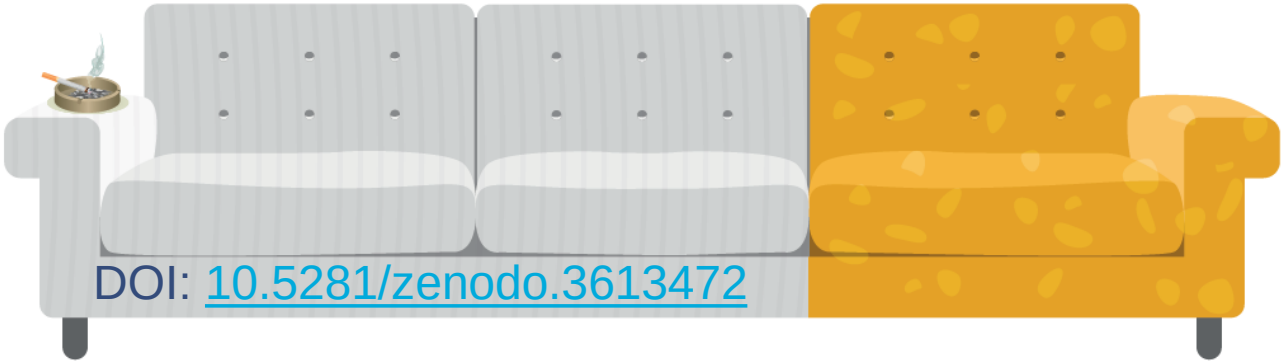
Environmental Cheminformatics:

Case study of Thirdhand Smoke in House Dust.



DOI: 10.5281/zenodo.3613472

Emma L. Schymanski[1], Sonia Torres[2], Noelia Ramirez[2]

[1]FNR ATTRACT Fellow; Luxembourg Centre for Systems Biomedicine, University of Luxembourg
[2]Metabolomics Core, IISPV-University Rovira i Virgili, Tarragona, Spain
emma.schymanski@uni.lu Twitter: @ESchymanski @soniatorres @noeliaramz

…plus many other colleagues who have contributed over the years!

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

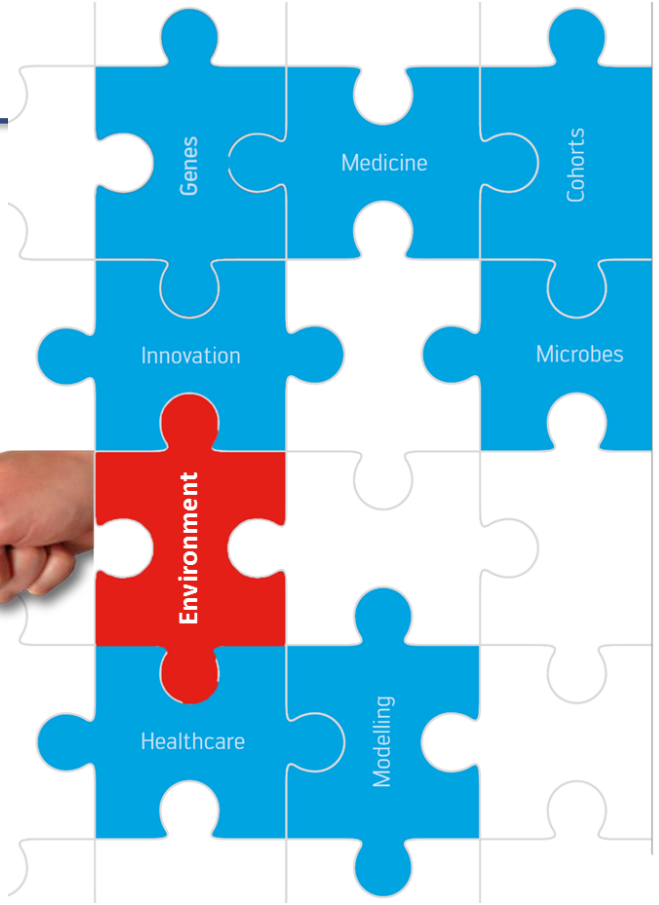
Lorenzo Favilli @lorenzo.favilli
Given access 2 months ago

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

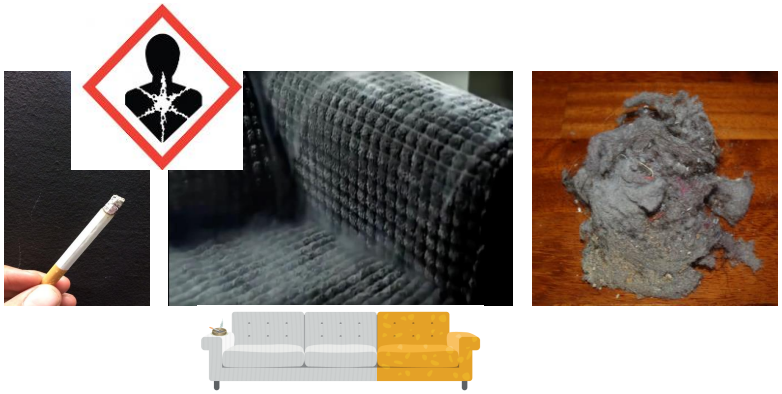


Luxembourg National Research Fund

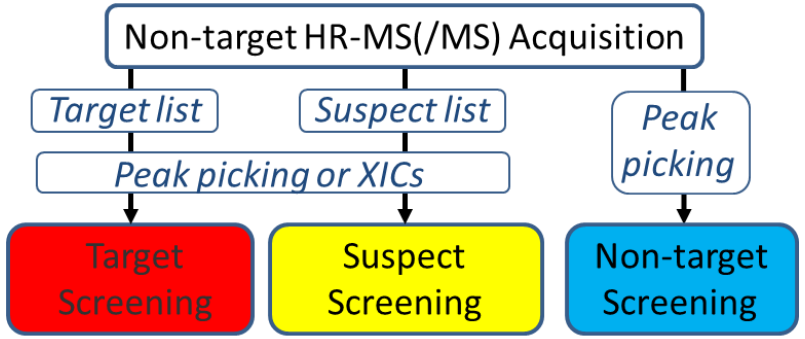
UNIVERSITÉ DU LUXEMBOURG

Environmental Cheminformatics: Case Study of

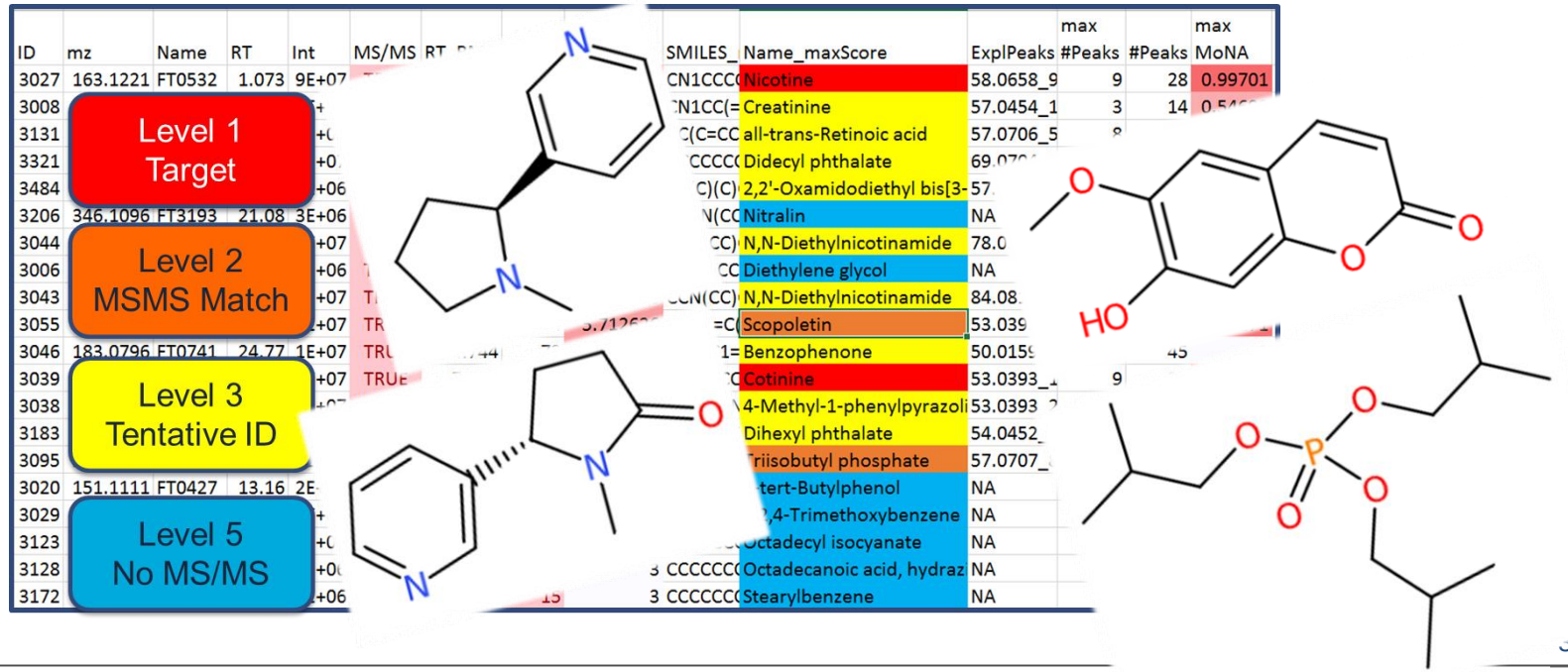
Thirdhand Smoke (THS) in House Dust



+

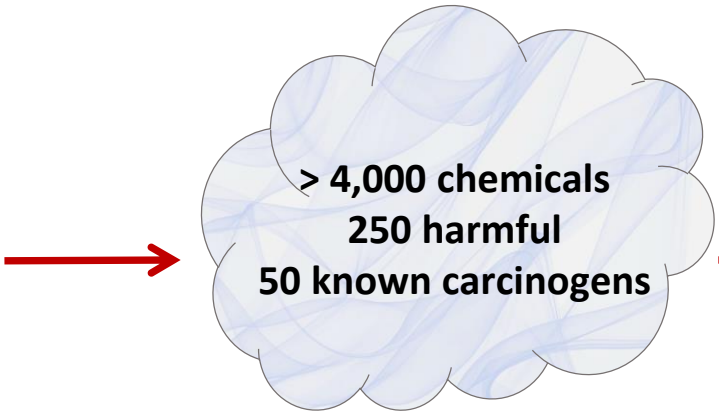


=

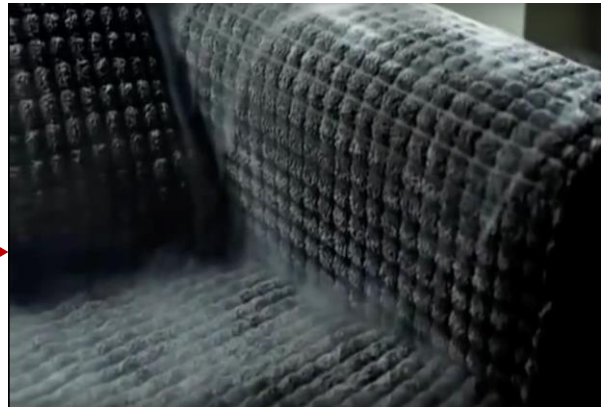


What is Thirdhand Smoke (THS)?

- **Thirdhand smoke** is the long-lasting residue resulting from second-hand smoke that accumulates in dust, in objects and on surfaces in indoor environments where tobacco has been smoked (WHO, 2017).



Tobacco smoke (second hand smoke)

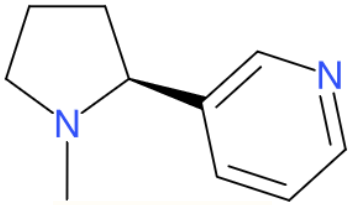


Thirdhand Tobacco Smoke (THS)

- Poorly studied, underestimated exposure route

Tobacco and its environmental impact: An overview. Geneva: World Health Organization; 2017.

What is Third Hand Smoke (THS)?



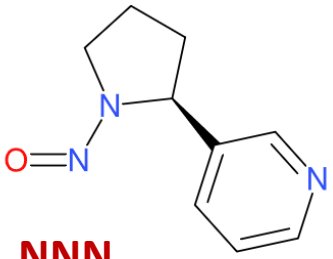
Nicotine

OXIDATION

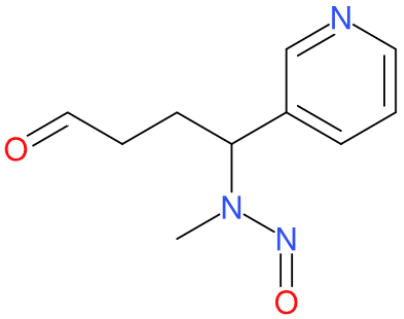
TOBACCO-SPECIFIC NITROSAMINES (TSNAs)

International Agency Research on Cancer

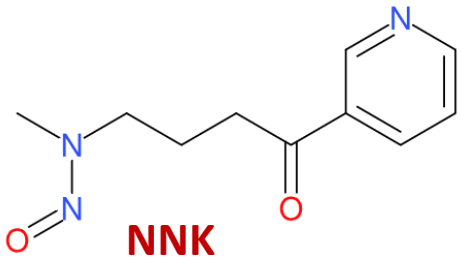
World Health Organization



NNN



NNA
(exclusive in THS)



NNK

Sleiman et al. PNAS (2010) DOI: 10.1073/pnas.0912820107

Main Exposure Pathways to THS

Dermal Absorption

Non-dietary Ingestion

Inhalation

Children < 5 years

Images c/o N. Ramirez

The challenge (for NTS)? Smoke is everywhere …

ROME (2011)

TPM (ng/m³)
Nicotine 1,700-4,800

Cecinato et al. Environ. Pollut. (2012)

TARRAGONA (2013)

TPM (ng/m³)
Nicotine 4 (max. 12.5)

Aragón et al. JCA (2013)

LONDON (2012)

$PM_{2.5}$ (ng/m³)
Nicotine 21 (max. 118)
TSNAs 1.2 (max. 6.2)

Farren et al. Environ. Sci. Technol. (2015)

SAN FRANCISCO (2016)

TPM (ng/m³)
Nicotine 339 (13% RSD)
TSNAs 0.8 (20% RSD)

Peyton et al. Chem. Res. Toxicol. (2017)

THS in Tarragona, Spain

SMOKER'S HOMES (n=22)

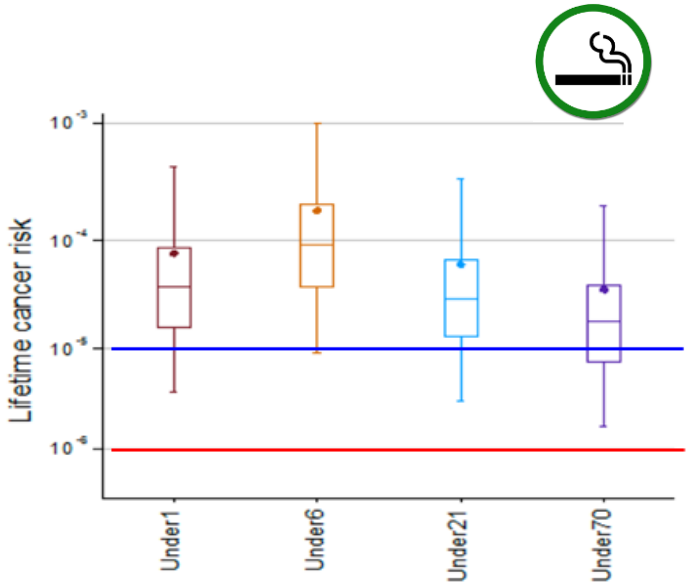
Nicotine 26 µg/g

TSNAs 0.5 µg/g

NON-SMOKER'S HOMES (n=24)

Nicotine 3.3 µg/g

TSNAs 0.09 µg/g

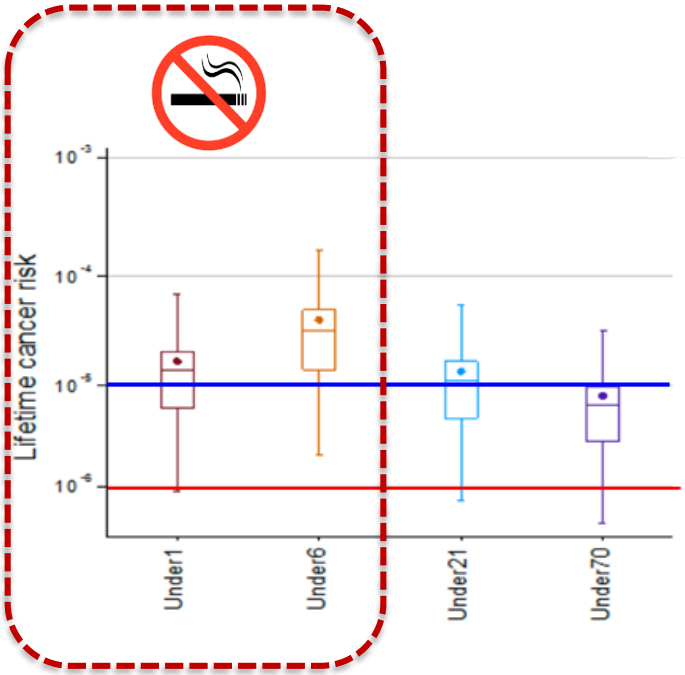


Lifetime cancer risk by age group.

