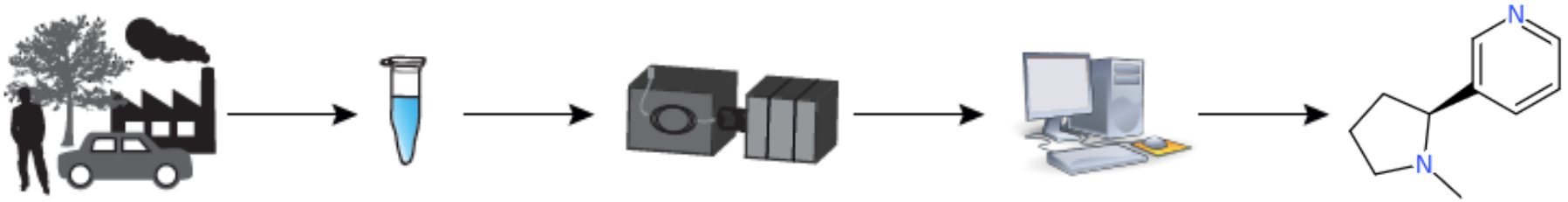
Measuring the Environmental Exposome



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

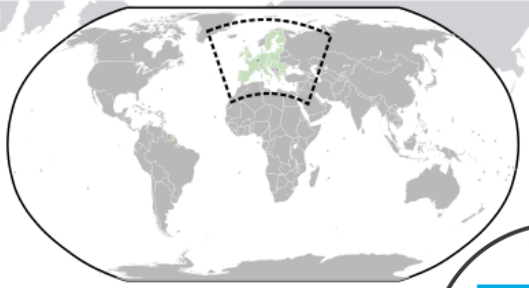
Plus ECI-LSCB, NORMAN, PubChem, IPB and many other colleagues who contributed to our science over the years!

Talk available under DOI: 10.5281/zenodo.4032909

Luxembourg National Research Fund

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

PubChem



751 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.luti.ns

KEMI Kemikalieinspektionen

ETH zürich

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

COLUMBIA MAILMAN SCHOOL OF PUBLIC HEALTH

IPB

WIKIDATA

Stockholm University

National and Kapodistrian UNIVERSITY OF ATHENS

elixir LUXEMBOURG

Maastricht University

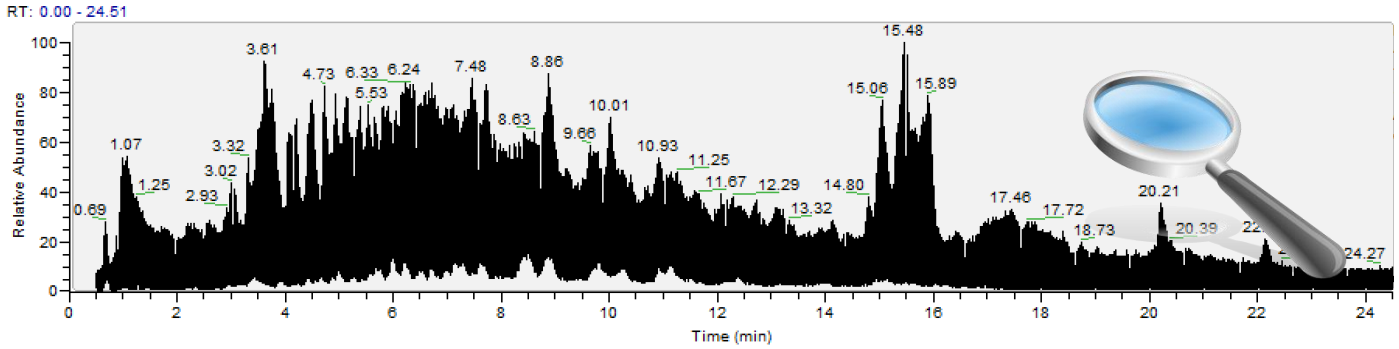
UNIVERSITY OF COPENHAGEN

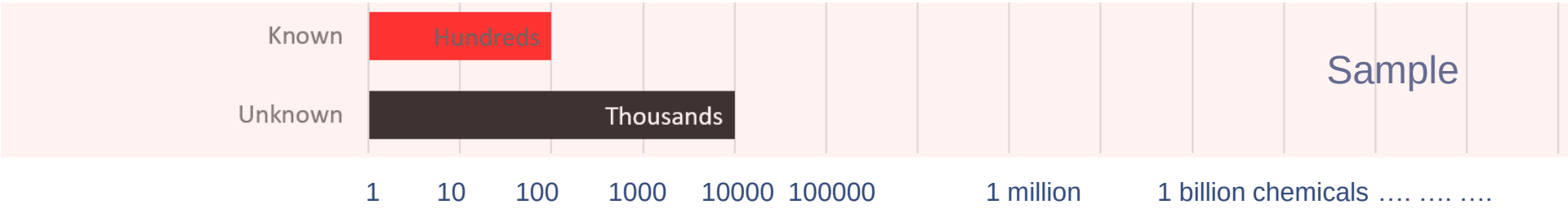
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

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

Our (Community) Challenge: Identifying Chemicals

High resolution mass spectrometry



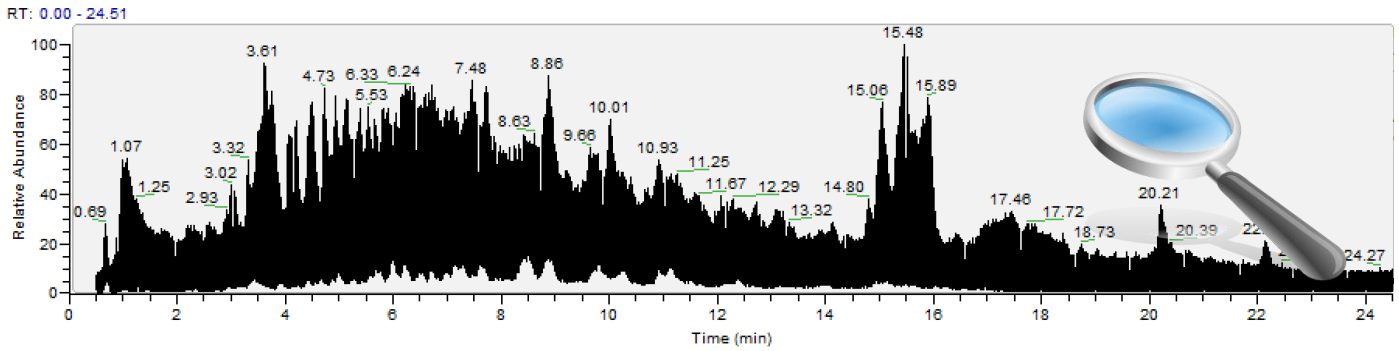


Schymanski *et al* (2014) *ES&T*, DOI: 10.1021/es4044374; Vermeulen *et al* (2020) Science. DOI: 10.1126/science.aay3164

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 …. …. ….

Chemicals

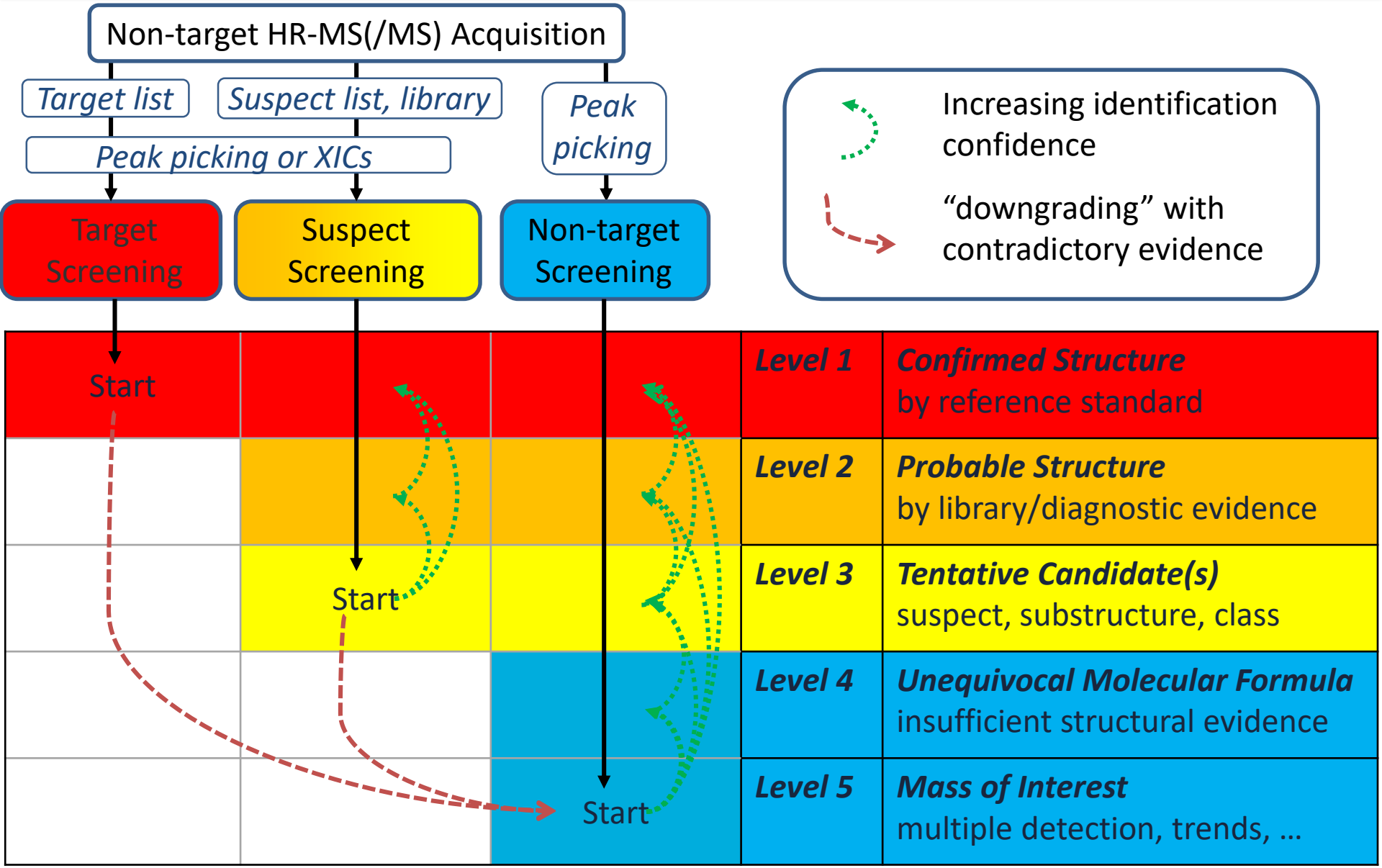
Suspect Chemicals	100s-1000s — norman suspects
Mass Spectral Libraries	~80,000 — MassBank.eu
CompTox Dashboard	882,000 — EPA Chemicals
PubChem Compound	>103 million — PubChem
1st Gen. PubChem Metabolites	>2 billion — PubChem
Generated Structures	Millions of billions

Schymanski *et al* (2014) *ES&T*, DOI: 10.1021/es4044374; Vermeulen *et al* (2020) Science. DOI: 10.1126/science.aay3164

NORMAN Collaborative Trial (2015)

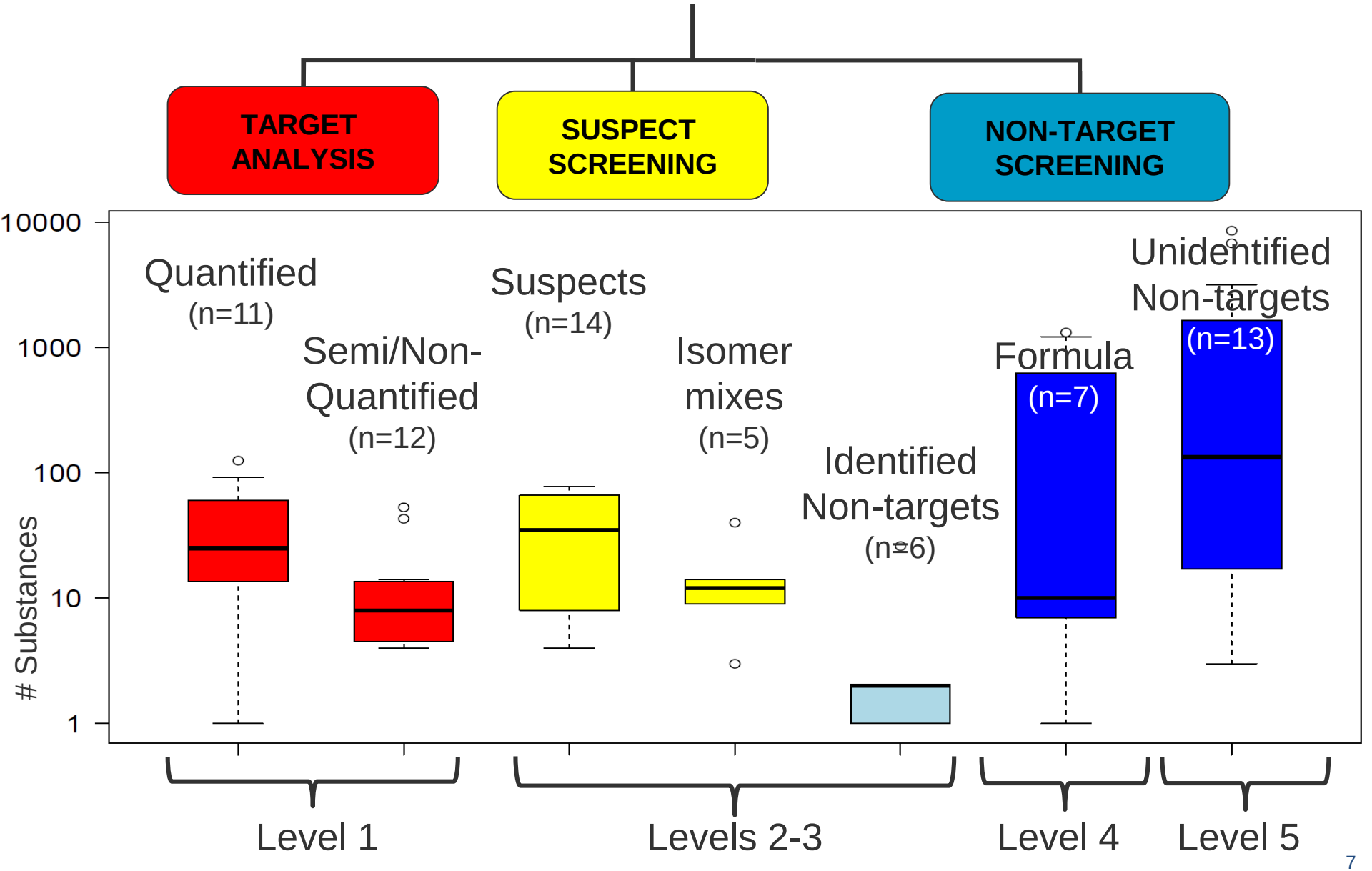
Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

