

## MassBank Europe – a success story of community efforts

Tobias Schulze, UFZ, Germany  
Rene Meier, IPB, Germany  
Steffen Neumann, IPB, Germany  
Herbert Oberacher, MUI, Austria  
Emma Schymanski, LCSB, Luxembourg

ASMS 2020 Reboot, June 3, 2020  
Live Webinar Workshop on Compound  
Identification by Mass Spectral Library Searching

DOI: 10.5281/zenodo.3865489

MassBank High Quality Mass Spectral Database

### MassBank Europe

Home Search Record Index Data Privacy Imprint

**Search**

Compound Name	CAS	SMILES
...	...	...

**Record Index**

Molecular Weight	Retention Time	Scan Number
...	...	...

Lipid Index

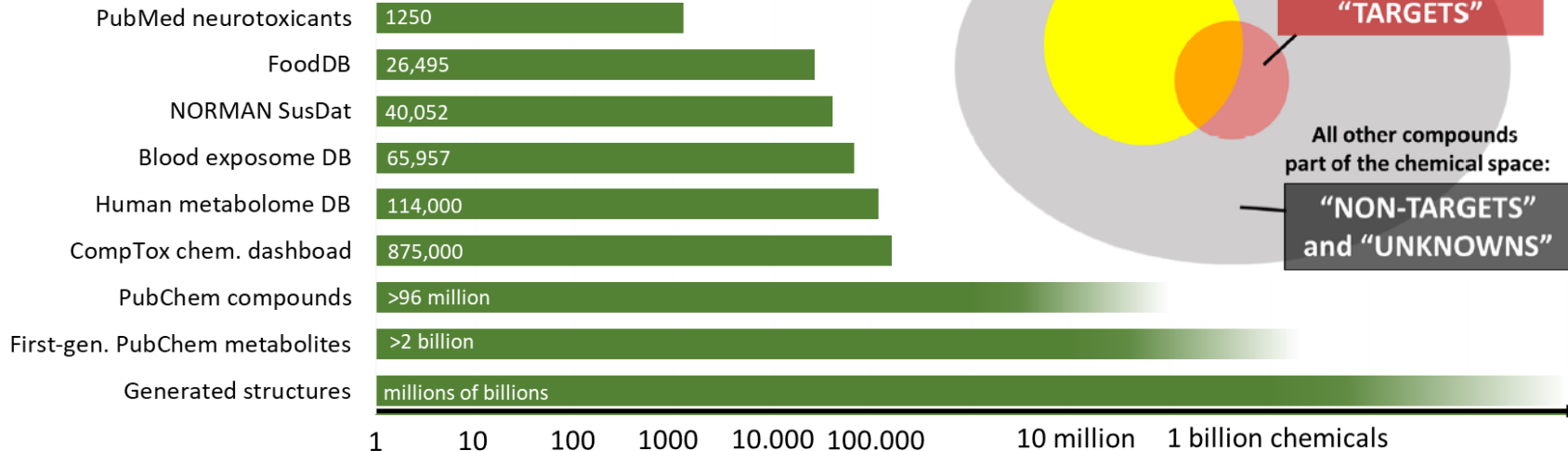
# Our (community) challenge: Identifying compounds



## Typical HRMS sample



## Selected exposomics, chemical data sources



Compounds of interest:

**"SUSPECTS"**

Compounds included in tandem mass spectral databases:

**"TARGETS"**

All other compounds part of the chemical space:

**"NON-TARGETS"**  
and **"UNKNOWN"**

DOI: 10.1126/science.aay3166 (adapted)