Threshold values

WHO: 10^{-5}

USEPA: 10^{-6}



Ramírez *et al.* Environment International (2014). DOI: 10.1016/j.envint.2014.06.012

Motivation for NTS on THS …

SMOKER'S HOMES (n=22)

Nicotine 26 µg/g

TSNAs 0.5 µg/g

NON-SMOKER'S HOMES (n=24)

Nicotine 3.3 µg/g

TSNAs 0.09 µg/g

Motivation for NTS: What else is in there?

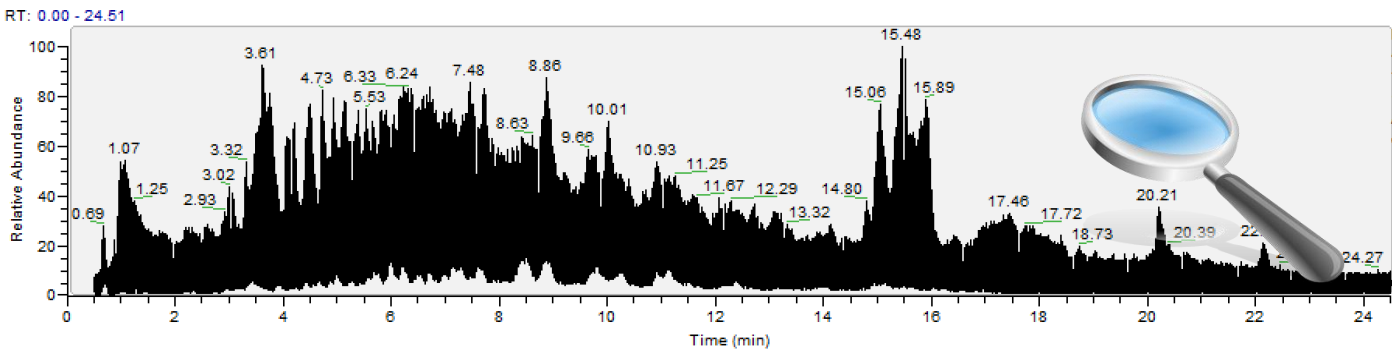
THS-RESEARCH TEAM

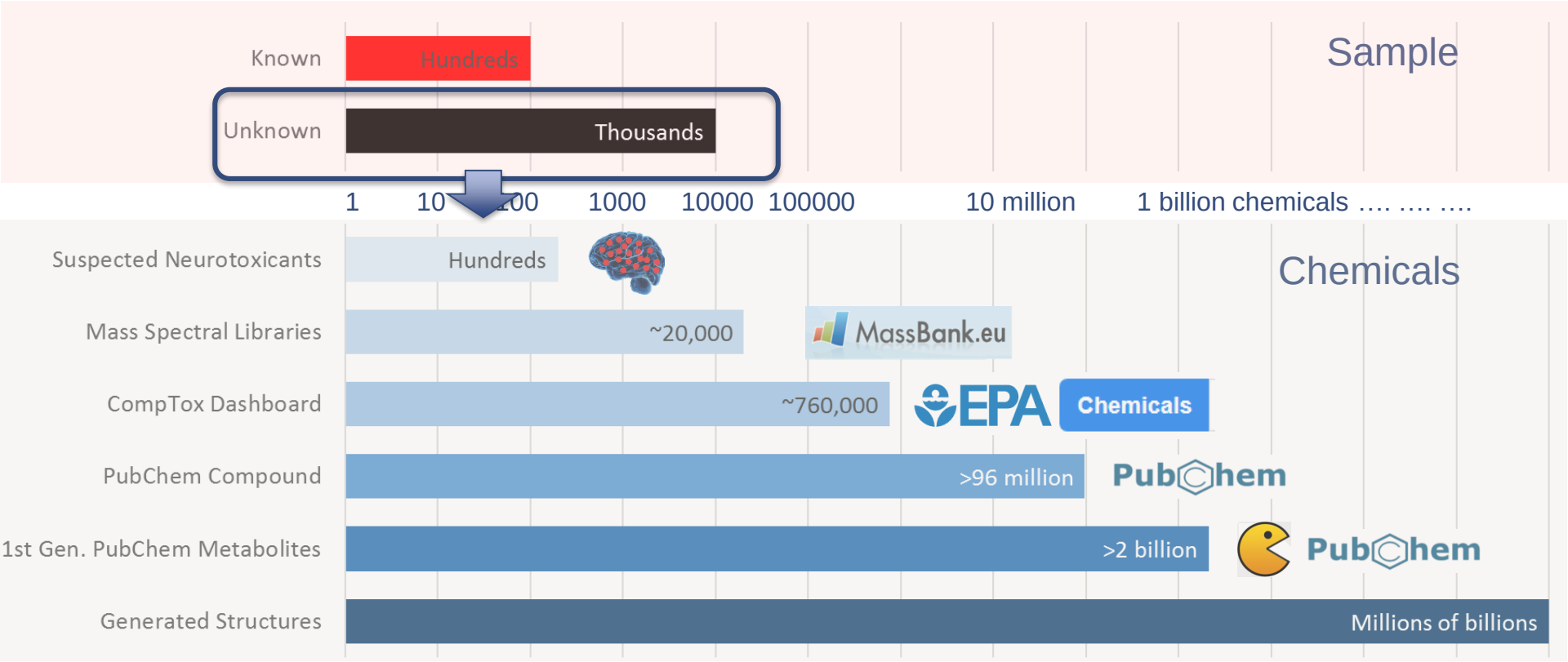
LCSB

Ramírez *et al.* Environment International (2014). DOI: 10.1016/j.envint.2014.06.012

The Challenge in Identifying Chemicals

High resolution mass spectrometry
+
Cheminformatics approaches
=
ECI@LCSB





Data: Schymanski *et al* 2014, Environ. Sci. Technol. DOI: 10.1021/es4044374; Hollender *et al* 2017 DOI: 10.1021/acs.est.7b02184

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

Gather evidence (nontarget, ReSOLUTION, RMassBank)

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

solutions

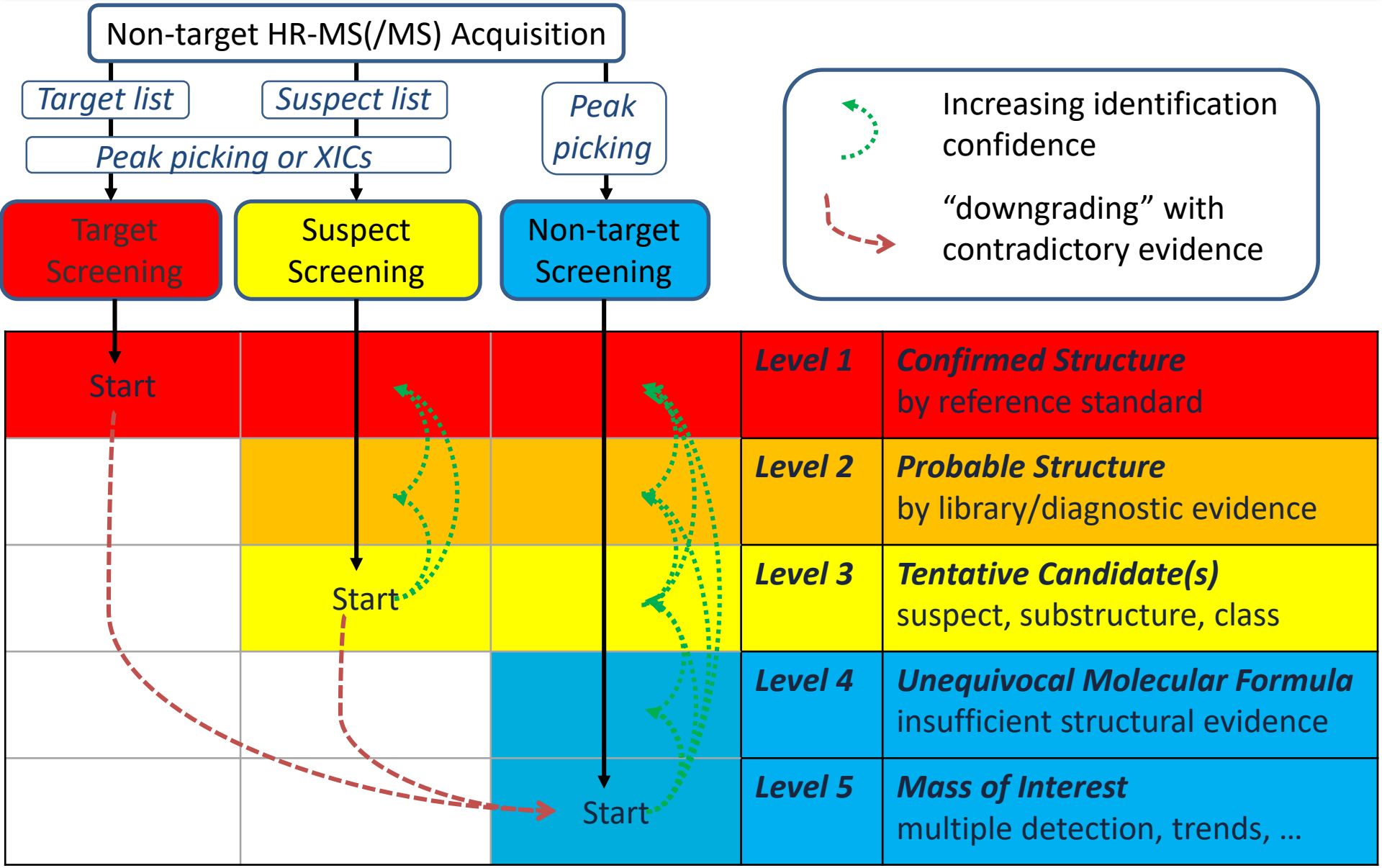
Terminology: Target, Suspect, Non-target/Unknown

TARGET: Known chemical, reference standard available *in house*

SUSPECT: Chemical suspected to be present in the sample, std. not (necessarily) available

NON-TARGET/ UNKNOWN: Mass/feature of interest detected in the sample, identity unknown

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

THS & Target, Suspect, Non-target/Unknown

TARGET:

Known chemical, reference standard available *in house*

SUSPECT:

Chemical suspected to be present in the sample, std. not (necessarily) available

NON-TARGET/ UNKNOWN:

Mass/feature of interest detected in the sample, identity unknown

Extraction (QuEChERS) and Target Analysis

300 mg dust

Add 2 mL water

Add 2 mL solvent

Shake 1 min

Addition of 0.5g extraction salts

1. Vortex 1 min
2. Centrifugation 4000rpm 5min

Two phase separation!

Transfer supernatant

Add 50 mg sorbent

1. Vortex 1 min
2. Centrifugation 4000rpm 5min

Filter supernatant into a vial

UHPLC-QQQ Analysis

Nicotine, Cotinine, NNAL, NNN, NNA, NNK + 5 d-IS

Target results essential to verify the non-target methods!

Torres, Schymanski, Ramirez *(in prep.).* Schematic c/o Sonia Torres

THS & Target, Suspect, Non-target/Unknown

TARGET: Known chemical, reference standard available *in house*

SUSPECT: Chemical suspected to be present in the sample, std. not (necessarily) available

NON-TARGET/ UNKNOWN: Mass/feature of interest detected in the sample, identity unknown

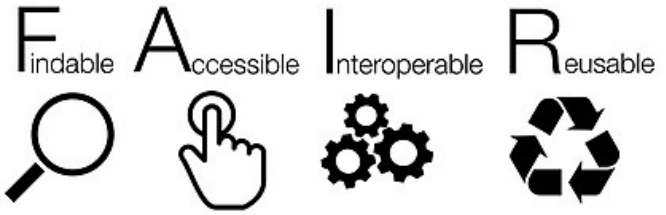
Compiled THS-specific Suspect List (n=95)

	Compound	Chemical name	CAS No.	EPA ID	Molecular Formula	Monoisotopic Mass
Tobacco Specific Nitrosamines TSNAs	Nicotine	3-(1-methylpyrrolidin-2-yl)pyridine	54-11-5	DTXSID1020930	C10H14N2	162.115698
	Cotinine	1-methyl-5-(pyridin-3-yl)pyrrolidin-2-one	486-56-6	DTXSID1047576	C10H12N2O	176.094963
	NNN	N'-Nitrosonornicotine	16543-55-8	DTXSID4021476	C9H11N3O	177.090212
	3-hydroxycotinine	3-Hydroxy-1-methyl-5-(3-pyridinyl)-2-pyrrolidinone	34834-67-8	DTXSID30873224	C10H12N2O2	192.089878
	NNA	4-(methylnitrosamino)-4-(3-pyridyl)butanal	64091-90-3	DTXSID00897139	C10H13N3O2	207.233
	NNK	4-(Methylnitrosoamino)-1-(3-pyridinyl)-1-butanone	64091-91-4	DTXSID3020881	C10H13N3O2	207.100777
	NNAL	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol	76014-81-8	DTXSID8020880	C10H15N3O2	209.116427
	NAB	N′-nitrosoanabasine	37620-20-5	DTXSID3021019	C10H13N3O	191.23
	NAT	N′-nitrosoanatabine	887407-16-1	DTXSID40868005	C10H11N3O	189.21
Nicotine Alkaloids	Myosmine	3-(3,4-Dihydro-2H-pyrrol-5-yl)-pyridin	532-12-7	DTXSID70891866	C9H10N2	146.19
	β-nicotyrine	Pyridine, 3-(1-methyl-1H-pyrrol-2-yl)-	487-19-4	DTXSID3075048	C10H10N2	158.084398
	2,3'-Bipyridine	2,3'-Bipyridine	581-50-0	DTXSID00206823	C10H8N2	156.068748
	N-Formylnornicotine	2-(Pyridin-3-yl)pyrrolidine-1-carbaldehyde	3000-81-5	DTXSID30336006	C10H12N2O	176.094963
	Nicotelline	Nicotelline	494-04-2	DTXSID40197781	C15H11N3	233.095297
Secondary Products of Nicotine heterogeneous nitrosation	Methyl nicotinate	3-Pyridinecarboxylic acid, methyl ester	93-60-7	DTXSID7044471	C7H7NO2	137.047678
	N-methylnicotinamide	N-Methylpyridine-3-carboxamide	114-33-0	DTXSID00870467	C7H8N2O	136.15

- Add these to “inclusion list” for DDA-MS/MS
- Now added to CompTox Dashboard & NORMAN-SLE & Zenodo

Torres, Schymanski, Ramirez *(in prep.).* Dataset DOI: 10.5281/zenodo.2669467

THSMOKE on Zenodo



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

zenodo

Search

Upload

Communities

emma.schymanski@uni.lu

NORMAN Suspect List Exchange

Recent uploads

Search NORMAN Suspect List Exchange

91 views

55 downloads

See more details...

