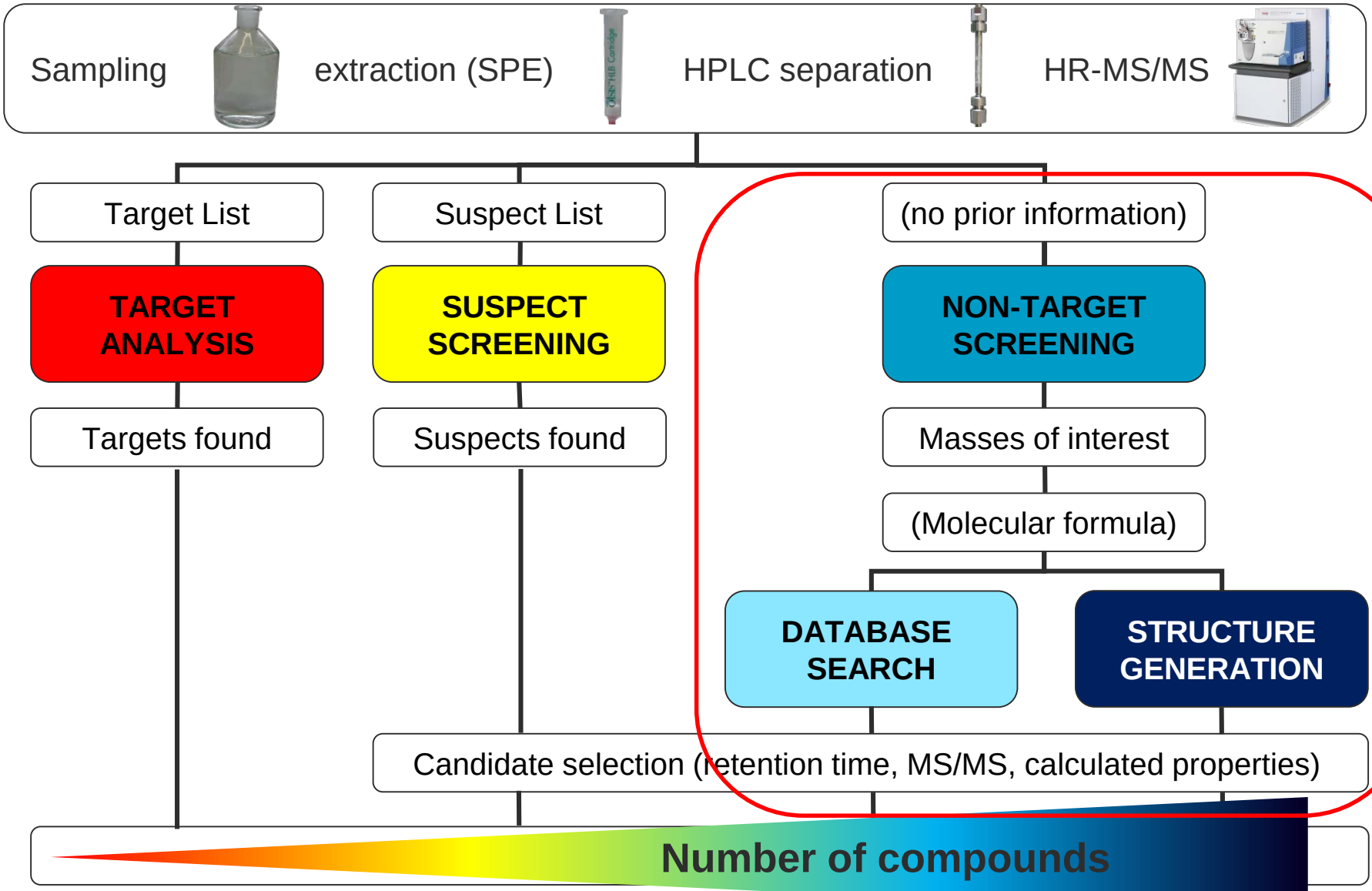
Stravs *et al*. (2013), *J. Mass Spectrom,* 48(1):89-99. DOI: 10.1002/jms.3131

Swiss Wastewater: Top 30 Peaks (ESI-)



Schymanski *et al*. *(*2014), *ES&T,* 48: 1811-1818. DOI: 10.1021/es4044374

What about Non-Target Screening?



Sampling  extraction (SPE)  HPLC separation  HR-MS/MS 

Conversion (Proteowizard) and Peak Picking (enviPick, xcms, MZmine, …)

Detection of blank/blind/noise/internal standards; time trend analysis (enviMass)

Target List

TARGET ANALYSIS

(enviMass, vendor software)

Suspect List (e.g. NORMAN, LMC, Eawag-PPS, ReSOLUTION)

SUSPECT SCREENING

NON-TARGET SCREENING

Componentization (nontarget)

Prioritization (enviMass)

Masses of interest

Molecular formula determination (enviPat, GenForm)

MS/MS Extraction (RMassBank)

Non-target identification (MetFrag2.3, ReSOLUTION)

Interpretation, confirmation, peak inventory, confidence and reporting

Altenburger *et al*, 2019, Env. Sci. Europe. DOI: 10.1186/s12302-019-0193-1

MetFrag in 2010

eawag aquatic research solutions

mz [M-H]⁻
213.9637
± 5 ppm

ChemSpider
Search and share chemistry

or

PubChem OPEN CHEMISTRY DATABASE

5 ppm
0.001 Da

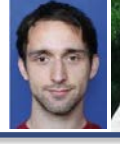
MetFrag

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

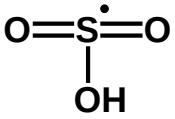
MetFrag – Adding RT + MetaData



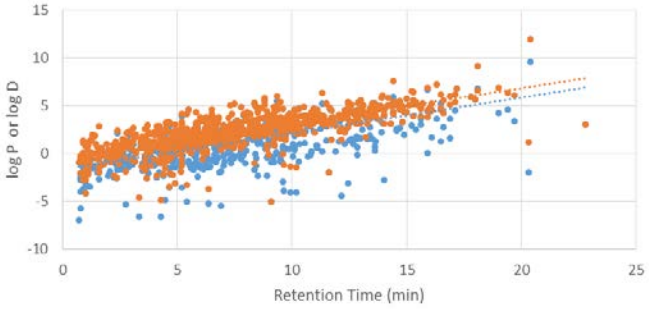
2010 => 2016

mz [M-H]⁻
213.9637
± 5 ppm

Elements: C,N,S
5 ppm
0.001 Da



RT: 4.54 min
355 InChI/RTs



ChemSpider
Search and share chemistry

or

PubChem OPEN CHEMISTRY DATABASE

References
External Refs
Data Sources
RSC Count
PubMed Count

Literature Patents

MetFrag

Suspect Lists

TOFF IDENT
Chemistry Dashboard

MoNA
MassBank of North America
MassBank.eu

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 Evaluation


Test set of 473 Eawag Target Substances

	MetFrag 2010	MetFrag2.3 Fragments only	***MetFrag2.3 +References +Retention time***
ChemSpider[1]			
Top 1 Ranks	73	105	***420***
% Top 1 Ranks	15 %	22 %	***89 %***
PubChem[2]			
Top 1 Ranks	-	30	***336***
% Top 1 Ranks	-	6 %	***71 %***