Identification Strategies and Confidence



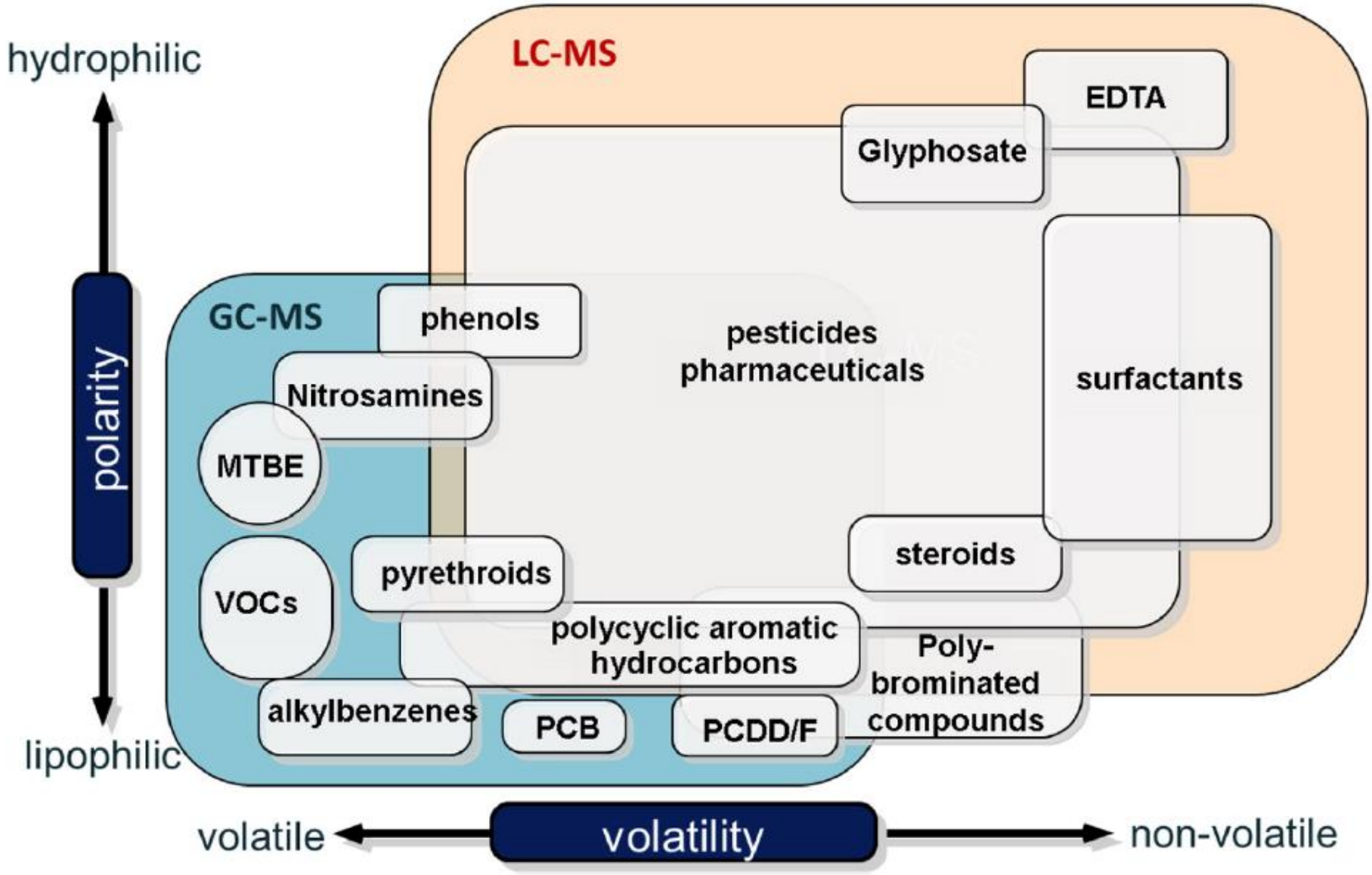
Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

NORMAN Collaborative Trial (2015): Results



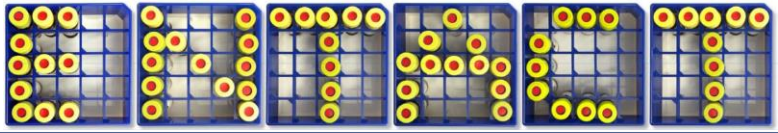
Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

How to measure? LC vs GC?

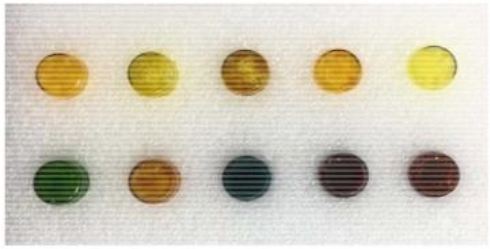


Brack *et al* 2016. DOI: 10.1016/j.scitotenv.2015.11.102

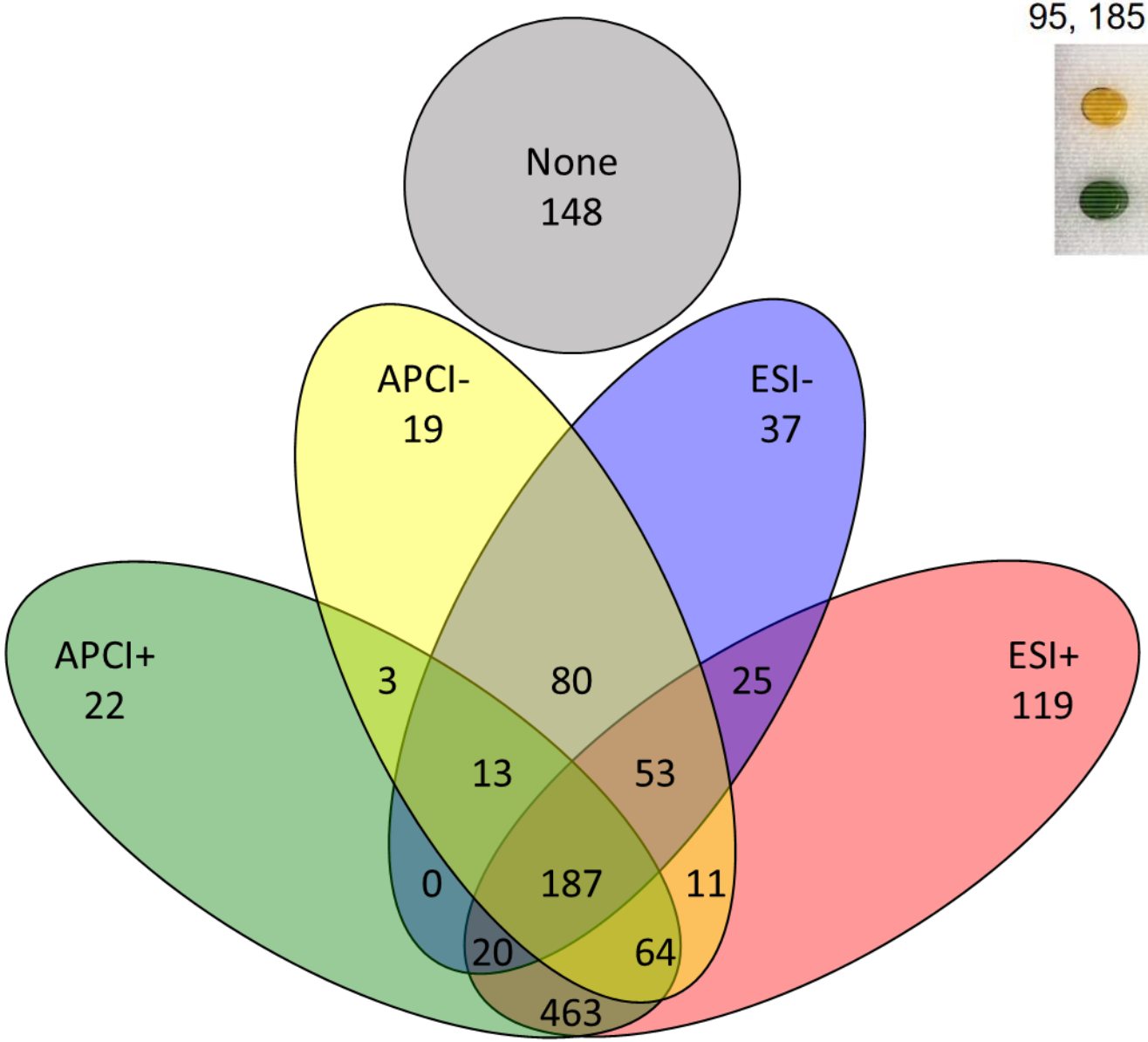
How to measure? Ionisation



95, 185 or 365 substances/mixture

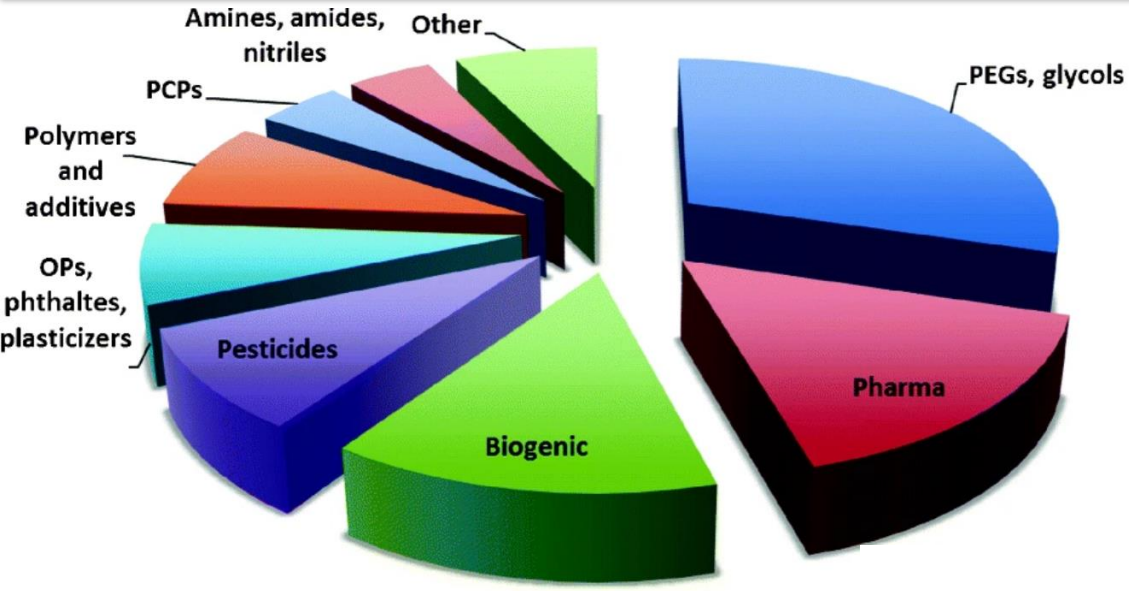


n=1264



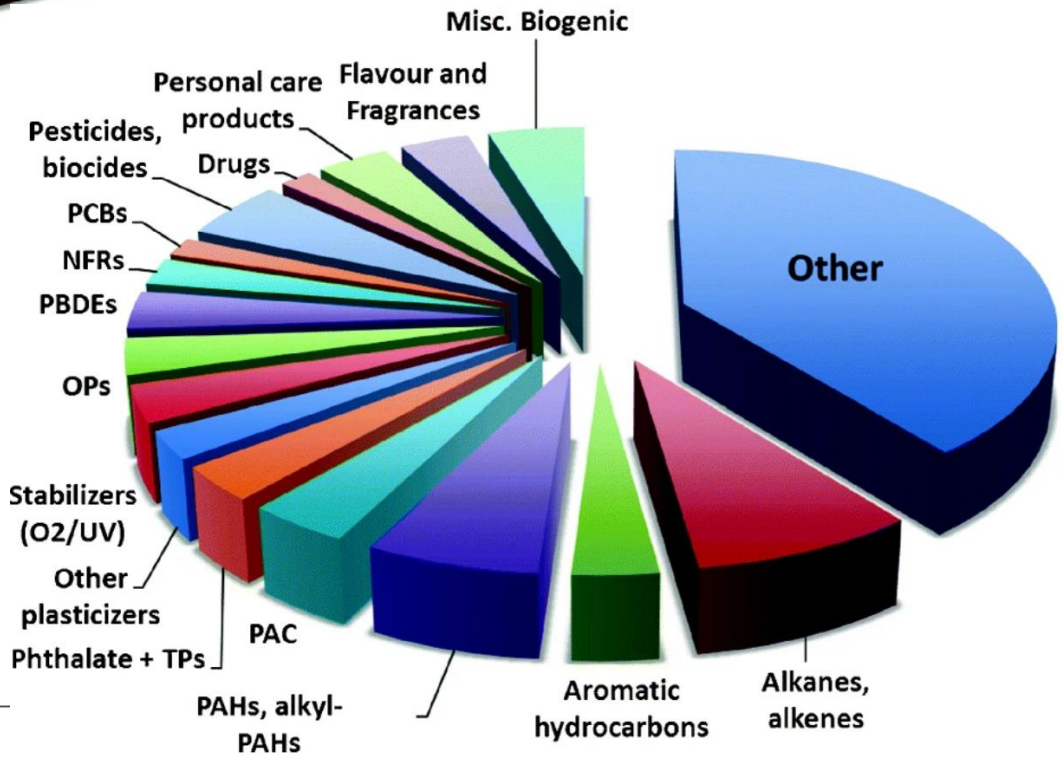
Singh, RR *et al* (2020) DOI: 10.1007/s00216-020-02716-3 ; Ulrich, EM *et al* (2019) DOI: 10.1007/s00216-018-1435-6

What we measure (in dust): LC vs GC



LC-MS
(n=969)

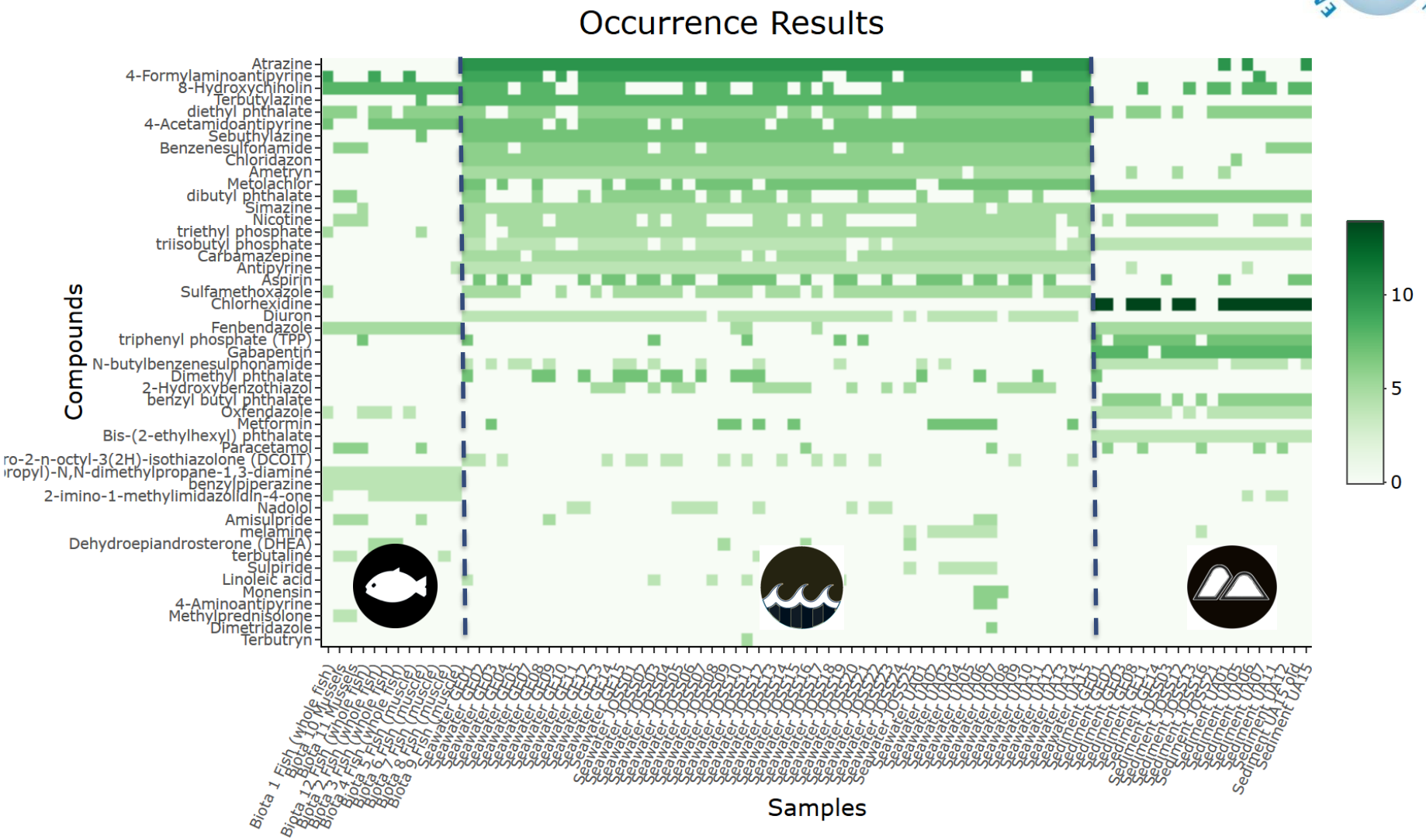
GC-MS
(n=592)



Rostkowski *et al* 2019. DOI: 10.1007/s00216-019-01615-6

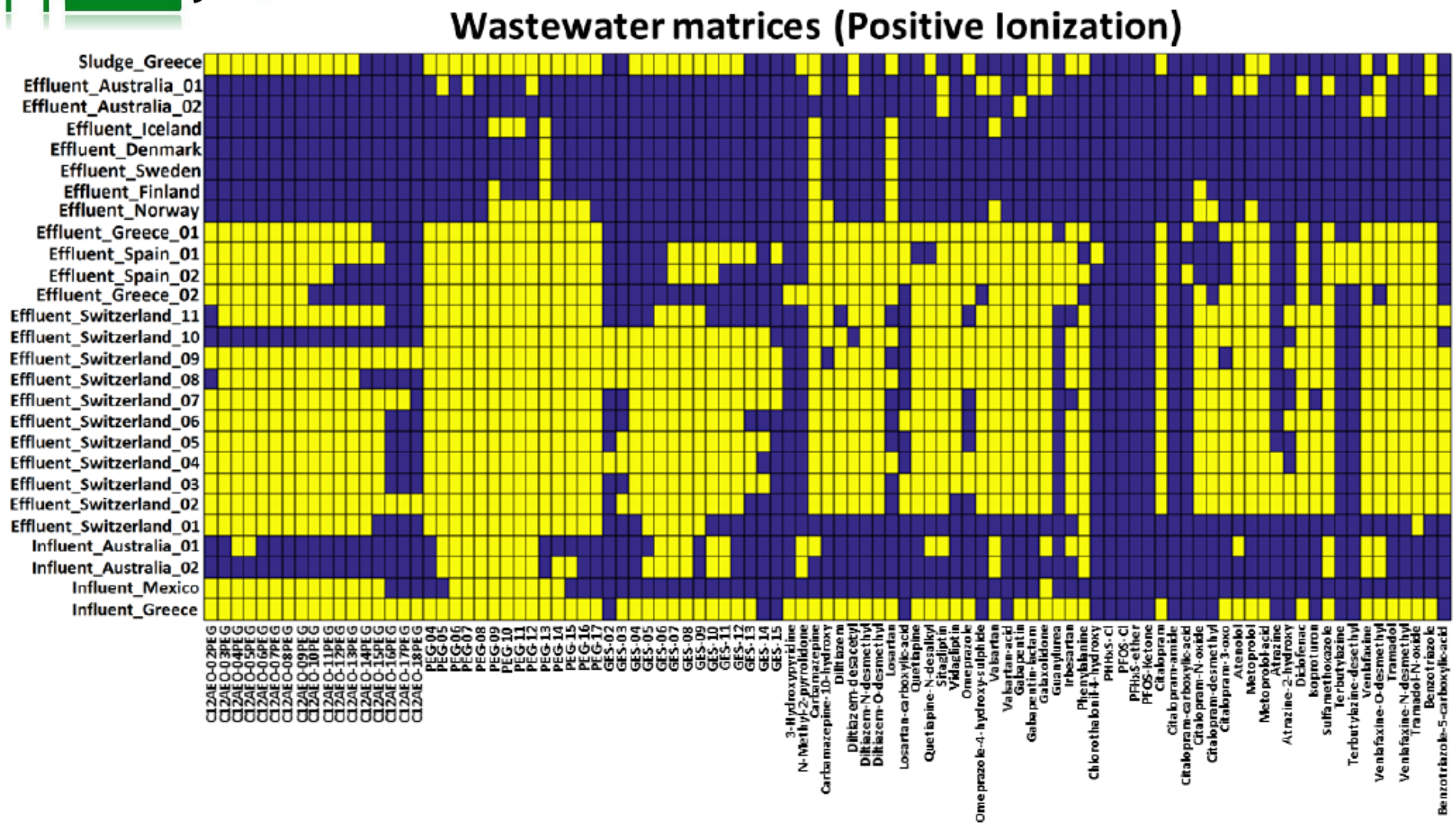
What we measure in biota/water/sediment

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



Alygizakis NA *et al* (2019) DOI: 10.1016/j.trac.2019.04.008 Interactive heatmap at http://norman-data.eu/NORMAN-REACH

Emerging contaminants in different countries

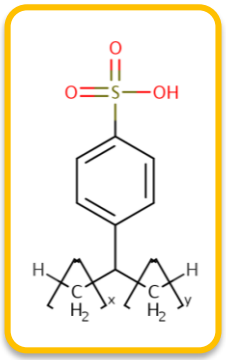


Alygizakis NA *et al* (2018) DOI: 10.1021/acs.est.8b00365 and DOI: 10.5281/zenodo.2623815

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

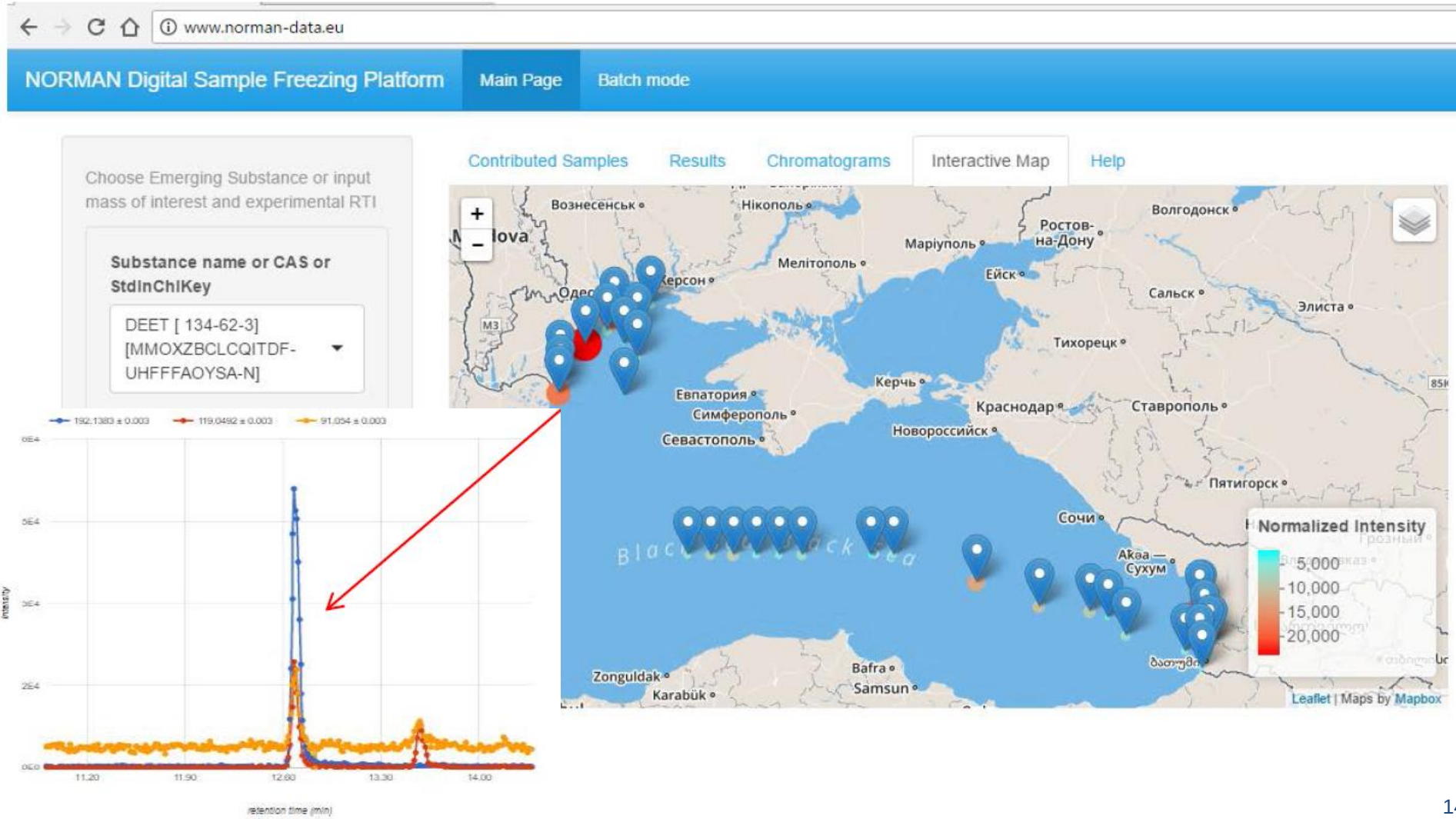
LCSB

uni.lu UNIVERSITÉ DU LUXEMBOURG

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

NORMAN Digital Sample Freezing Platform

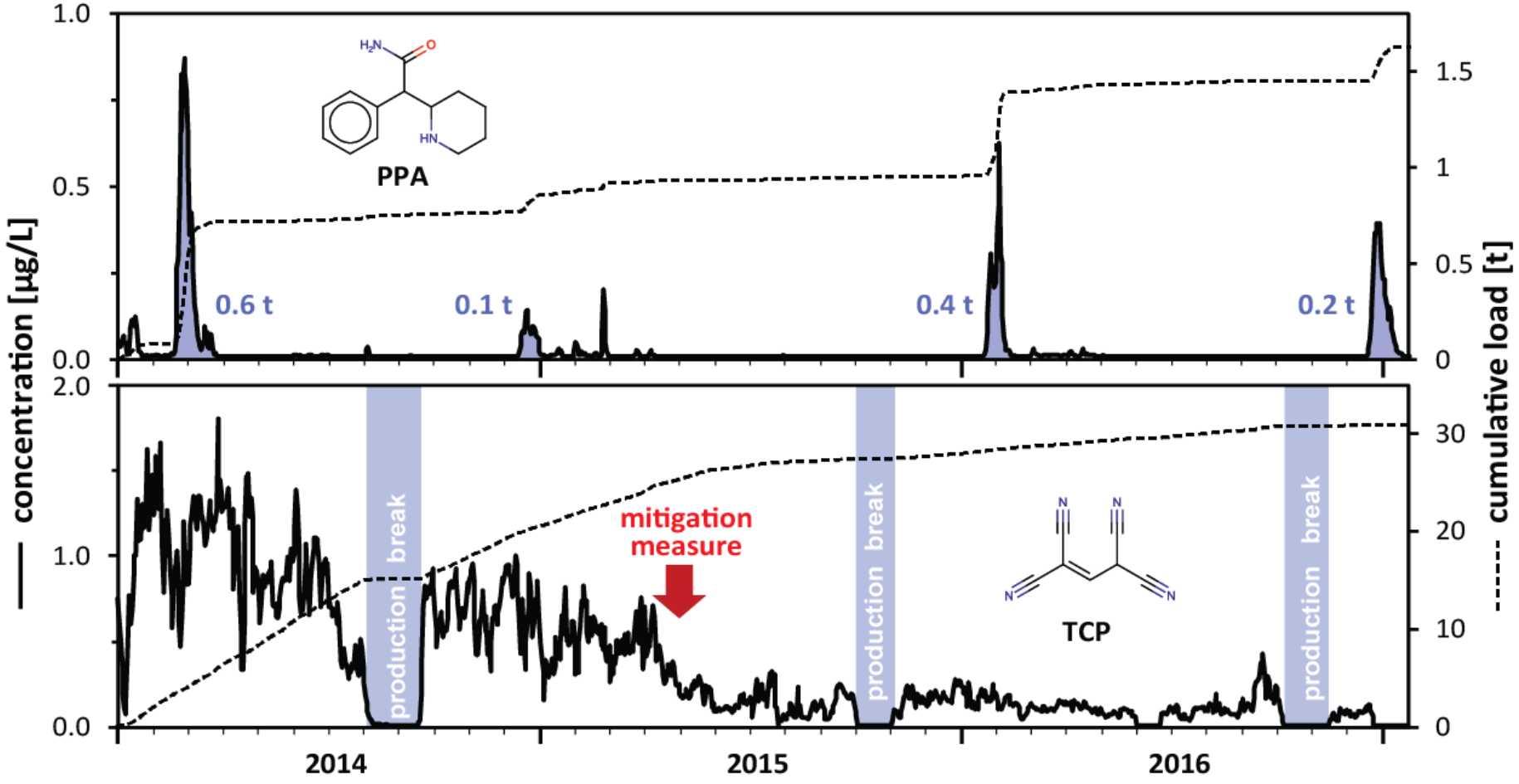
"Live" retrospective screening of known and unknown chemicals in European samples (various matrices)



http://norman-data.eu/ and Alygizakis NA *et al* (2019) DOI: 10.1016/j.trac.2019.04.008.

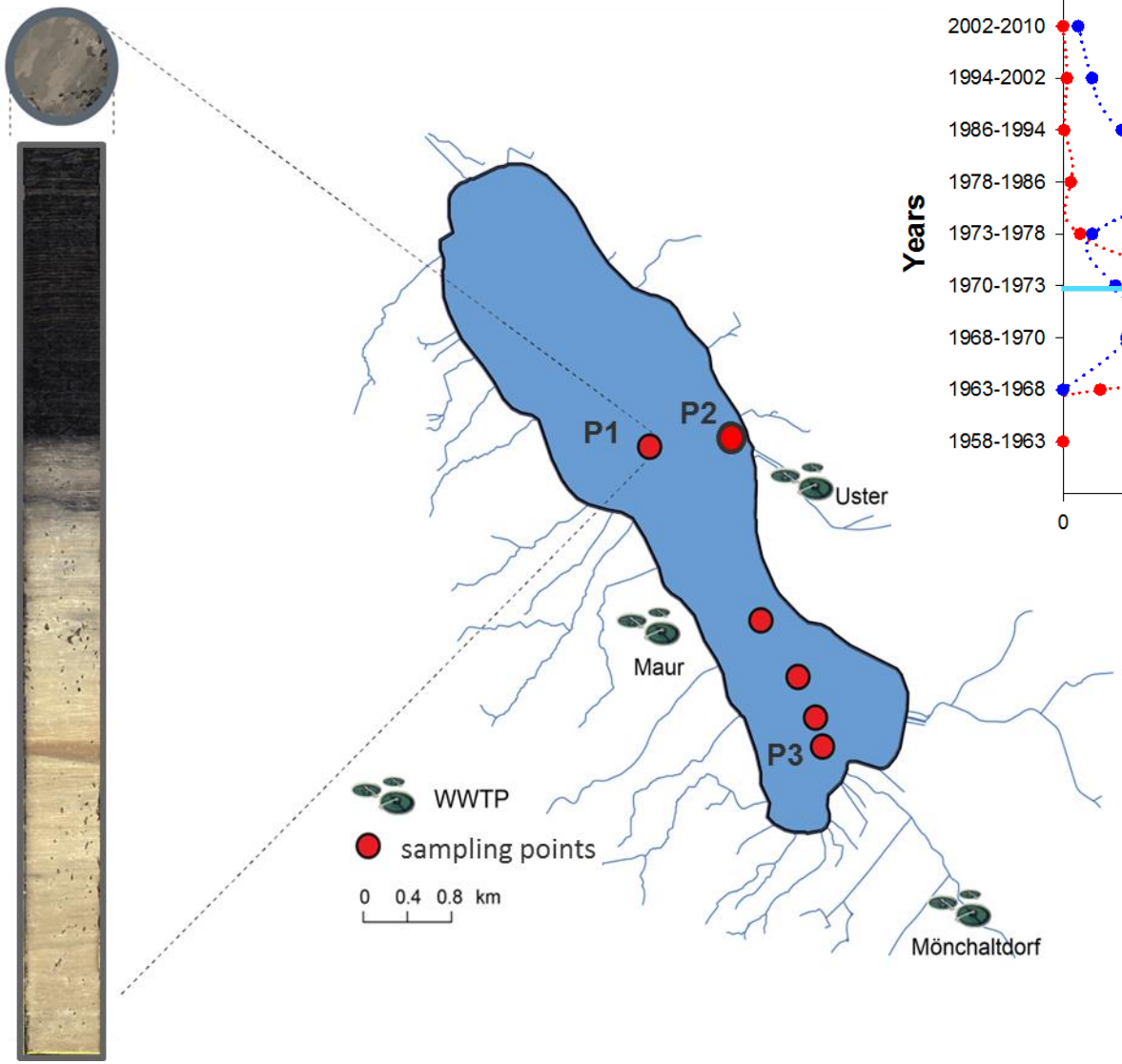
Real-time Monitoring of the Rhine River

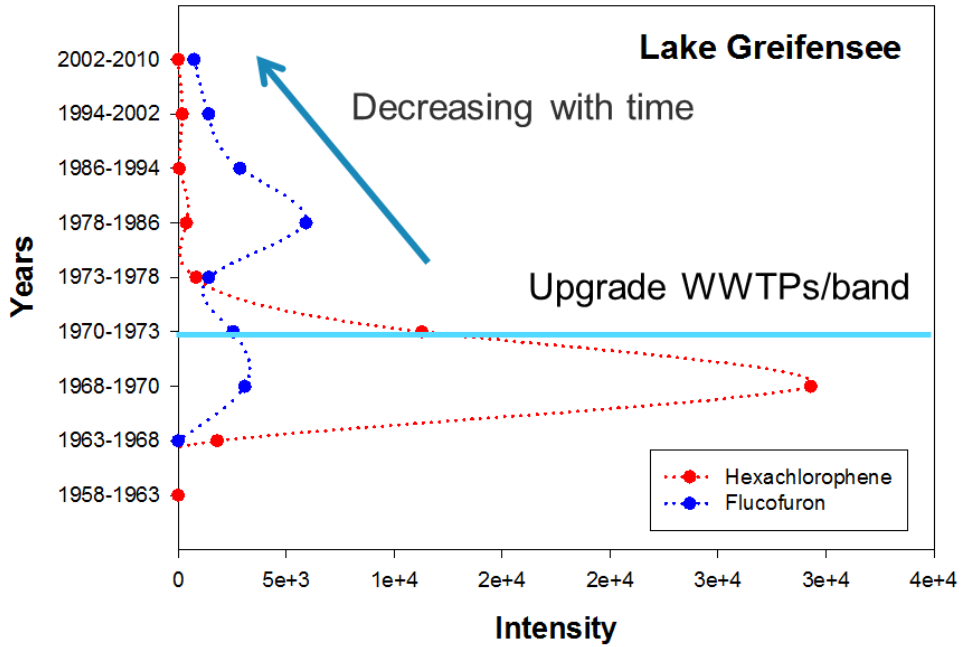
Previously unknown chemicals detected due to “stand-out” patterns



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

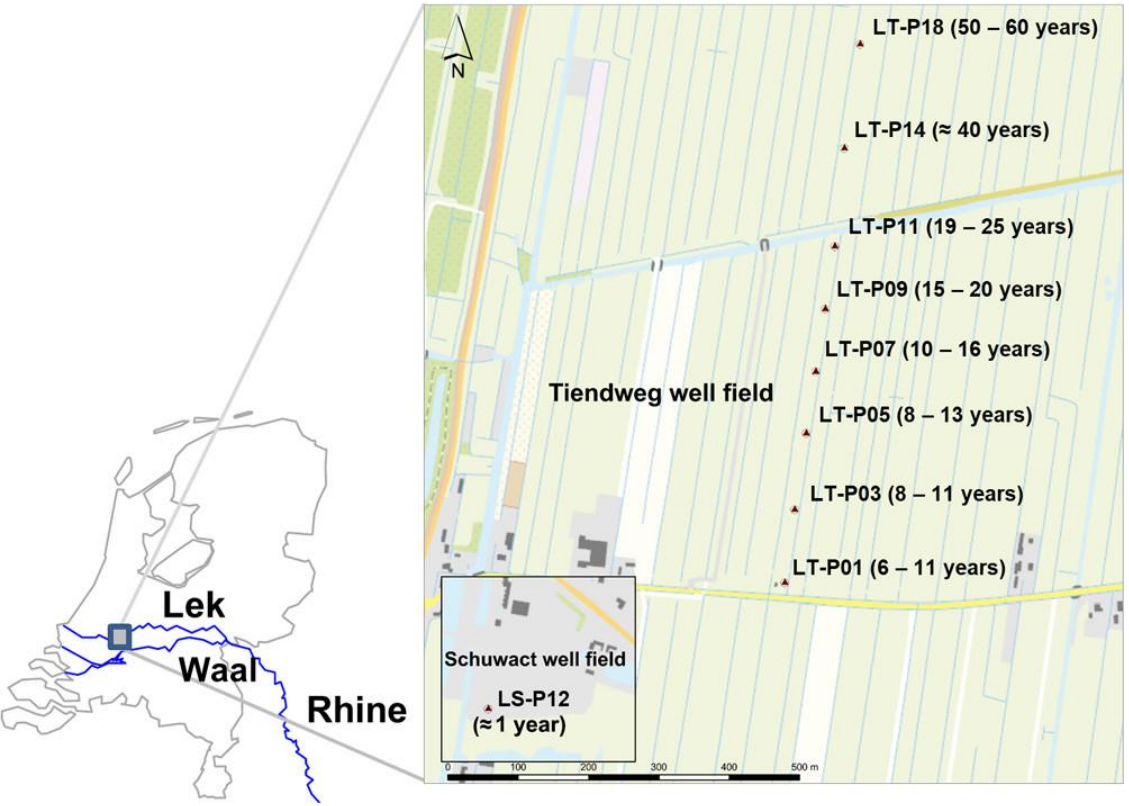
Historical Contamination in Lake Sediments

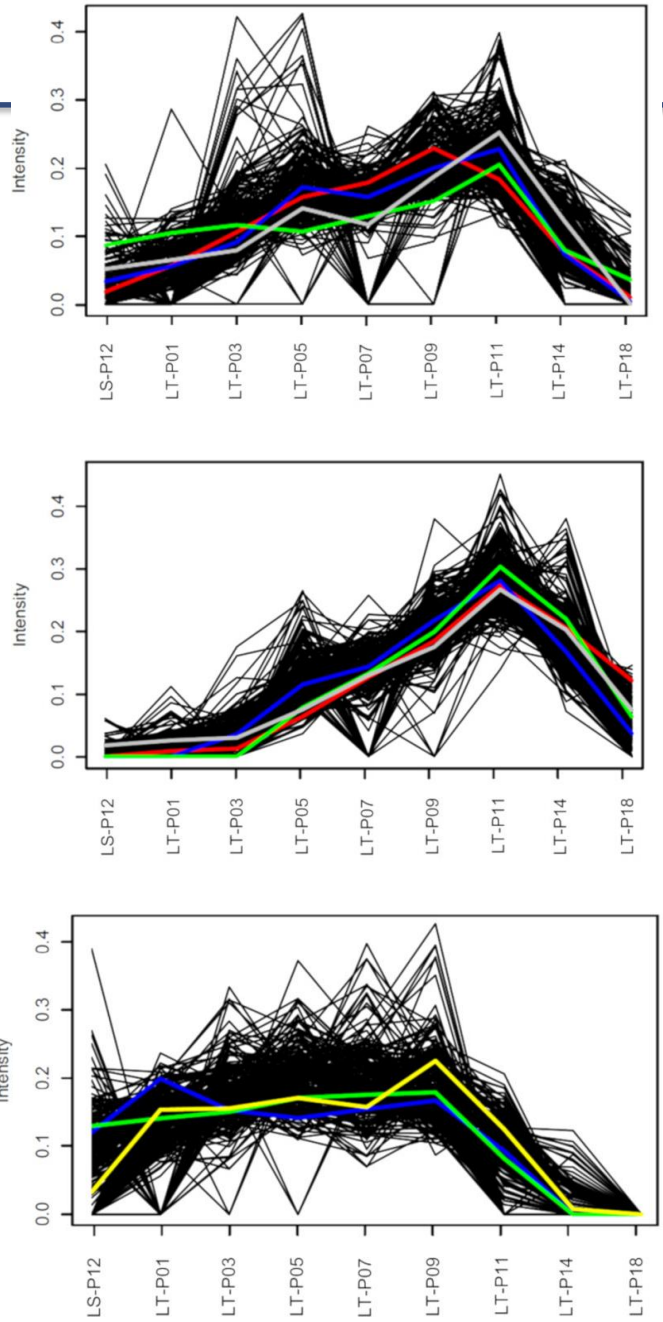




Chiaia-Hernandez *et al.* Anal. Bioanal. Chem. 2014, 406 (28), pp 7323-7335. DOI: 10.1007/s00216-014-8166-0

Micropollutant Time Trends in Riverbank Filtration Systems



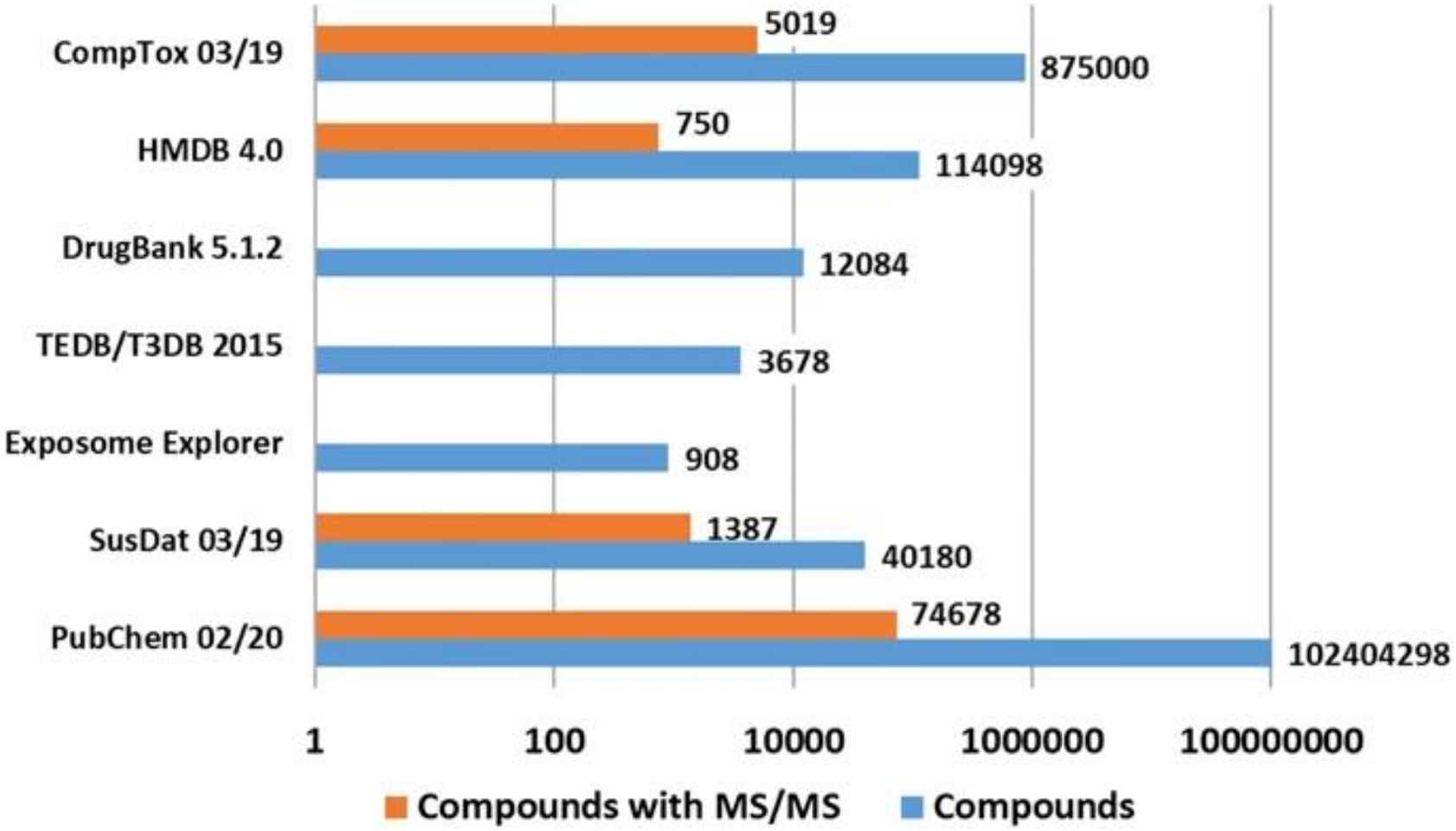


Albergamo *et al*. 2019 *Environ. Sci. Technol.* 53 (13), 7584-7594, DOI: 10.1021/acs.est.9b01750

Challenge 1: Scarcity of MS/MS Spectra

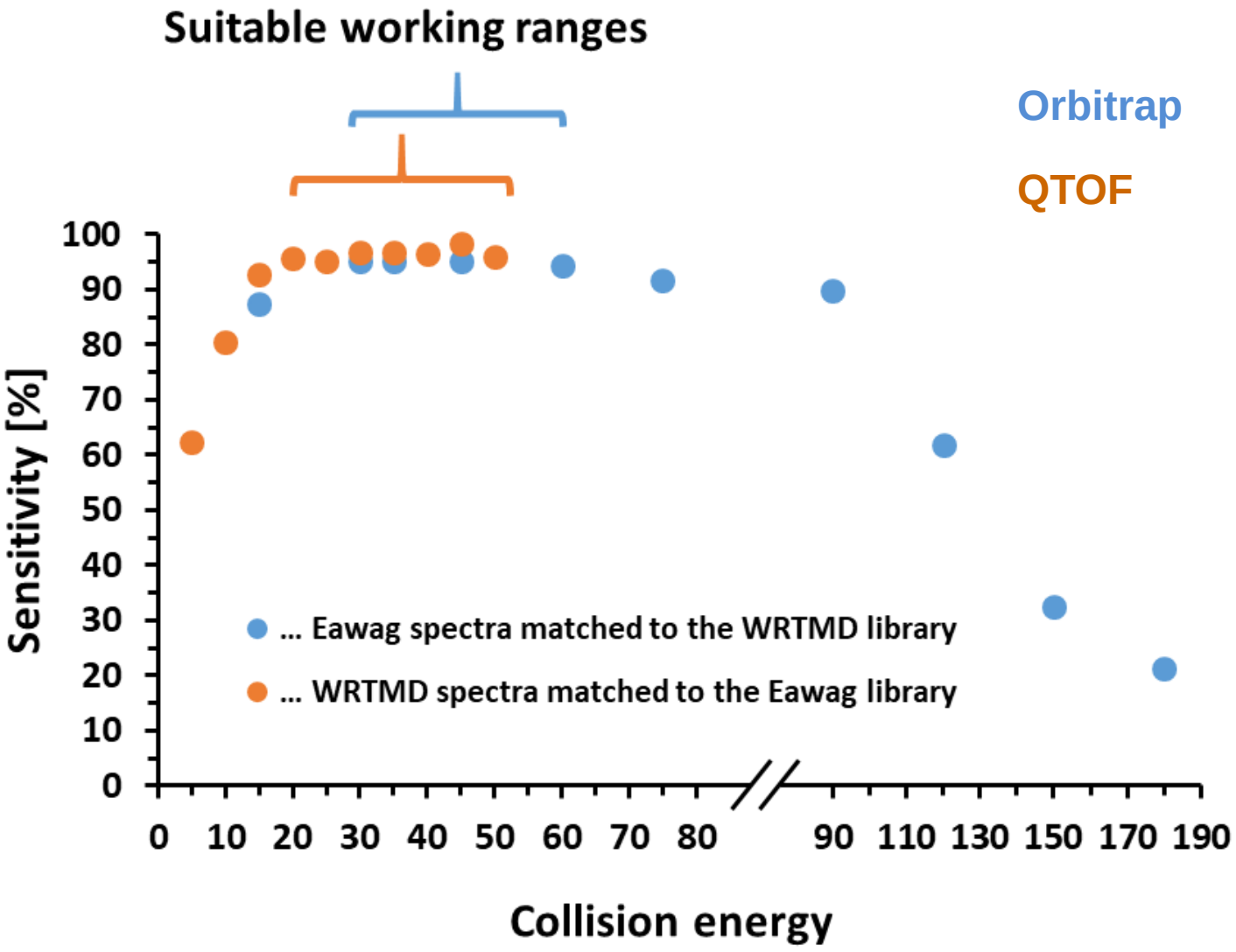
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

- Only available for ~0.1-4 % of Exposomics-relevant resources



H. Oberacher *et al.* (2020) Environmental Sciences Europe 32: 43. DOI: 10.1186/s12302-020-00314-9

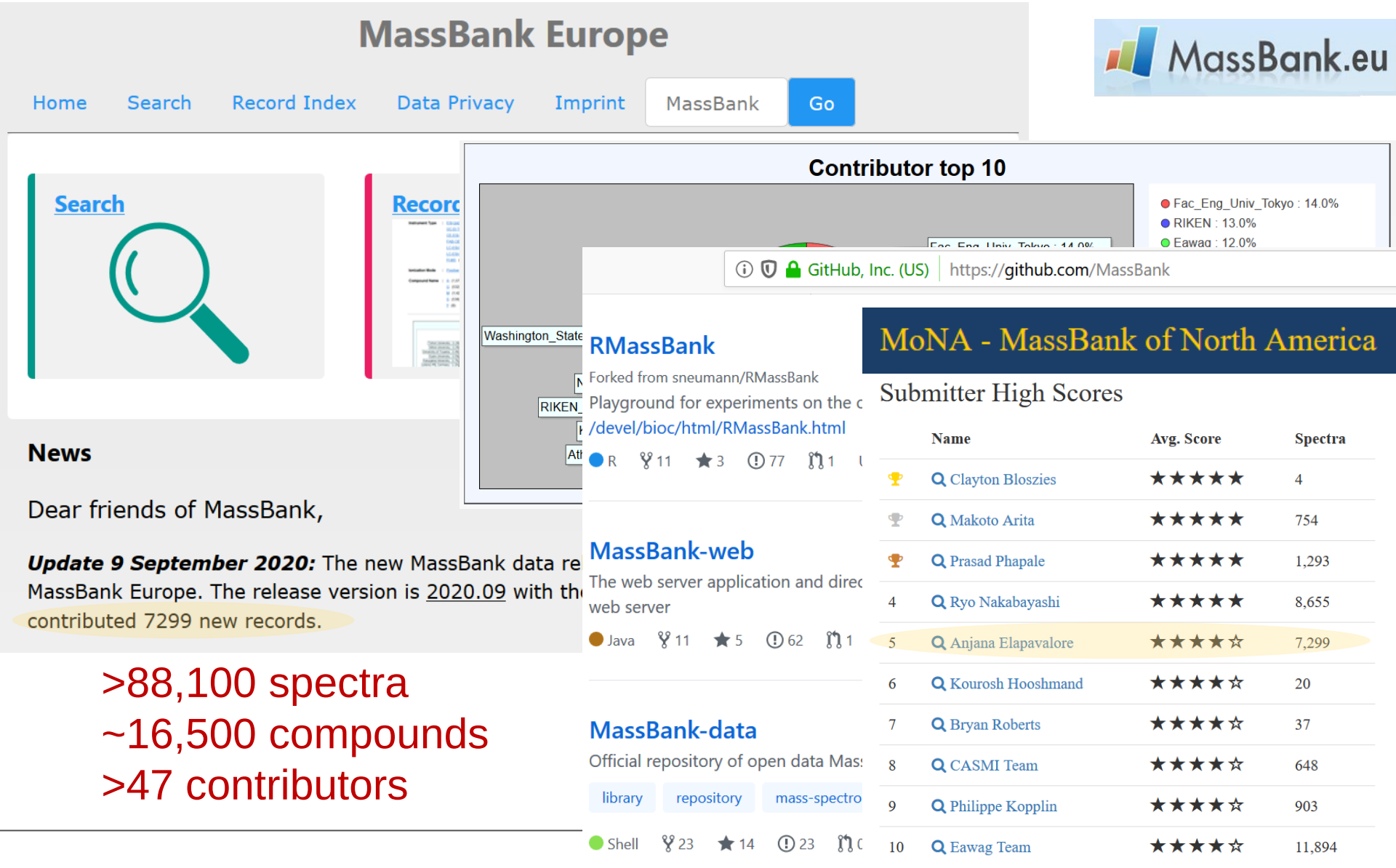
Comparability of MS/MS Spectra: QTOF v Orbitrap



Oberacher *et al*. 2019, *Metabolites, 9*(1), 3; https://doi.org/10.3390/metabo9010003

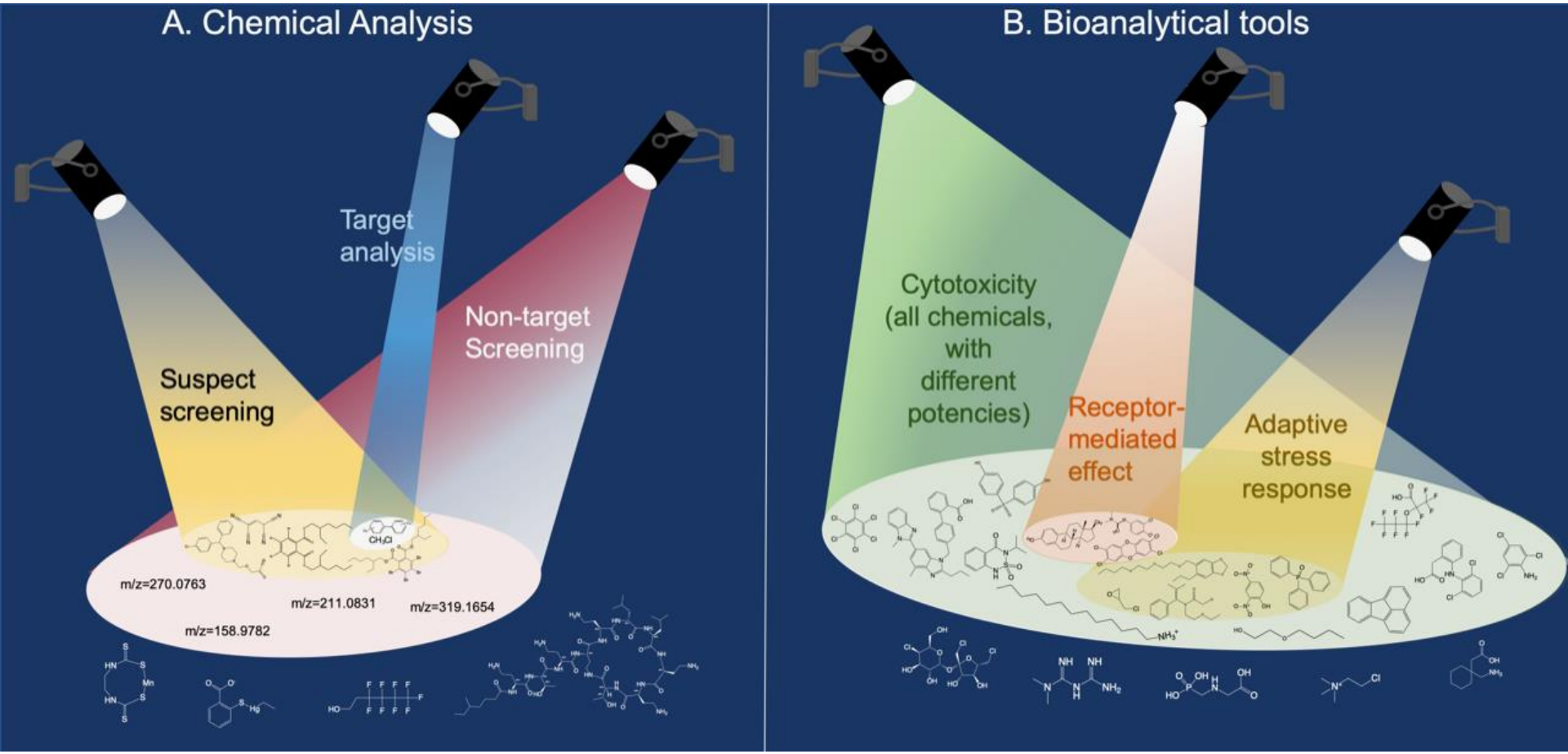
MassBank EU: Capturing Environmental Spectra

http://massbank.eu/MassBank und https://github.com/MassBank/



>88,100 spectra
~16,500 compounds
>47 contributors

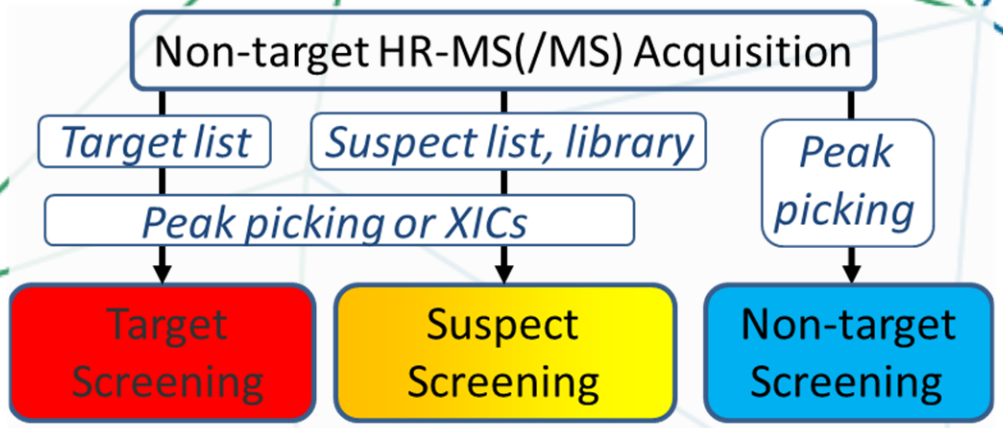
Challenge 2: Connection to Effects!



Modified from Escher, Stapleton, Schymanski (2020). Science. DOI: 10.1126/science.aay6636

How can we digest all this information in Exposomics?

collaborative



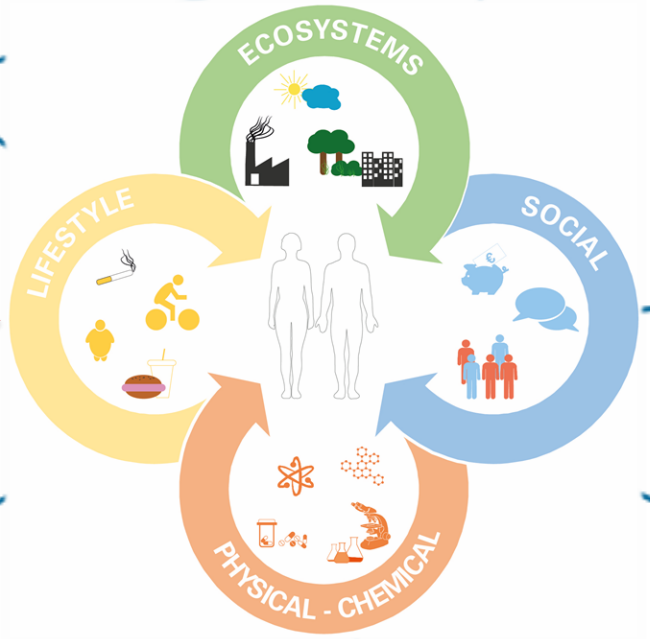
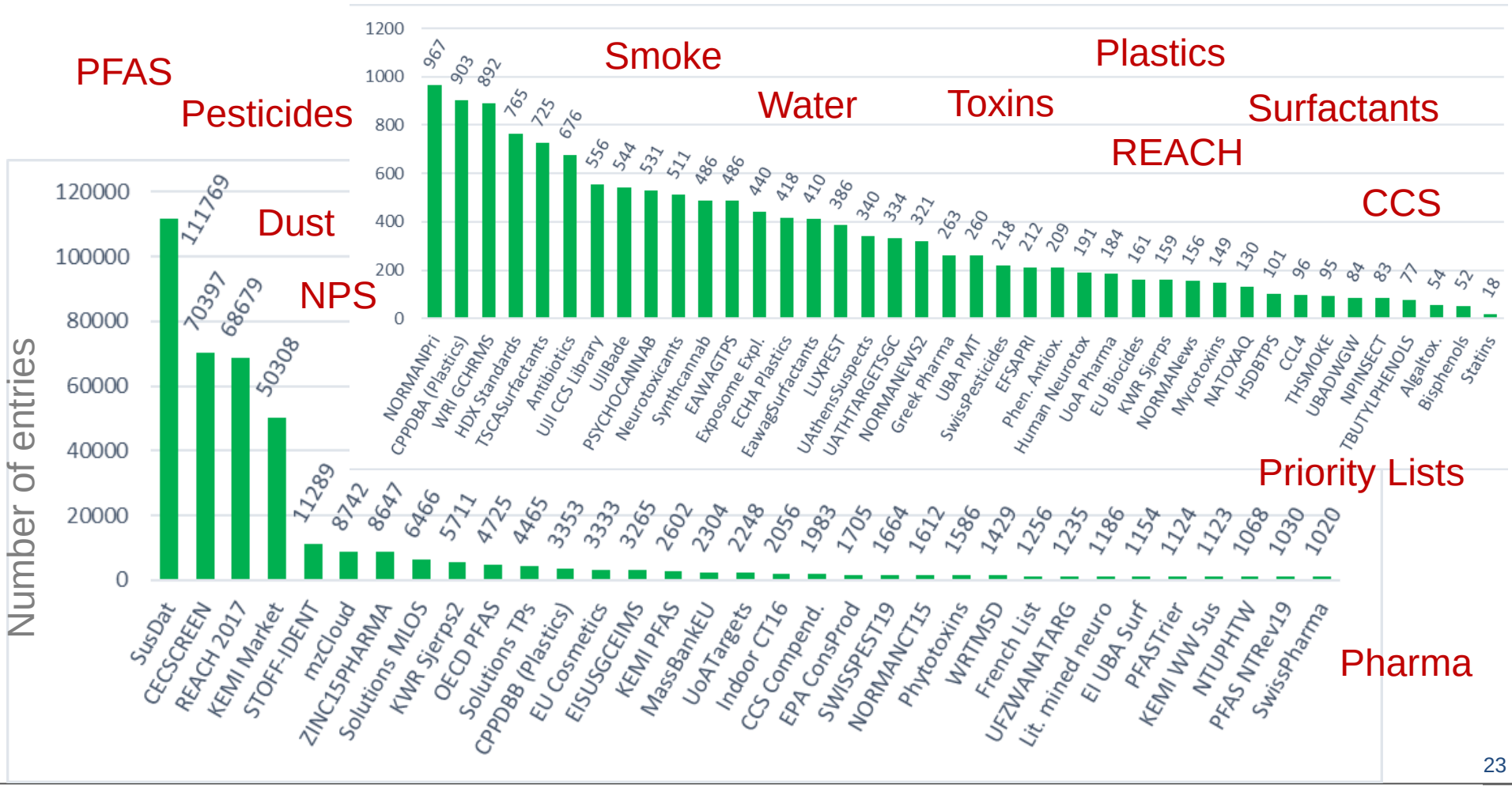


Image: https://www.wikidata.org/wiki/Wikidata:Main_Page; Vermeulen *et al* (2020) Science. DOI: 10.1126/science.aay3164

NORMAN Suspect List Exchange: Capturing Knowledge

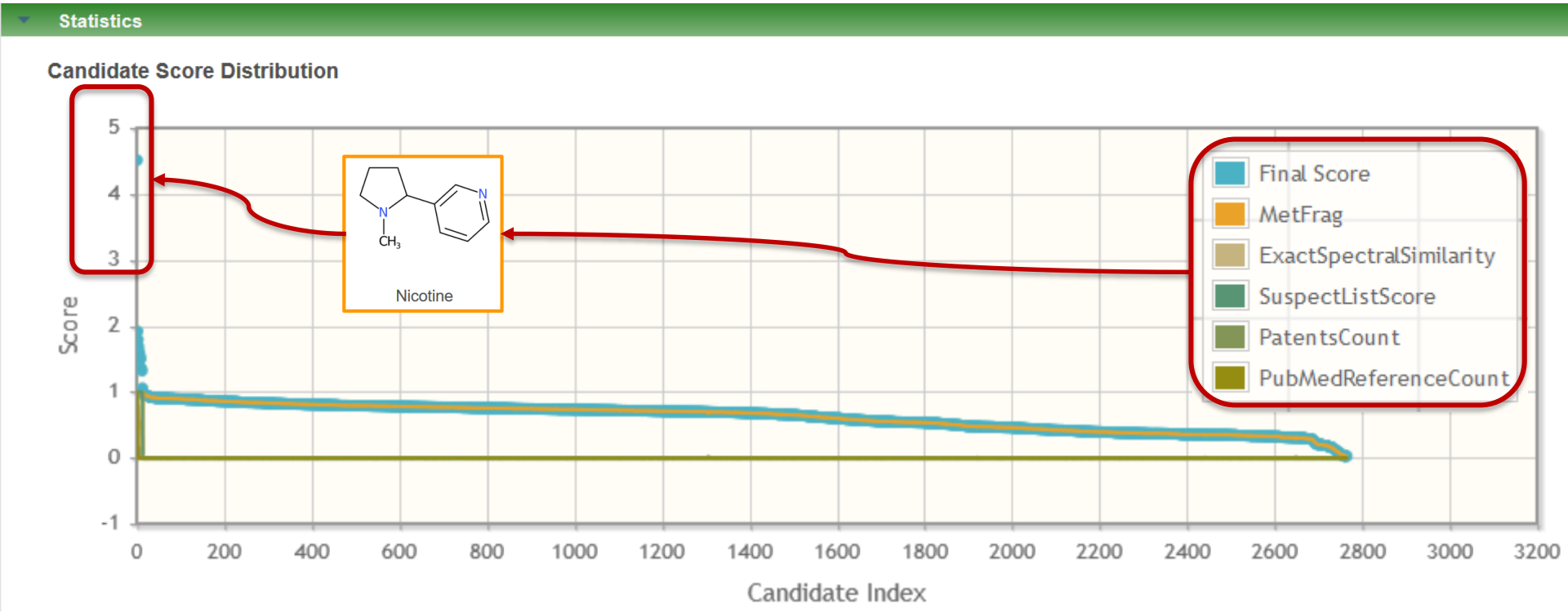
- https://www.norman-network.com/nds/SLE/
- https://zenodo.org/communities/norman-sle
- https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

>73 lists!



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

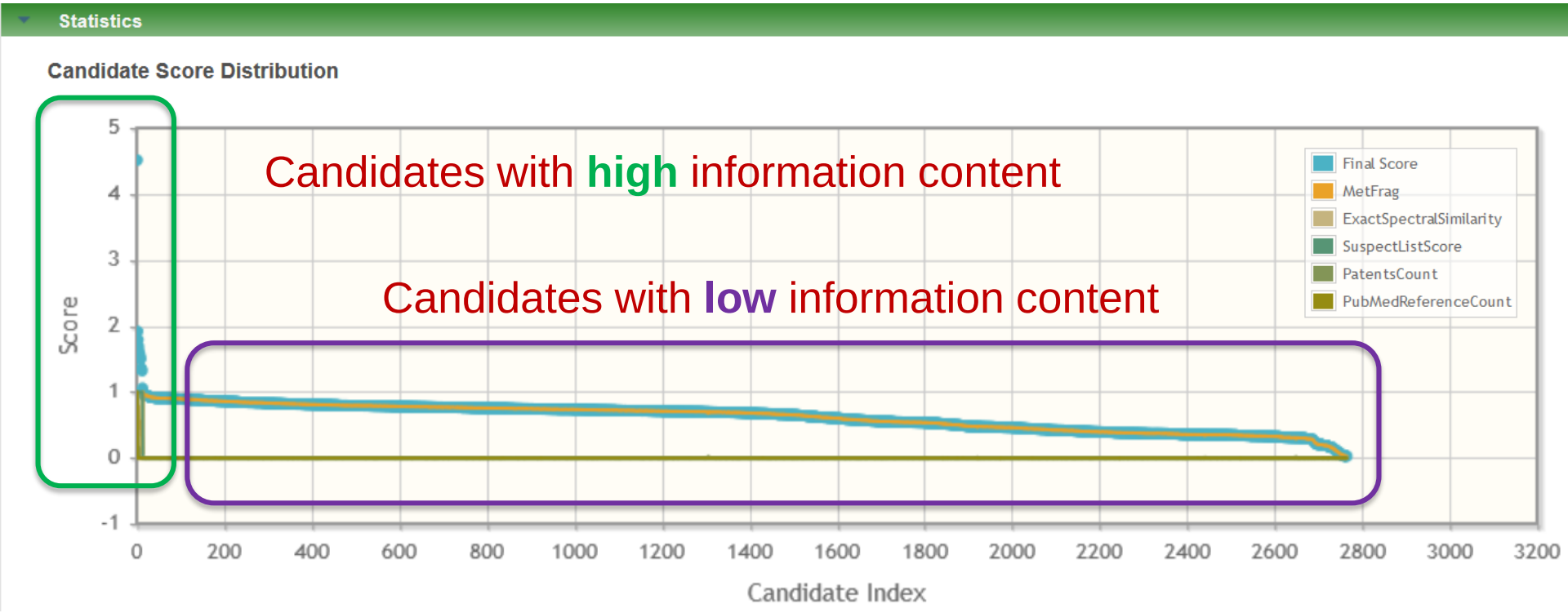
Connecting multiple lines of evidence for identification



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

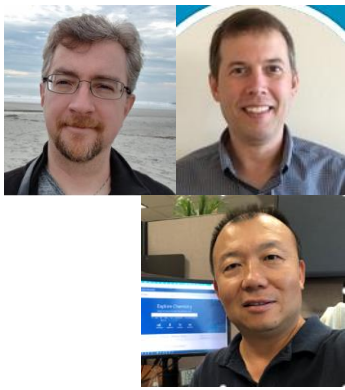
- PubChem Compound TOC ? 33,765,953
  - Agrochemical Information ? 2,002
  - Biologic Description ? 1,539,532
  - Biological Test Results ? 3,467,416
  - Biomolecular Interactions and Pathways ? 109,610
  - Chemical and Physical Properties ? 237,729
  - Classification ? 18,569,627
  - Diseases ?
  - Drug and Medication Information ? 15,955
  - Food Additives and Ingredients ? 7,447
  - Identification ? 5,746
  - 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

Edit

New version

Communities

LCSB Environmental Cheminformatics Group — Remove

503 views — 482 downloads

See more details...

Database Settings

Database: PubChemLite_14Jan2020

Neutral Mass: 162.11576 Search ppm: 5

Formula: C10H14N2

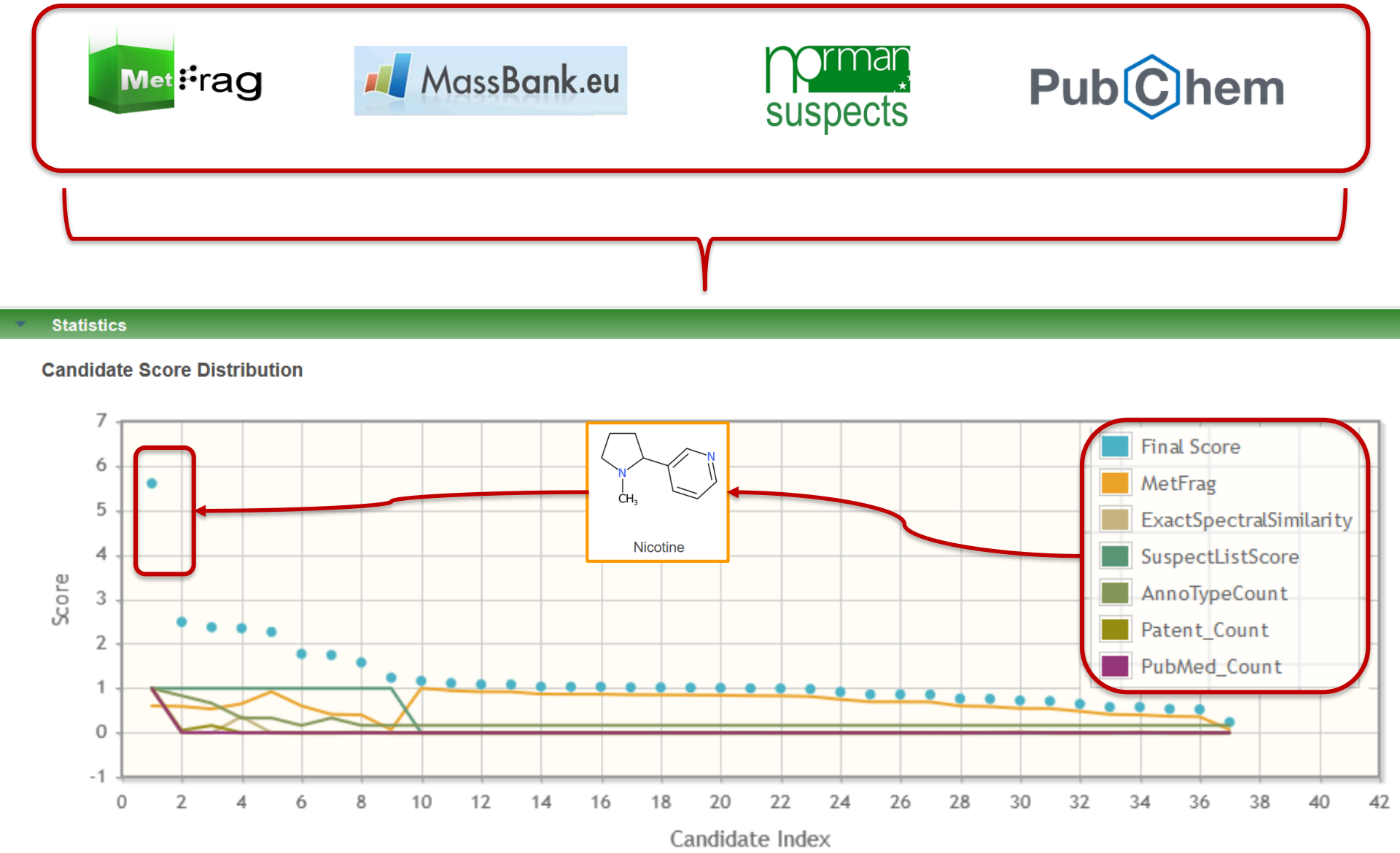
Identifiers:

Retrieve Candidates — 37 Candidates

MetFrag

Bolton & Schymanski (2020). PubChemLite tier0 and tier1 (Version PubChemLite.0.2.0) DOI: 10.5281/zenodo.3611238

PubChemLite: tailor-made database + metadata

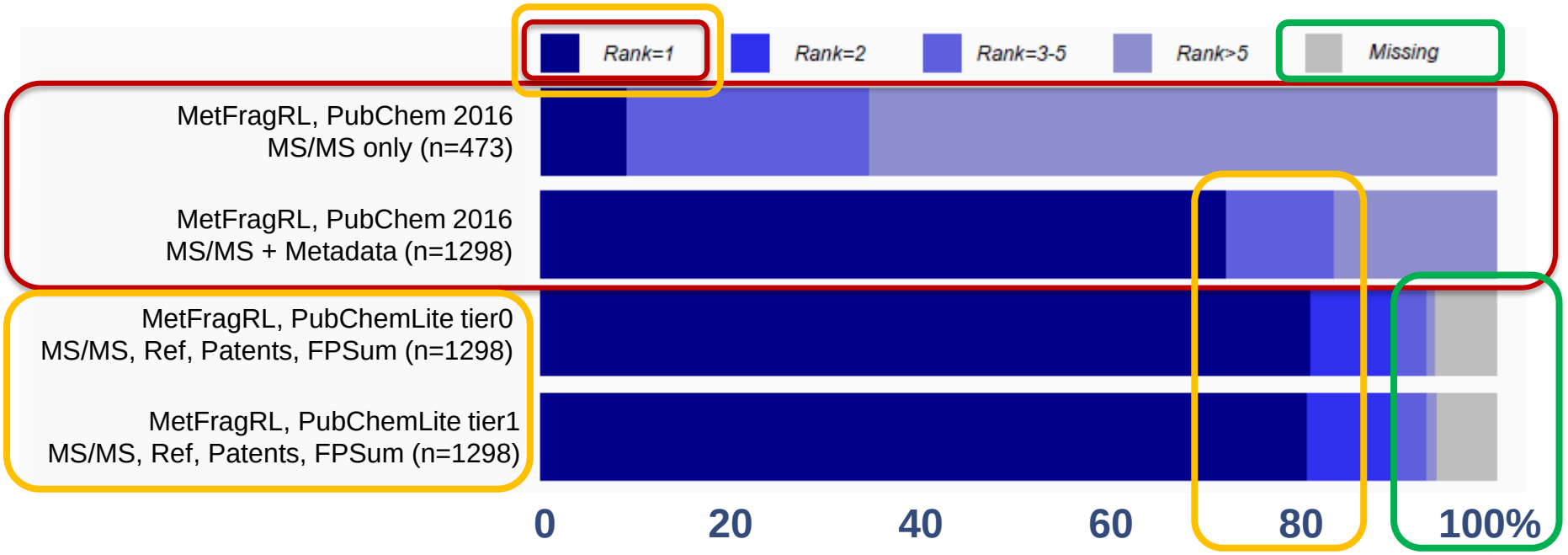


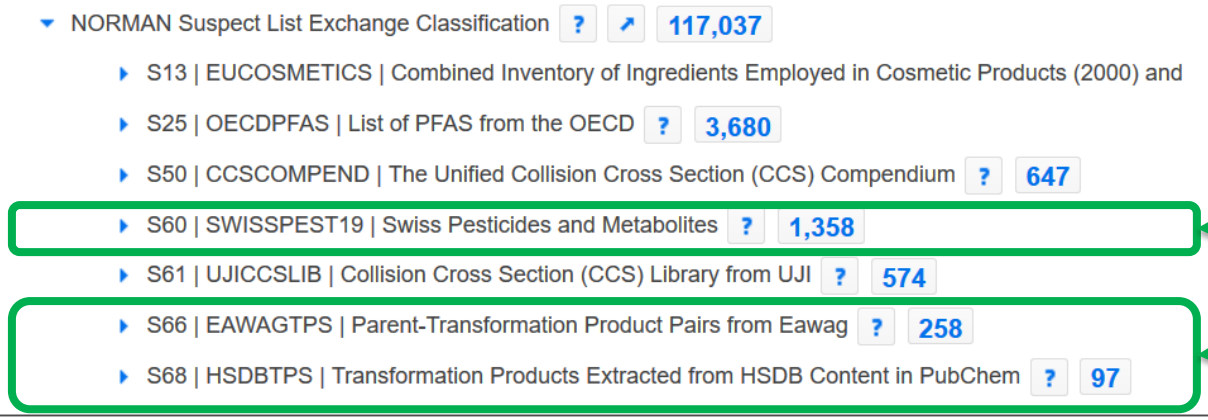
MetFrag+PubChemLite+Formula+MoNA+SusDat+Pat+Refs+Anno + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

How does PubChemLite perform?

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

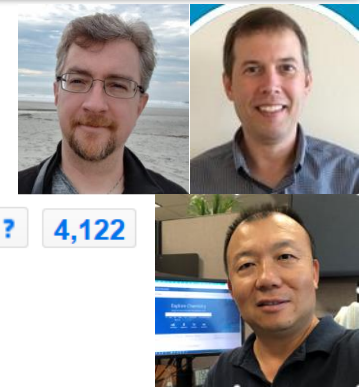
- 103 M => 300 K … how does this influence performance?





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

Transformation Products: Filling the Data Gaps!



PubChem NORMAN Suspect List Exchange

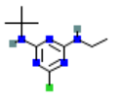
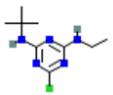
- NORMAN Suspect List Exchange Classification ? 117,037
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 4,122
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (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

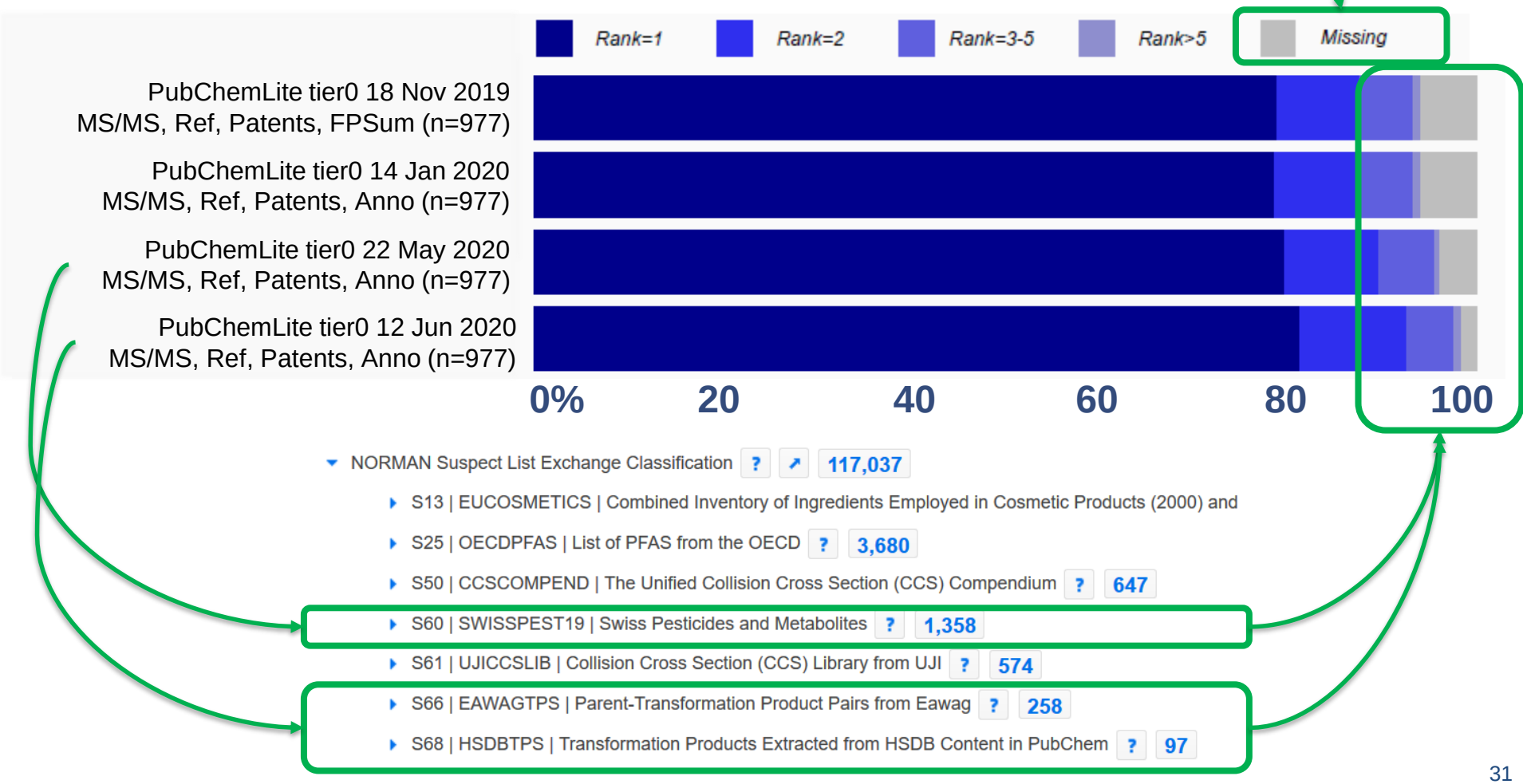
Terbutylazine has known environmental transformation products that include CSAA036479, CSAA04949, CSCD648241, CSCD692760, GS31398, MT1, GS 26379, MT13, GS 23158, Terbutylazine metabolite MT14, Terbutylazine metabolite MT23, and Terbutylazine metabolite MT24.

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

Filling in the “Chemical Space”

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

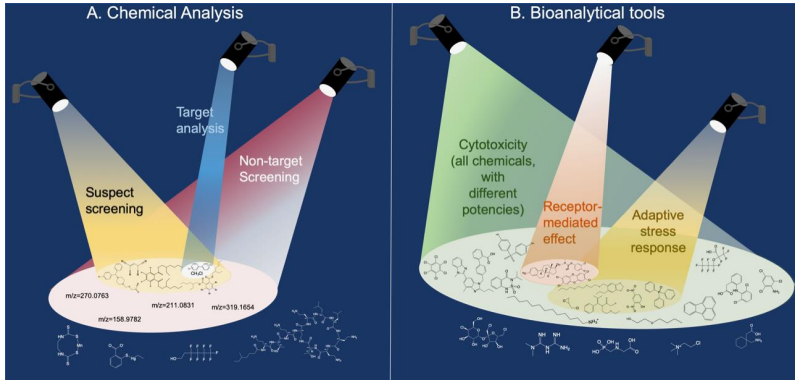
- Assessing the missing entries …



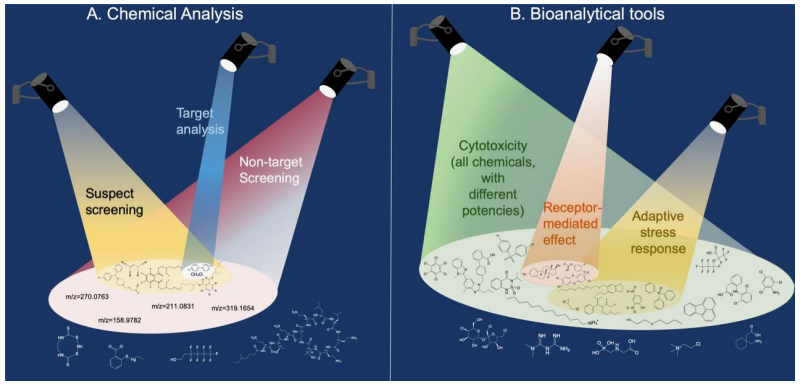
Ruttkies et al. (2016) DOI: 10.1186/s13321-016-0115-9, Bolton & Schymanski (2020), DOI: 10.5281/zenodo.3611238 & in prep.

Take Home Messages

- Challenges:
  Selecting ***which method(s)***
  Increasing ***% identified***
  Improving ***interpretation***



Take Home Messages

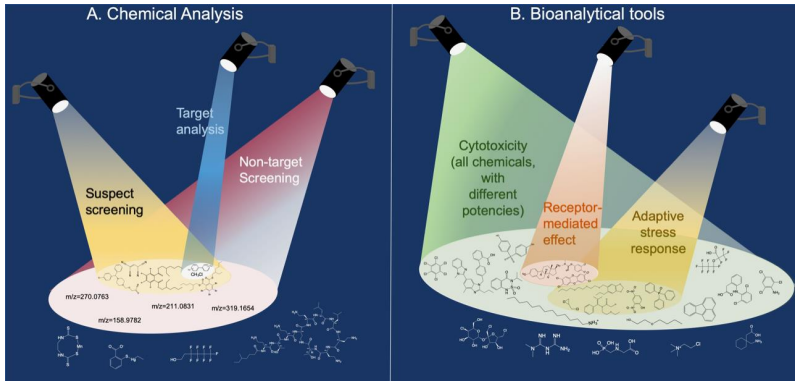


- Challenges:
  Selecting ***which method(s)***
  Increasing ***% identified***
  Improving ***interpretation***
- ***Annotating unknowns*** requires data and evidence from ***many different sources***
  - Many excellent workflows available to collate this information
  - Incorporation of ***all available metadata*** is critical to success!

MetFrag MassBank.eu norman suspects PubChem EPA Chemicals

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

Take Home Messages



- Challenges:
  Selecting ***which method(s)***
  Increasing ***% identified***
  Improving ***interpretation***
- ***Annotating the exposome*** 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!

MetFrag · MassBank.eu · NORMAN suspects · PubChem · EPA Chemicals

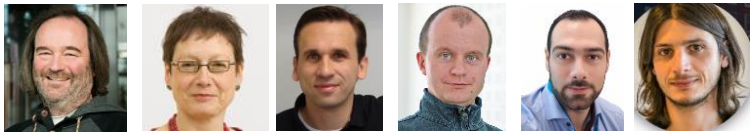
- Finally … information in the public domain helps ***everybody***
  - You never know when it will help you ☺

NORMAN NEWS

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

Acknowledgements

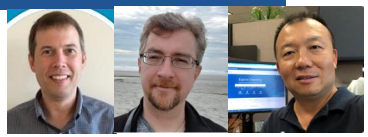
Luxembourg National Research Fund



Beate Escher
Spotlights!

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

Christoph Ruttkies,
Steffen Neumann,
MetFrag Team



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

Tobias Schulze
MassBank Team

Gary Miller
All things Exposome

emma.schymanski@uni.lu and @ESchymanski

Further Information:

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

https://zenodo.org/communities/lcsb-eci/

Slides available under DOI: 10.5281/zenodo.4032909

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