Tweeted by 5

3

See more details

New upload

Suspect List

lic repository (under
) for suspect lists currently
the NORMAN Suspect List
tps://www.norman-
n/?q=suspect-list-exchange

Read more

icy:
This community will collect data that is
relevant and public (or to be public) for

May 6, 2019 (NORMAN-SLE-S52.0.1.0) Dataset Open Access

S52 | THSMOKE | Thirdhand Smoke (THS) Compounds

Torres, Sonia; Schymanski, Emma; Ramirez, Noelia;

This is the collection associated with list S52 THSMOKE on the NORMAN Susp
https://www.norman-network.com/?q=suspect-list-exchange S52 THSMOKE T
(THS) Compounds THSMOKE XLSX , CSV (06/05/2019) CompTox THSMOKE I
InChIKeys (06/05/

Uploaded on May 6, 2019

Publication date:
May 6, 2019

DOI:
DOI 10.5281/zenodo.2669467

Keyword(s):
Suspect Screening Thirdhand smoke

Communities:
NORMAN Suspect List Exchange

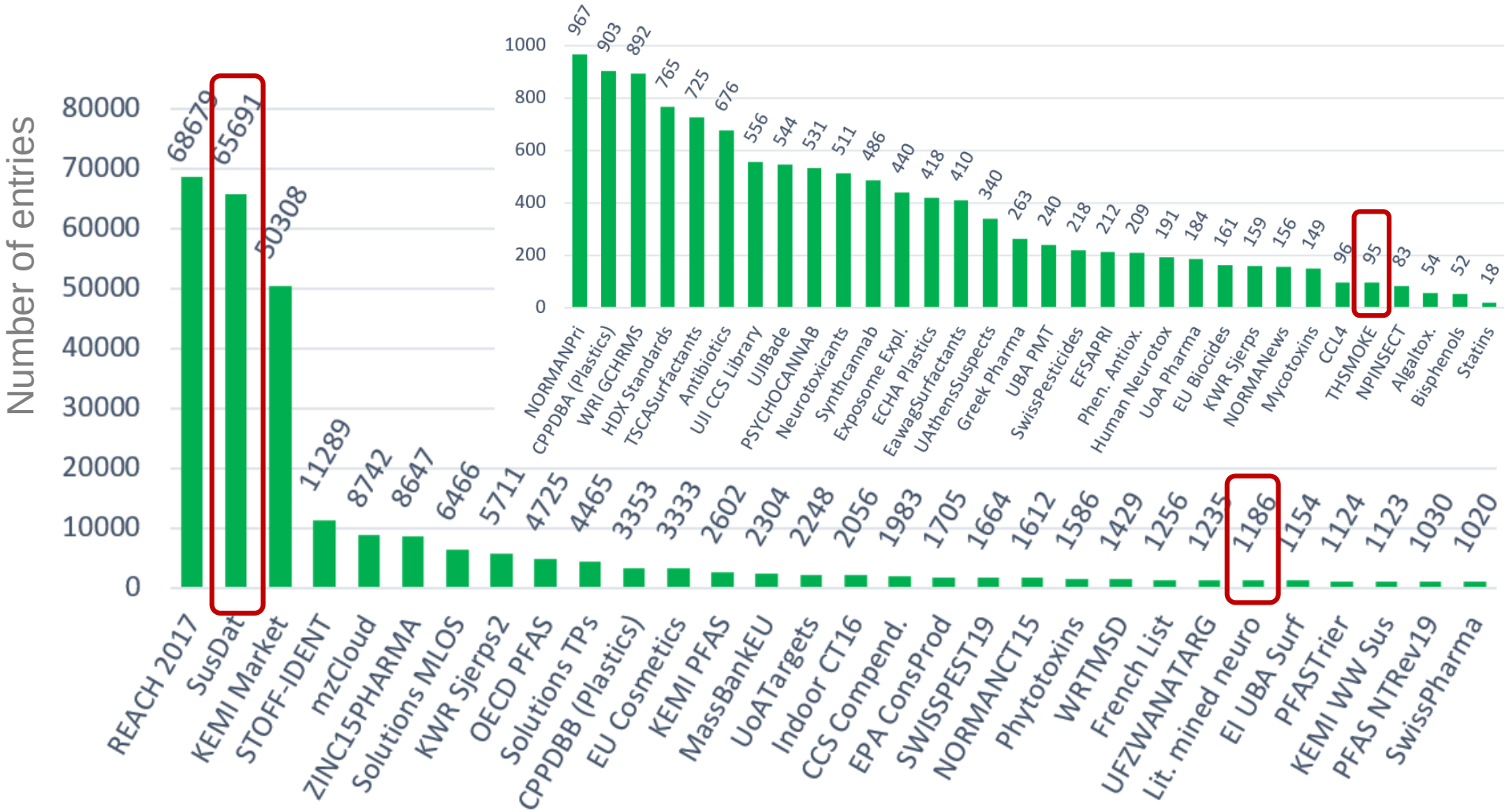
License (for files):
Creative Commons Attribution 4.0 International

March 1, 2018 (NORMAN-SLE-S0.0.1.0) Dataset Open Access

S0 | SUSDAT | Merged NORMAN Suspect List: SusDat

NORMAN Suspect List Exchange (SLE)

- https://www.norman-network.com/nds/SLE/ …now 62 lists!
- https://zenodo.org/communities/norman-sle ... with DOI



Schymanski *et al*. (in prep.) and https://www.norman-network.com/nds/SLE/ and https://zenodo.org/communities/norman-sle

THS & Target, Suspect, Non-target/Unknown

TARGET:

Known chemical, reference standard available *in house*

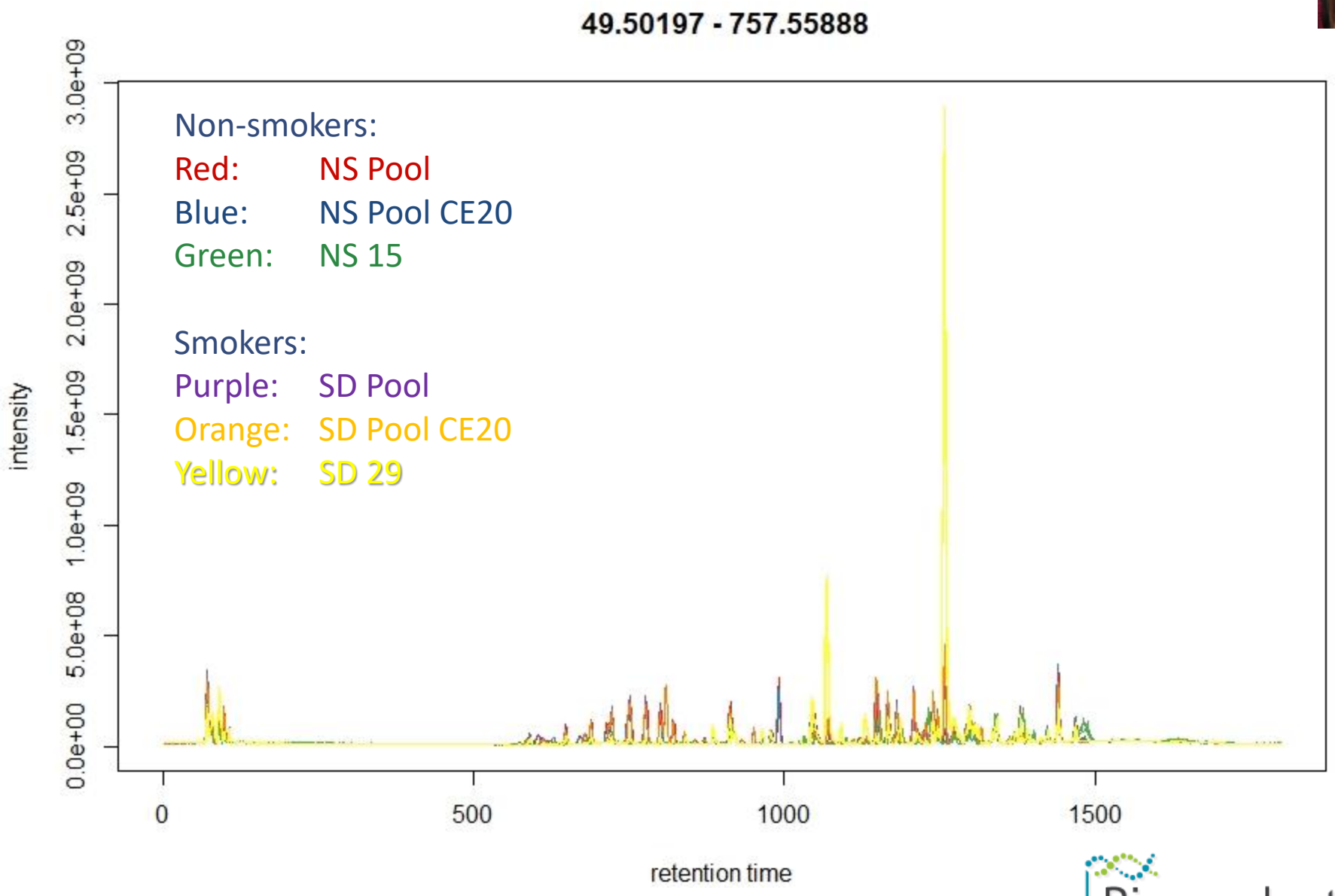
SUSPECT:

Chemical suspected to be present in the sample, std. not (necessarily) available

NON-TARGET/ UNKNOWN:

Mass/feature of interest detected in the sample, identity unknown

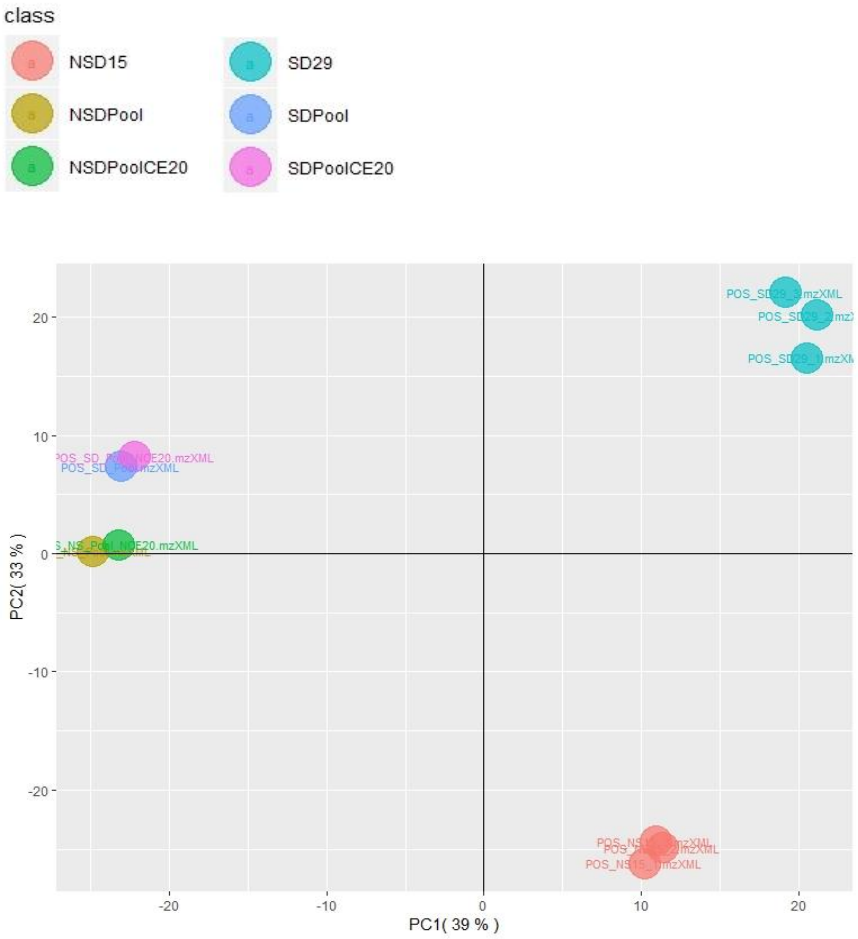
IISPV-URV XCMS-based Metabolomics Workflow

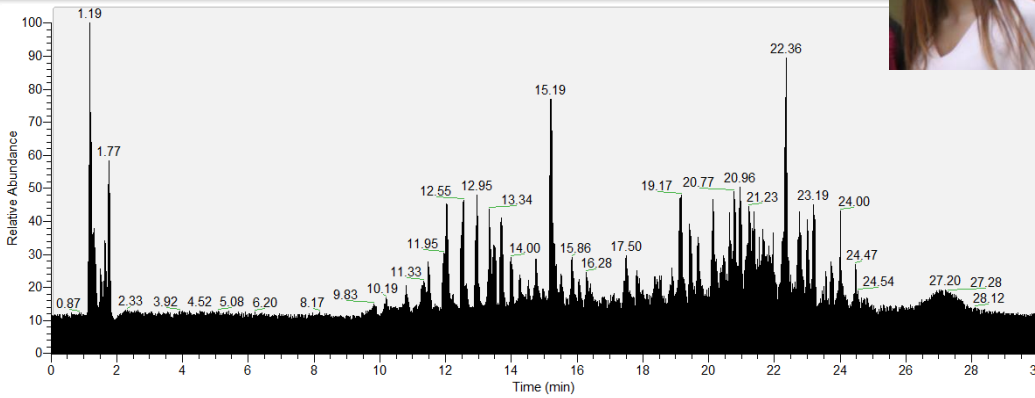


http://bioconductor.org/packages/xcms/ and DOI: 10.18129/B9.bioc.xcms

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

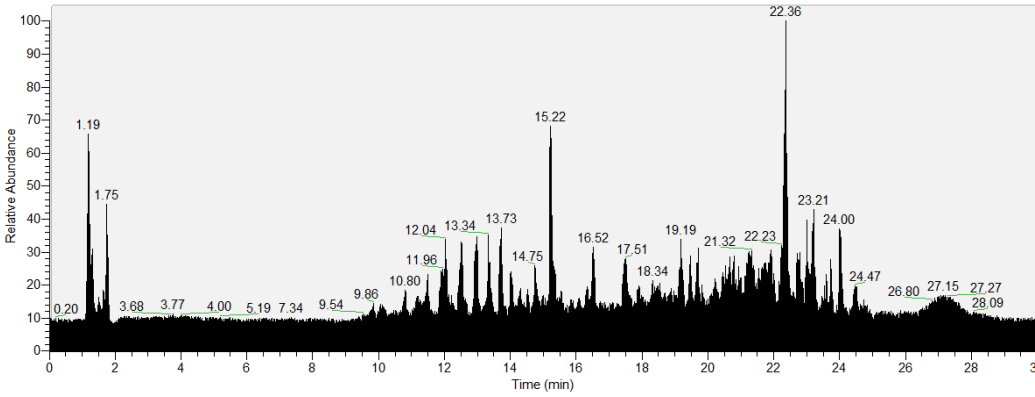
IISPV-URV XCMS-based Metabolomics Workflow





Smoker Pool

Non Smoker Pool



Images c/o Sonia Torres

Sampling  Extraction (AcN+ QuEChERS+ Z−Sep)  UHPLC separation  HR-MS/MS 

Conversion (Proteowizard)

TSNA+ Target List

THS Suspect List

SUSPECT SCREENING

Masses of interest

NON-TARGET SCREENING

Peak Picking (xcms)

Prioritization (xcms)

Statistics (in house)

Masses of interest

RMassBank

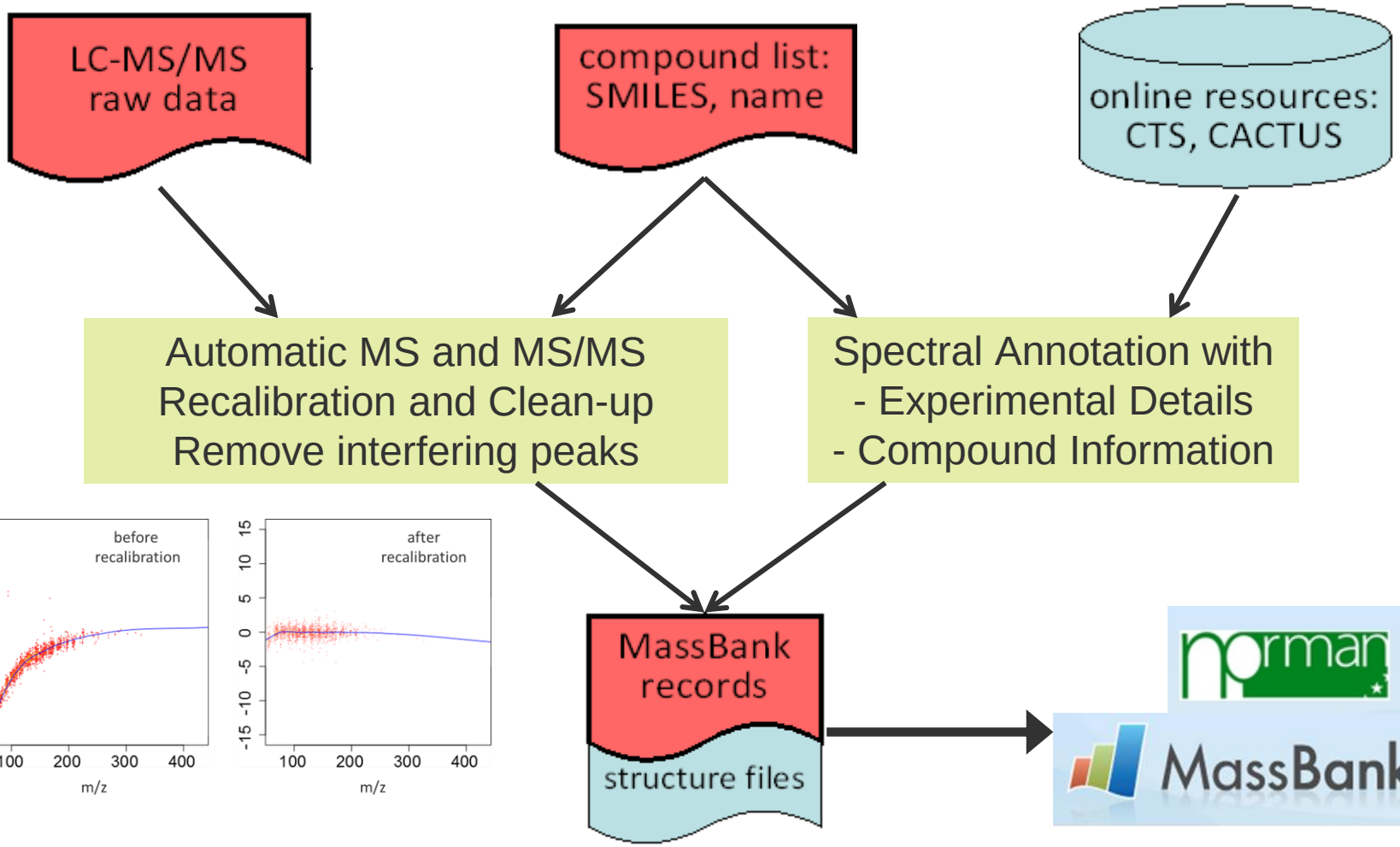
MS/MS Extraction (RMassBank)