DOI: 10.1186/s12302-020-00314-9

# MassBank – library coverage

## Record Index

Home Search Record Index Data Privacy Imprint MassBank I

<b>Contributor</b>	<a href="#">AAFC</a> (950) <a href="#">Athens_Univ</a> (5252) <a href="#">BGC_Munich</a> (903) <a href="#">BS</a> (1318) <a href="#">Boise_State_Univ</a> (4) <a href="#">CASMI_2012</a> (26) <a href="#">CASMI_2016</a> (622) <a href="#">Chubu_Univ</a> (2563) <a href="#">Eawag</a> (11191) <a href="#">Eawag_Additional_Specs</a> (895)	<a href="#">Env_Anal_Chem_U_Tuebingen</a> (128) <a href="#">Fac_Eng_Univ_Tokyo</a> (12379) <a href="#">Fiocruz</a> (898) <a href="#">Fukuyama_Univ</a> (340) <a href="#">GL_Sciences_Inc</a> (174) <a href="#">HBM4EU</a> (1925) <a href="#">IPB_Halle</a> (677) <a href="#">ISAS_Dortmund</a> (339) <a href="#">JEOL_Ltd</a> (44) <a href="#">KWR</a> (207)	<a href="#">Kazusa</a> (273) <a href="#">Keio_Univ</a> (4780) <a href="#">Kyoto_Univ</a> (184) <a href="#">Literature_Specs</a> (39) <a href="#">MPI_for_Chemical_Ecology</a> (691) <a href="#">MSSJ</a> (328) <a href="#">MetaboLights</a> (58) <a href="#">Metabolon</a> (149) <a href="#">NAIST</a> (621) <a href="#">NaToxAq</a> (3756)	<a href="#">Nihon_Univ</a> (706) <a href="#">Osaka_MCHRI</a> (20) <a href="#">Osaka_Univ</a> (449) <a href="#">PFOS_research_group</a> (41) <a href="#">RIKEN</a> (11935) <a href="#">RIKEN_IMS</a> (1140) <a href="#">RIKEN_NPDepo</a> (1956) <a href="#">RIKEN_ReSpecT</a> (4642) <a href="#">Tottori_Univ</a> (16) <a href="#">UFZ</a> (1242)
<b>Instrument Type</b>	<a href="#">APCI-ITFT</a> (69) <a href="#">APCI-ITTOF</a> (1) <a href="#">APCI-Q</a> (3) <a href="#">CE-ESI-TOF</a> (20) <a href="#">CI-B</a> (796) <a href="#">CI-Q</a> (8) <a href="#">EI-B</a> (11810) <a href="#">EI-EBER</a> (12) <a href="#">ESI-ITFT</a> (62)	<a href="#">ESI-ITTOF</a> (4) <a href="#">ESI-QIT</a> (1) <a href="#">ESI-QTOF</a> (252) <a href="#">ESI-TOF</a> (128) <a href="#">FAB-B</a> (26) <a href="#">FAB-BE</a> (15) <a href="#">FAB-EB</a> (5) <a href="#">FAB-EBER</a> (172) <a href="#">FD-B</a> (41)	<a href="#">FI-B</a> (1) <a href="#">GC-APCI-QTOF</a> (94) <a href="#">GC-EI-Q</a> (45) <a href="#">GC-EI-QQ</a> (19) <a href="#">GC-EI-TOF</a> (954) <a href="#">GC-FI-TOF</a> (6) <a href="#">LC-APCI-ITFT</a> (50) <a href="#">LC-APCI-Q</a> (12) <a href="#">LC-APCI-QTOF</a> (633)	<a href="#">LC-APPI-QQ</a> (287) <a href="#">LC-ESI-IT</a> (664) <a href="#">LC-ESI-ITFT</a> (16884) <a href="#">LC-ESI-ITTOF</a> (262) <a href="#">LC-ESI-Q</a> (2738) <a href="#">LC-ESI-QFT</a> (7136) <a href="#">LC-ESI-QIT</a> (413) <a href="#">LC-ESI-QQ</a> (8867) <a href="#">LC-ESI-QQQ</a> (1956)
<b>MS Type</b>	<a href="#">MS</a> (19982)	<a href="#">MS2</a> (59680)	<a href="#">MS3</a> (929)	<a href="#">MS4</a> (70)
<b>Ion Mode</b>	<a href="#">NEGATIVE</a> (23266)	<a href="#">POSITIVE</a> (57395)		

80661 unique spectra (MS2, MS1, some MS<sub>n</sub>; ESI, EI, APCI...)