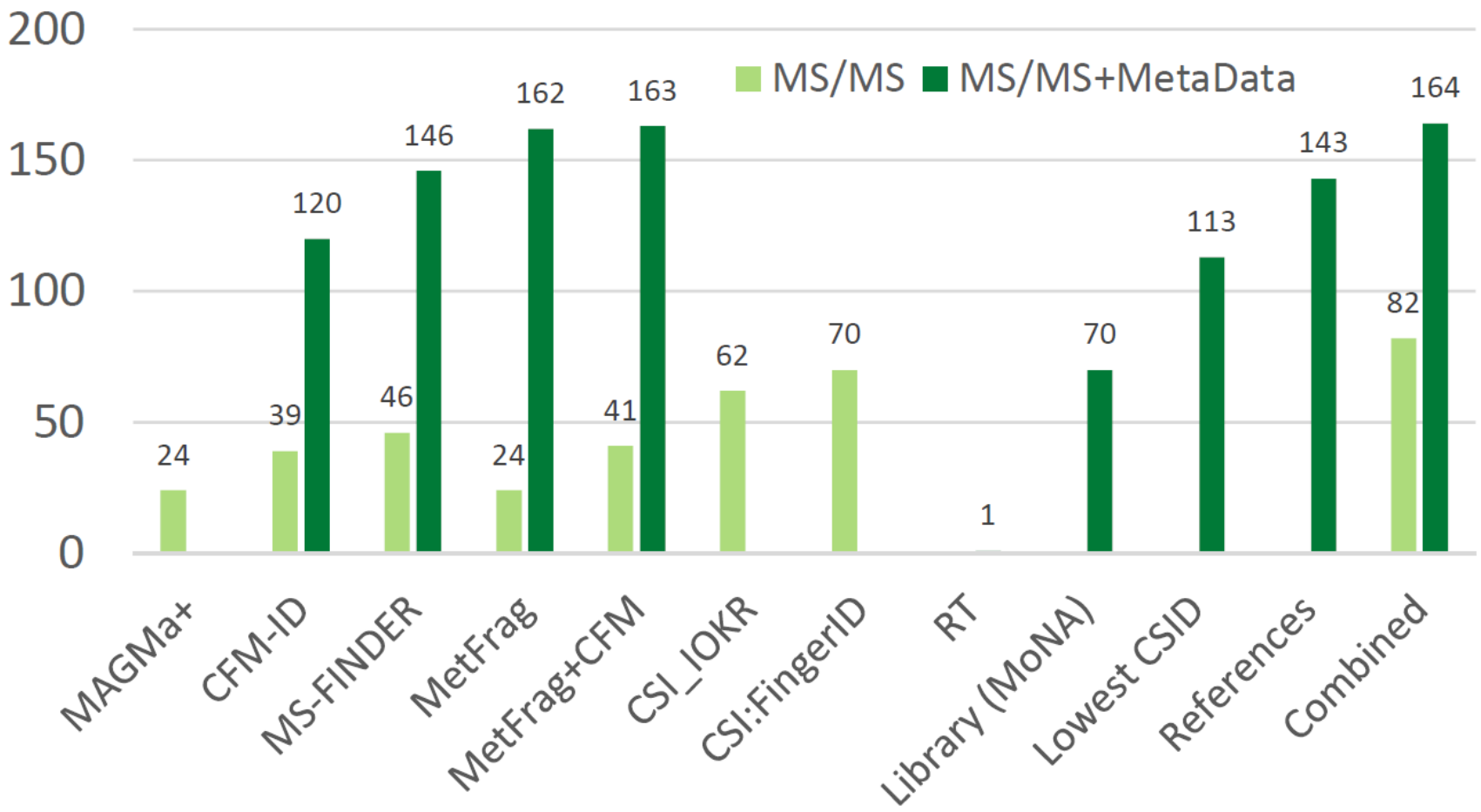
Similar results with 3 independent datasets of 310, 289 and 225 substances from Eawag and UFZ (www.massbank.eu)

[1]www.chemspider.com; ~34 million entries
[2]https://pubchem.ncbi.nlm.nih.gov/; ~74 million entries

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

State of the Art in Small Molecule Identification

Metadata is critical to improving annotation of known unknowns!