RMassBank

Torres, Schymanski, Ramirez, *in prep.*

Creating High-Quality Mass Spectra

RMassBank



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

Extracting Mass Spectra for NTS

RMassBank

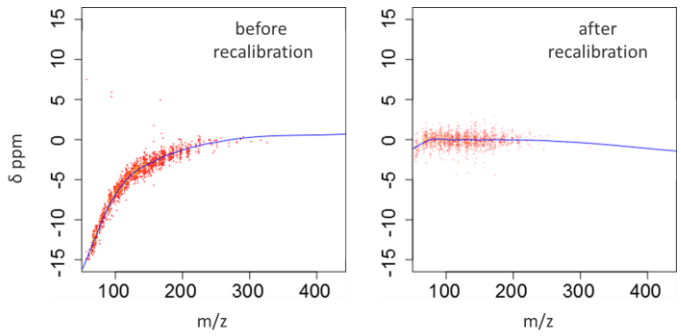
LevelCode	LevelKeyword	COMMENT.CONFIDENCE
1	standard	Reference Standard (Level 1)
1a	standard	Reference Standard (Level 1)
1b	parent	Parent Substance with Reference Standard (Level 1)
1c	confirmed	Identification confirmed with Reference Standard (Level 1)
2	probable	Probable structure, tentative identification (Level 2)
2a	probableLibrary	Probable structure via library match, tentative identification (Level 2a)
2b	probableDiagnostic	Probable structure via diagnostic evidence, tentative identification (Level 2b)
3	tentative	Tentative identification only (Level 3)
3a	tentativeStructure	Tentative identification: most likely structure (Level 3)
3b	tentativeIsomer	Tentative identification: isomers possible (Level 3)
3c	tentativeTPClass	Tentative identification: substance class known (Level 3)
3d	tentativeBestMatch	Tentative identification: best match only (Level 3)
4	formula	Tentative identification: molecular formula only (Level 4)
5	unknown	Tentative identification: structure and formula unknown (Level 5)
5	exactMass	Tentative identification: structure and formula unknown (Level 5)

LC-MS/MS raw data

online resources: CTS, CACTUS

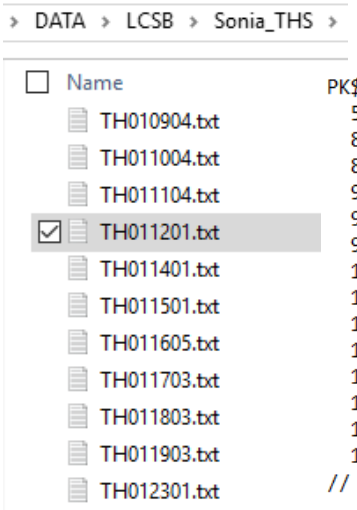
Automatic MS and MS/MS Recalibration and Clean-up Remove interfering peaks

Spectral Annotation with
- Experimental Details
- Compound Information



MassBank records

structure files




```
PK$PEAK: m/z int. rel.int
  58.0654 4282.6 30
  80.0495 62017.7 445
  84.0808 80870.8 581
  91.0543 3360.7 24
  92.0495 6060 43
  94.0651 2986.8 21
  102.0915 5082.2 36
  105.0695 2255.8 16
  106.0651 91919.1 660
  115.0543 7079.6 50
  117.0574 108403.4 778
  120.0809 26838 192
  130.0652 139047.5 999
  132.0808 87200.2 626
//
```

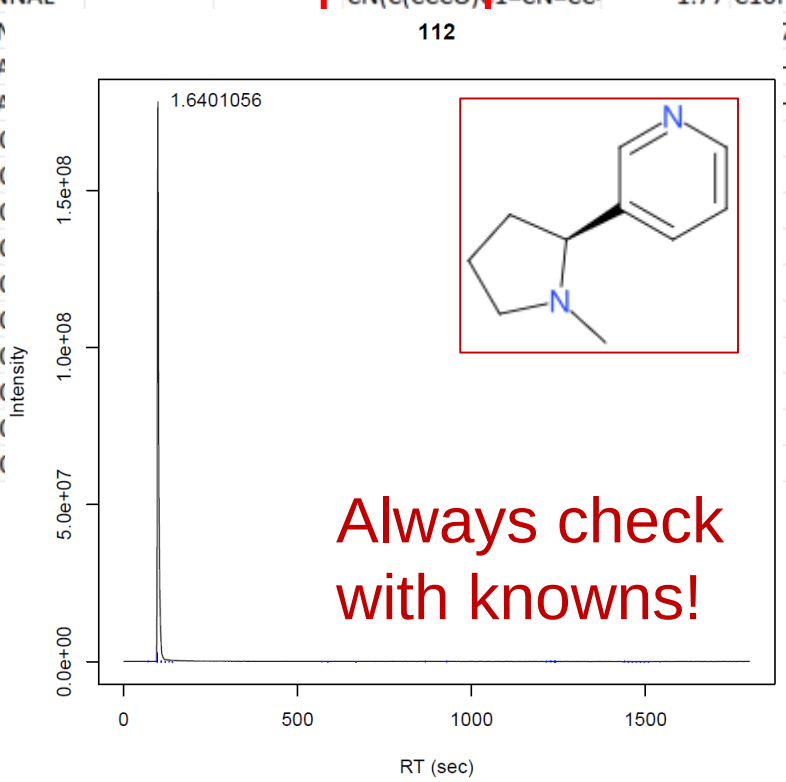
https://github.com/MassBank/RMassBank/
http://bioconductor.org/packages/RMassBank/

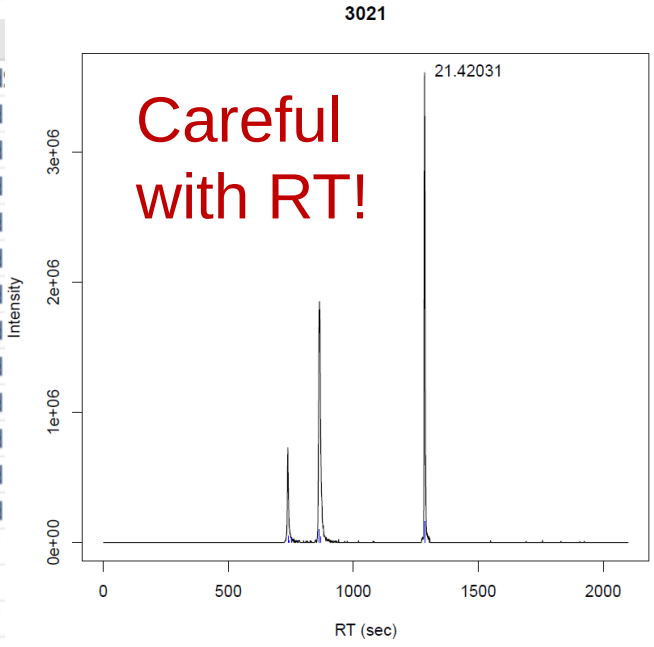
Stravs, Schymanski, Singer and Hollender, 2013, Journal of Mass Spectrometry, 48, 89–99. DOI: 10.1002/jms.3131

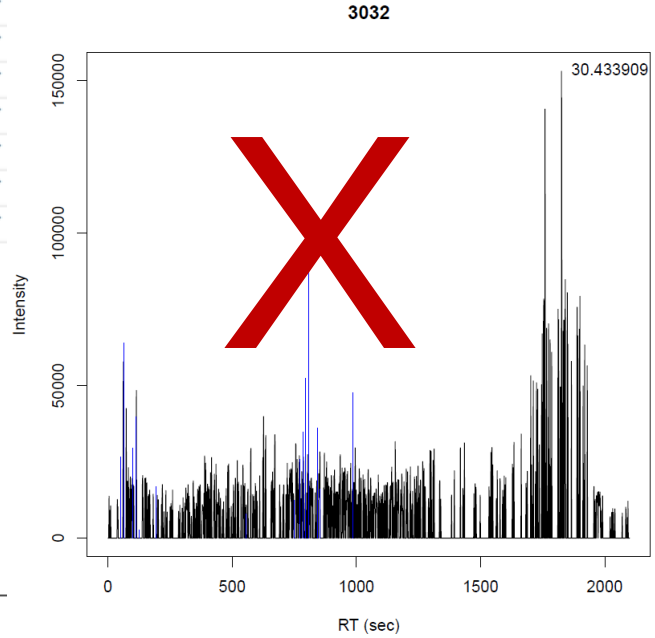
MS/MS Extraction with RMassBank

RMassBank

	A	B	C	D	E	F	G	H	I	
	ID	Name	Name2	DTXSID	SMILES	CAS	RT	Formula	mz	Std
112		Nicotine (3	Nicotine	DTXSID10 2	CN1CCC[C@	54-11-5	1.62	C10H14N2		Std
114		Nicotine-d3	DL-Nicotine	DTXSID80 4	[2H]C([2H])	69980-24-1	1.62	C10H11D3N2		Std
117		Cotinine (1	Cotinine	DTXSID10 4	CN1[C@@H	486-56-6	1.77	C10H12N2O		Std
119		NNN (N'-N	N'-Nitrosor	DTXSID40 2	O=NN1CCC	16543-55-8	1.79	C9H11N3O		Std
122		Cotinine-d3	Cotinine-d3	DTXSID40 4	[2H]C([2H])	110952-70	1.77	C10H9D3N2O		Std
134		NNK (4-(M	4-(N-Methy	DTXSID30 2	CN(CCCC(=O	64091-91-4	11.33	C10H13N3O2		Std
135		NNA (4-(m	4-(methyln	DTXSID00 8	CN(N=O)C(C	64091-90-3	1.77	C10H13N3O2		Std
136		NNAL (4-(n	4-(Methyln	DTXSID80 2	CN(CCCC(O	76014-81-8	1.77	C10H15N3O2		Std
157		isoNNAL			CN(C(CCCO)	C1=CN=CC	1.77	C10H15N3O2		Std
158		NNN						7D4N3O		Std
159		NNA						H10D3N3O2		Std
160		NNA						H12D3N3O2		Std
1001		FT0(							60.04523	NT
1002		FT0(							61.04228	NT
1003		FT0(							90.97725	NT
1004		FT0(							104.1072	NT
1005		FT0(							109.1013	NT
1006		FT0(							111.0216	NT
1007		FT0(							114.0666	NT
1008		FT0(							121.0398	NT
1009		FT0(							125.0392	NT
1010		FT0(							125.0959	NT

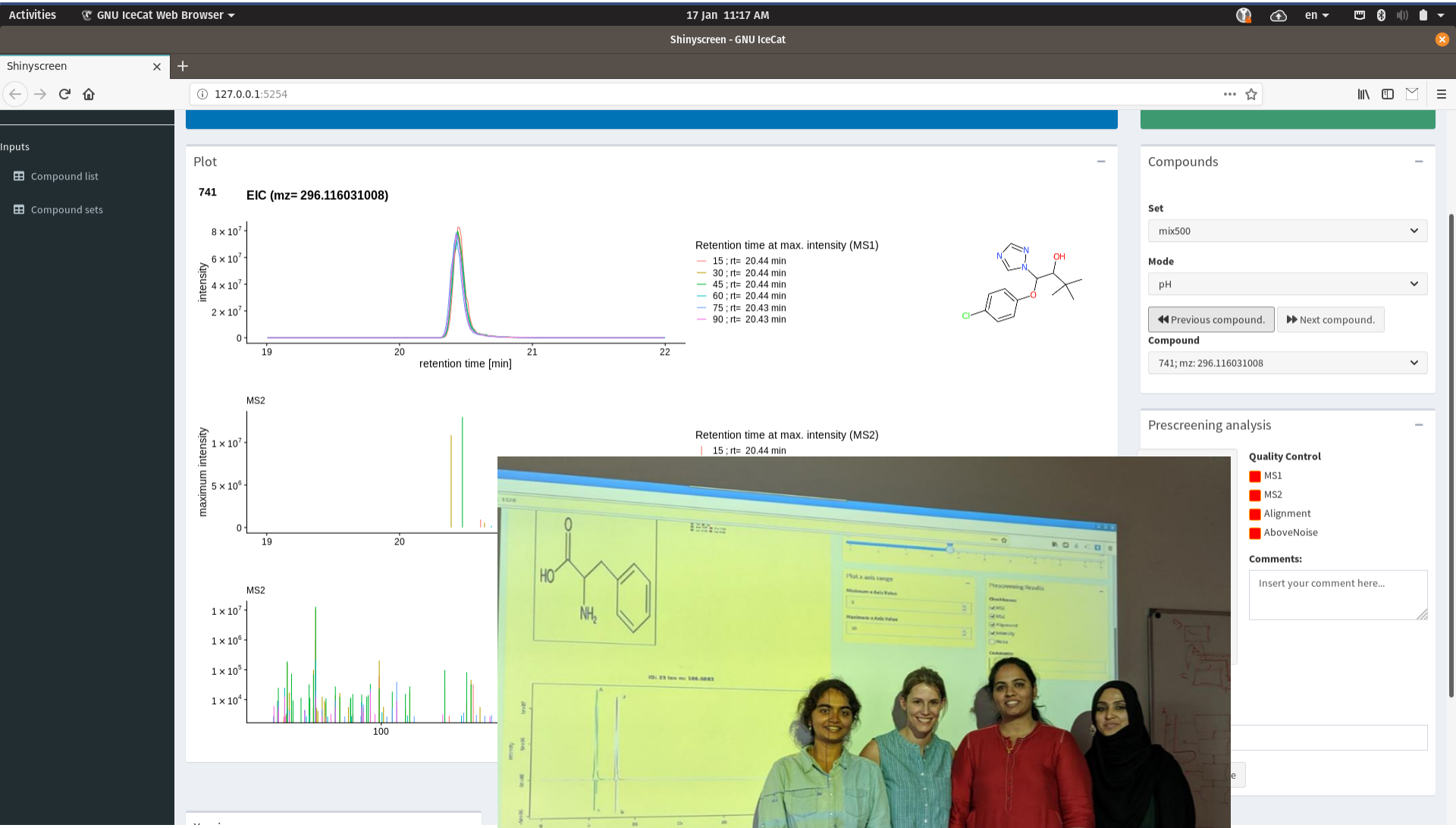






In development: Shinyscreen: Kondic *et al* – huge thanks to the ECI Interns!

Coming soon … pre-screening with Shinyscreen!