14382 unique compounds (anthropogenic, metabolites, TPs)

**Search** → mass c

[Home](#) [Search](#) [Record Index](#) [Data Privacy](#) [Imprint](#)

**Search for:**

Compound Information

**Compound name**

**Exact Mass**  **Tolerance**

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

Similarity search:

$$W_i = [\text{Intensity of peak}_i]^m [m/z \text{ of peak}_i]^n$$

DOI: 10.1002/jms.1777

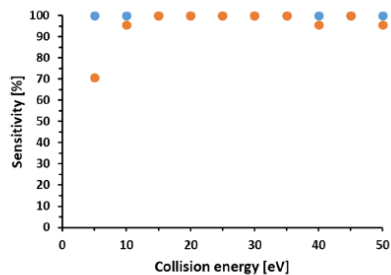
DOI: 10.1016/1044-0305(94)87009-8

DOI: 10.1038/nbt.3689

Integration in tools such as MetFrag, MASST, MZmine,  
NIST MS Search possible by download from MoNA

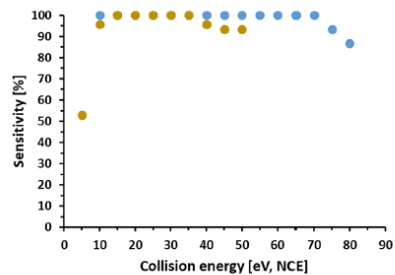
# Reliability of library searches

**a QqTOF**



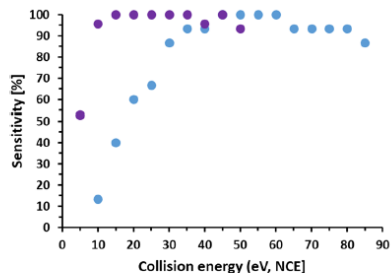
● ... TripleTOF spectra matched to the WRTMD library  
● ... WRTMD spectra matched to the TripleTOF hybrid library

**b Q-Orbitrap**



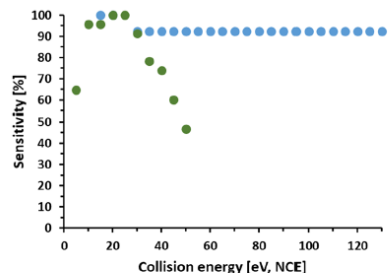
● ... Orbitrap-HCD spectra matched to the WRTMD library  
● ... WRTMD spectra matched to the Orbitrap-HCD hybrid library

**c LIT-Orbitrap with HCD**

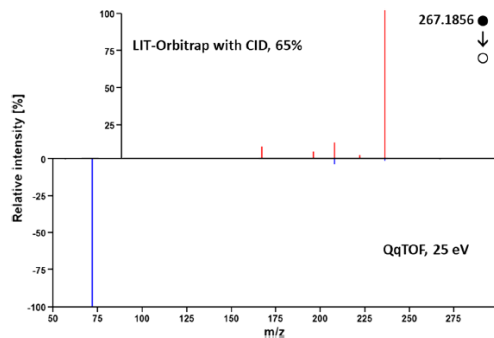


● ... Orbitrap-HCD spectra matched to the WRTMD library  
● ... WRTMD spectra matched to the Orbitrap-HCD hybrid library

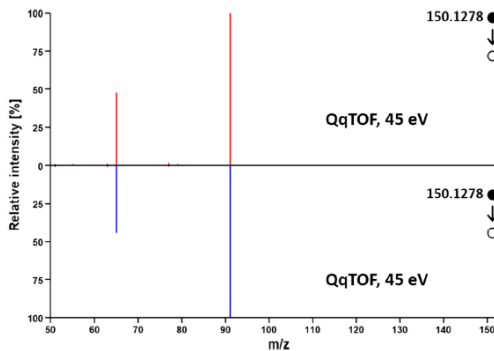
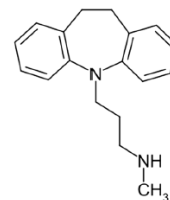
**d LIT-Orbitrap with CID**