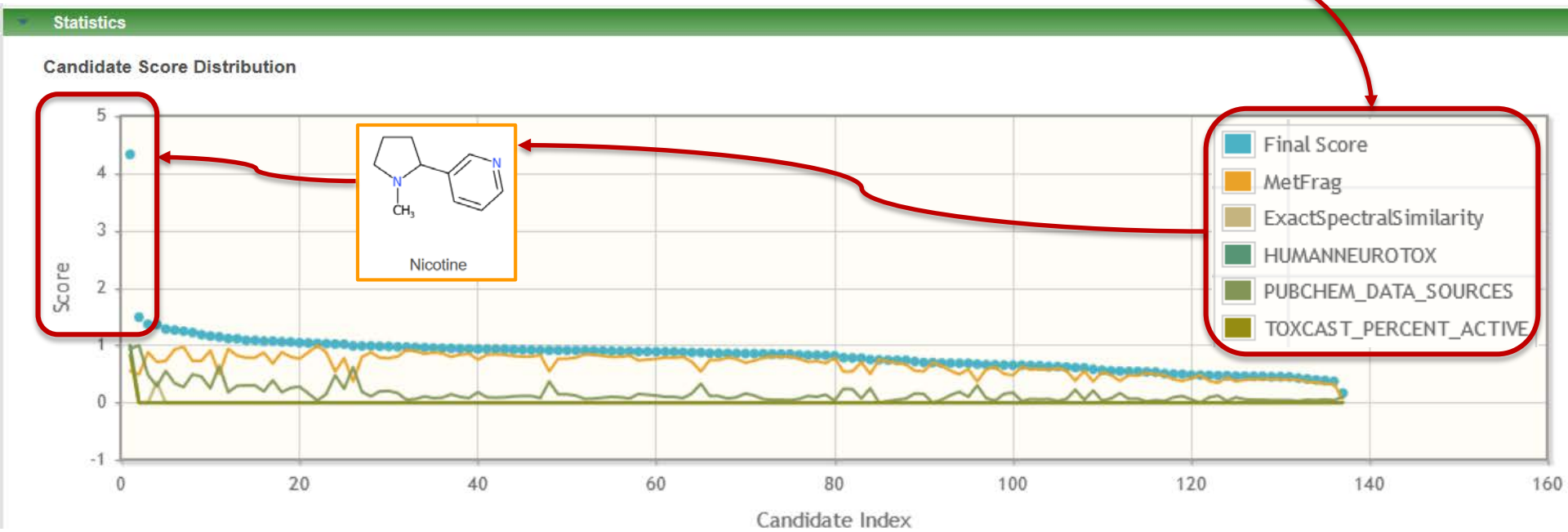


Schymanski *et al*, 2017, J Cheminf., DOI: 10.1186/s13321-017-0207-1

www.casmi-contest.org

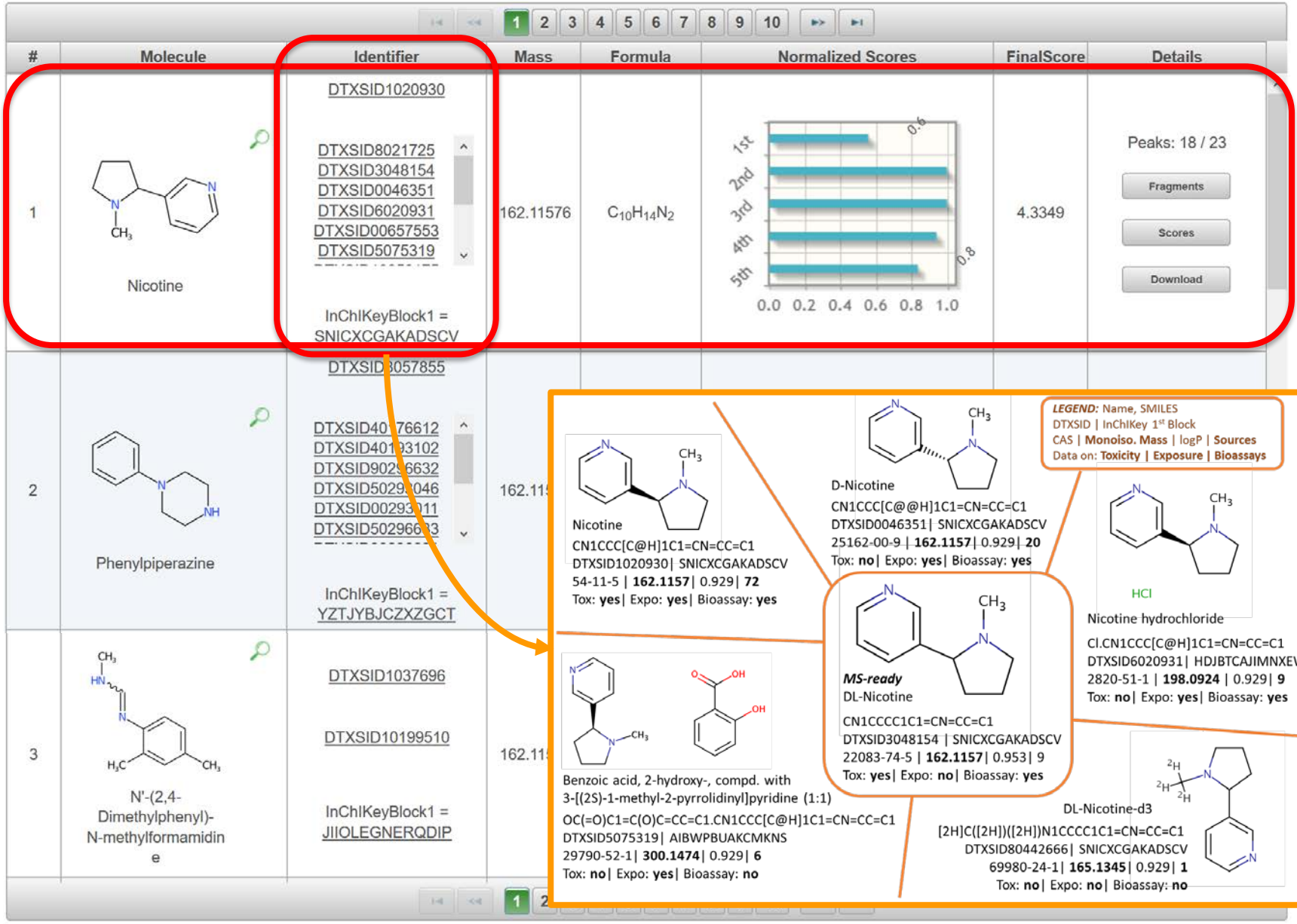
Connecting Resources in MetFrag

Combined evidence clearly highlights potential neurotoxicant among chemical candidates



https://msbi.ipb-halle.de/MetFragBeta/ ; https://comptox.epa.gov/dashboard/ ; https://massbank.eu ; http://normandata.eu/

“MS-ready” Form for MetaData in MetFrag



McEachran *et al*. 2018, DOI: 10.1186/s13321-018-0299-2; Schymanski & Williams, 2017 *ES&T* DOI: 10.1021/acs.est.7b01908

“MS-ready” Form for MetaData in MetFrag

MetFrag

https://msbi.ipb-halle.de/MetFragMSready/

Search

Results

1 2

#	Molecule	Identifier	Mass	Formula	Normalized Scores
1	(C10-C16) Alkylbenzenesulfonic acid	DTXSID2028723 DTXSID7059696 DTXSID5041647 InChIKeyBlock1 = UASQKKHYUPBQJR	298.16027	$C_{16}H_{26}O_3S$	1st 0.0
2	Alkylbenzenesulfonate, linear	DTXSID3020041 DTXSID70881146 InChIKeyBlock1 = KIIODTIGNUGDLO	298.16027	$C_{16}H_{26}O_3S$	1st 0.6, 2nd 0.0, 3rd 0.0, 4th 0.0
7	Benzenesulfonic acid, isodecyl-, compd. with 2,2'-iminobis[ethanol] (1:1)	DTXSID2070762 InChIKeyBlock1 = CKNPXKLNOLAXDF	298.16027	$C_{16}H_{26}O_3S$	1st 0.1, 2nd 0.0, 3rd 0.0, 4th 0.3, 5th 0.0

0.0 0.2 0.4 0.6 0.8 1.0

(C10-C16) Alkylbenzenesulfonic acid
68584-22-5 | DTXSID2028723

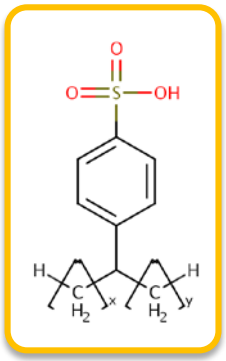
4-Decylbenzenesulfonic acid
140-60-3 | DTXSID7059696

Sodium 4-decylbenzenesulfonate
2627-06-7 | DTXSID5041647

Connecting and Enhancing Open Resources

- Sharing knowledge is a win-win situation

2014



MassBank.eu

NORMAN suspects

EPA Chemicals

2015: found in waters across Europe

2016: 1 datapoint cross-annotates 3072 in GNPS

Hits in GNPS MassIVE datasets:

Surfactants: http://goo.gl/7sY9Pf

2017: Early-Warning System is born

NORMAN NEWS

2018: Highlighted in

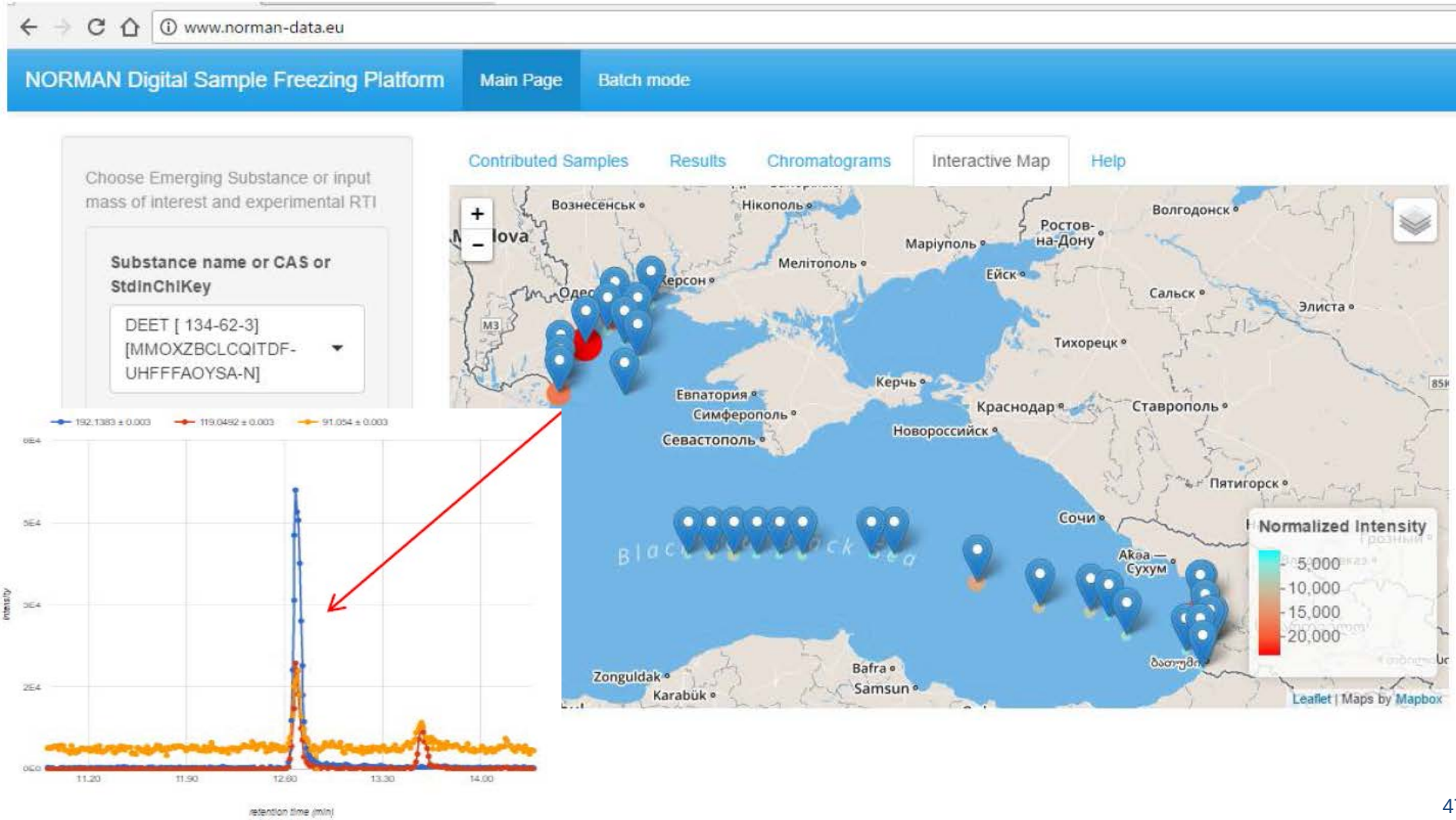
CHEMISTRY

Early warning about emerging contaminants

https://www.slideshare.net/EmmaSchymanski/small-molecules-in-big-data-analytica-munich

NORMAN Digital Sample Freezing Platform

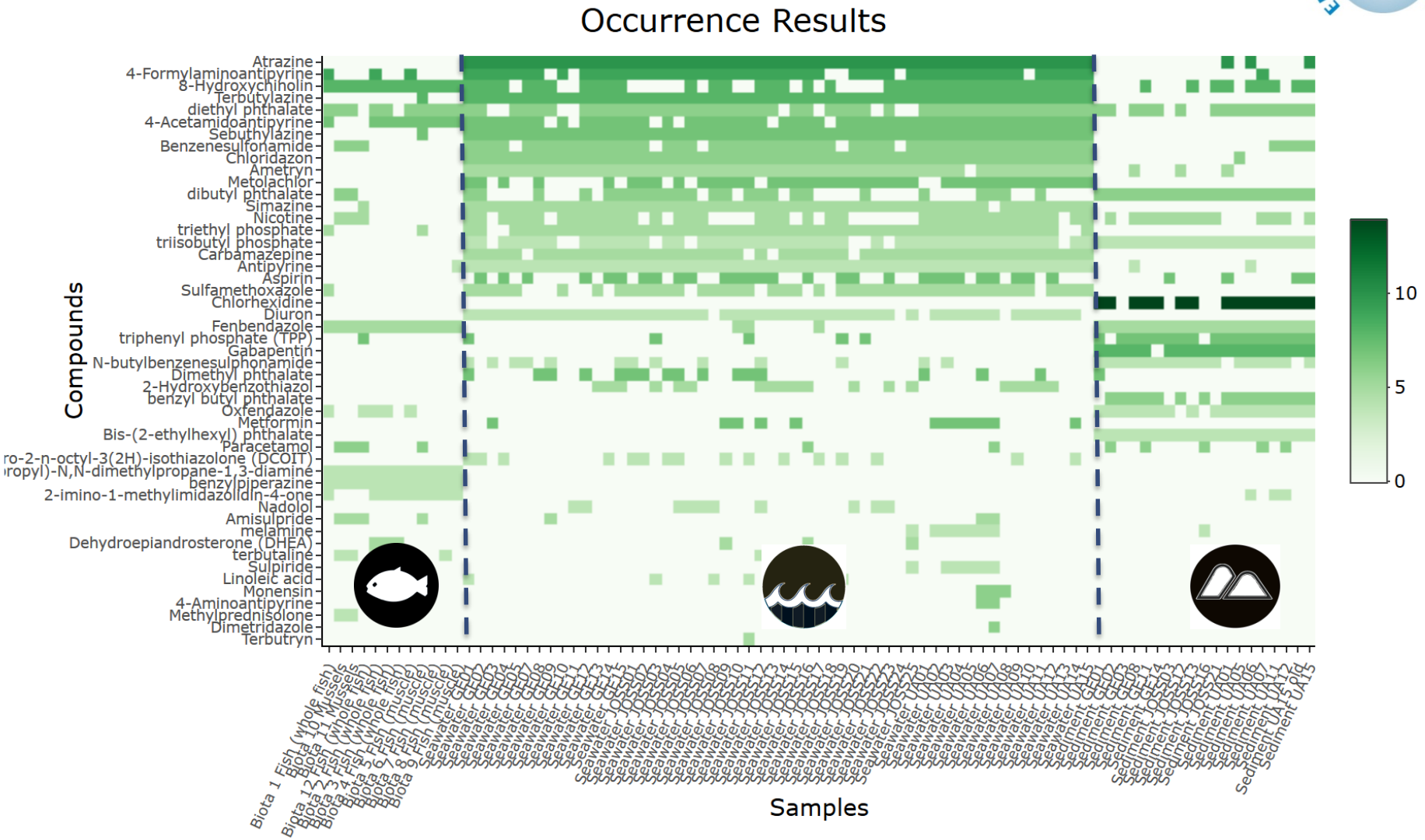
"Live" retrospective screening of known and unknown chemicals in European samples (various matrices)



http://norman-data.eu/ and Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008