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

Figure source: Todor Kondic

Sampling  Extraction (AcN+ QuEChERS+ Z−Sep)  UHPLC separation  HR-MS/MS 

Conversion (Proteowizard)

TSNA+ Target List

THS Suspect List

SUSPECT SCREENING

Masses of interest

NON-TARGET SCREENING

Peak Picking (xcms) → Prioritization (xcms)

Statistics (in house) → Masses of interest

RMassBank

MS/MS Extraction (RMassBank)

RMassBank

MetFrag

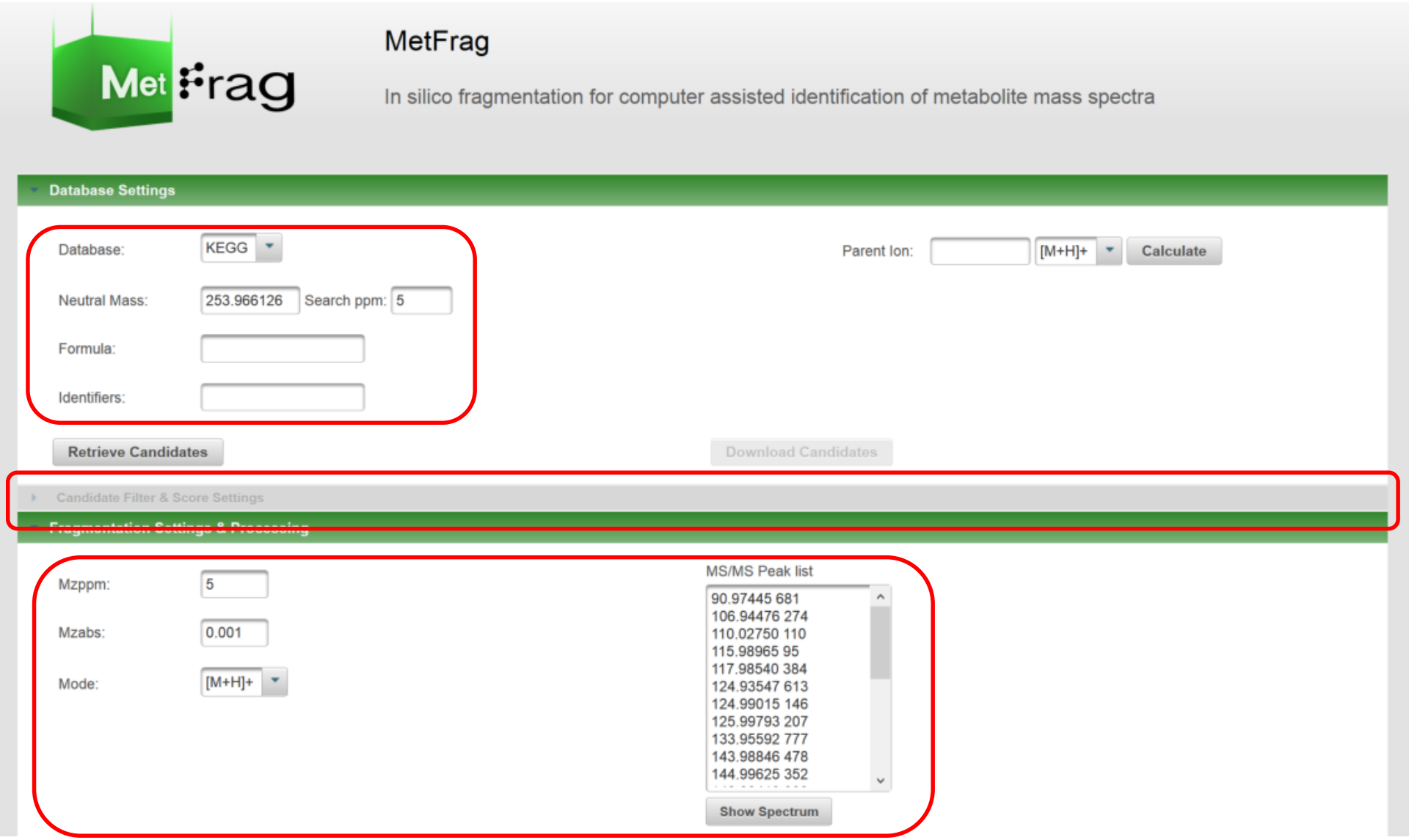
Non-target identification (MetFrag, ReSOLUTION)

MetFrag

Torres, Schymanski, Ramirez, *in prep.*

Tentative Identification with MetFrag

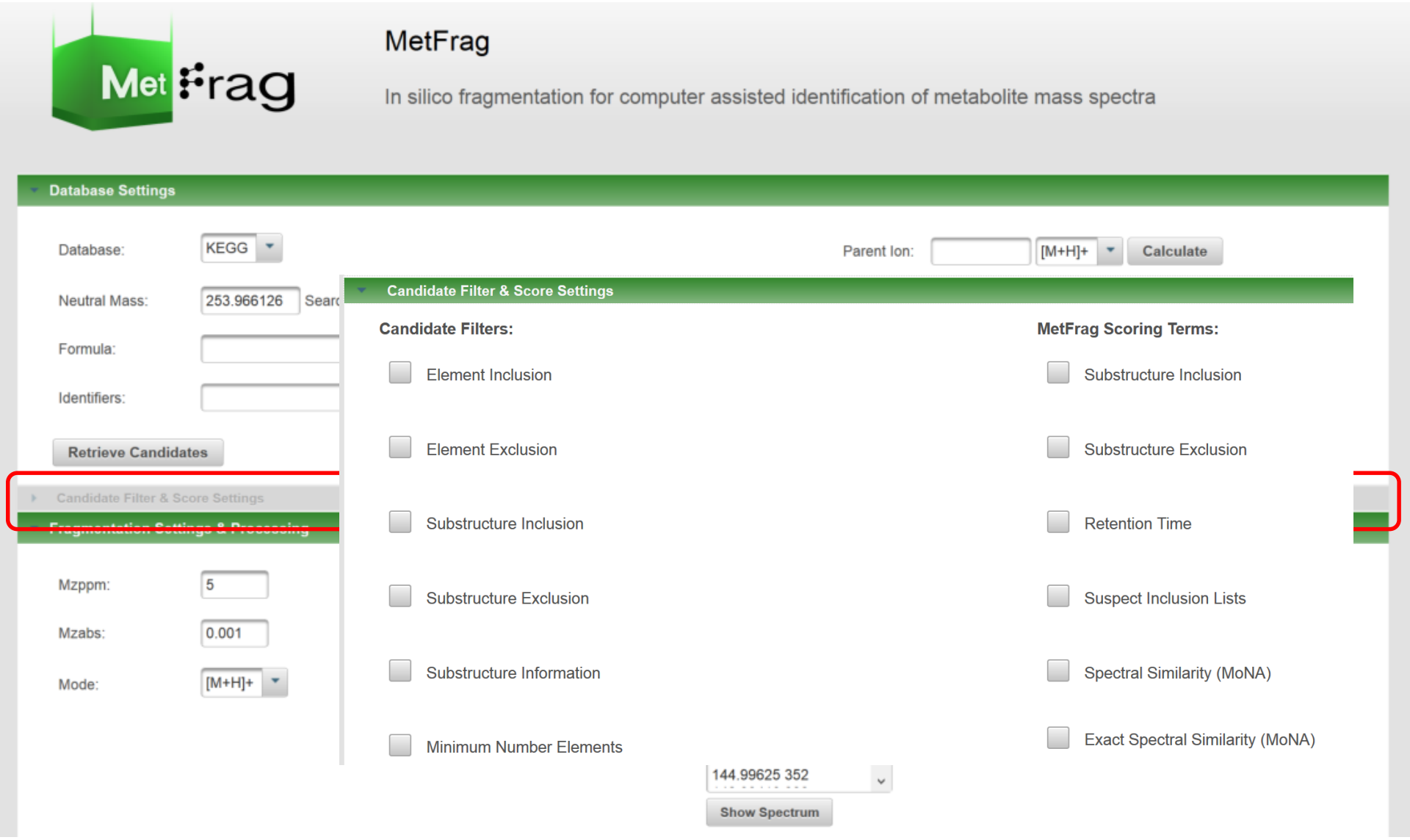
Web Interface: http://msbi.ipb-halle.de/MetFrag/



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

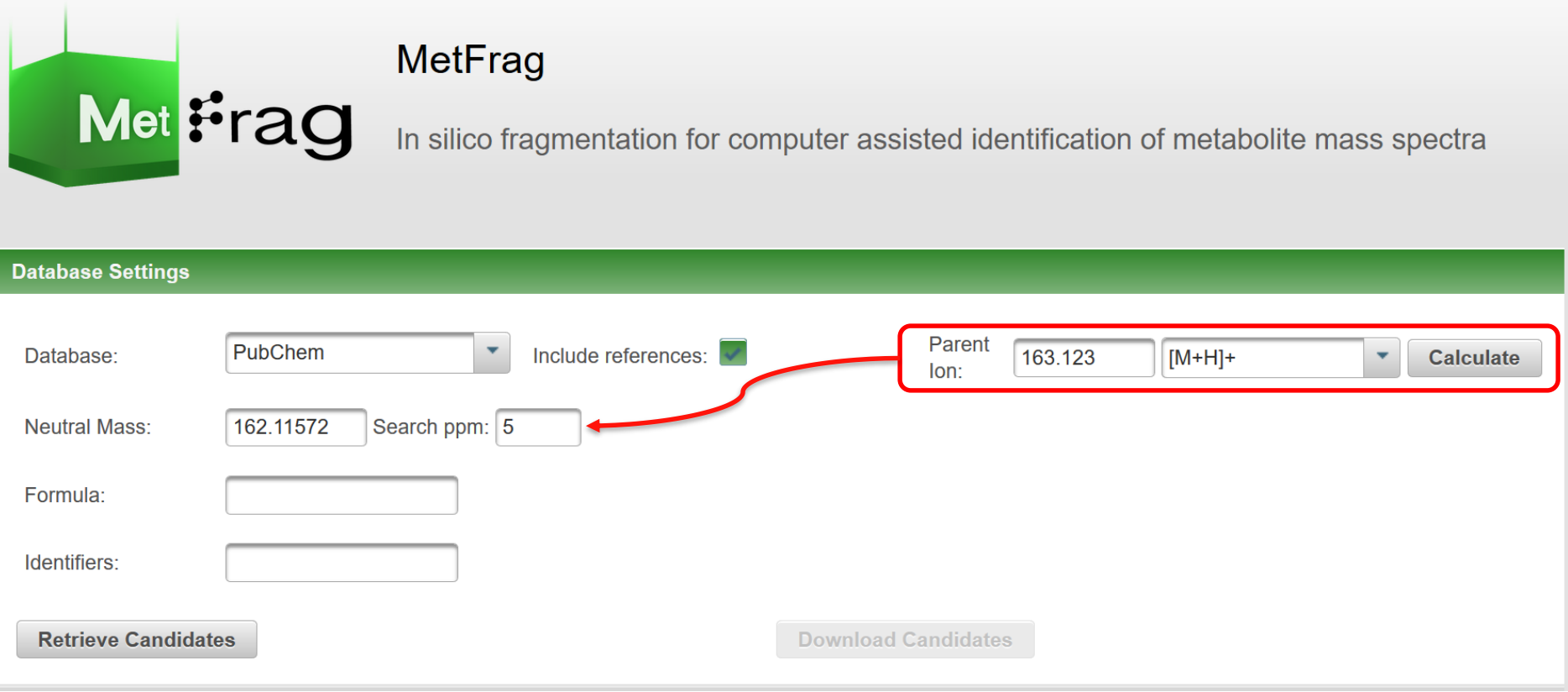
Tentative Identification with MetFrag

Web Interface: http://msbi.ipb-halle.de/MetFrag/



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

MetFrag – Example: CompTox + Nicotine I



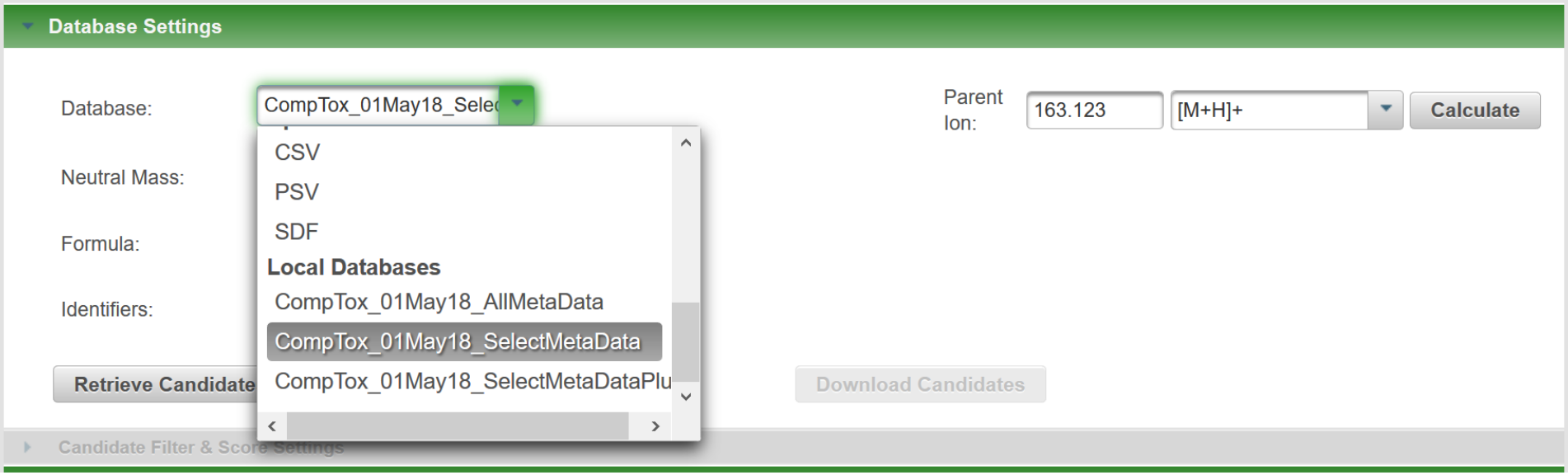
https://comptox.epa.gov/dashboard/

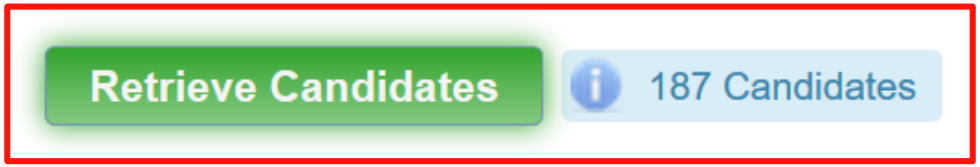
Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, 2016, DOI: 10.1186/s13321-016-0115-9

MetFrag – Example: CompTox + Nicotine II

MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra





MetFrag – Example: CompTox + Nicotine III

Candidate Filter & Score Settings

Candidate Filters

- [ ] Element Inclusion
- [ ] Element Exclusion
- [ ] Substructure Inclusion
- [ ] Substructure Exclusion
- [ ] Substructure Information
- [ ] Minimum Number Elements

MetFrag Scoring Terms

- [ ] Substructure Inclusion
- [ ] Substructure Exclusion
- [ ] Retention Time
- [ ] Suspect Inclusion Lists
- [ ] Spectral Similarity (MoNA)
- [x] Exact Spectral Similarity (MoNA)
- [ ] Statistical Scoring

MoNA
MassBank of North America
MassBank.eu

Database Scoring Terms

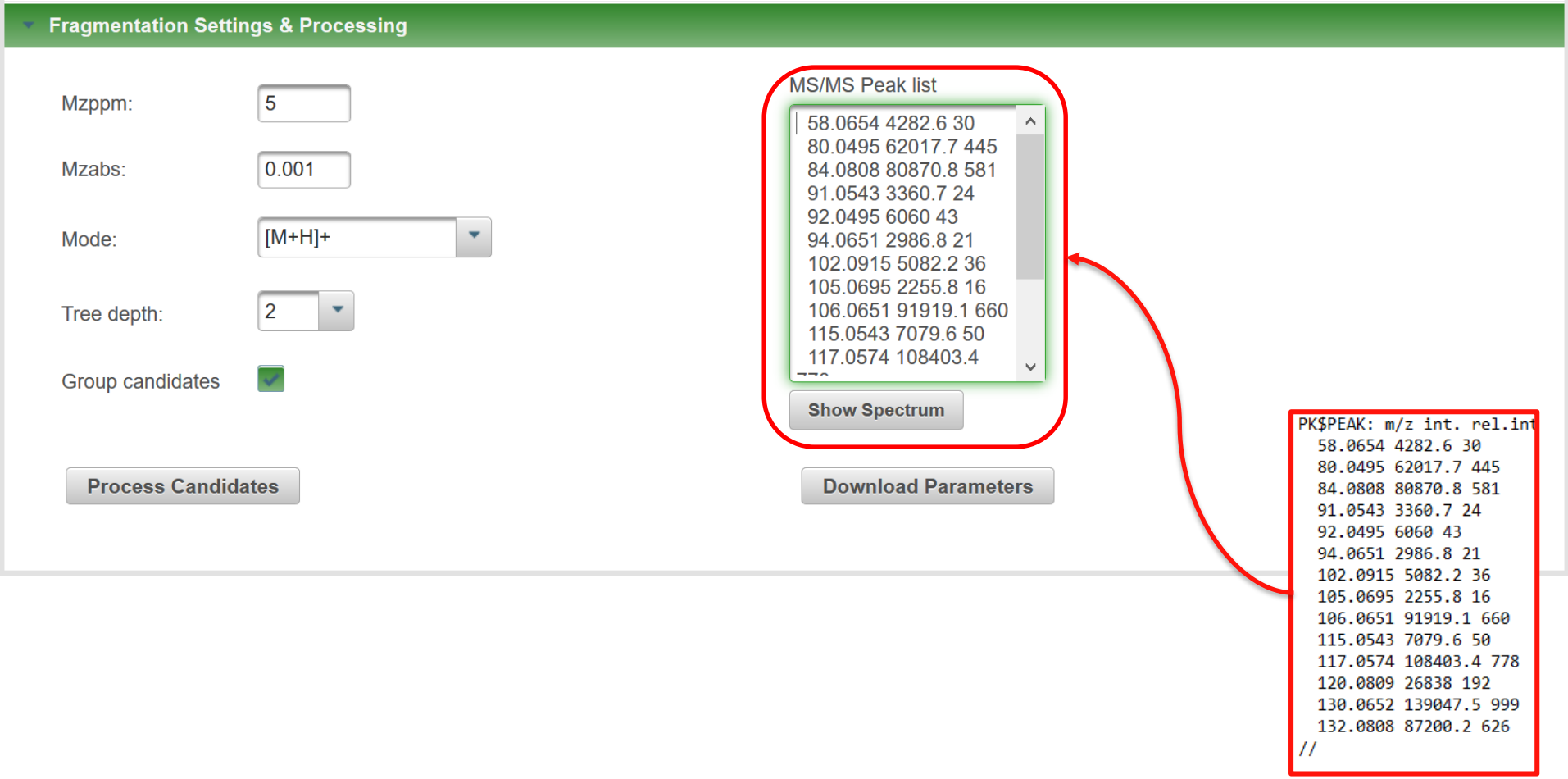
Select Item(s) — 4 of 11 item(s) selected