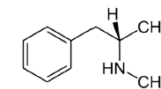
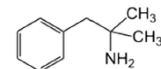
● ... Orbitrap CID spectra matched to the WRTMD library  
● ... WRTMD spectra matched to the Orbitrap-CID hybrid library



**Desipramine**



**Phentermine**

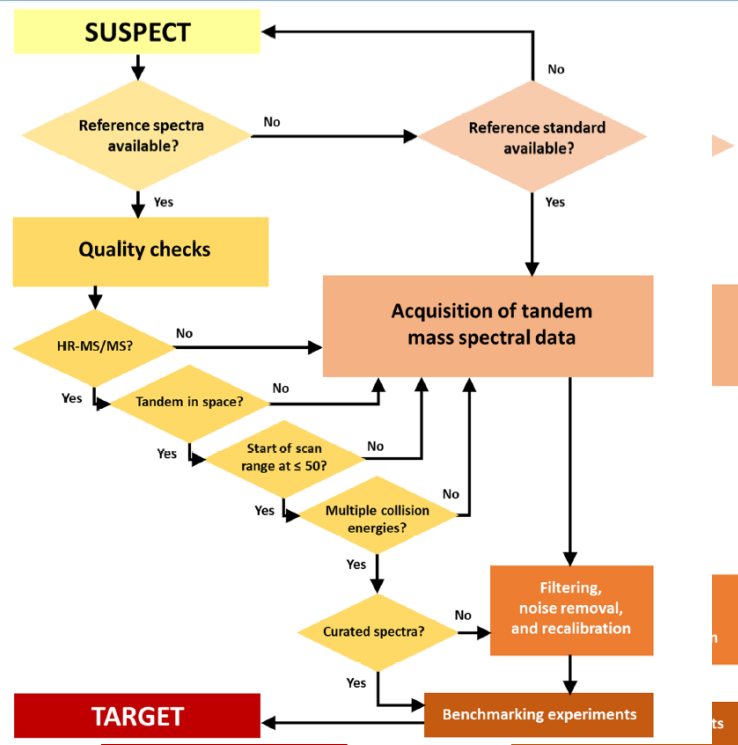


**Methamphetamine**

# Smart strategy to grow community libraries



- Community MS libraries cannot grow as fast as commercial databases
- Smart collections of suspect lists (priority compounds) could help to focus on relevant compounds for different communities
- Shared MS acquisition will help to populate community MS libraries with relevant content more quickly



# MassBank – data community



MassBank records

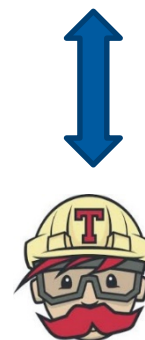
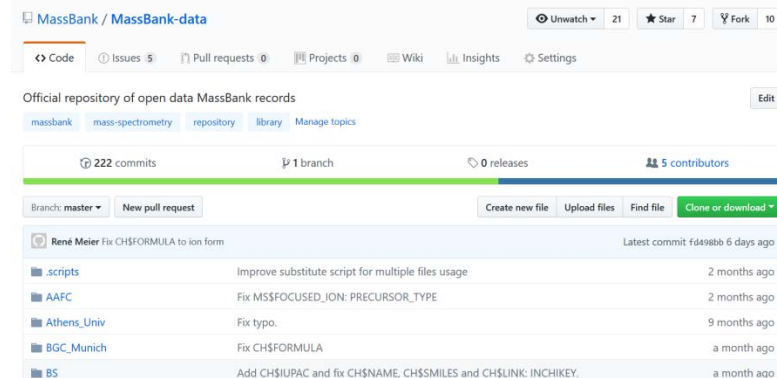
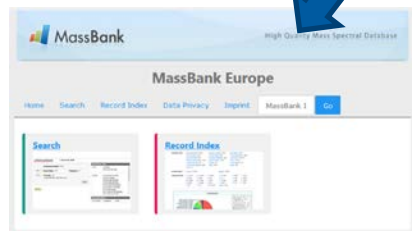
User upload to GitHub