NORMAN Digital Sample Freezing Platform

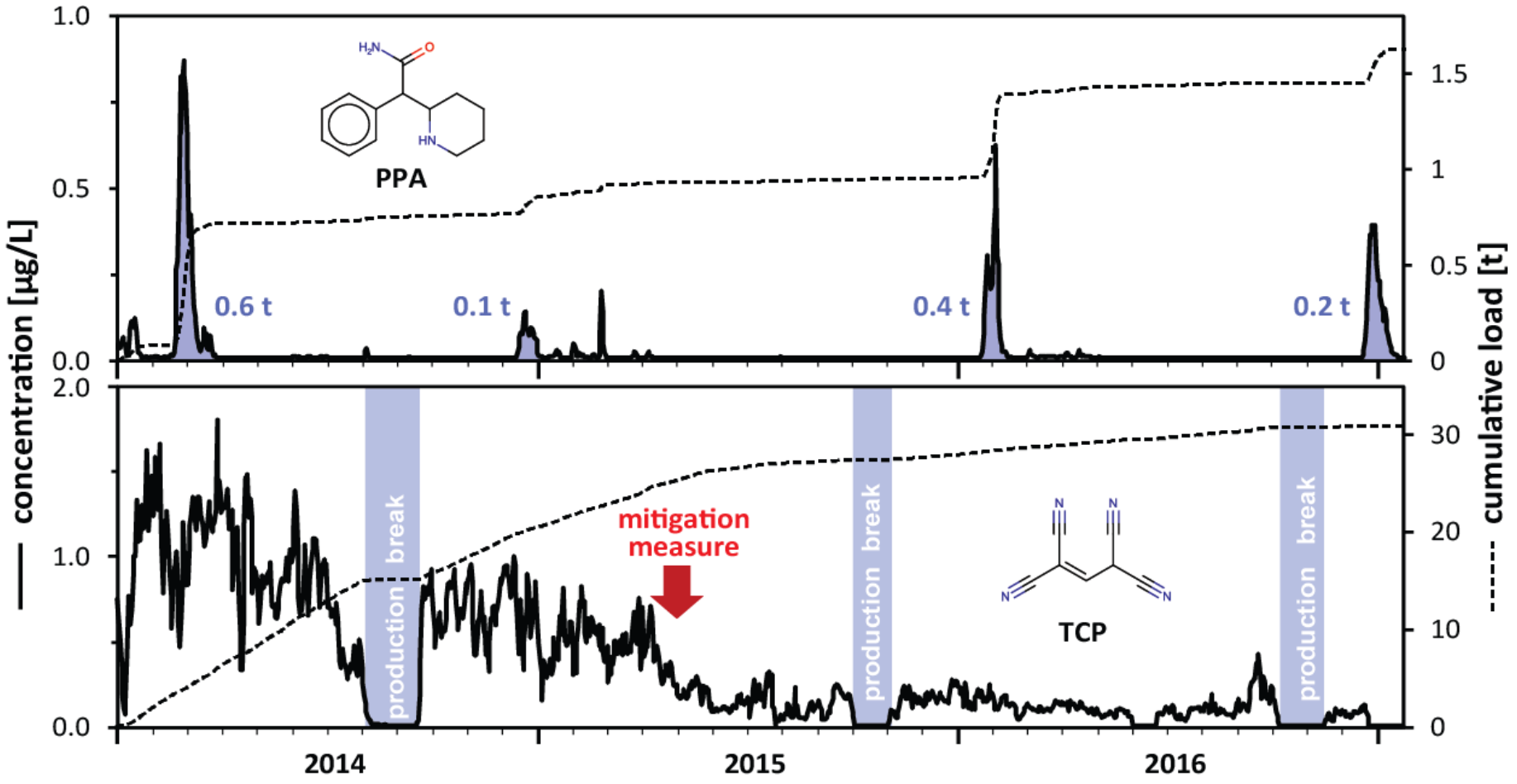
Retrospective screening of REACH chemicals in Black Sea samples (various matrices)



Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008. Interactive heatmap @ http://norman-data.eu/NORMAN-REACH

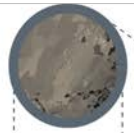
Real-time Monitoring of the Rhine River

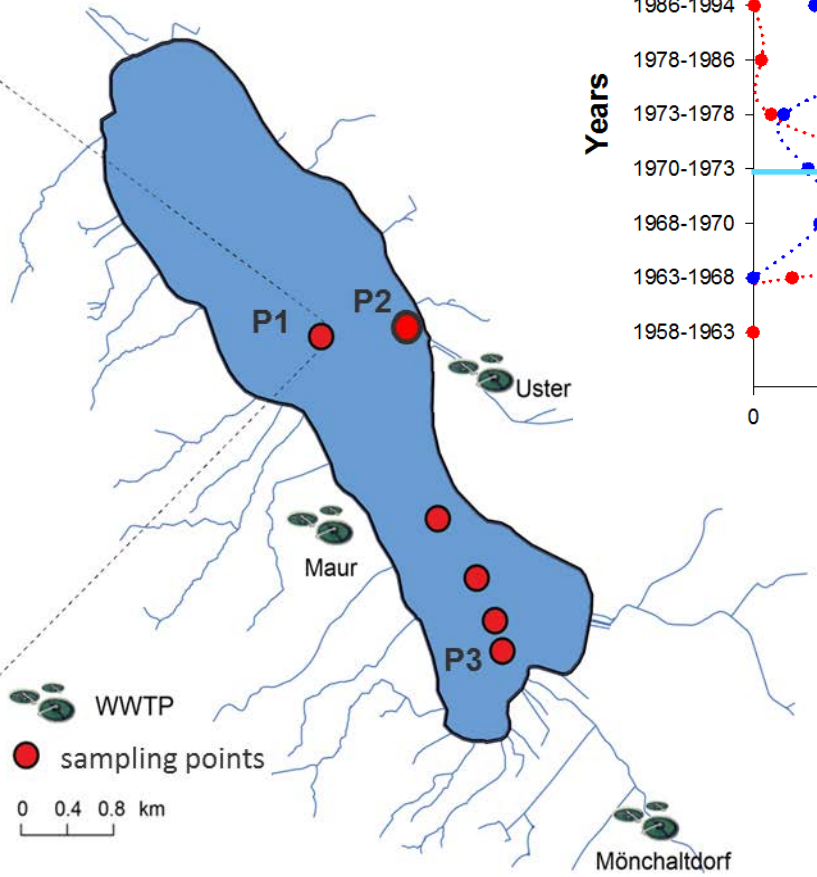
Previously unknown chemicals detected due to “stand-out” patterns