- [ ] MASSBANKEU
- [x] NORMANSUSDAT
- [x] NUMBER_OF_PUBMED_ARTICLES
- [ ] PUBCHEM_DATA_SOURCES
- [ ] TOX21SL
- [ ] TOXCAST
- [x] TOXCAST_PERCENT_ACTIVE

Select Item(s) — 4 of 11 item(s) selected

MetFrag – Example: CompTox with Nicotine IV

- Fragmentation Settings and Processing



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

MetFrag – Example: CompTox + Nicotine V

Results

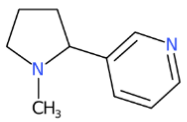
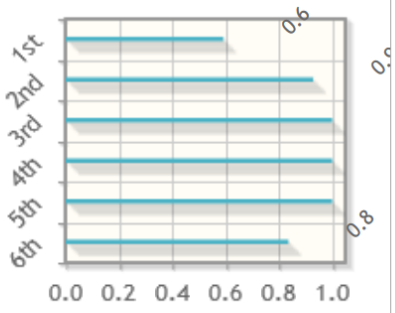
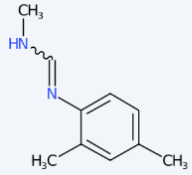
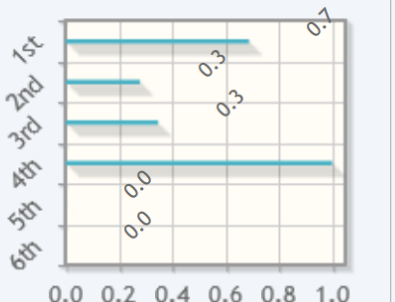
Weights		
MetFrag (1st)	100	%
ExactSpectralSimilarity (2nd)	100	%
DATA_SOURCES (3rd)	100	%
NORMANSUSDAT (4th)	100	%
NUMBER_OF_PUBMED_ARTICLES (5th)	100	%
TOXCAST_PERCENT_ACTIVE (6th)	100	%

Download Results

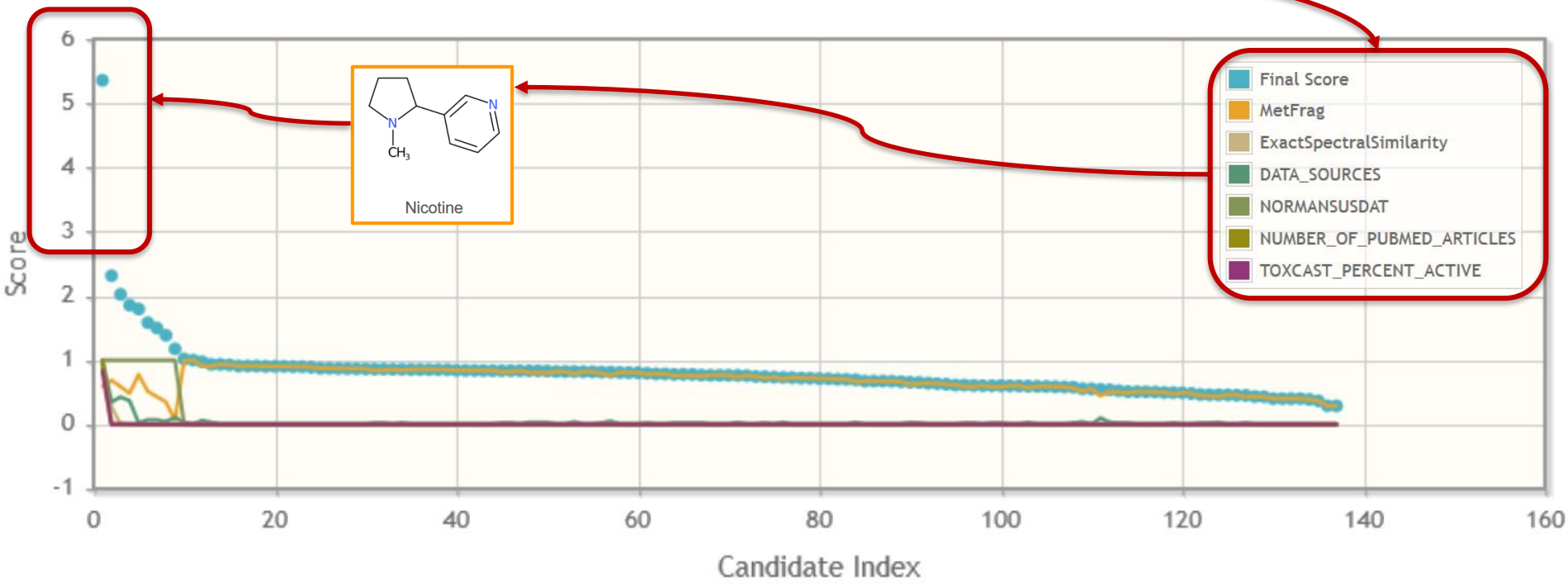
Filter Candidates by explained MS/MS Peaks

MS/MS Peaks | Filter Candidates

1 2 3 4 5 6 7 8 9 10

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 Nicotine	DTXSID1020930 DTXSID8021725 DTXSID3048154 DTXSID0046351 DTXSID6020931 DTXSID5075319 DTXSID3088421 InChIKeyBlock1 = SNICXCGAKADSCV	162.11576	$C_{10}H_{14}N_2$		5.3553	Peaks: 10 / 14 Fragments Scores Download
2	 N'-(2,4-Dimethylphenyl)-N-methylformamidi ne	DTXSID1037696 DTXSID10199510 InChIKeyBlock1 = JIIOLEGNERQDIP	162.11576	$C_{10}H_{14}N_2$		2.3137	Peaks: 9 / 14 Fragments Scores Download

Connecting Resources in MetFrag



https://msbi.ipb-halle.de/MetFrag/ ; https://comptox.epa.gov/dashboard/ ; https://massbank.eu ; https://www.norman-network.com/nds/susdat/

NTS with MetFrag and ReSOLUTION

- Create config files (MetFragConfig)
- Run MetFrag (runMetFrag)

```
176     # now, run MetFrag and extract results for reporting into cmpd_info.
177
178     results_filename <- paste0(run_name,"_",cmpdID_char,"_",as.character(i))
179     if (isPos) {
180       config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
181                                    results_filename = results_filename,
182                                    peaklist_path = MetFrag_msms, base_dir = results_run_dir,
183                                    DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
184                                    IsPosMode = TRUE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
185     } else {
186       config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
187                                    results_filename = results_filename,
188                                    peaklist_path = MetFrag_msms, base_dir = results_run_dir,
189                                    DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
190                                    IsPosMode = FALSE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
191     }
192
193     runMetFrag(config_file, MetFrag_dir, CL_name = "MetFrag2.4.4-msready-CL.jar")
194
195     results_file <- paste0(results_run_dir,"/results/",results_filename,".xls")
```


http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

NTS with MetFrag and ReSOLUTION

- Extract results and summarize

```
198     #extract results we need:
199     MetFrag_res <- read_excel(results_file)
200
201     cmpd_info$num_poss_IDs[i] <- length(MetFrag_res$Score)
202     cmpd_info$poss_IDs[i] <- paste(MetFrag_res$Name,collapse=";")
203     cmpd_info$poss_ID_scores[i] <- paste(MetFrag_res$Score,collapse=";")
204     cmpd_info$max_Score[i] <- max(MetFrag_res$Score)
205     cmpd_info$n_Score_GE3p5[i] <- length(which(MetFrag_res$Score>=3.5))
206     cmpd_info$n_Score_GE3[i] <- length(which(MetFrag_res$Score>=3))
207     cmpd_info$n_Score_GE2p5[i] <- length(which(MetFrag_res$Score>=2.5))
208     cmpd_info$poss_DTXSIDs[i] <- paste(MetFrag_res$DTXSID,collapse=";")
209     cmpd_info$poss_CAS[i] <- paste(MetFrag_res$CAS,collapse=";")
210     cmpd_info$MoNAScore[i] <- paste(MetFrag_res$OfflineIndividualMoNAScore,collapse=";")
211     cmpd_info$MaxMoNAScore[i] <- max(MetFrag_res$OfflineIndividualMoNAScore)
212
213   }
214
215
216   write.csv(cmpd_info,paste0(results_summary_dir,"/MetFragResultSummary_",run_name,".csv"),row.names = F)
```


http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

CSV Summary Output for Results Interrogation

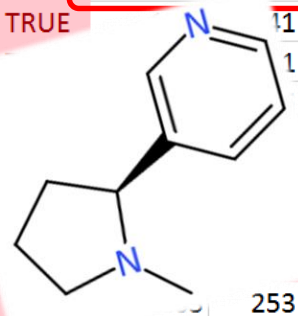
ID	mz	Name	RT	Int	MS/MS	RT_RMB	#Cand	MaxScore	SMILES_	Name_maxScore	ExplPeaks	max #Peaks	#Peaks	max MoNA
3027	163.1221	FT0532	1.073	9E+07	TRUE	1.052	137	7.270015	CN1CCC(	Nicotine	58.0658_9	9	28	0.99701
3008				+07	TRUE		11	7.22864	CN1CC(=	Creatinine	57.0454_1	3	14	0.54699
3131				+06			1	7.028138	CC(C=CC	all-trans-Retinoic acid	57.0706_5	8		7
3321				+				6.093	CCCCCC(	Didecyl phthalate	69.0704_			
3484				+(				6	CC(C)(C)	2,2'-Oxamidodiethyl bi				
3206	346.1096	FT3193	21.08	3E+0(				5.973483	CCCN(CC	Nitralin				
3044				+07				5.960103	CCN(CC)	N,N-Diethylnicotinamid				
3006				+06				5.94264	OCCOCC	Diethylene glycol				0
3043				+07			253	5.721703	CCN(CC)	N,N-Diethylnicotinamide		7	24	0.00209
3055				+07	T	11.202	114	5.712626	COC1=C(	Scopoletin	0393_1	9	28	0.94271
3046	183.0796	FT0741	24.77	1E+07	TRUE	24.744	72	5.689798	O=C(C1=	Benzophenone	50.0159_2	5	45	0
3039				+07	TRUE	7.635	219	5.467659	CN1C(CC	Cotinine	53.0393_1	9	23	0.9987
3038				+07	TRUE	8.017		36	CC1CN(N	4-Methyl-1-phenylpyrazoli	53.0393_2	9	40	0.66408
3183				+07	TRUE			2	CCCCCC(	Dihexyl phthalate	54.0452_2	9	51	0
3095				+08					CC(C)CO	Triisobutyl phosphate	57.0707_8		15	0.99613
3020	151.1111	FT0427	13.16	2E+06					CC(C)(C)	4-tert-Butylphenol	NA		0	0
3029				+05	FA				COC1=C(	1,2,4-Trimethoxyb			0	0
3123				+06	FA			5	CCCCCC(	Octadec				0
3128				+06	FAL		13	3	CCCCCC(	Octadeca				0
3172				+06	FAL		15	3	CCCCCC(	Stearylbe				0

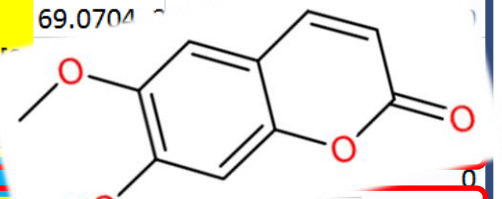
Level 1 Target

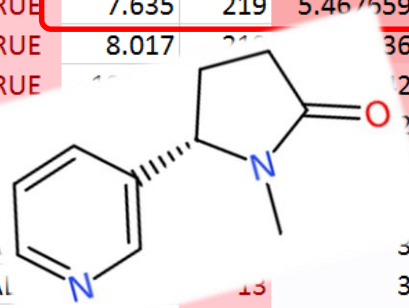
Level 2 MSMS Match

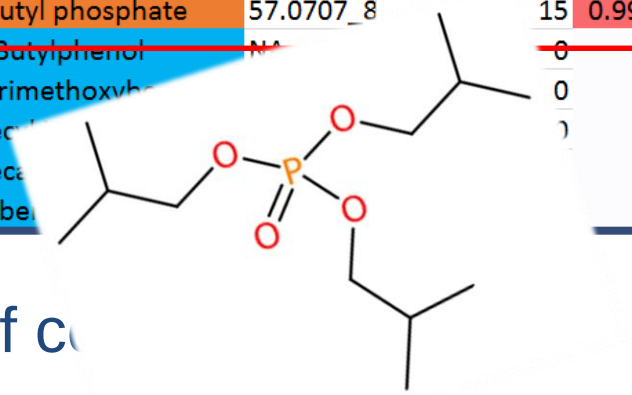
Level 3 Tentative ID

Level 5 No MS/MS









⇒ Approaching automatic assignment of c

⇒ Quick, high throughput prioritization for d cquisition

TIBP well documented in dust, e.g. Rostkowski *et al* 2019 DOI: 10.1007/s00216-019-01615-6 and DOI: 10.5281/zenodo.2653206

Sampling  Extraction (AcN+ QuEChERS+ Z−Sep)  UHPLC separation  HR-MS/MS 

Conversion (Proteowizard)

TSNA+ Target List

THS Suspect List

SUSPECT SCREENING

Masses of interest

NON-TARGET SCREENING

Peak Picking (xcms)

Prioritization (xcms)

Statistics (in house)

Masses of interest

RMassBank

MS/MS Extraction (RMassBank)

RMassBank

MetFrag

Non-target identification (MetFrag, ReSOLUTION)