Share with friends

Automated deployment in MassBank

Automated validation and curation with Travis CI

<https://www.travis-ci.org>



## Full machine readable representation of each record and open access integration

```
"identifier": "AU248503",
"url": "https://massbank.eu/RecordDisplay?id=AU248503",
"name": "(Methoxymethyl)triphenylphosphonium",
"alternateName": ["(Methoxymethyl)triphenylphosphonium", "methoxymethyl(triphenyl)phosphanium"],
"molecularFormula": "[C20H20OP]+",
"monoisotopicMolecularWeight": "307.12517689999998538041836582124233245849609375",
"inChI": "InChI=1S/C20H20OP/c1-21-17-22(18-11-5-2-6-12-18,19-13-7-3-8-14-19)20-15-9-4-10-16-20/h2-16H,17H2,1H3/q+1",
"smiles": "COC[P+](C1=CC=CC=C1)(C1=CC=CC=C1)C1=CC=CC=C1",
"@context": "http://schema.org",
"@type": "MolecularEntity"
},
{
"identifier": "AU248503",
"url": "https://massbank.eu/RecordDisplay?id=AU248503",
"headline": "(Methoxymethyl)triphenylphosphonium; LC-ESI-QTOF; MS2; CE: 30 eV; R=35000; [M+H]+",
"name": "(Methoxymethyl)triphenylphosphonium",
"datePublished": "2019-04-08",
"license": "https://creativecommons.org/licenses/",
"citation": "null",
"comment": ["CONFIDENCE standard compound", "INTERNAL_ID 2485"],
"alternateName": ["(Methoxymethyl)triphenylphosphonium", "methoxymethyl(triphenyl)phosphanium"],
"@context": "http://schema.org",
"@type": "Dataset"
```

<https://bioschemas.org>

<https://www.assembleplus.eu>

<https://elixir-europe.org/about-us/who-we-are/nodes/germany>

[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/MassBank\\_Accession\\_ID](https://www.wikidata.org/wiki/Wikidata:Property_proposal/MassBank_Accession_ID)

### Wikidata:Property proposal/MassBank Accession ID

[Wikidata:Property proposal](#)

#### MassBank Accession ID [\[ edit \]](#)

Originally proposed at [Wikidata:Property proposal/Natural science](#)

<span style="color: green;">■</span> Done: MassBank Accession ID (P6689) <a href="#">(Talk and documentation)</a>	
<b>Description</b>	Accession number for entries in the MassBank database (records of mass spectrometry).
<b>Represents</b>	MassBank (Q24088019)
<b>Data type</b>	External identifier
<b>Domain</b>	chemical compound (Q11173)
<b>Allowed values</b>	[A-Z][2][A-Z0-9][0-9]{5}
<b>Example 1</b>	caffeine (Q60235) → EA030311 <a href="#">ⓘ</a>
<b>Example 2</b>	Kanamycin A (Q27094615) → SM100011
<b>Example 3</b>	acetaminophen (Q57055) → AU112601
<b>Source</b>	<a href="https://github.com/MassBank/MassBank-web/blob/master/Documentation/MassBankRecordFormat.md#2.1.1">https://github.com/MassBank/MassBank-web/blob/master/Documentation/MassBankRecordFormat.md#2.1.1</a> <a href="#">ⓘ</a>
<b>Planned use</b>	Link Wikidata the MassBank (Q24088019) database entries.
<b>Number of IDs in source</b>	53390
<b>Expected completeness</b>	always incomplete (Q21873886)
<b>Formatter URL</b>	<a href="https://massbank.eu/MassBank/RecordDisplay.jsp?id=\$1">https://massbank.eu/MassBank/RecordDisplay.jsp?id=\$1</a>
<b>Robot and gadget jobs</b>	Accession IDs will be added from a public CCZero dataset <a href="#">(to be compiled)</a> with QuickStatements



# MassBank – compound database integration

## MassBank Europe



MassBank Europe (MassBank.EU) was created in 2011 as an open access database of mass spectra of emerging substances to support identification of unknown substances within the NORMAN Network (<https://www.norman-network.com/>). MassBank.EU is the partner project of MassBank.JP, hosted at the Helmholtz Center for Environmental Research (HZE) Leipzig and jointly maintained by UCL, UCLB (University of Luxembourg) and IFE Halle.