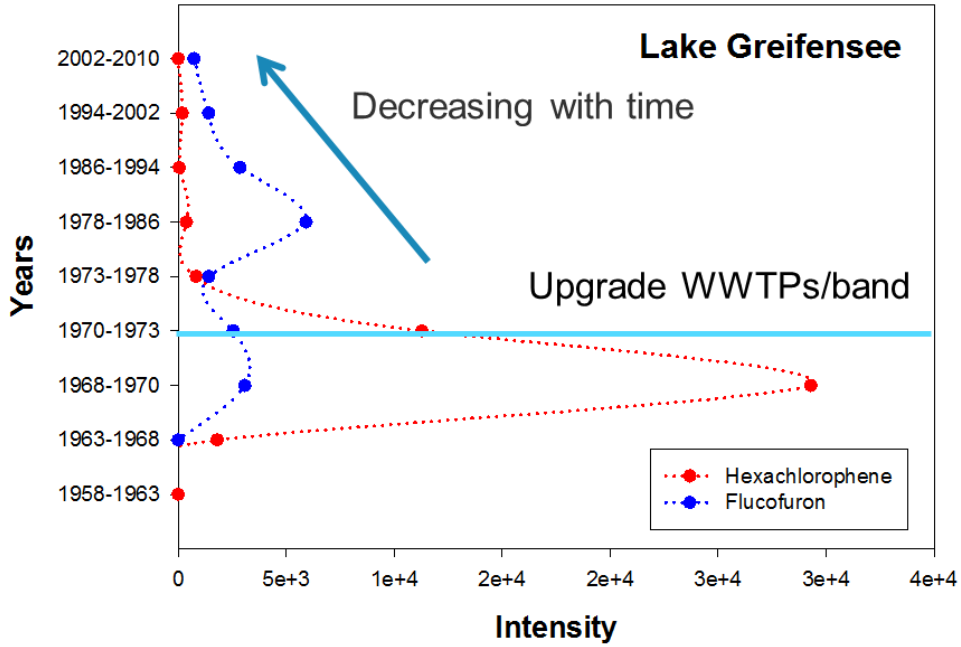


Hollender, Schymanski, Singer & Ferguson, 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Historical Contamination in Lake Sediments

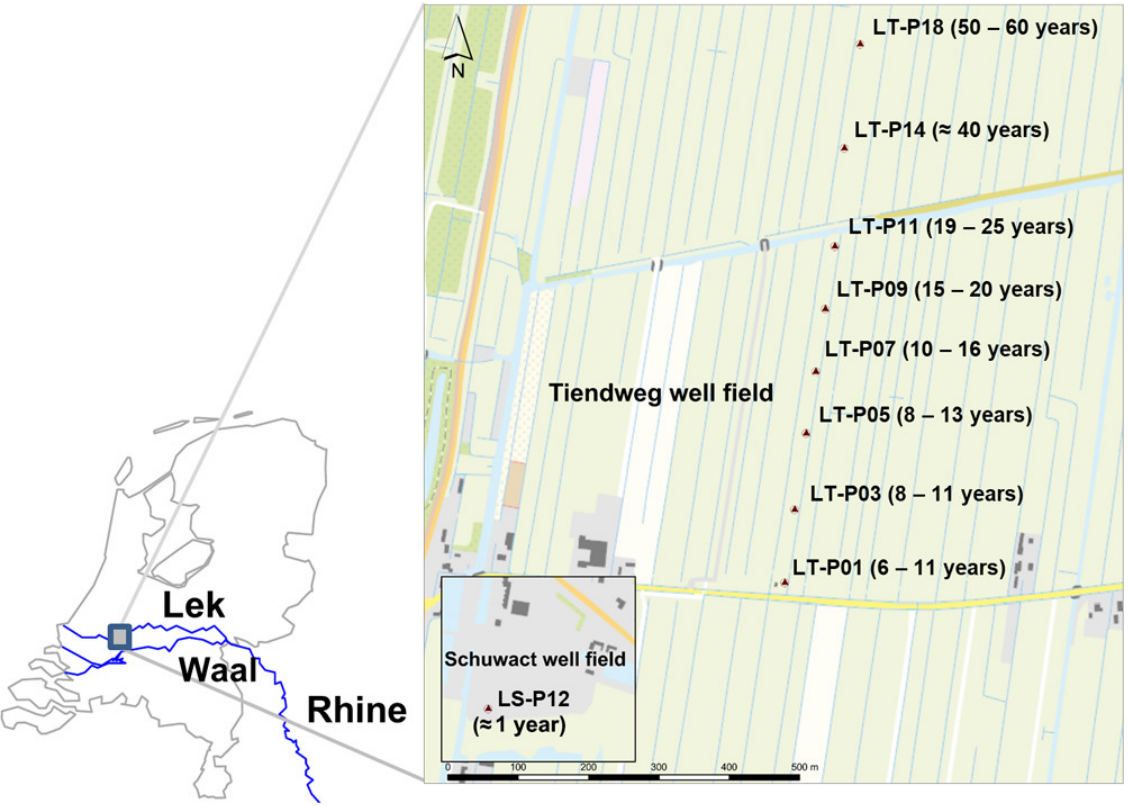


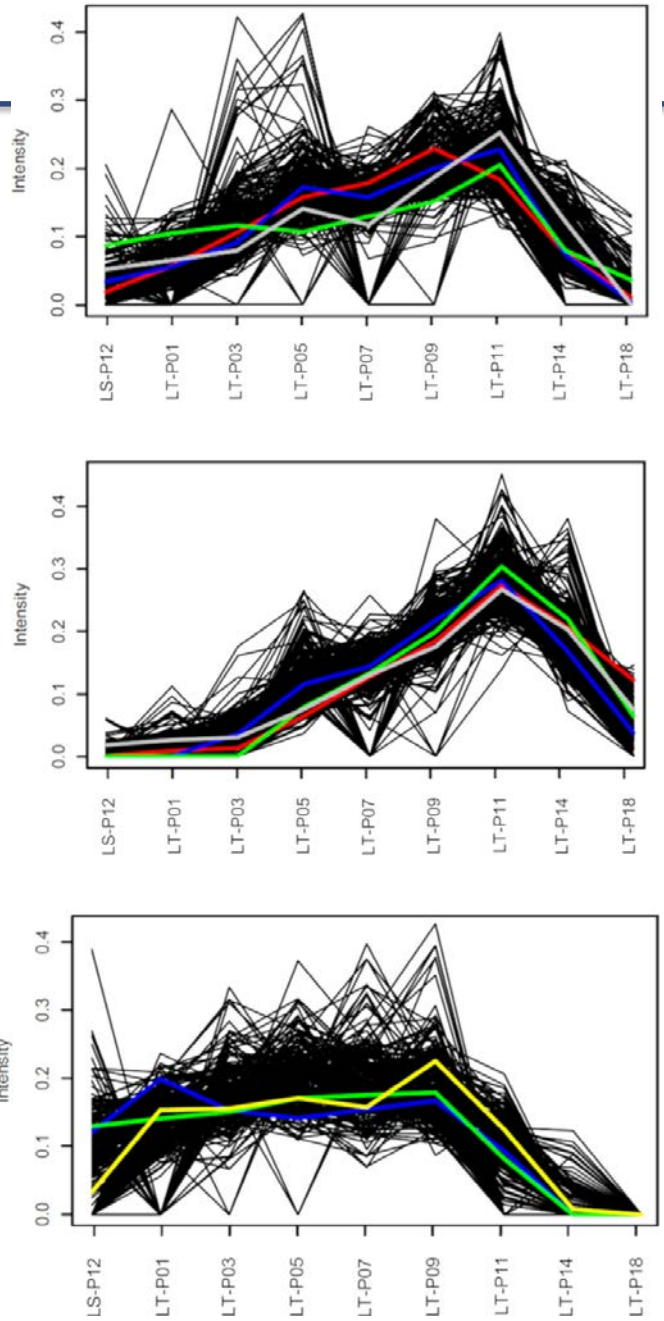




Chiaia-Hernandez *et al.* Anal. Bioanal. Chem. 2014, 406 (28), pp 7323-7335. DOI: 10.1007/s00216-014-8166-0