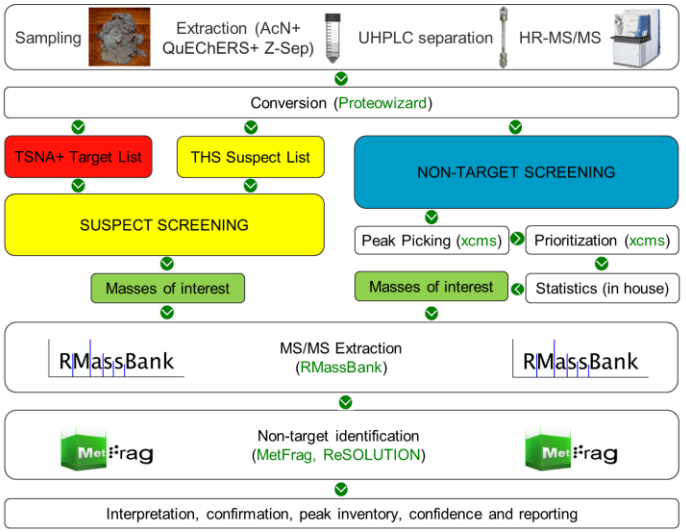
MetFrag

Interpretation, confirmation, peak inventory, confidence and reporting

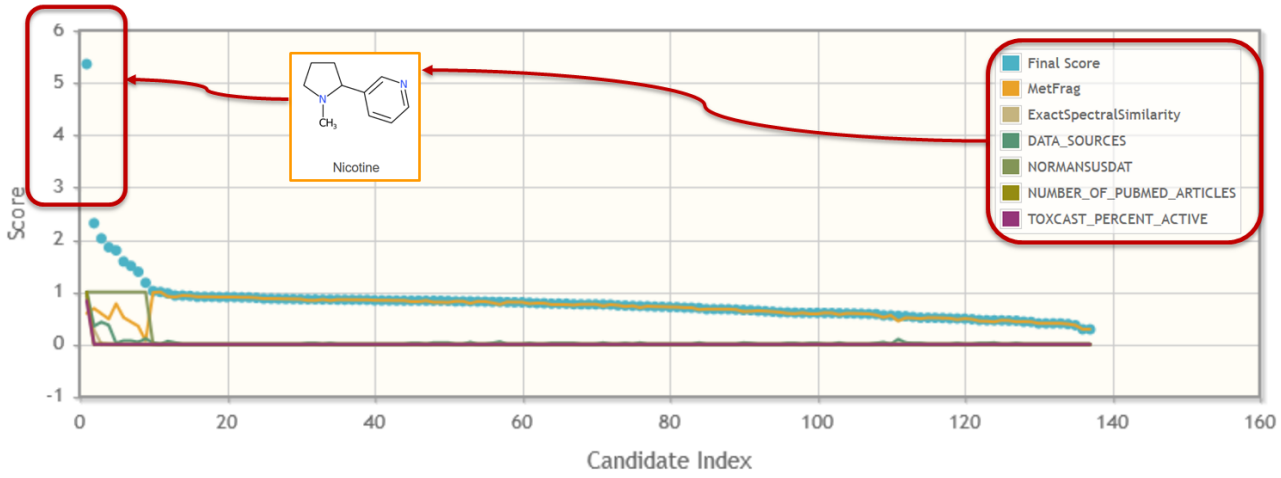
Torres, Schymanski, Ramirez, *in prep.*

Perspectives: Identifying Small Molecules in NTS

- Many comprehensive workflows
  - *I have presented just one of them!*



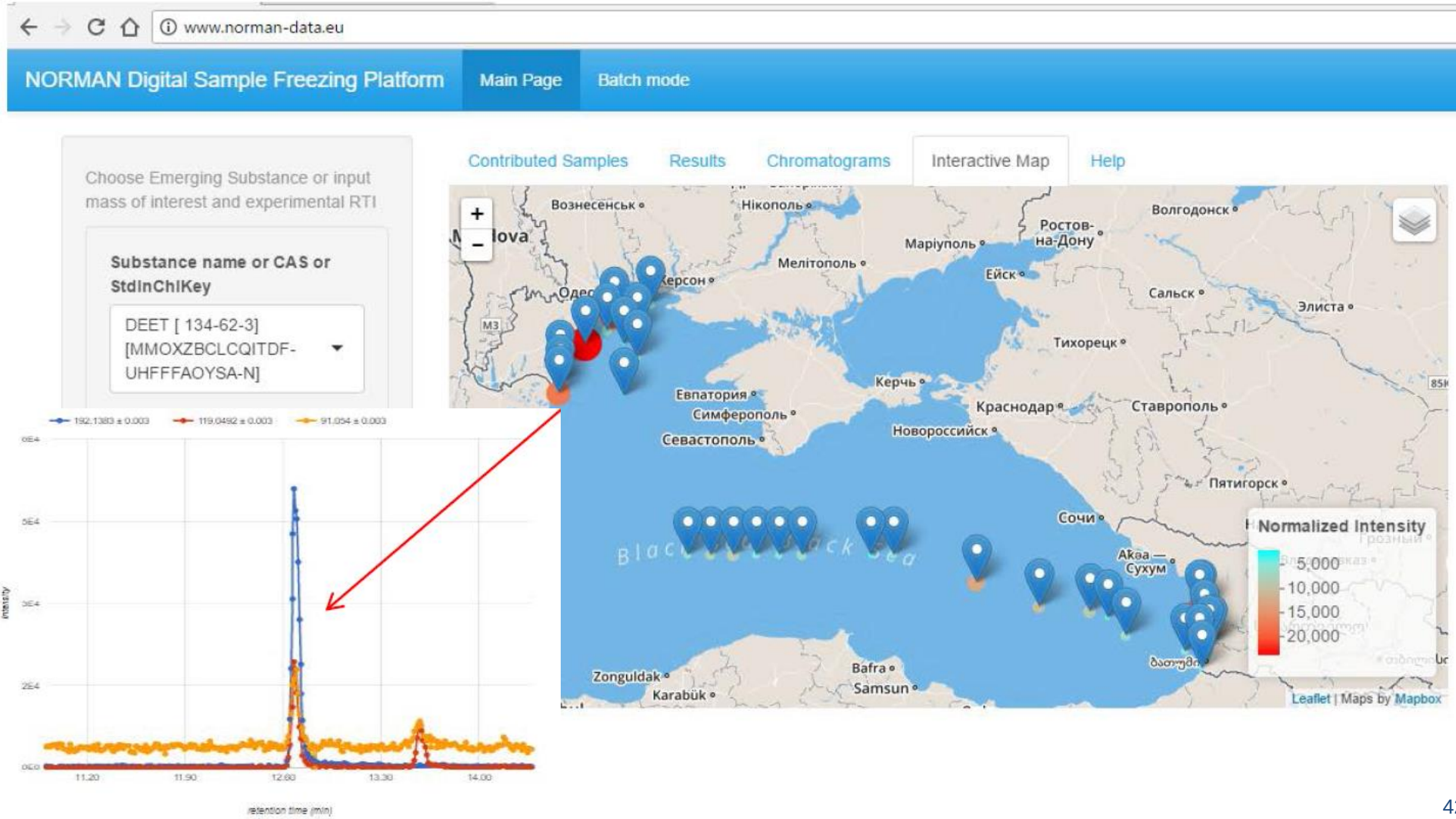
- Annotation of “known unknowns” is now relatively “quick”:
  - *Especially with well-chosen suspect screening and metadata*



- The bottleneck is still in expt. design and interpretation
  - *But in the meantime we can do some pretty neat things!*

NORMAN Digital Sample Freezing Platform

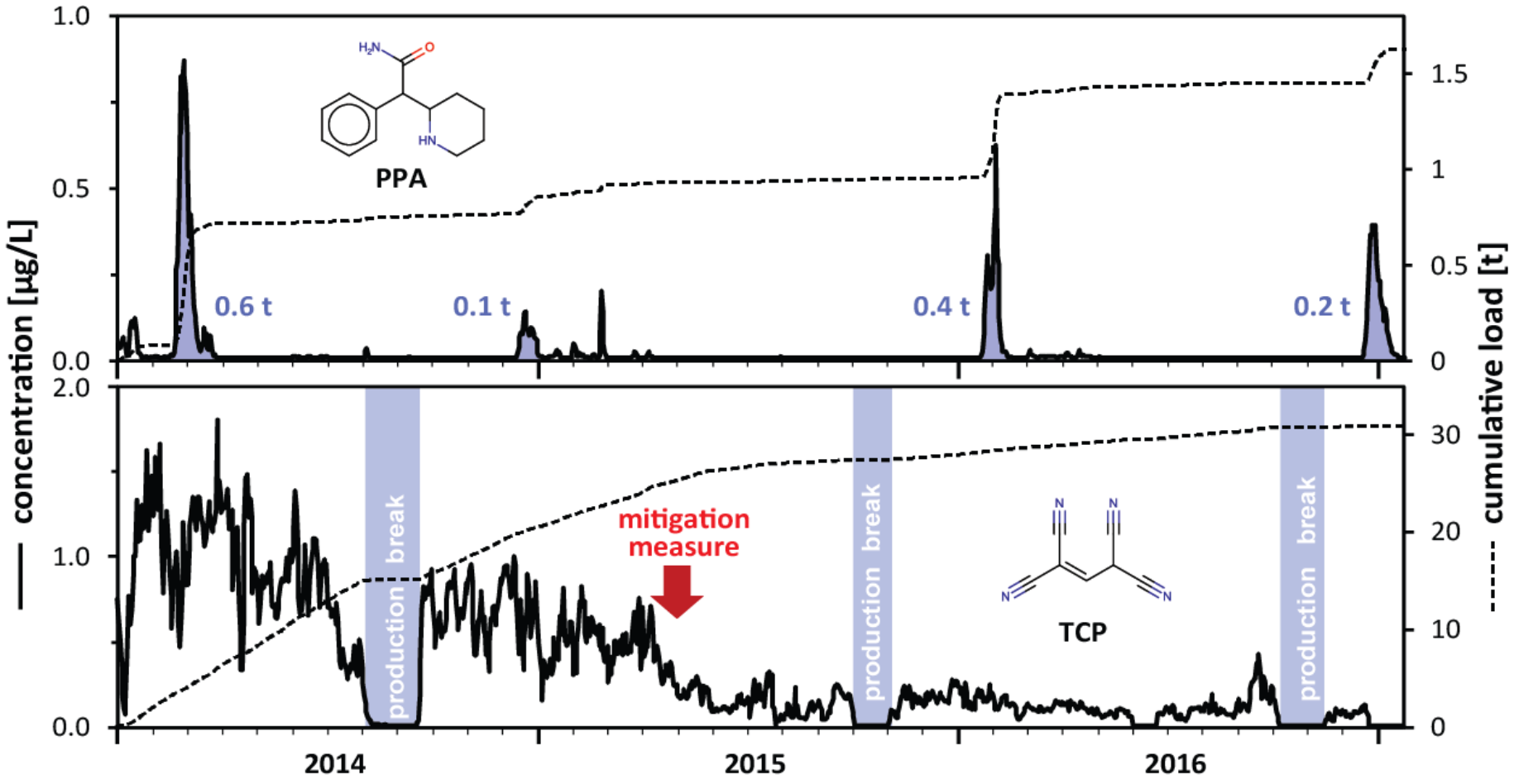
"Live" retrospective screening of known and unknown chemicals in European samples (various matrices)



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

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

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.* (Perspective) Environ. Sci.: Processes Impacts, 2019, DOI: 10.1039/C9EM00068B [cover image]

The 102 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

- 102 million …
  OR
- Most relevant/annotated
  => PubChemLite
  tier0: 316 K
  tier1: 360 K

zenodo Search Upload Communities

January 14, 2020 Dataset Open Access

PubChemLite tier0 and tier1

Bolton, Evan; Schymanski, Emma

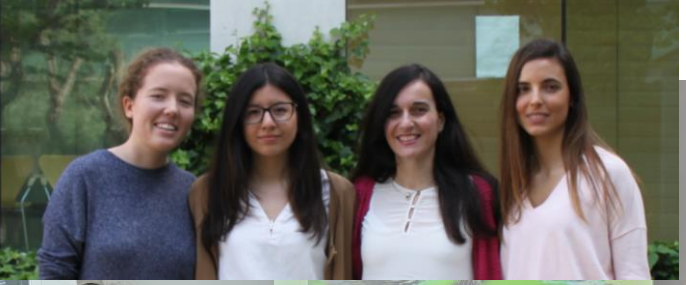
PubChemLite is a subset of PubChem (https://pubchem.ncbi.nlm.nih.gov/) selected from major categories of the Table of Contents page at the PubChem Classification Browser (https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72). So far we are providing two "flavours":

tier0 is 316,810 compounds (14 Jan 2020) compiled from 7 categories: AgroChemInfo, DrugMedicInfo, FoodRelated, PharmacoInfo, SafetyInfo, ToxicityInfo, KnownUse

tier1 is 363,911 compounds (14 Jan 2020) compiled from 8 categories (tier0 + BioPathway): AgroChemInfo, BioPathway, DrugMedicInfo, FoodRelated, PharmacoInfo, SafetyInfo, ToxicityInfo, KnownUse

PubChemCIDs have been collapsed by InChIKey first block, reporting the structure from the most annotated CID, plus related CIDs. Entries that will be ignored by MetFrag (salts, disconnected substances) or cause errors (e.g. transition metals) have been removed. The Patent and PubMed ID counts are extracted from files on the PubChem FTP site. The "AnnoTypeCount" term counts how many of the categories are represented, the subsequent column (named per category) counts the number of annotation categories available in the next sub-category of the TOC entry.

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



THS-RESEARCH TEAM

ECI@LCSB

Acknowledgements

Luxembourg National Research Fund

emma.schymanski@uni.lu
@ESchymanski @soniatorres @noeliaramz

Further Information:

DOI: 10.5281/zenodo.3613472

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

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

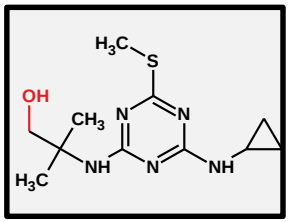
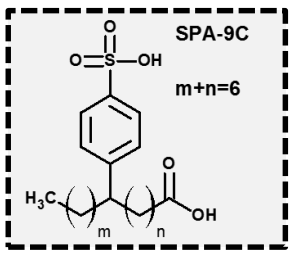
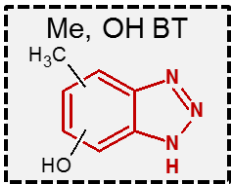
https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics

THS & ECI@LCSB teams, M Stravs, J Hollender, S Neumann, C Ruttkies, T Schulze, N Alygizakis, R Aalizadeh, A Williams, E Bolton


Community Efforts!
modelkey
solutions
MassBank consortium
norman
ART2 CURE
© SCHLOSS DAGSTUHL – LZI GMBH
licensed under Creative Commons License CC BY-NC-ND

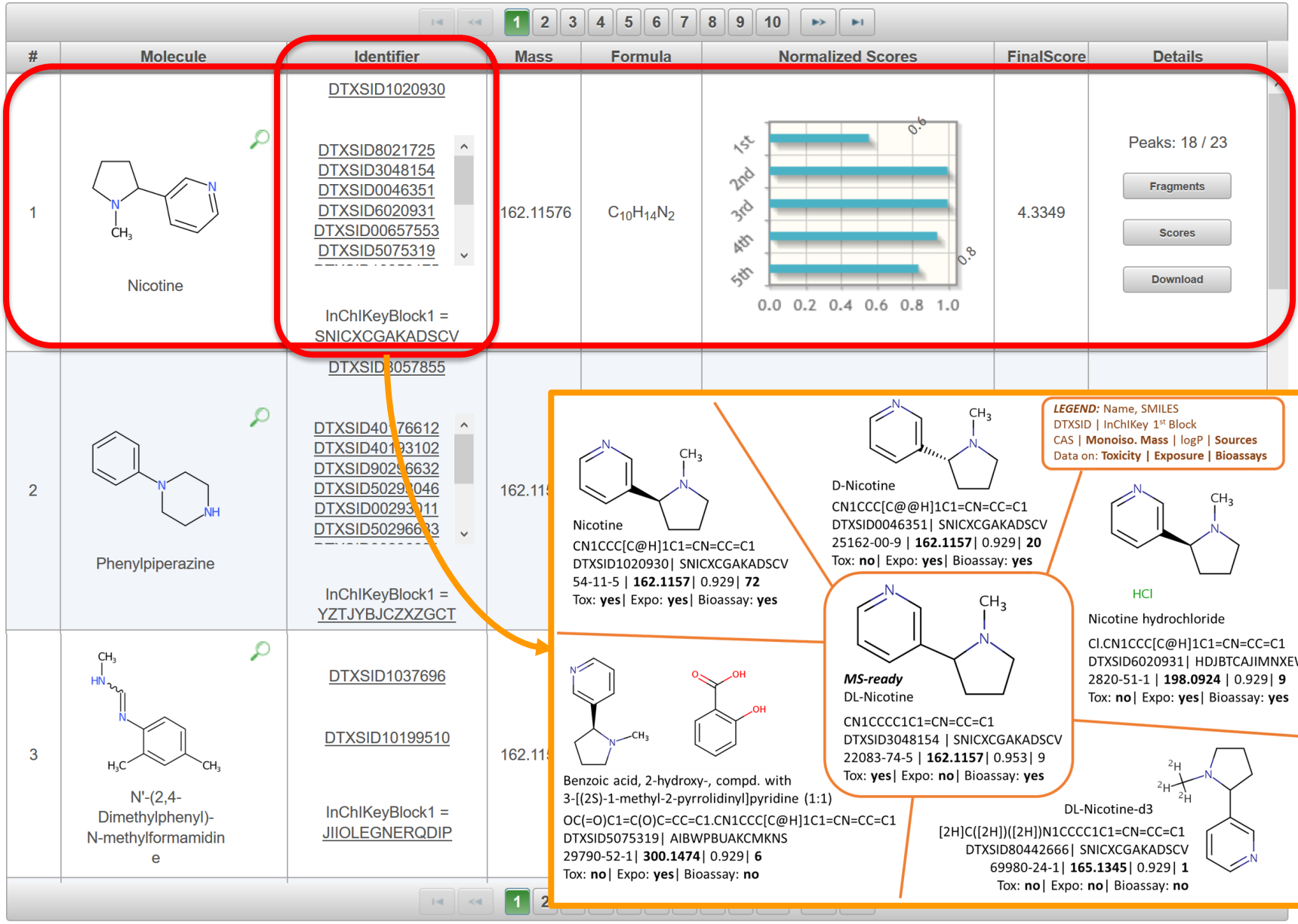
Supporting Info

Confidence Levels for Tentative Structures

Example	*Identification confidence*	*Minimum data requirements*
	***Level 1: Confirmed structure*** by reference standard	MS, MS^2, RT, Reference Std.
	Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS^2, Library MS^2 MS, MS^2, Exp. data
 	***Level 3: Tentative candidate(s)*** structure, substituent, class	MS, MS^2, Exp. data
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Schymanski *et al*. (2014) ES&T, 48 (4), 2097-2098. DOI: 10.1021/es5002105