Organization: MassBank Europe  
Category: Research and Development  
URL: <https://massbank.eu/MassBank/>  
License URL: <https://github.com/MassBank/MassBank-eu/blob/master/LICENCE>  
Contact Name: MassBank Europe  
Address: Fernowstrasse 15, Leipzig, Germany, D-04318  
Data Source ID: Z3999  
Data in PubChem: 16,824 Live Substances  
Last updated: 2020/03/06



PubChem Atrazine (Compound)

MolView ID	SM841401
MS Category	Experimental
MS Type	Chromatography identified as LC-MS
MS Level	MS2
Precursor Type	[M+H] <sup>+</sup>
precursor m/z	216.101
Instrument	Q Exactive Plus Orbitrap Thermo Scientific
Instrument Type	LC-ESI-QFT
Ionization	ESI
Ionization Mode	positive
Collision Energy	35 (nominal)
Retention Time	9.326 min
Spinach	<a href="https://pubchem.ncbi.nlm.nih.gov/compound/Atrazine">https://pubchem.ncbi.nlm.nih.gov/compound/Atrazine</a>

Thumbnail



## MASSPECDB: MassBank Reference Spectra Collection



Search MASSPECDB Chemicals

Identifier substring search

### List Details

Select all Download Send to Batch Search Default 1267 chemicals

DTXSID CASRN TOXCAST Hide chemicals that are: [dropdown] [button]

- Acetaminophen**  
DTXSID:DTXSID0020006  
CASRN:102-90-3  
TOXCAST:5486
- Acifluorfen**  
DTXSID:DTXSID0020022  
CASRN:5094-66-9  
TOXCAST:37/803
- Alfatoxin B1**  
DTXSID:DTXSID0020035  
CASRN:1142-65-8  
TOXCAST:-
- Albendazole**  
DTXSID:DTXSID0020043  
CASRN:97-59-6  
TOXCAST:3/402
- Gabapentin**  
DTXSID:DTXSID0020074  
CASRN:60142-96-3  
TOXCAST:2/211
- Amtrazole**  
DTXSID:DTXSID0020076  
CASRN:61-82-5  
TOXCAST:2/409
- Aniline**  
DTXSID:DTXSID0020090  
CASRN:62-53-3  
TOXCAST:1/211
- Anthranic acid**  
DTXSID:DTXSID0020094  
CASRN:119-92-3  
TOXCAST:4/211

<https://pubchem.ncbi.nlm.nih.gov/source/23999>  
[https://comptox.epa.gov/dashboard/chemical\\_lists/MASSBANKREF](https://comptox.epa.gov/dashboard/chemical_lists/MASSBANKREF)



- Open data mass spectral libraries are a great benefit for science
- Creating open data libraries is a big community effort, but they are growing
- Sharing smartly collated suspect lists will help to populate community mass spectral libraries with relevant content for users (in progress)
- Open data and strict implementation of FAIR principles guarantees full interoperability with external mass spectral libraries, compound databases and tools (in progress)

# Acknowledgements



- Hendrik Treudler (IPB)
- Juliane Hollender, Michael Stravs, Heinz Singer (Eawag)
- Werner Brack, Martin Krauss, Erik Müller, Marc Stöhr, and the IT team (UFZ)
- Takaaki Nishioka (MassBank Japan)
- Antony Williams (US EPA)
- Evan Bolton, Jeff Zhang, Paul Thiessen (NCBI/NLM/NIH)
- Egon Willighagen (Maastricht University)
- Sajjan Singh Mehta, Gert Wohlgemuth, Oliver Fiehn (UC Davis)
- Ming Wang, Pieter Dorrestein (UC San Diego)
- David Wishart (University of Alberta)
- Xiaoyu Yang, Steven Stein (NIST)
- MassBank community



MoNA - MassBank of North America



MEDICAL  
UNIVERSITY  
INNSBRUCK



WIKIDATA

NIST



**Contact:**

**tobias.schulze@ufz.de**  
**emma.schymanski@uni.lu**  
**sneumann@ipb-halle.de**