Micropollutant Time Trends in Riverbank Filtration Systems





Albergamo *et al*. 2019 *Environ. Sci. Technol.* 53 (13), 7584-7594, DOI: 10.1021/acs.est.9b01750

New MetaData: Disease-Specific Reference Counts

https://comptox.epa.gov/dashboard/chemical_lists/litminedneuro

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
Lithium	7439-93-2	DTXSID5036761	689	30	50	9	22	5	36	13	25	6	93	12	15
Valproic Acid	76584-70-8	DTXSID70227388	666	32	10	3	65	6	10	18	45	5	18	4	2
1-Methyl-4-phe	28289-54-5	DTXSID8040933	638	1	24	0	11	0	6	289	0	0	5	0	1
Vincristine	2068-78-2	DTXSID8044331	567	17	59	125	15	5	1	1	5	3	2	1	8
Phenytoin	57-41-0	DTXSID8020541	560	37	24	25	16	9	3	1	9	3	8	4	6
Haloperidol	52-86-8	DTXSID4034150	555	6	6	1	10	6	153	51	4	4	11	1	0
Cocaine	50-36-2	DTXSID2038443	530	151	16	0	8	0	2	3	3	8	6	12	11
Aspirin	50-78-2	DTXSID5020108	489	8	3	0	3	2	2	0	9	4	1	0	5
Paclitaxel	33069-62-4	DTXSID9023413	485	4	43	217	9	14	0	0	0	0	0	1	2
Aluminum	7429-90-5	DTXSID3040273	477	13	41	1	105	4	0	0	1	0	1	13	12
Lidocaine	6108-05-0	DTXSID80209953	464	150	26	15	3	2	0	0	8	4	6	2	10
Methotrexate	59-05-2	DTXSID4020822	451	17	25	1	79	4	0	1	5	0	1	9	18
Mercury	7439-97-6	DTXSID1024172	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, 2019, Advanced Manuscript

New MetaData: Disease-Specific Reference Counts

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

Search

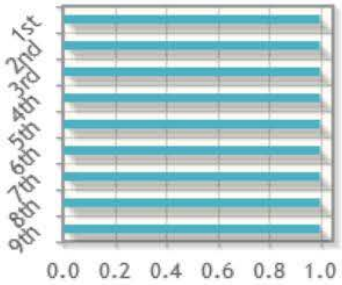
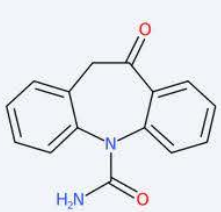
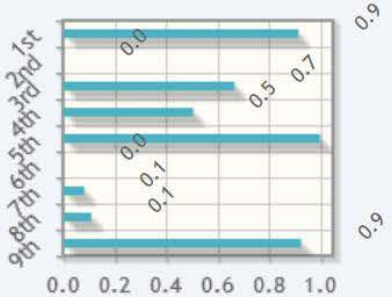
Weights		
MetFrag (1st)	100	%
ExactSpectralSimilarity (2nd)	100	%
Basal Ganglia Diseases (3rd)	100	%
DATA_SOURCES (4th)	100	%
Parkinson Disease, Secondary (5th)	100	%
Peripheral Nervous System Diseases (6th)	100	%

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks | Filter Candidates

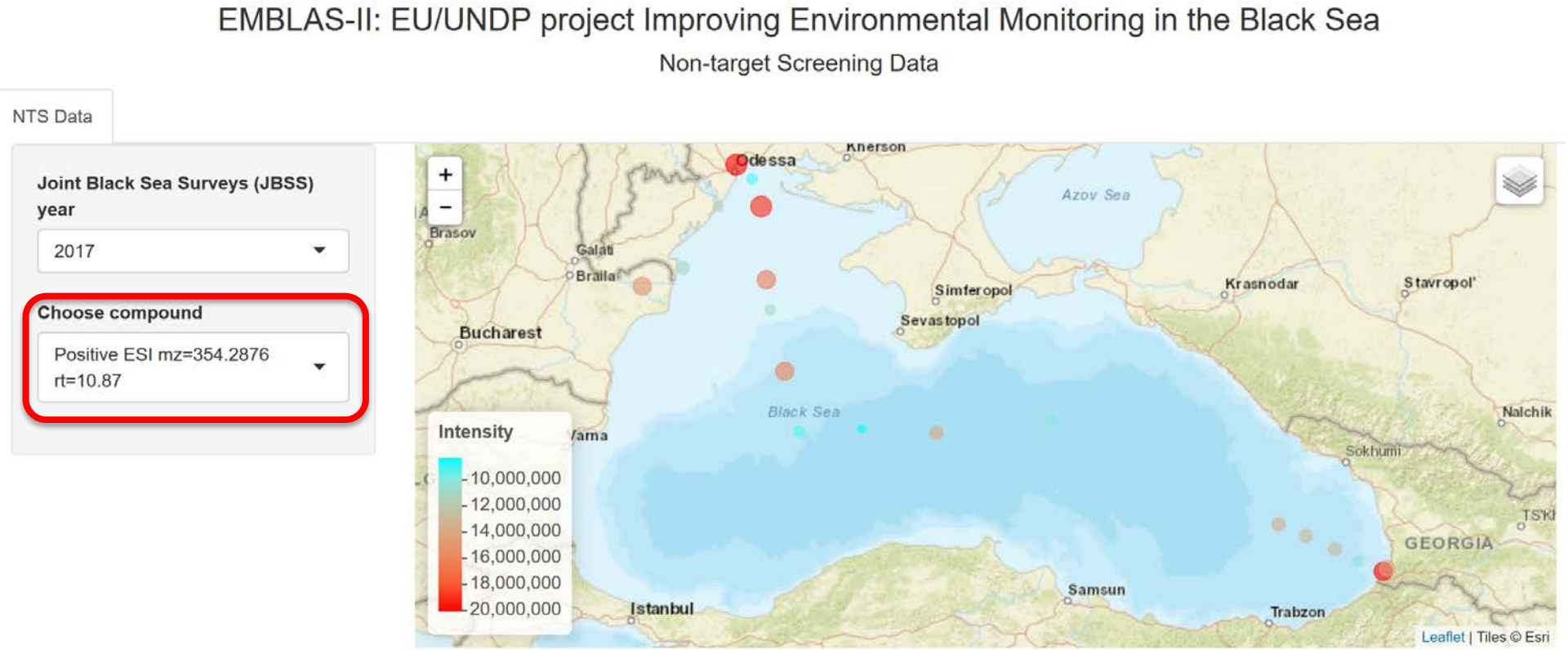
1

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 Phenytoin	DTXSID8020541 InChIKeyBlock1 = CXOFVDLJLONNDW	252.08992	$C_{15}H_{12}N_2O_2$		9.0	Peaks: 9 / 14 Fragments Scores Download
2	 oxcarbazepine	DTXSID0045703 InChIKeyBlock1 = CTRLABGOLIVAIY	252.08992	$C_{15}H_{12}N_2O_2$		4.2006	Peaks: 11 / 14 Fragments Scores Download