“MS-ready” Form for MetaData in MetFrag

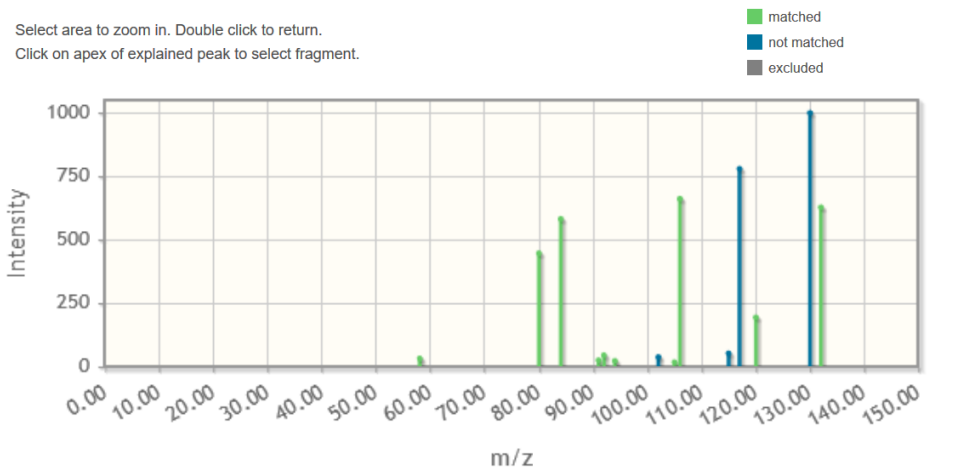


McEachran *et al*. 2018, DOI: 10.1186/s13321-018-0299-2; Schymanski & Williams, 2017 *ES&T* DOI: 10.1021/acs.est.7b01908

MetFrag – Example with Nicotine VI

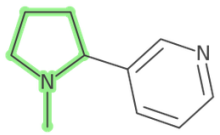
Fragments View

Select area to zoom in. Double click to return.
Click on apex of explained peak to select fragment.



Fragments

Fragment 3



Peak m/z: 84.0808

Fragment Mass: 84.08082 Da

Fragment Formula: $[C_5H_{10}N]^+$

Scores View

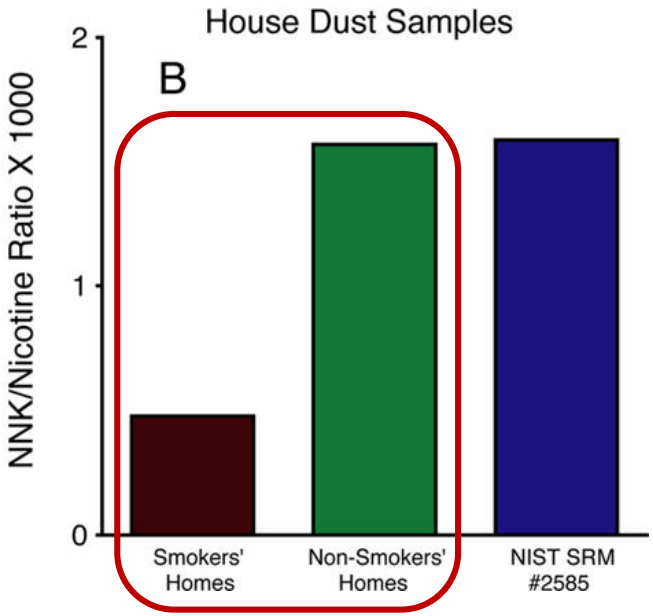
Candidate Name: 3-(1-methylpyrrolidin-2-yl)pyridine
Candidate Identifier: 942

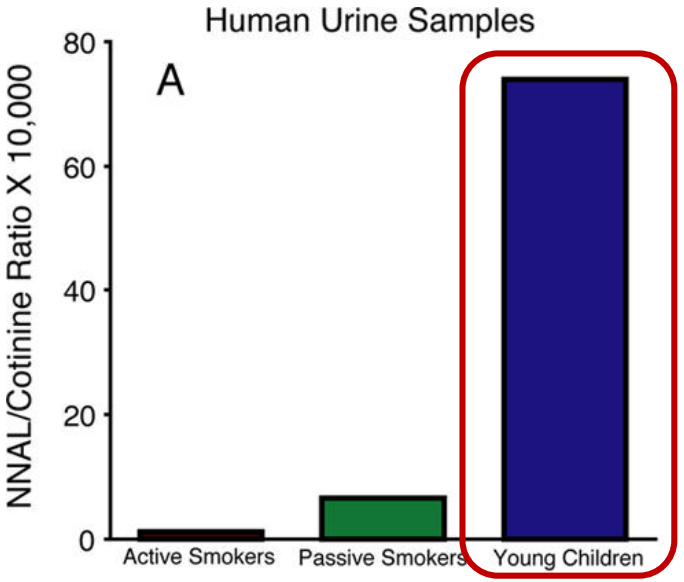
	Name	Normalized Value	Raw Value
▶	MetFrag	0.5197	332.8153
▶	SpectralSimilarity	0.9712	6.4214
▶	ExactSpectralSimilarity	0.9284	0.9284
▶	PatentsCount	0.9991	71329.0
▶	PubMedReferenceCount	1.0	18271.0

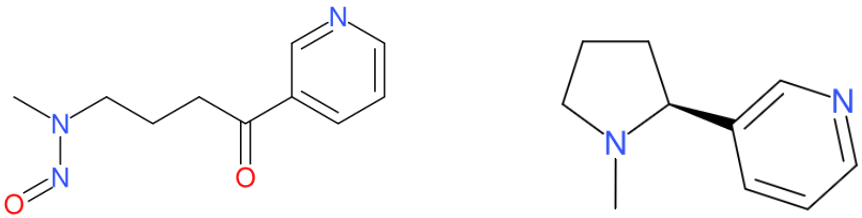
Why THS is different to SHS

Ratio
NNK : Nicotine
in house dust

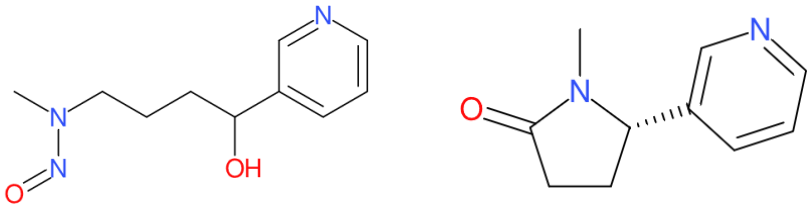
Ratio of
NNAL : Cotinine
in urine







NNK : Nicotine



NNAL : Cotinine

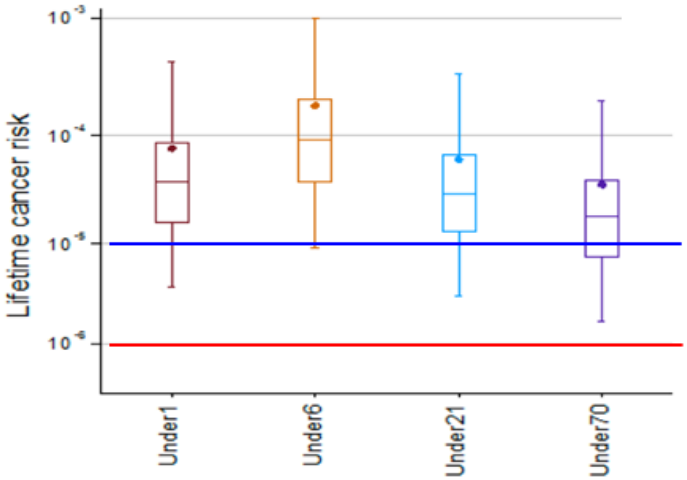
Jacob et al. Chem. Res. Toxicol. (2016). DOI 10.1021/acs.chemrestox.6b00343

THS in Tarragona, Spain

UNIVERSITAT ROVIRA i VIRGILI

Nicotine exposure (ng/kg-day)

Age (years old)	Mean	Max.	Mean	Max.
< 1	129	1637	11	25
1-5	136	1729	12	27
6-21	83	1048	7	16
22-70	80	1030	7	16

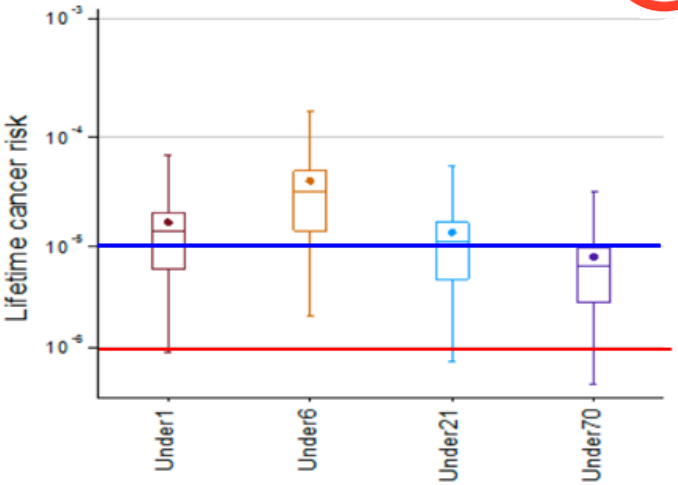


Lifetime cancer risk by age group.

Threshold values

WHO: 10^{-5}

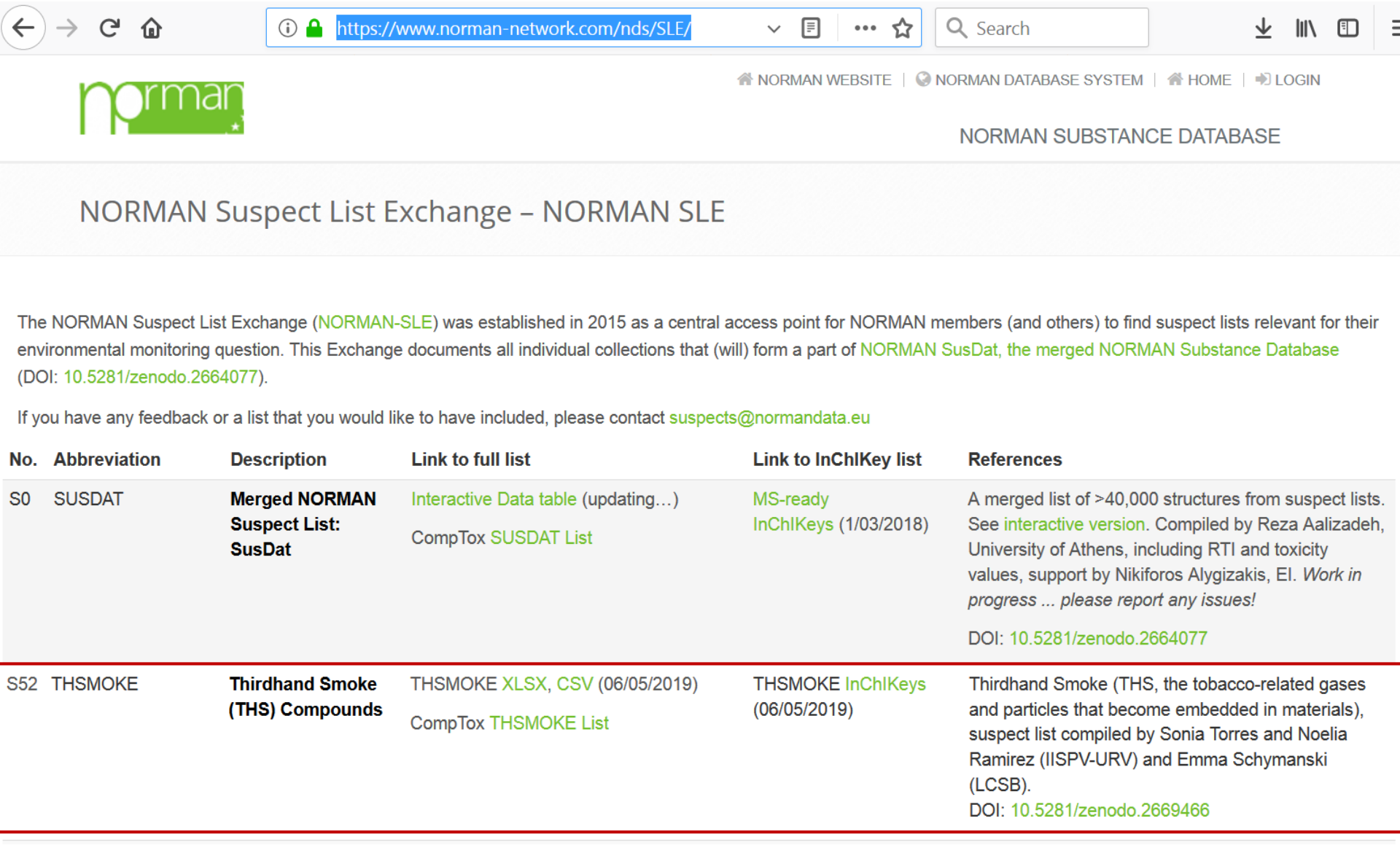
USEPA: 10^{-6}



Ramírez *et al.* Environment International (2014). DOI: 10.1016/j.envint.2014.06.012

THSMOKE on NORMAN SLE

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



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

Search

NORMAN WEBSITE | NORMAN DATABASE SYSTEM | HOME | LOGIN

NORMAN SUBSTANCE DATABASE

NORMAN Suspect List Exchange – NORMAN SLE

The NORMAN Suspect List Exchange (NORMAN-SLE) was established in 2015 as a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring question. This Exchange documents all individual collections that (will) form a part of NORMAN SusDat, the merged NORMAN Substance Database (DOI: 10.5281/zenodo.2664077).

If you have any feedback or a list that you would like to have included, please contact suspects@normandata.eu

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
S52	THSMOKE	**Thirdhand Smoke (THS) Compounds**	THSMOKE XLSX, CSV (06/05/2019) CompTox THSMOKE List	THSMOKE InChIKeys (06/05/2019)	Thirdhand Smoke (THS, the tobacco-related gases and particles that become embedded in materials), suspect list compiled by Sonia Torres and Noelia Ramirez (IISPV-URV) and Emma Schymanski (LCSB). DOI: 10.5281/zenodo.2669466

THSMOKE on CompTox Chemicals Dashboard

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

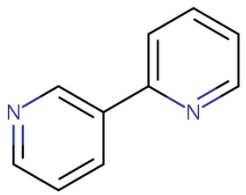
NORMAN: Thirdhand Smoke (THS) Compounds: Suspect List

List Details

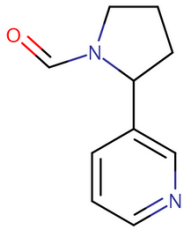
Description: A collection of compounds for mass spectrometry suspect screening of Thirdhand Smoke (THS, the tobacco-related gases and particles that become embedded in materials), compiled by S. Torres, N. Ramirez, Institut D'investigacio Sanitaria Pere Virgili at Universitat Rovira i Virgili (IISPV-URV) and E. Schymanski (Luxembourg Center for Systems Biomedicine, LCSB)
Number of Chemicals: 95

95 chemicals

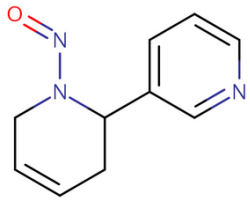
Select all | Download | Send to Batch Search | Default | DTXSID | Mass | Sources | Hide chemicals that are:



2,3'-Bipyridine
DTXSID:DTXSID00206823
Mass:156.068748
Sources:16



Nornicotine, N-formyl
DTXSID:DTXSID30336006
Mass:176.094963
Sources:3



1-Nitroso-1,2,3,6-tetrahydro-2,3'-bipyrid...
DTXSID:DTXSID40868005
Mass:189.090212
Sources:4