Schymanski *et al.* DOI: 10.1039/C9EM00068B (Perspective) and https://twitter.com/ESchymanski/status/1150374899552280576

NORMAN Digital Sample Freezing Platform

*Future work: use results of **unknowns** to drive prioritization efforts*



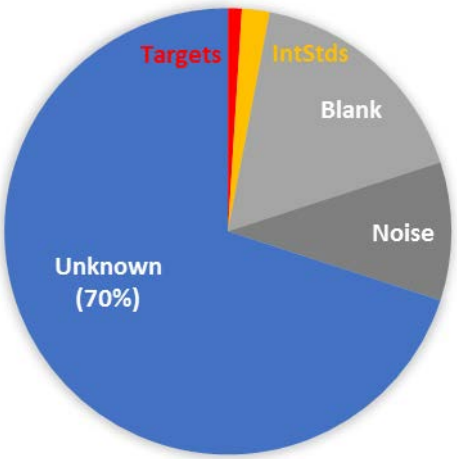
http://norman-data.eu/ and Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008.

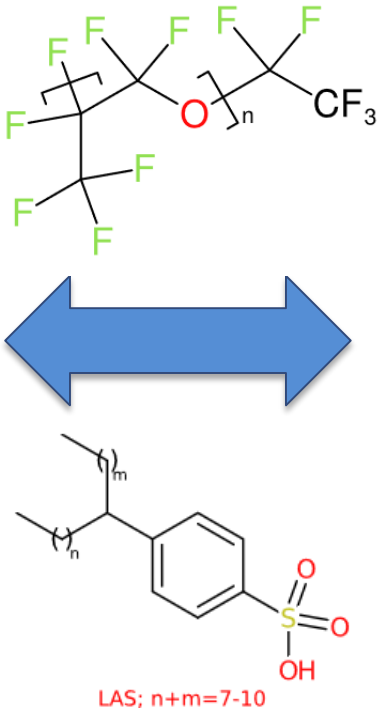
Take Home Messages

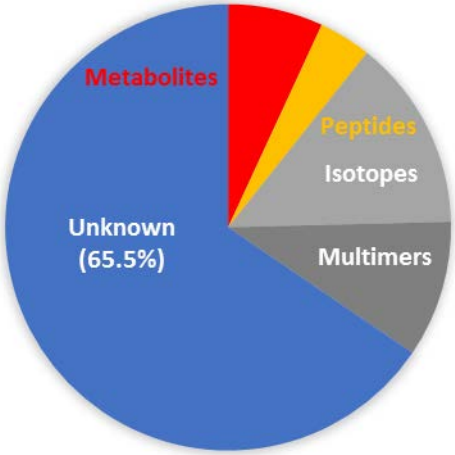
Unknowns and High Resolution Mass Spectrometry

- Over 60 % of HR-MS peaks are potentially ***relevant*** but ***unknown***

Environment





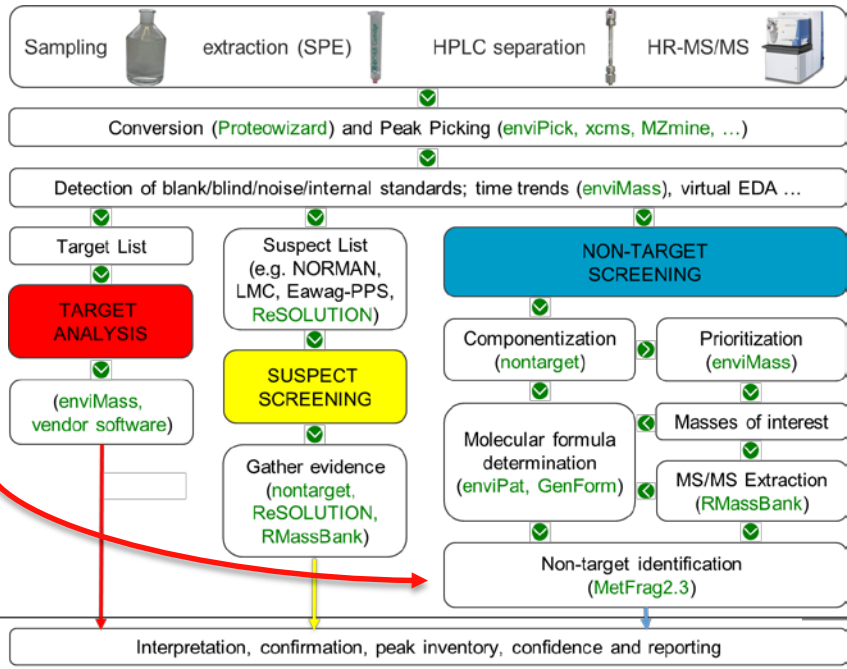


Cells

Take Home Messages

Unknowns and High Resolution Mass Spectrometry

- Over 60 % of HR-MS peaks are potentially ***relevant*** but ***unknown***
- ***Annotating unknowns*** requires data and evidence from ***many different sources***
  - Many excellent workflows available to collate this information
  - Incorporation of ***all available metadata*** is critical to success!
  - *E.g.* MetFrag has greatly improved the speed and success of tentative identification of “known unknowns”: ***15 % => 89 % Ranked Number 1***
    - https://ipb-halle.github.io/MetFrag/



Take Home Messages

Unknowns and High Resolution Mass Spectrometry

- Over 60 % of HR-MS peaks are potentially ***relevant*** but ***unknown***
- ***Annotating unknowns*** requires data and evidence from ***many different sources***
- Exchange expert knowledge worldwide
  - Community efforts contribute greatly to improved cross-annotation
  - Information in the public domain helps everyone!
    - You never know when it will help you ☺

Schymanski et al. 2015, *ABC*, DOI: 10.1007/s00216-015-8681-7; Alygizakis *et al*. 2018 *ES&T*, DOI: 10.1021/acs.est.8b00365

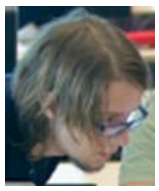
Acknowledgements



















MassBank.eu

Luxembourg National Research Fund

solutions

EU Grant 603437

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

EPA

emma.schymanski@uni.lu

@ESchymanski

Further Information:

https://massbank.eu/MassBank/

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

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

https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics