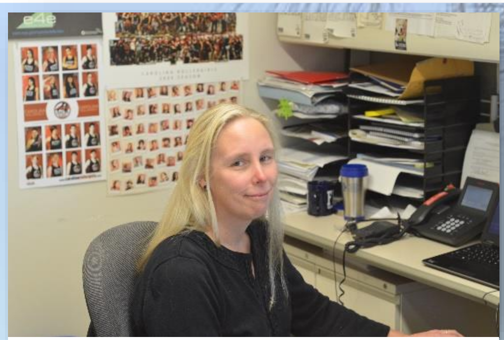
United States Environmental Protection Agency

Building a Non-Targeted Analysis Research Program at the U.S. EPA

Jon R. Sobus, Ph.D. & the EPA/ORD NTA Team

Center for Computational Toxicology and Exposure
Research Triangle Park, NC

Current NTA Team


Elin ‘Da Boss’ Ulrich


Tony ‘Stark’ Williams

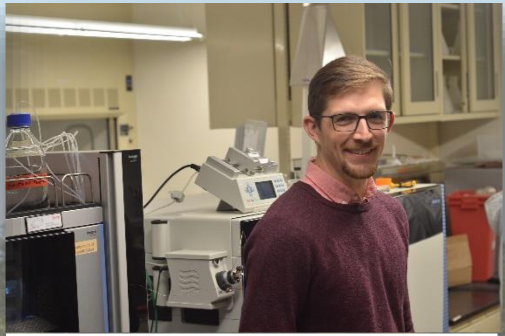

‘Dapper’ Charlie Lowe

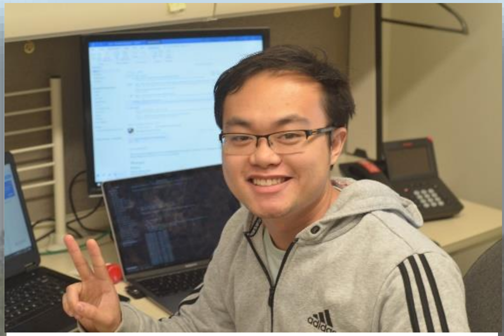

Scott ‘The Postman’ Clifton

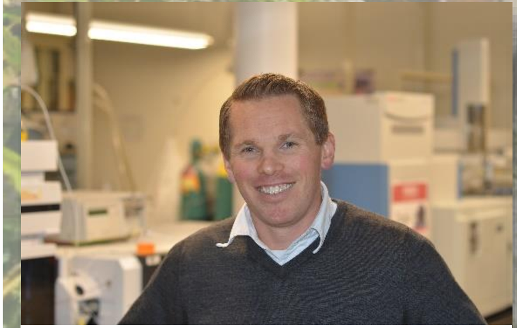

Jon ‘Nature Boy’ Sobus

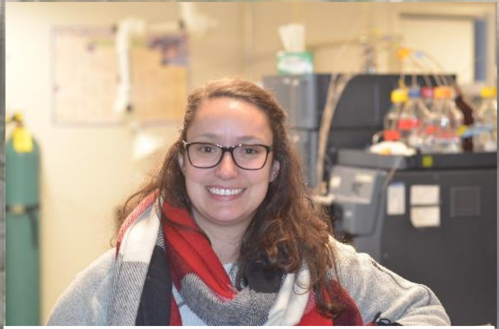

Alex ‘Can Do’ Chao


Mark ‘Blue Steel’ Strynar

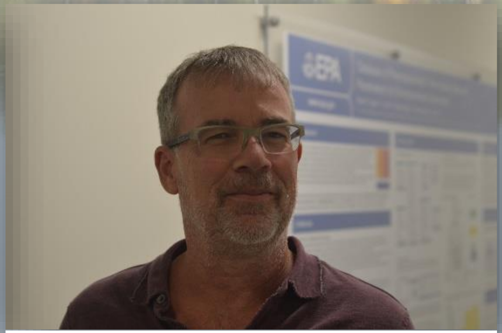

James ‘Shake-n-Bake’ McCord


Nelson ‘Prints’ Yeung


Seth ‘Nice guy’ Newton


Hannah ‘Dr. Cool’ Liberatore


The Unflappable Ariel Wallace

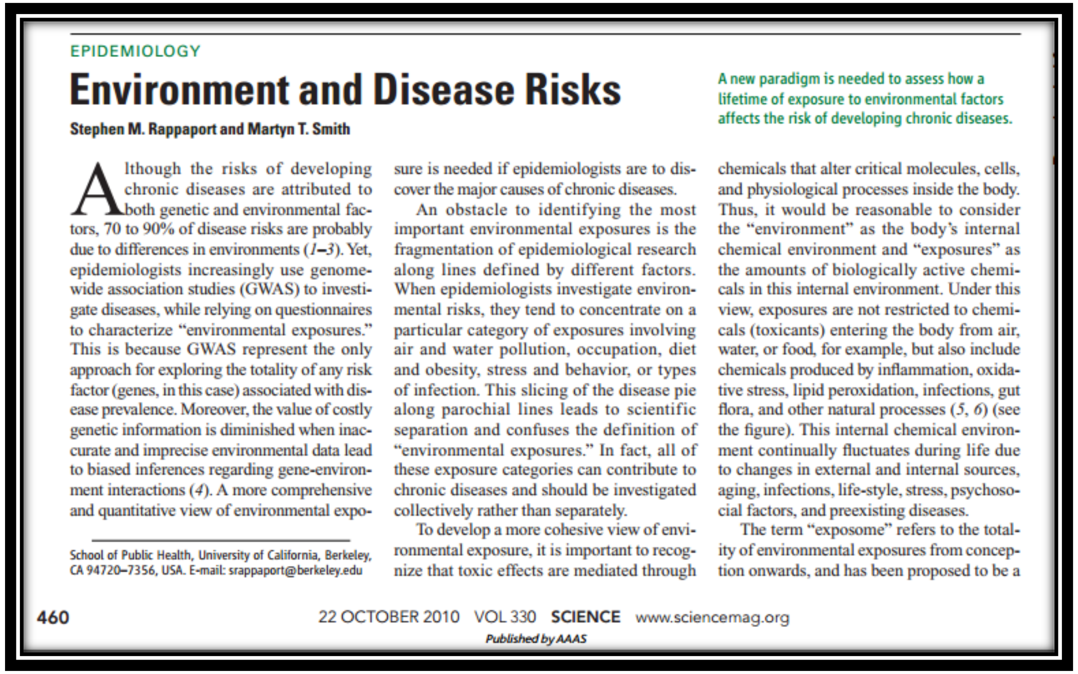

Tom ‘Mystery Man’ Purucker


‘Adventurin’ Jeff Minucci

Key Drivers for 21st Century Exposure Science

1) Understanding causes of disease

"…70-90% of disease risks are probably due to differences in environments"



EPIDEMIOLOGY

Environment and Disease Risks

Stephen M. Rappaport and Martyn T. Smith

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (*1*–*3*). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize "environmental exposures." This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (*4*). A more comprehensive and quantitative view of environmental exposure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of "environmental exposures." In fact, all of these exposure categories can contribute to chronic diseases and should be investigated collectively rather than separately.

To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxicants) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (*5*, *6*) (see the figure). This internal chemical environment continually fluctuates during life due to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

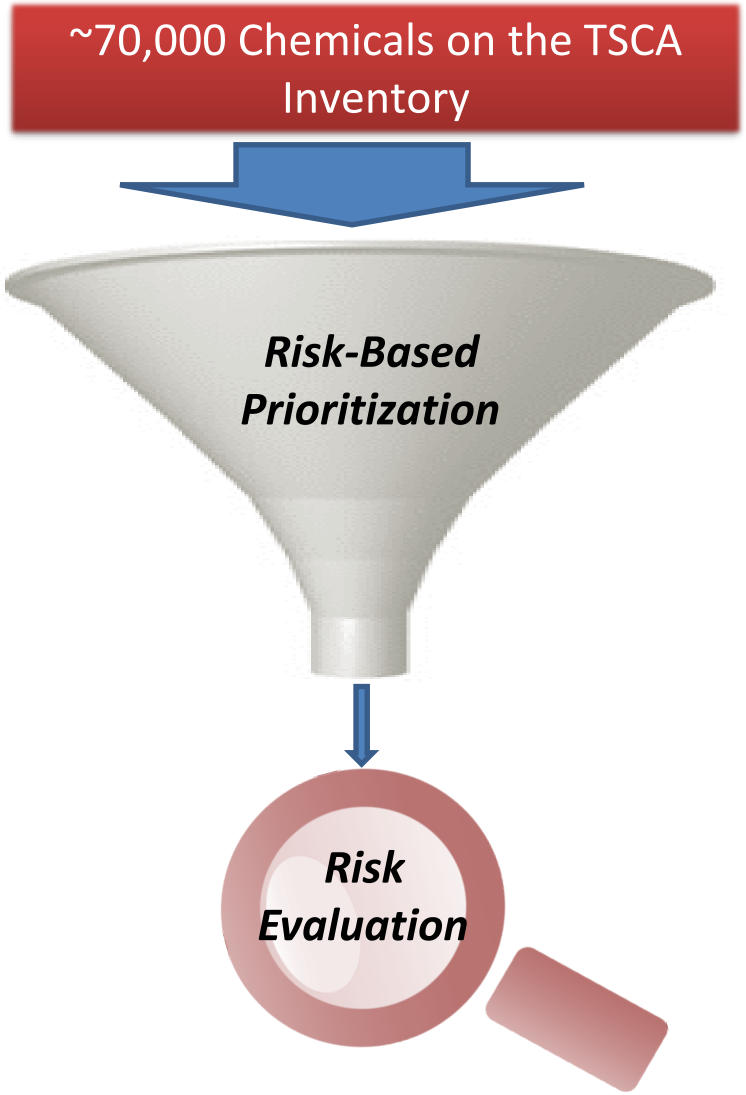
The term "exposome" refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

School of Public Health, University of California, Berkeley, CA 94720–7356, USA. E-mail: srappaport@berkeley.edu

2) Ensuring chemical safety

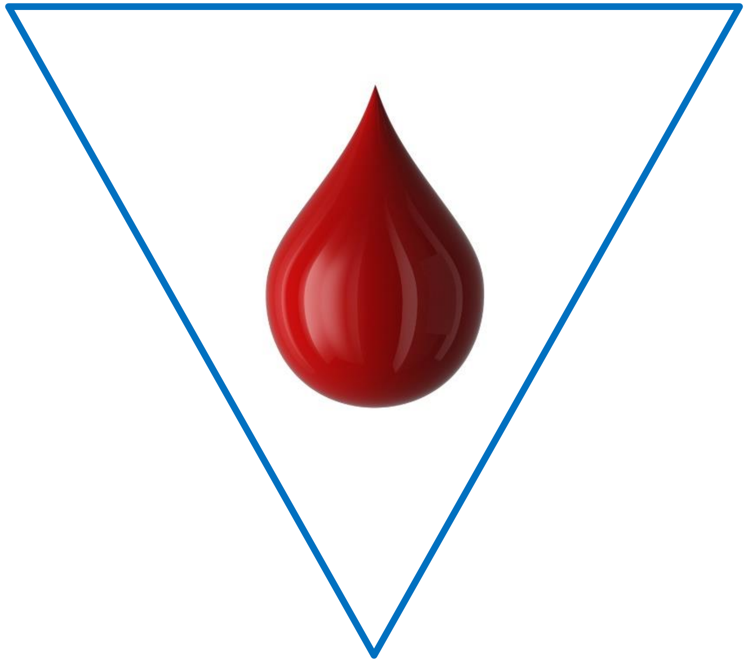
High-Throughput Risk Characterization

- Many industrial & commercial chemicals are covered by the Toxic Substances Control Act (TSCA), which is administered by EPA.
- TSCA updated in June 2016 to allow *risk-based* evaluation of existing and new chemicals.
- Characterization of risk requires exposure and hazard data.
- EPA's Office of Research and Development (ORD) is developing new approach methodologies (NAMs) for rapid risk characterization.
- <u>NTA is a promising NAM, but requires careful evaluation and implementation</u>

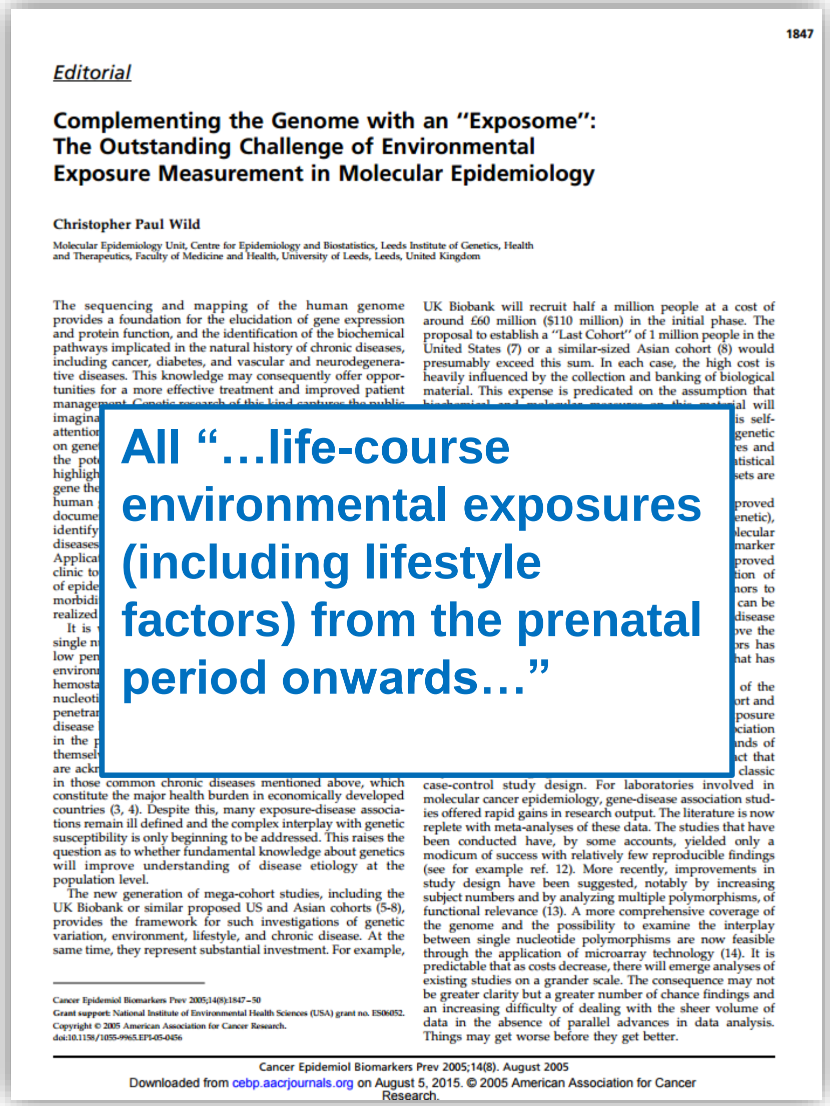


NTA Research Produces Critical Data

Top-Down Exposomics via NTA



Measure Important Exposures Within the Receptor



1847

Editorial

Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology

Christopher Paul Wild

Molecular Epidemiology Unit, Centre for Epidemiology and Biostatistics, Leeds Institute of Genetics, Health and Therapeutics, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

The sequencing and mapping of the human genome provides a foundation for the elucidation of gene expression and protein function, and the identification of the biochemical pathways implicated in the natural history of chronic diseases, including cancer, diabetes, and vascular and neurodegenerative diseases. This knowledge may consequently offer opportunities for a more effective treatment and improved patient management. …

… in those common chronic diseases mentioned above, which constitute the major health burden in economically developed countries (3, 4). Despite this, many exposure-disease associations remain ill defined and the complex interplay with genetic susceptibility is only beginning to be addressed. This raises the question as to whether fundamental knowledge about genetics will improve understanding of disease etiology at the population level.

The new generation of mega-cohort studies, including the UK Biobank or similar proposed US and Asian cohorts (5-8), provides the framework for such investigations of genetic variation, environment, lifestyle, and chronic disease. At the same time, they represent substantial investment. For example, UK Biobank will recruit half a million people at a cost of around £60 million ($110 million) in the initial phase. The proposal to establish a "Last Cohort" of 1 million people in the United States (7) or a similar-sized Asian cohort (8) would presumably exceed this sum. In each case, the high cost is heavily influenced by the collection and banking of biological material. This expense is predicated on the assumption that …

… case-control study design. For laboratories involved in molecular cancer epidemiology, gene-disease association studies offered rapid gains in research output. The literature is now replete with meta-analyses of these data. The studies that have been conducted have, by some accounts, yielded only a modicum of success with relatively few reproducible findings (see for example ref. 12). More recently, improvements in study design have been suggested, notably by increasing subject numbers and by analyzing multiple polymorphisms, of functional relevance (13). A more comprehensive coverage of the genome and the possibility to examine the interplay between single nucleotide polymorphisms are now feasible through the application of microarray technology (14). It is predictable that as costs decrease, there will emerge analyses of existing studies on a grander scale. The consequence may not be greater clarity but a greater number of chance findings and an increasing difficulty of dealing with the sheer volume of data in the absence of parallel advances in data analysis. Things may get worse before they get better.

Cancer Epidemiol Biomarkers Prev 2005;14(8):1847–50

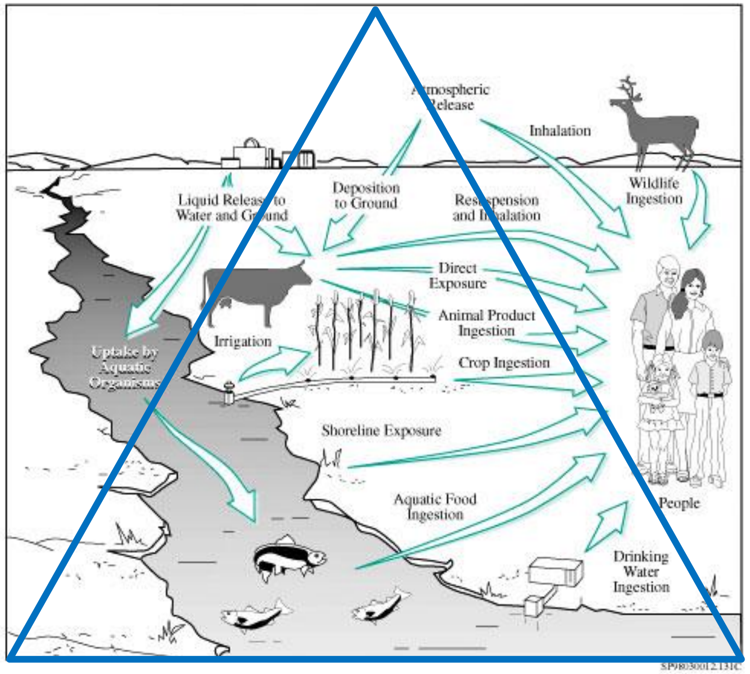
Grant support: National Institute of Environmental Health Sciences (USA) grant no. ES06052.

doi:10.1158/1055-9965.EPI-05-0456

Cancer Epidemiol Biomarkers Prev 2005;14(8). August 2005

All "…life-course environmental exposures (including lifestyle factors) from the prenatal period onwards…"

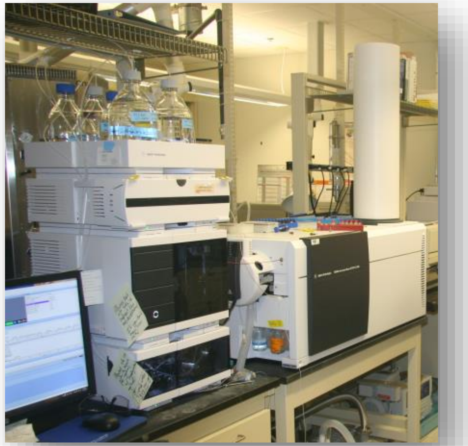
Bottom-Up Exposomics via NTA



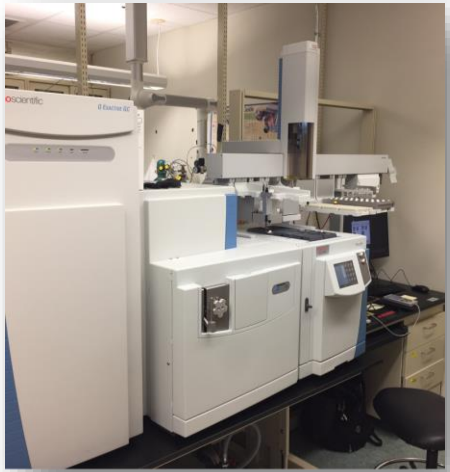
Measure Important Exposures in All Relevant Media

Figure adapted from: Rappaport SM. *J Expo Sci Environ Epidemiol*. 2011 Jan-Feb;21(1):5-9.

Our HRMS Tools of the Trade



Agilent 6530B
LC/Q-TOF



Thermo GC/Q Exactive
Hybrid Quad-Orbitrap

Thermo LC/Orbitrap
Fusion Tribrid

Agilent 7250 GC/Q-TOF

Coming Soon!!

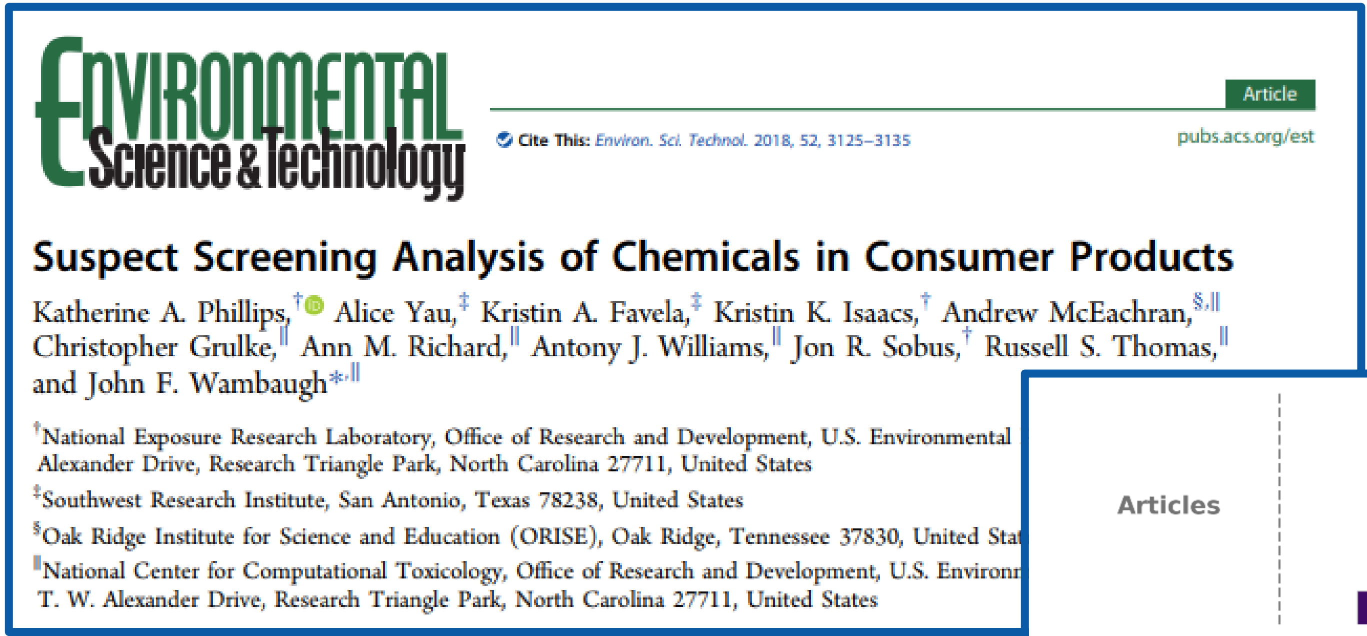
Agilent 6546 LC/Q-TOF

United States Environmental Protection Agency

NTA Applications at EPA

- **Exposure surveillance**
  - What chemicals are in water, products, dust, blood, etc.?
- **Chemical prioritization**
  - What are relevant chemicals & mixtures?
- **Exposure forensics**
  - What are chemical signatures of exposure sources?
- **Biomarker discovery**
  - What chemicals are associated with health impairment?

Exposure Surveillance for Consumer Products


Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2018, 52, 3125–3135

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

Katherine A. Phillips,[†] Alice Yau,[‡] Kristin A. Favela,[‡] Kristin K. Isaacs,[†] Andrew McEachran,[§,∥] Christopher Grulke,[∥] Ann M. Richard,[∥] Antony J. Williams,[∥] Jon R. Sobus,[†] Russell S. Thomas,[∥] and John F. Wambaugh*[,∥]

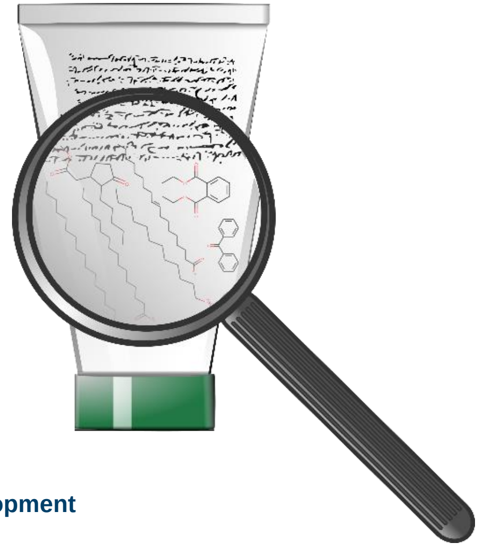
[†]National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental
Alexander Drive, Research Triangle Park, North Carolina 27711, United States

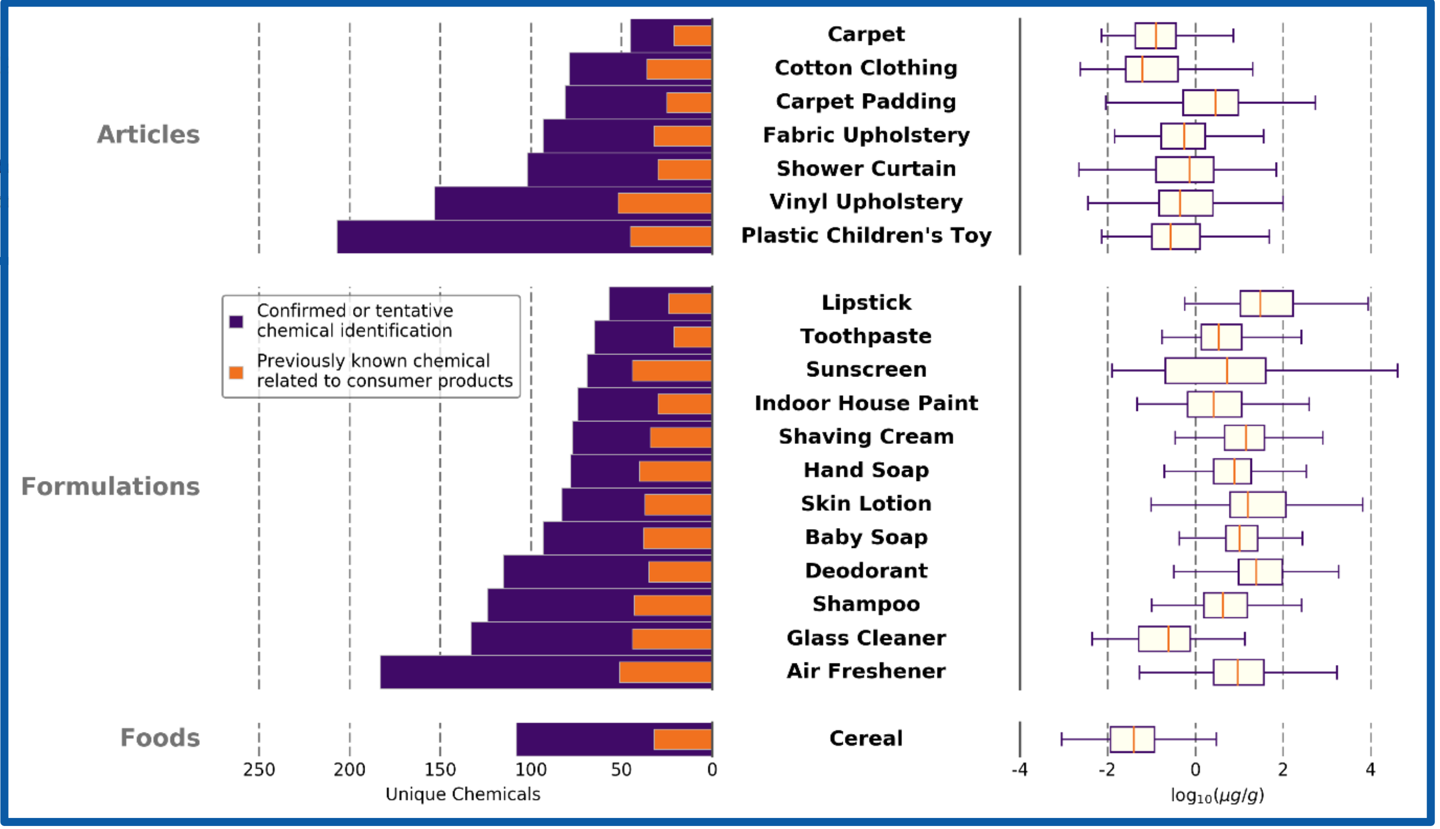
[‡]Southwest Research Institute, San Antonio, Texas 78238, United States

[§]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, Tennessee 37830, United Stat

[∥]National Center for Computational Toxicology, Office of Research and Development, U.S. Environ
T. W. Alexander Drive, Research Triangle Park, North Carolina 27711, United States

19% of chemicals identified by NTA are on consumer product chemical lists





Chemical Prioritization for Drinking Water

Environmental Pollution 234 (2018) 297–306

Contents lists available at ScienceDirect

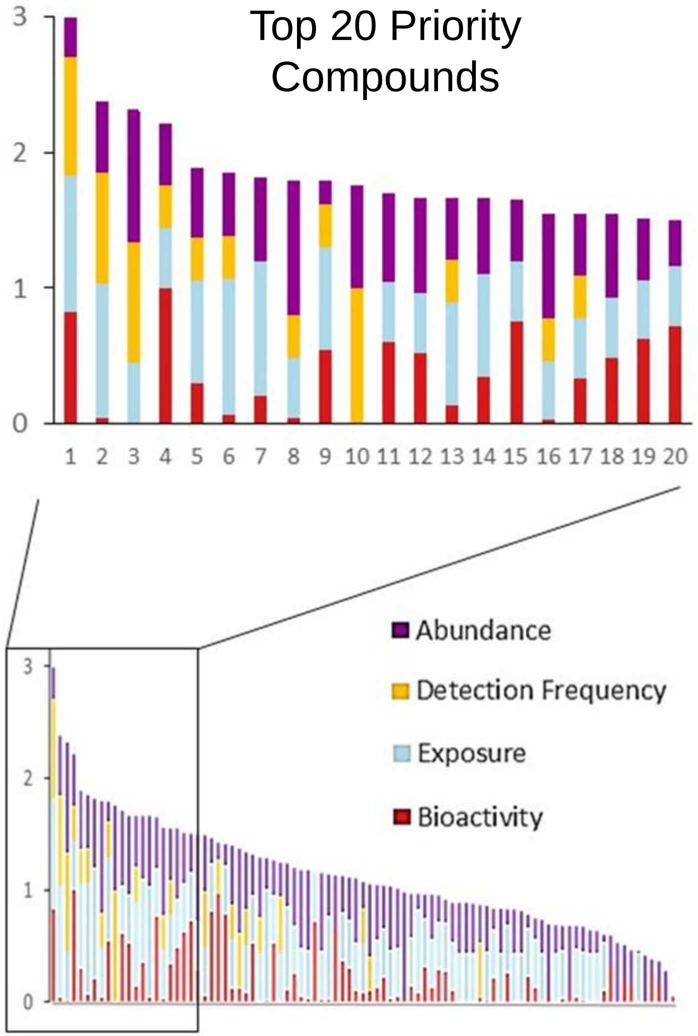
Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Suspect screening and non-targeted analysis of drinking water using point-of-use filters☆

Seth R. Newton [a,*], Rebecca L. McMahen [a,b], Jon R. Sobus [a], Kamel Mansouri [b,c,1], Antony J. Williams [c], Andrew D. McEachran [b,c], Mark J. Strynar [a]

[a] United States Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC 27709, United States
[b] Oak Ridge Institute for Science and Education Research Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[c] United States Environmental Protection Agency, National Center for Computational Toxicology, Research Triangle Park, NC 27709, United States



Top 100 Priority Compounds

#	Compound	ToxPi Score
1	1,2-Benzisothiazolin-3-one*	2.99
2	Diethyleneglycol	2.38
3	N-[3-(Dimethylamino)propyl] methacrylamide	2.32
4	Nonylparaben	2.22
5	Dipentyl phthalate	1.89
6	2-[2-(2-Butoxyethoxy) ethoxy]ethanol*	1.85
7	N,N-Dimethyldodecan-1-amine*	1.81
8	Sucralose	1.80
9	PFOS*	1.79
10	2-(2-Ethoxyethoxy) ethyl acetate*	1.76
11	TDCPP*	1.71
12	Zearalanol	1.67
13	PFOA*	1.66
14	Butylparaben	1.66
15	Noristerat	1.65
16	p-Synephrine	1.55
17	Alprostadil	1.55
18	Sclareol	1.55
19	PFDA*	1.51
20	Simvastatin	1.50

*Confirmed with standard

United States Environmental Protection Agency

Exposure Forensics for Recycled Products

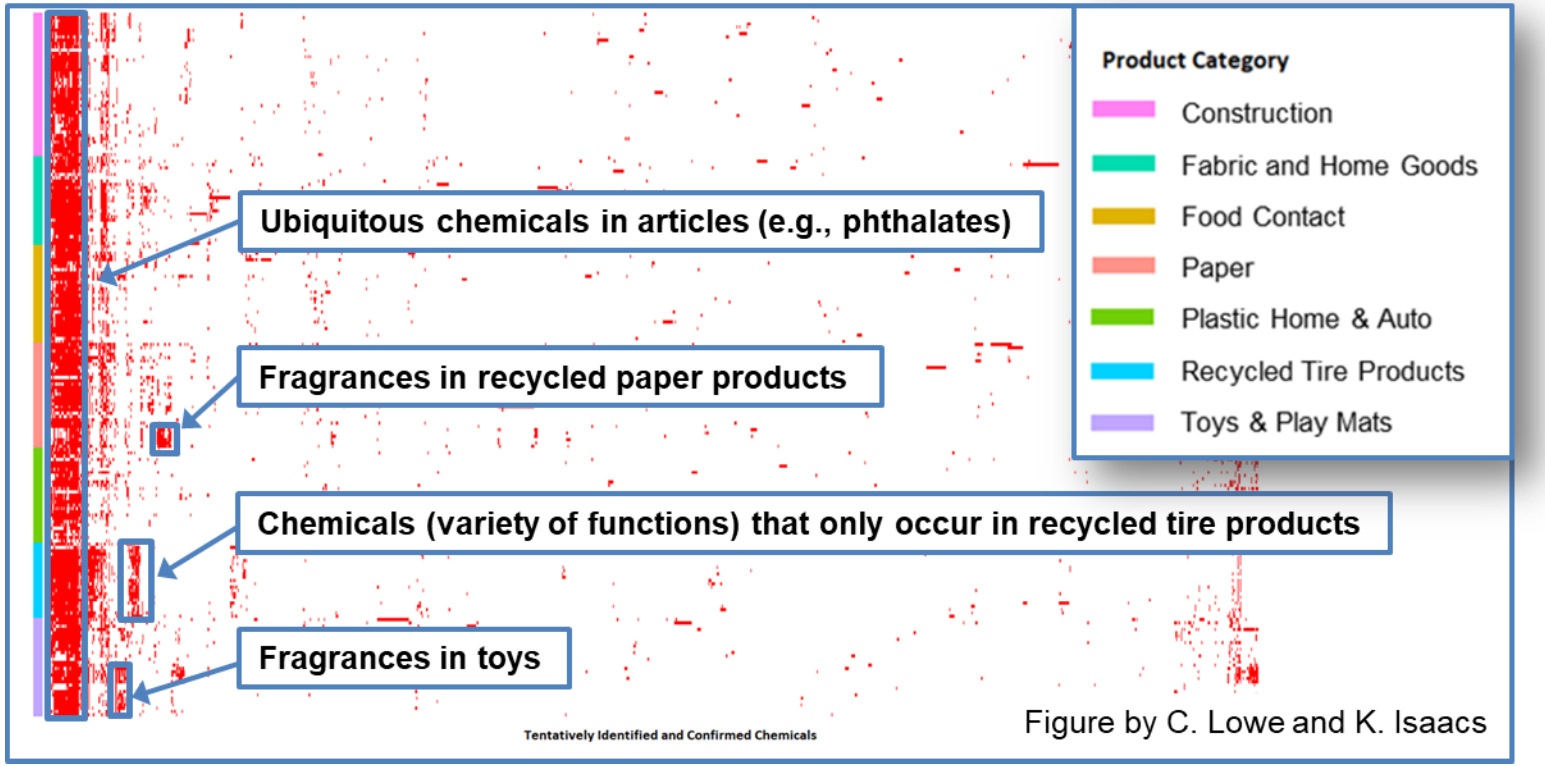
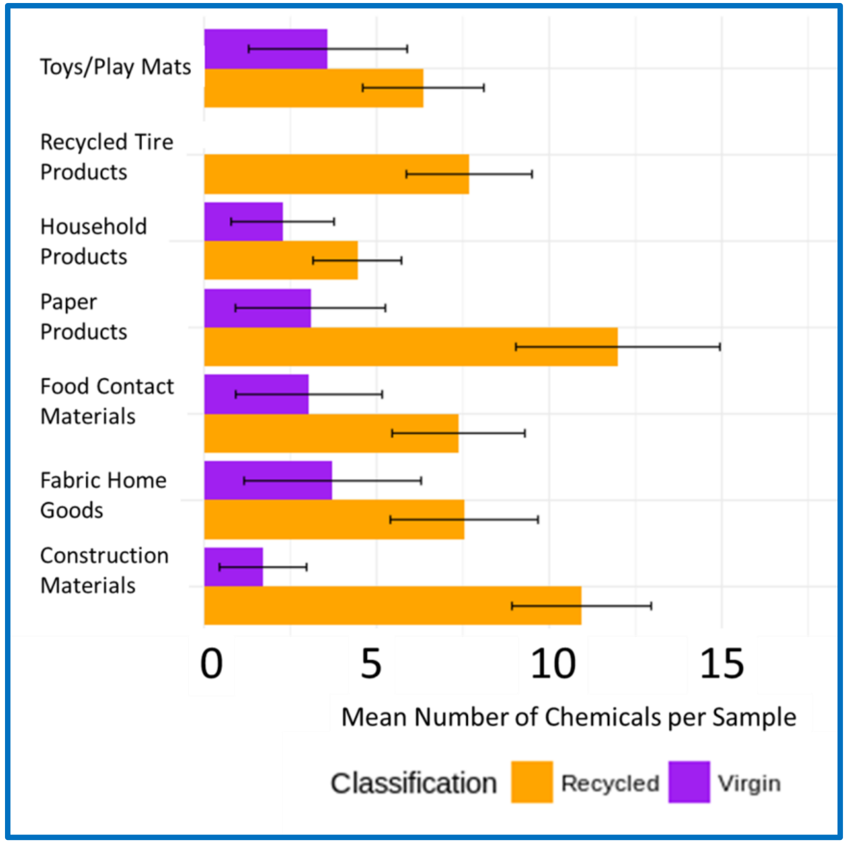


Figure by C. Lowe and K. Isaacs



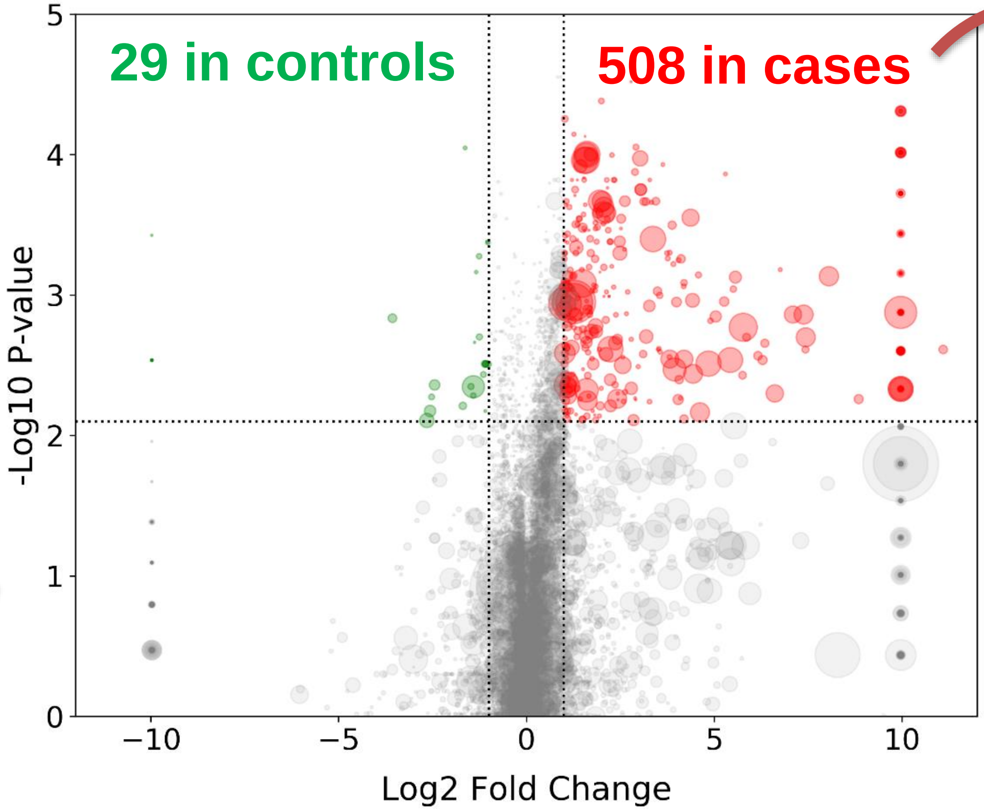
Biomarker Discovery for Placenta Samples

Different Environmental Exposures

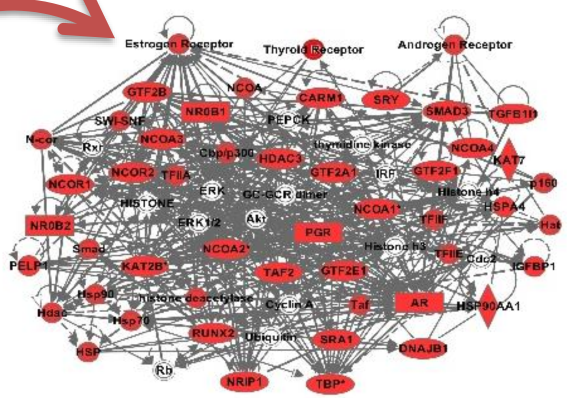




The Placental Exposome (via LC-HRMS)



Altered Cell Signaling



Impaired Angiogenesis

Preeclampsia

Collaboration with J. Rager (UNC Chapel Hill) and J. Grossman (Agilent)

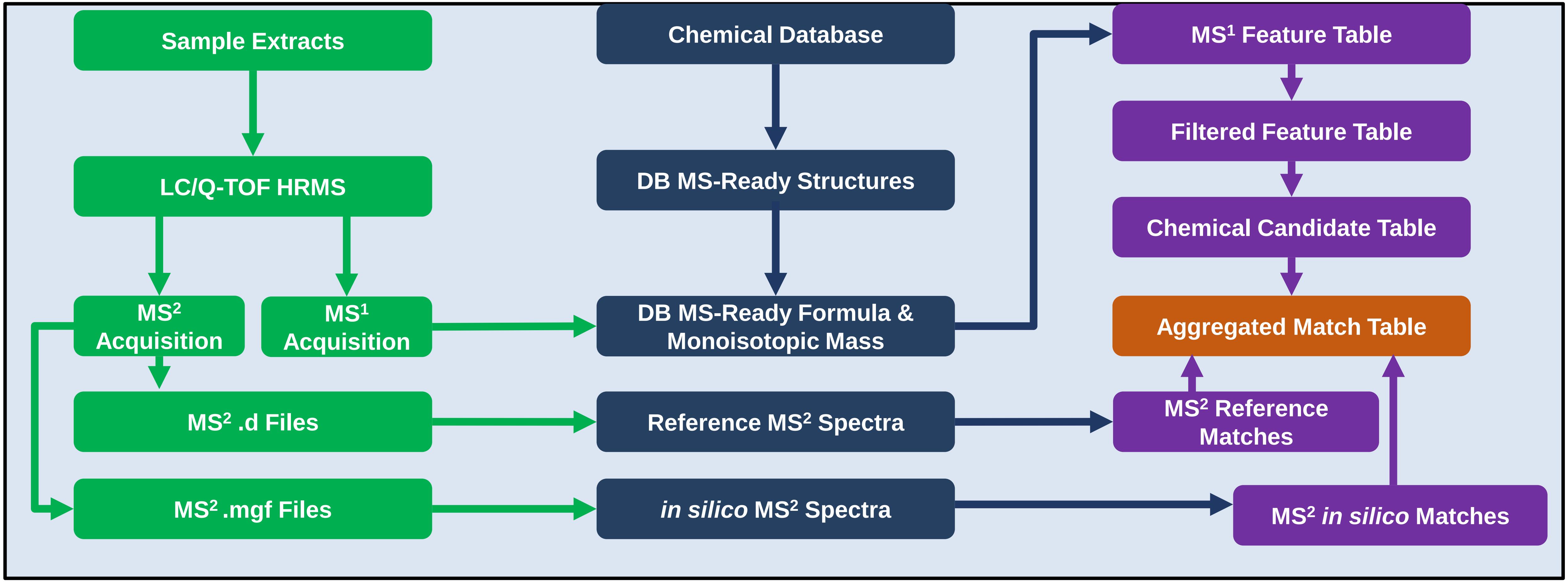
NTA Best Practices

Name	Example	Purpose
Tracers	Isotopically labeled standards: $^{13}C_3$-Atrazine, D_3-Thiamethoxam, $^{13}C_4$,$^{15}N_2$-Fipronil	Allows tracking of chromatographic performance and mass accuracy
Replication	Triplicate injections of same sample vial	Removes risk of “one hit wonder”
Run order randomization	8, 3, 7, 4, 2, 1, 10, 5, 8, 6, 9, 2, 5, 4, 1, 9, 4, 7, 3, 8, 1, 6, 10, 9, 6, 7, 5, 3, 2, 10	Minimizes/averages out batch or sample order effects (e.g., carryover, temp & instrument drift)
Pooled QC sample	Combine 5 mg/µL from each of 10 samples (total 50 mg/µL) prior to extract to create pooled QC	Separate confirmation of presence with different matrix, MS2 IDs
Blanks	Solvent, method, matrix, double blanks	Allows identification/subtraction/deletion of interferences introduced in lab processes
Multiple lines of evidence for ID	RT prediction/matching, spectra prediction/matching, data source ranking, functional/product uses, media occurrence	Improves confidence in identification when chemicals standards are unavailable

EPA United States Environmental Protection Agency

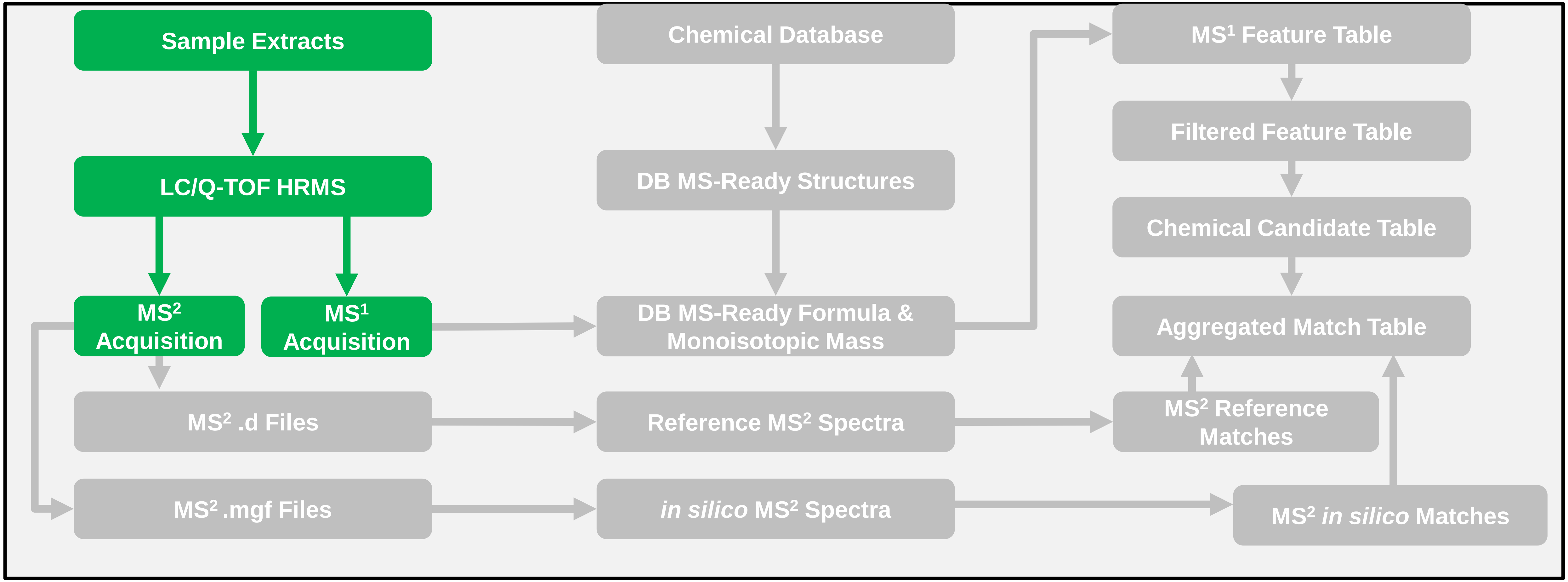
Agilent LC/Q-TOF Simplified Workflow

Experimental Acquisition | **DB & Library Matching** | **Data Analysis**



Agilent LC/Q-TOF Simplified Workflow

Experimental Acquisition | **DB & Library Matching** | **Data Analysis**



United States Environmental Protection Agency

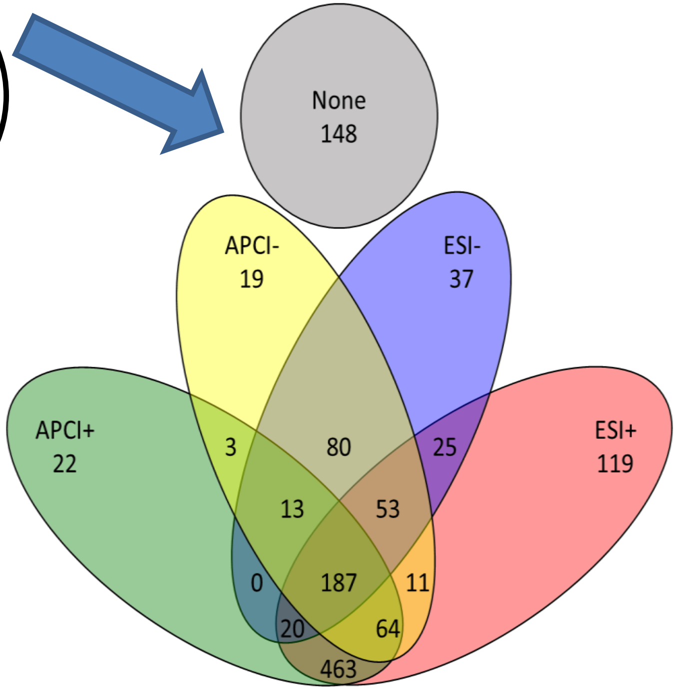
Experimental Acquisition

1,269 Substances in 10 Mixtures

Agilent 6530B Q-TOF

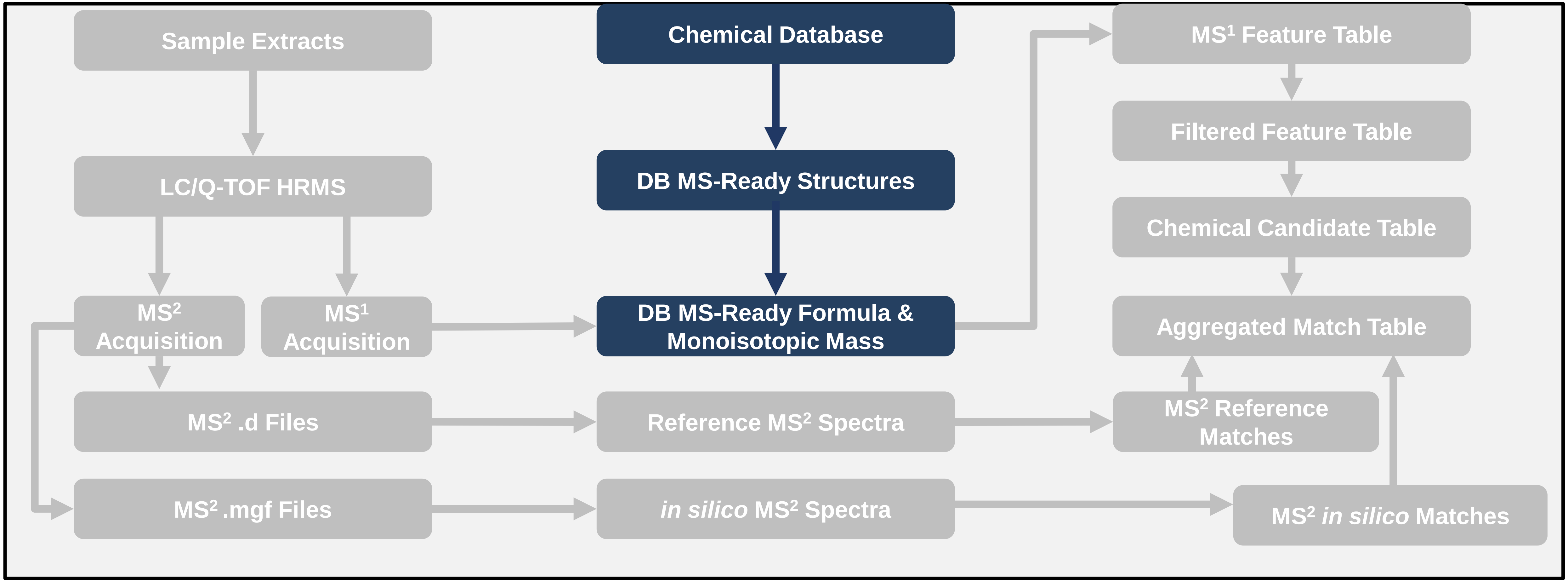
RP vs. HILIC

ESI vs. APCI

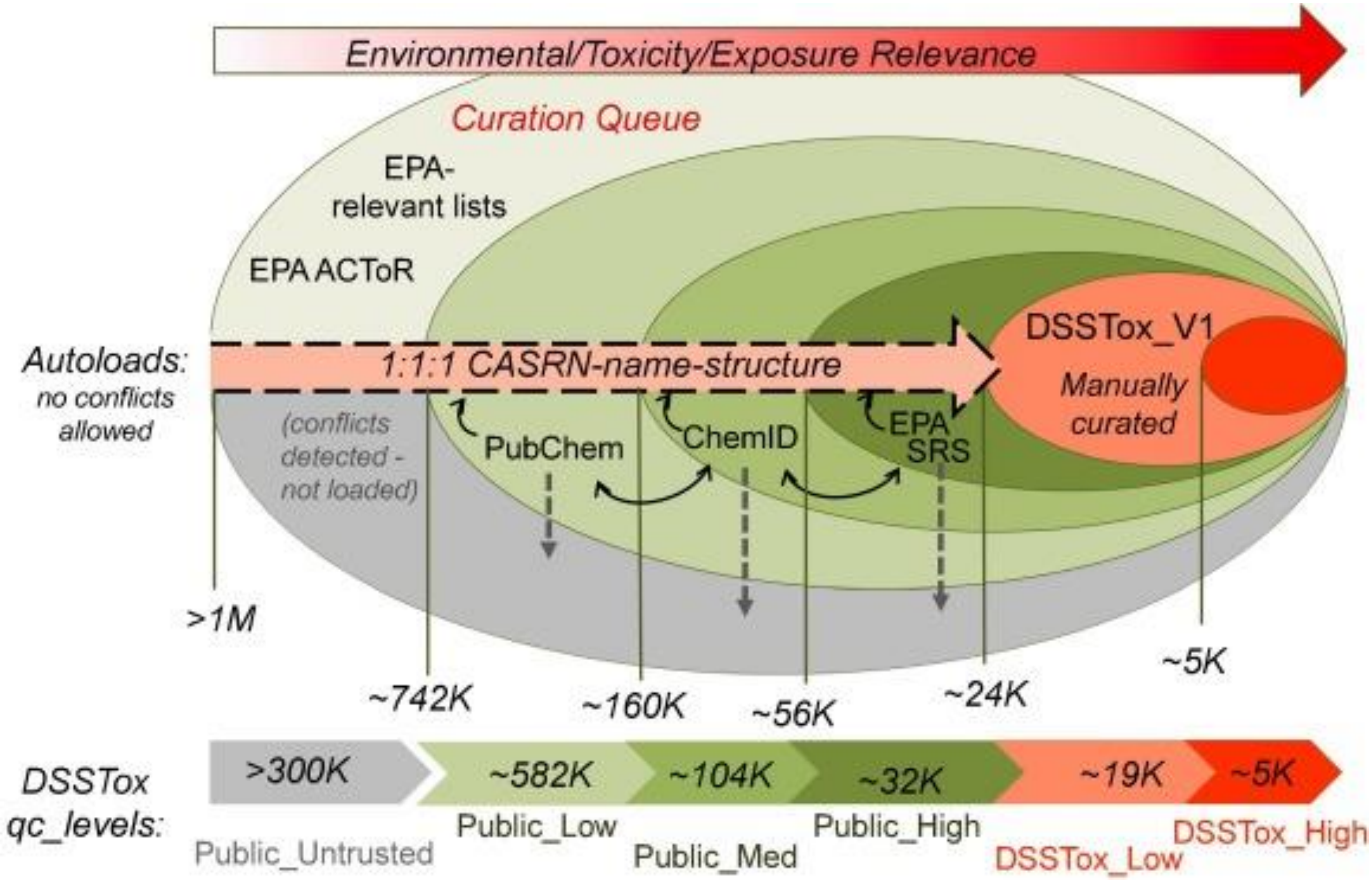


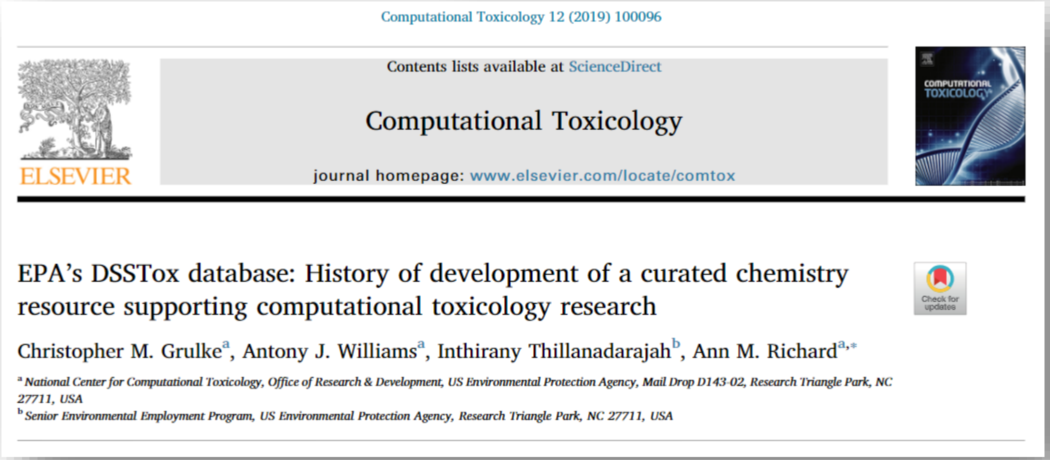
Agilent LC/Q-TOF Simplified Workflow

Experimental Acquisition | DB & Library Matching | Data Analysis



Chemical Database = DSSTox




Computational Toxicology 12 (2019) 100096

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

ELSEVIER

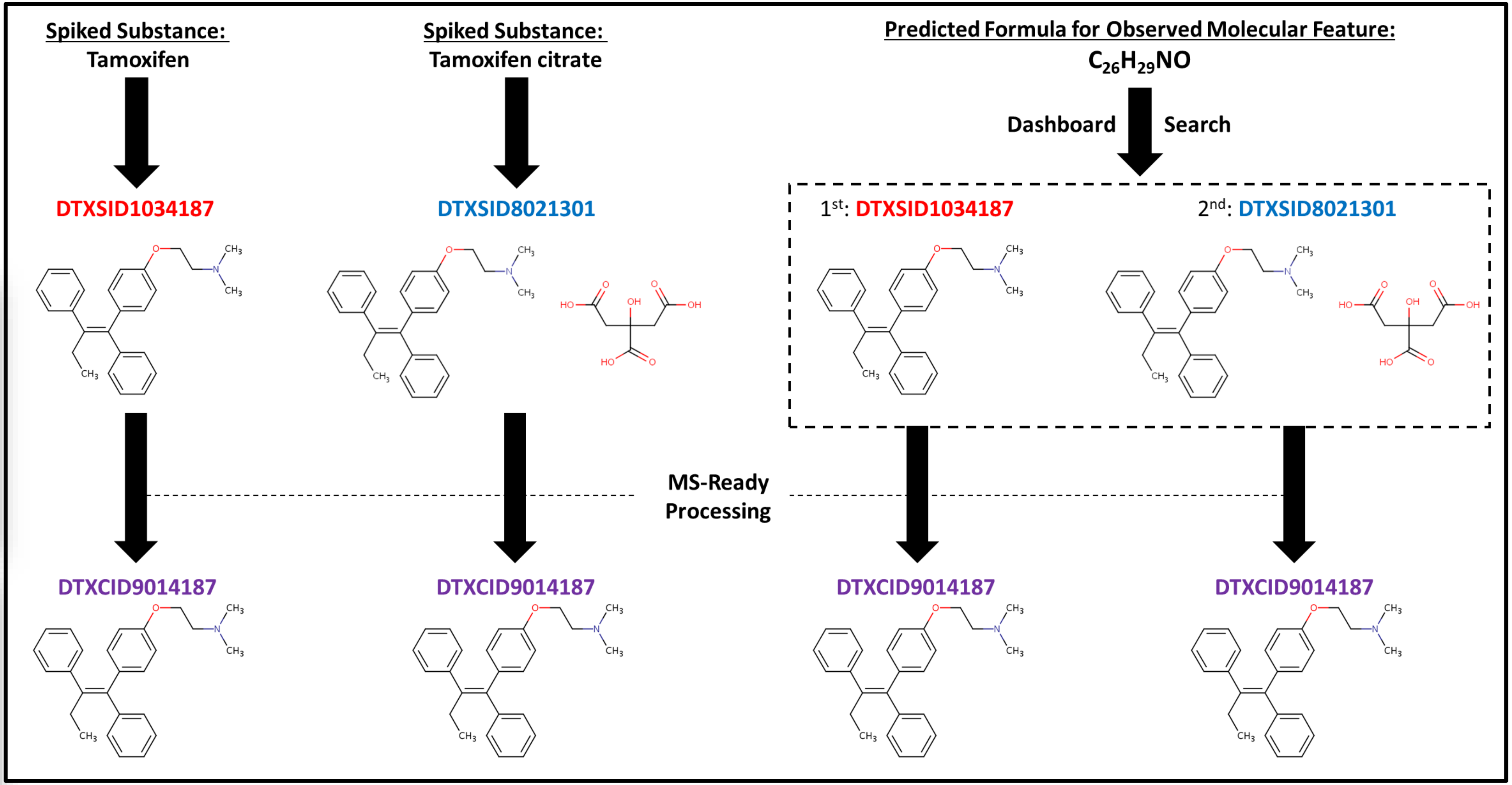
EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke[a], Antony J. Williams[a], Inthirany Thillanadarajah[b], Ann M. Richard[a,*]

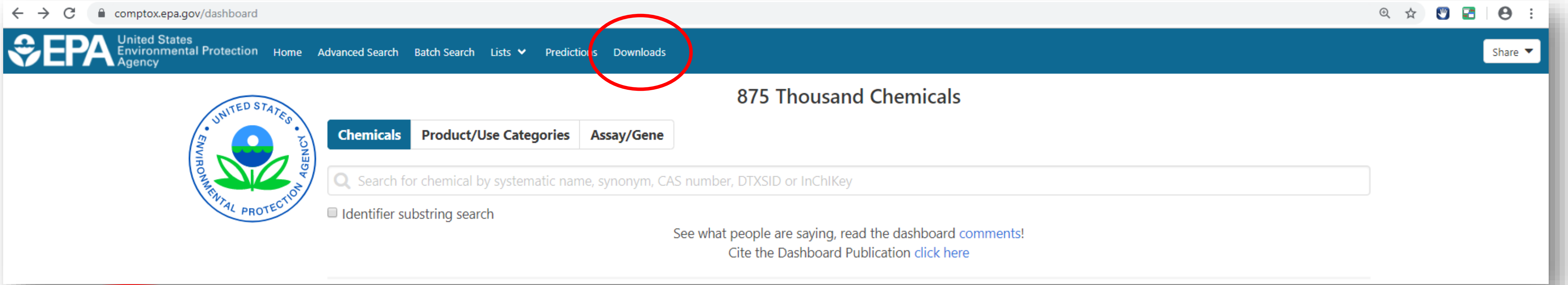
[a] National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Mail Drop D143-02, Research Triangle Park, NC 27711, USA
[b] Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

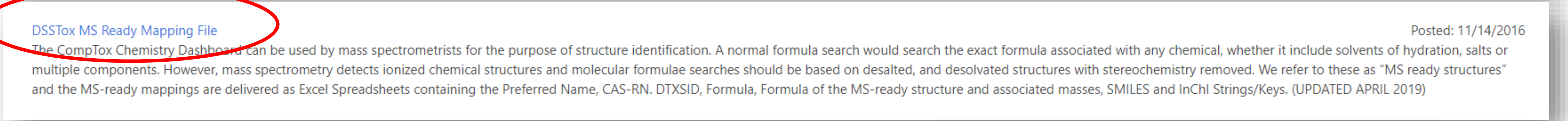
United States Environmental Protection Agency

MS-Ready Structures



Dashboard Access

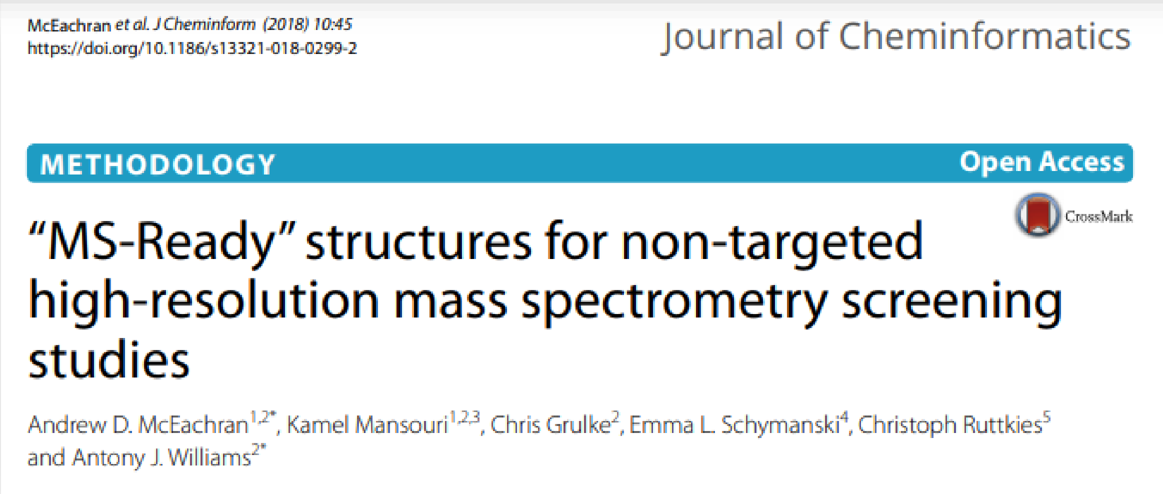




DSSTox MS Ready Mapping File

Posted: 11/14/2016

The CompTox Chemistry Dashboard can be used by mass spectrometrists for the purpose of structure identification. A normal formula search would search the exact formula associated with any chemical, whether it include solvents of hydration, salts or multiple components. However, mass spectrometry detects ionized chemical structures and molecular formulae searches should be based on desalted, and desolvated structures with stereochemistry removed. We refer to these as "MS ready structures" and the MS-ready mappings are delivered as Excel Spreadsheets containing the Preferred Name, CAS-RN. DTXSID, Formula, Formula of the MS-ready structure and associated masses, SMILES and InChI Strings/Keys. (UPDATED APRIL 2019)



McEachran *et al. J Cheminform (2018) 10:45*
https://doi.org/10.1186/s13321-018-0299-2

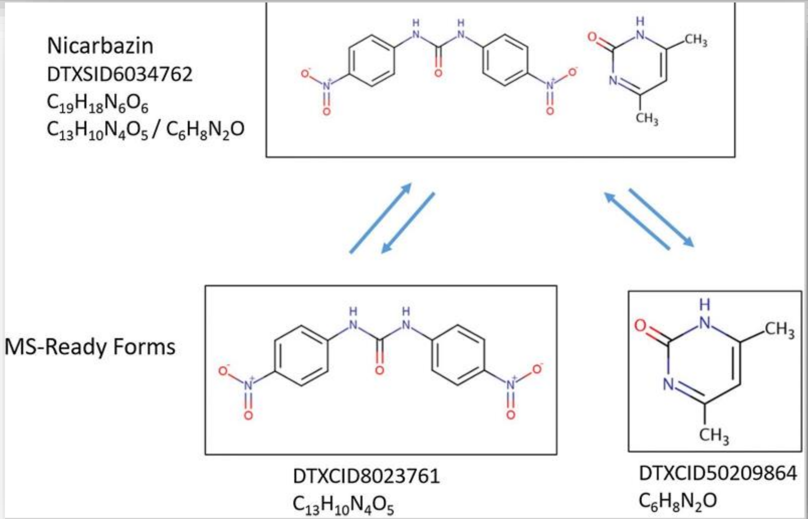
Journal of Cheminformatics

METHODOLOGY Open Access

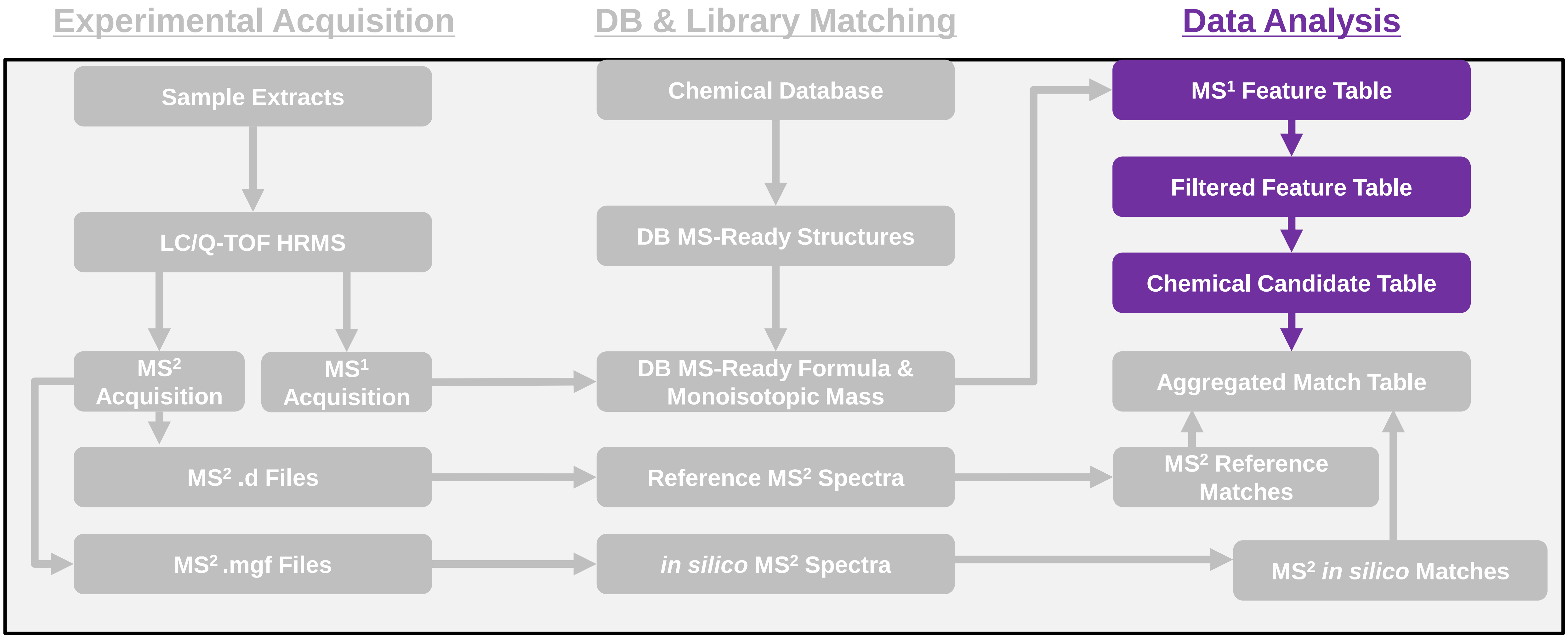
CrossMark

"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies

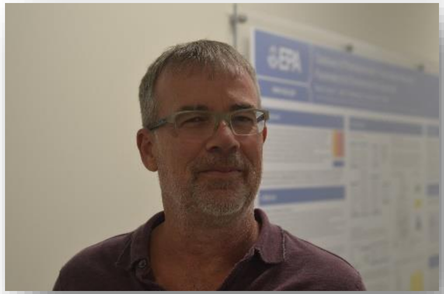
Andrew D. McEachran[1,2*], Kamel Mansouri[1,2,3], Chris Grulke[2], Emma L. Schymanski[4], Christoph Ruttkies[5] and Antony J. Williams[2*]



Agilent LC/Q-TOF Simplified Workflow



EPA NTA WebApp



Feature Removal:

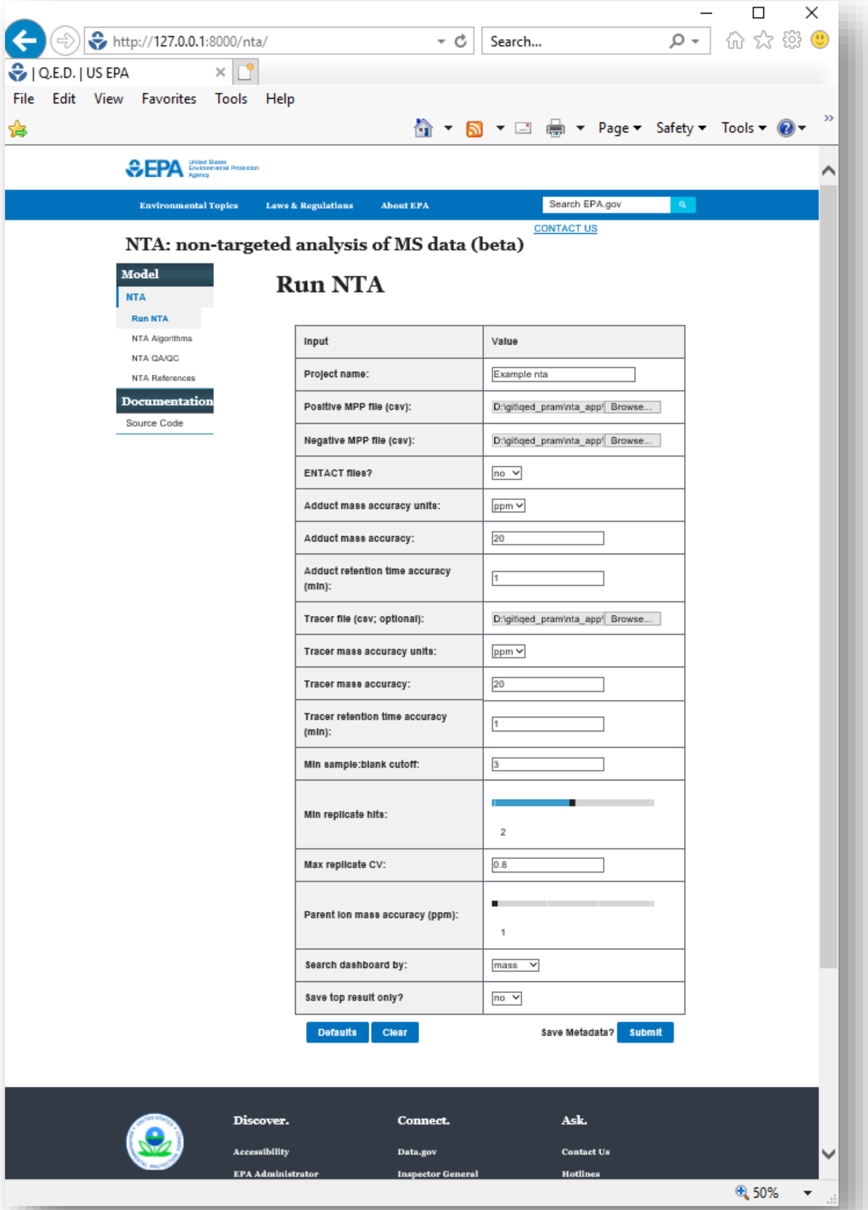
1) Duplicate features
2) Non-reproducible features
3) Blank features (sample:blank)
4) Non-responsive features (dilutions)

Feature Flagging:

1) Multi-mode hits (+ and -)
2) Meas. precision (CV threshold)
3) Formula match (score ≥ threshold)
4) Negative mass defect
5) Halogenation
6) Has/is adduct
7) Has/is neutral loss
8) Has/is multimer

Dashboard Integration:

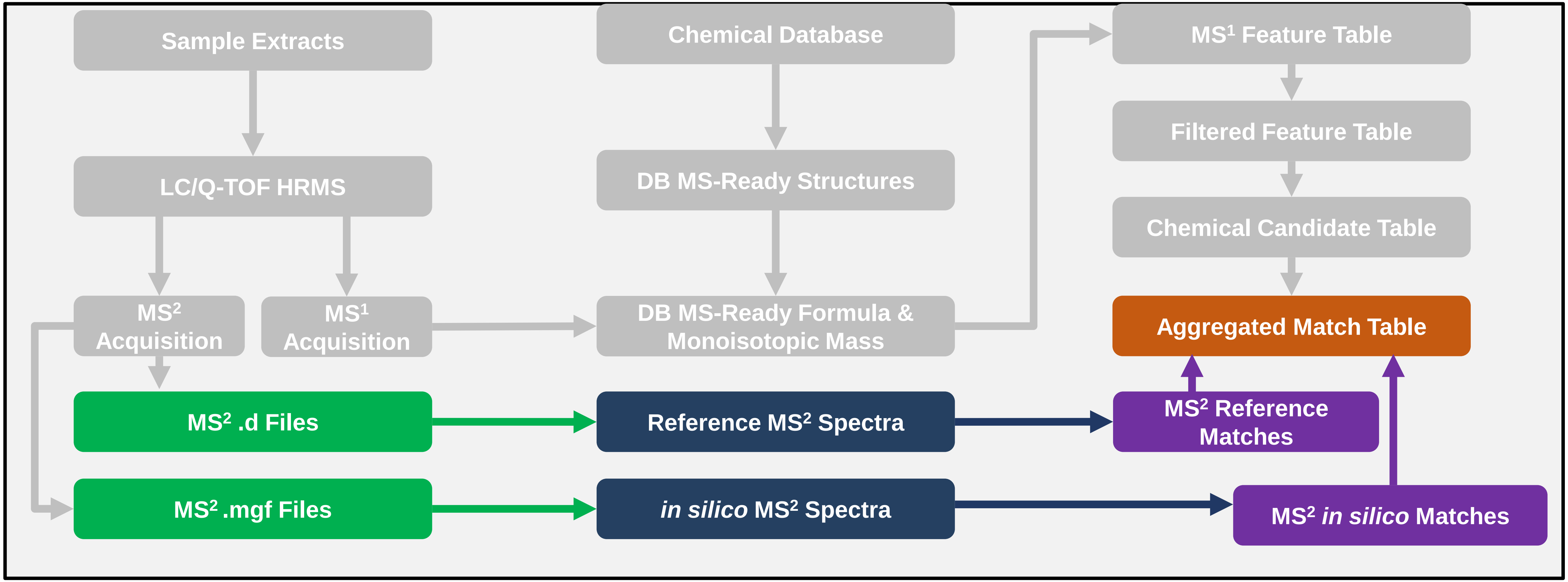
1) Data source & pub counts
2) Bioactivity & exposure levels
3) Presence on lists
4) Product & use categories



EPA United States Environmental Protection Agency

Agilent LC/Q-TOF Simplified Workflow

Experimental Acquisition | **DB & Library Matching** | **Data Analysis**



United States Environmental Protection Agency

Generation of *in silico* Spectra

CFM-ID v2.0

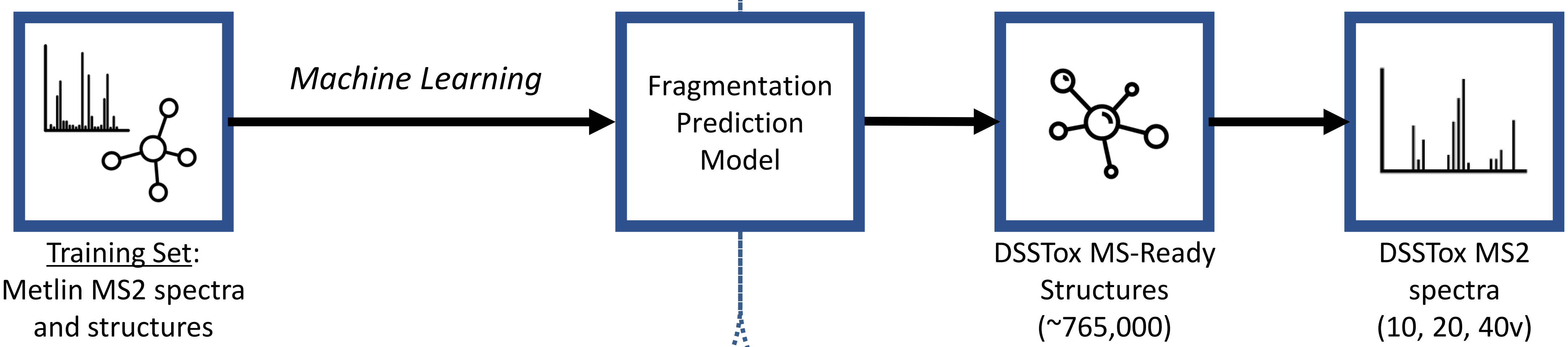
Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification

Authors   Authors and affiliations

Felicity Allen, Russ Greiner, David Wishart

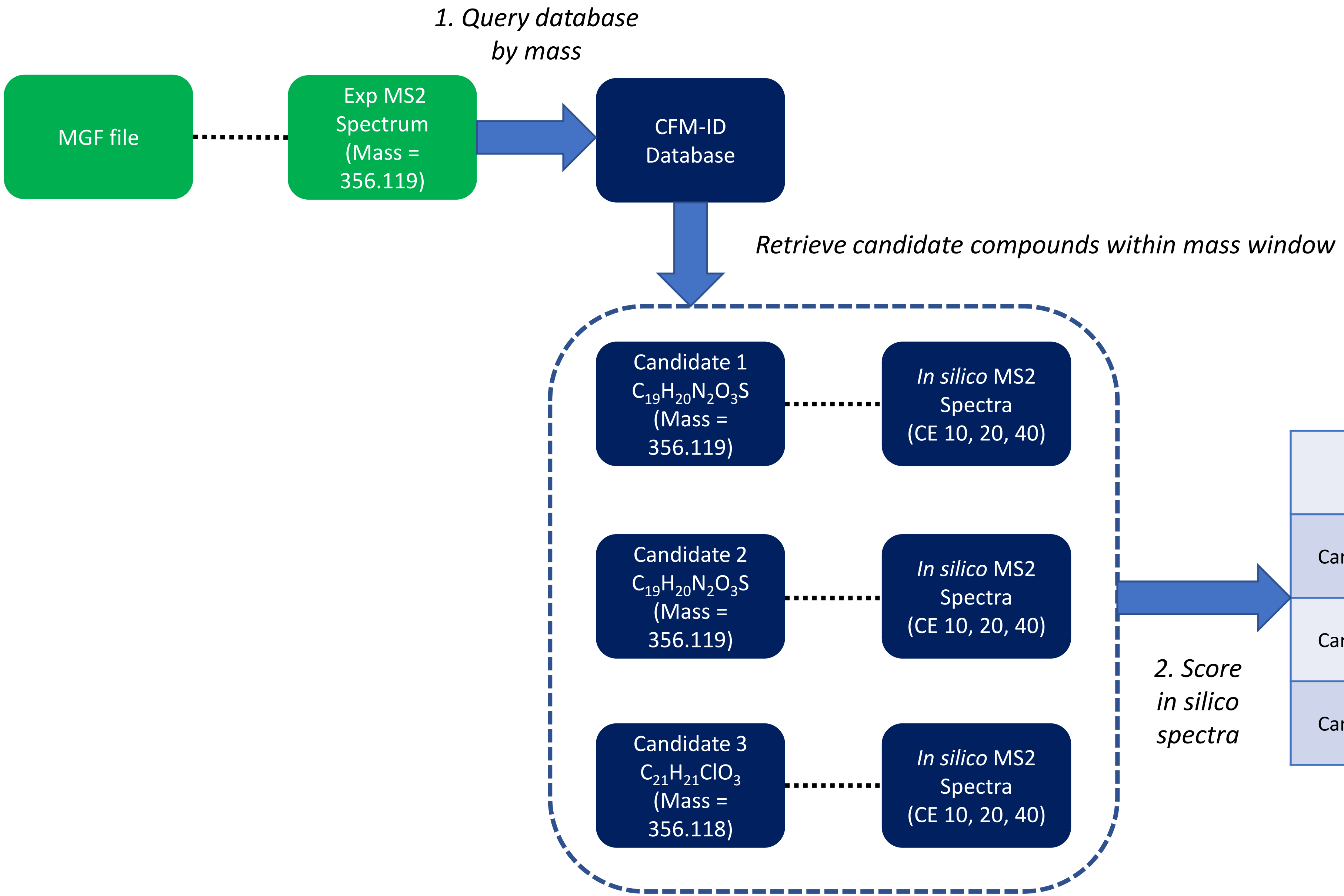
Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran, Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams



McEachran, Andrew D., et al. *Scientific data* 6.1 (2019): 1-9
Allen, Felicity, et al. *Metabolomics* 11.1 (2015): 98-110.

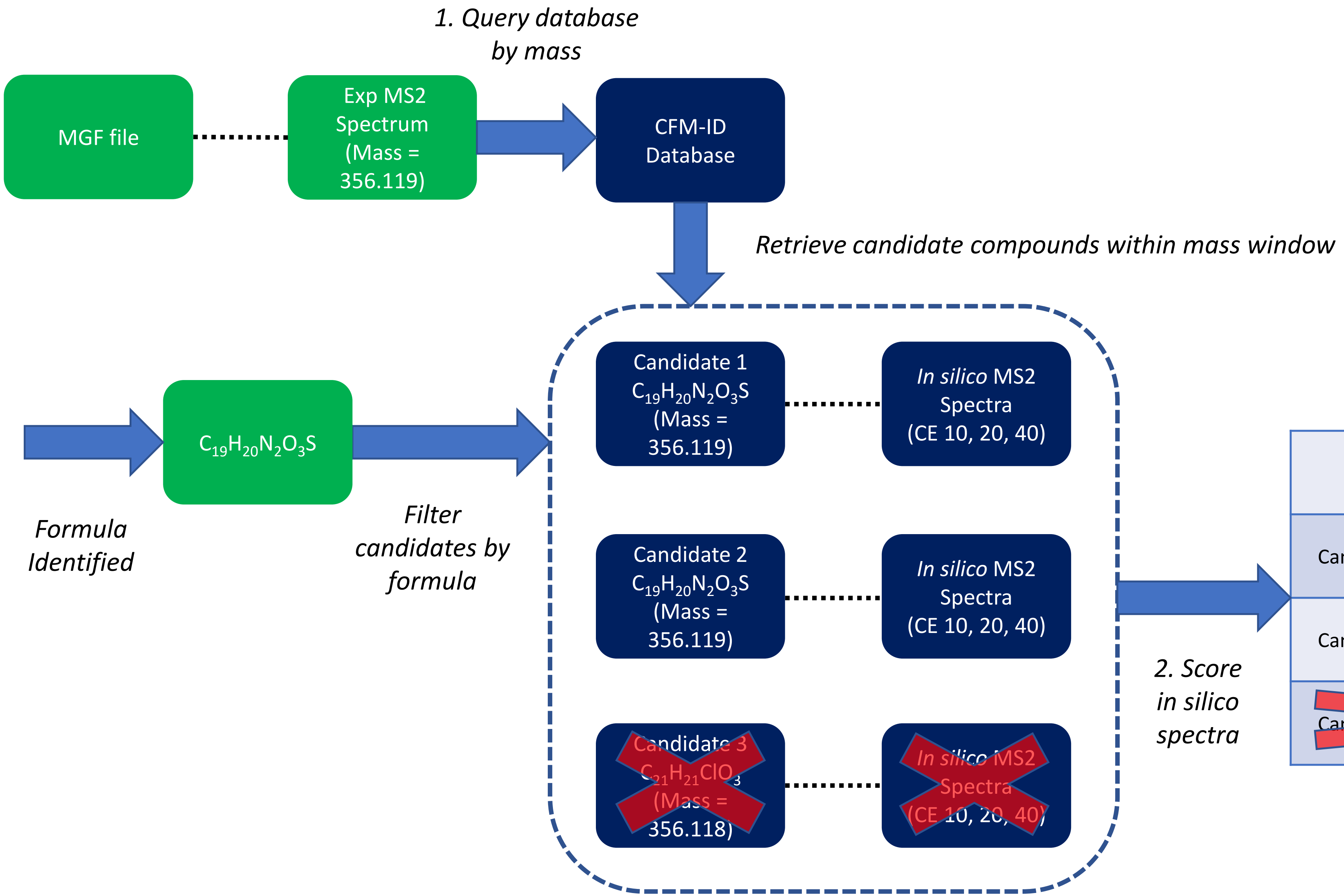
CFM-ID Database Matching



CFM-ID Scores

	in silico CE 10	*in silico* CE 20	*in silico* CE 40
Candidate 1	0.5	0.3	0.1
Candidate 2	0.2	0.1	0.02
Candidate 3	0.1	0.05	0.01

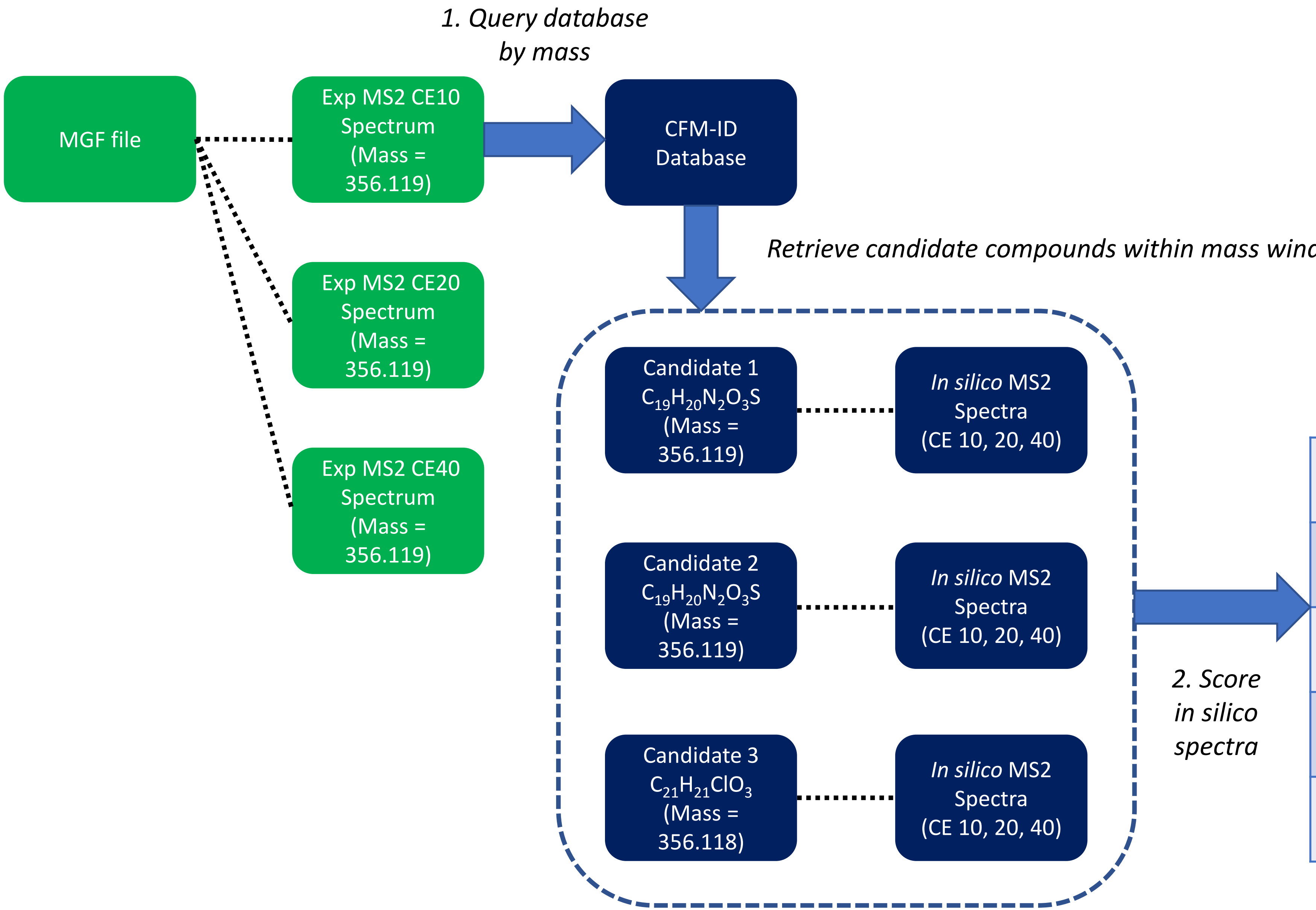
CFM-ID Database Matching (w/ Formula Information)



CFM-ID Scores

	in silico CE 10	*in silico* CE 20	*in silico* CE 40
Candidate 1	0.5	0.3	0.1
Candidate 2	0.2	0.1	0.02
Candidate 3	0.1	0.05	0.01

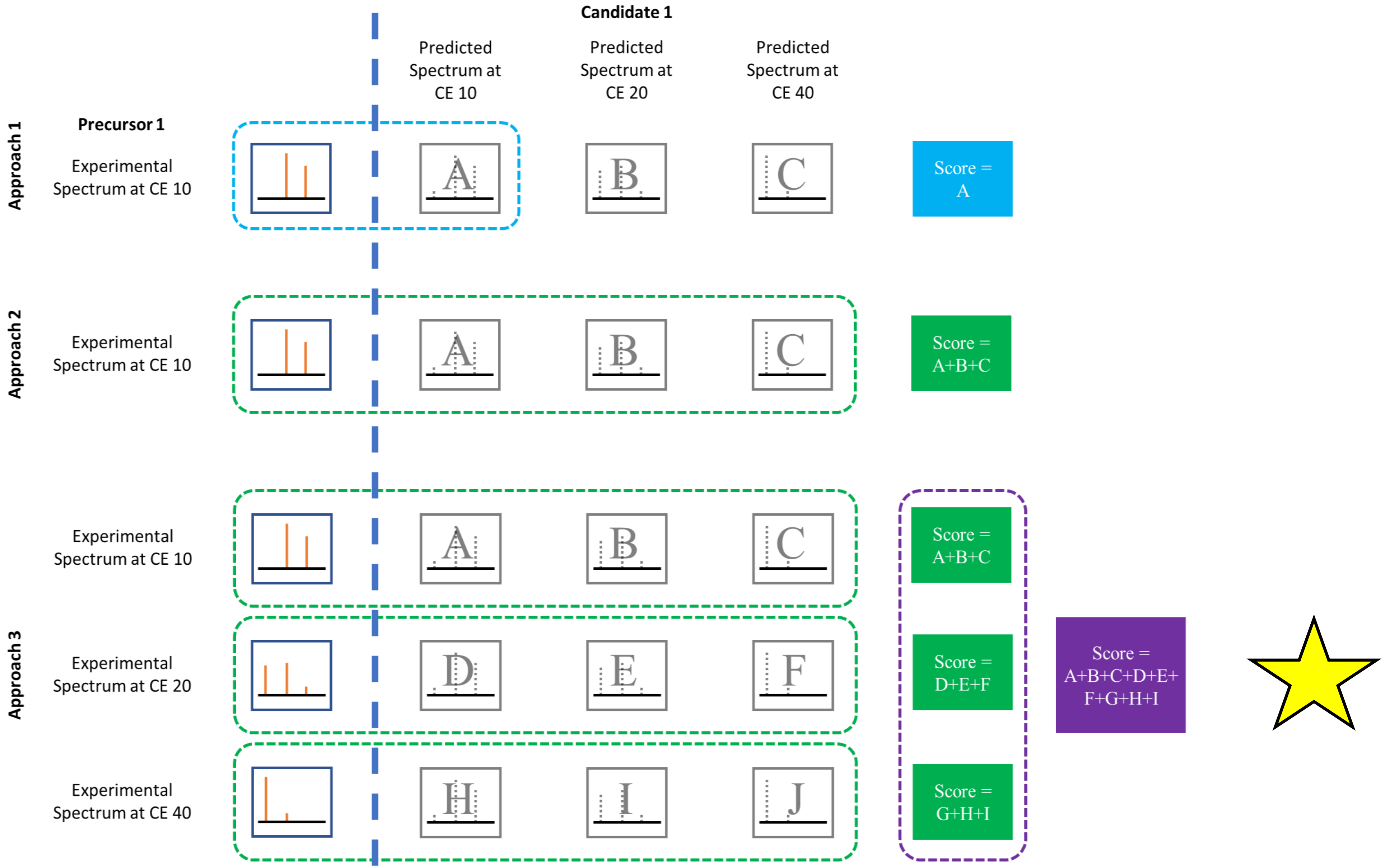
CFM-ID Database Matching (w/ Multiple $CE_{experimental}$)



CFM-ID Scores

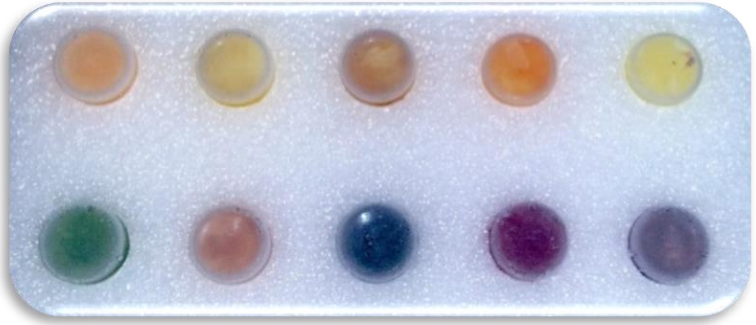
	in silico CE 10	*in silico* CE 20	*in silico* CE 40
Candidate 1 $CE_{exp} = 10$	0.5	0.3	0.1
Candidate 1 $CE_{exp} = 20$	0.4	0.5	0.12
Candidate 1 $CE_{exp} = 40$	0.05	0.1	0.2
…	…	…	…

CFM-ID Scoring Approaches



EPA'S Non-Targeted Analysis Collaborative Trial

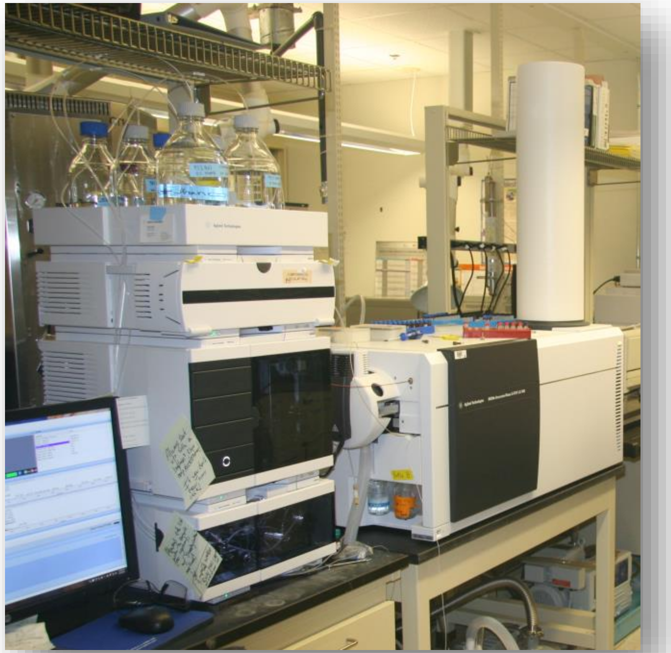
The Trial Mixtures:



10 Mixtures ranging from 95 to 365 compounds
(Total: 1,269 unique compounds)

"Pass" compounds = 377 with MS2 data

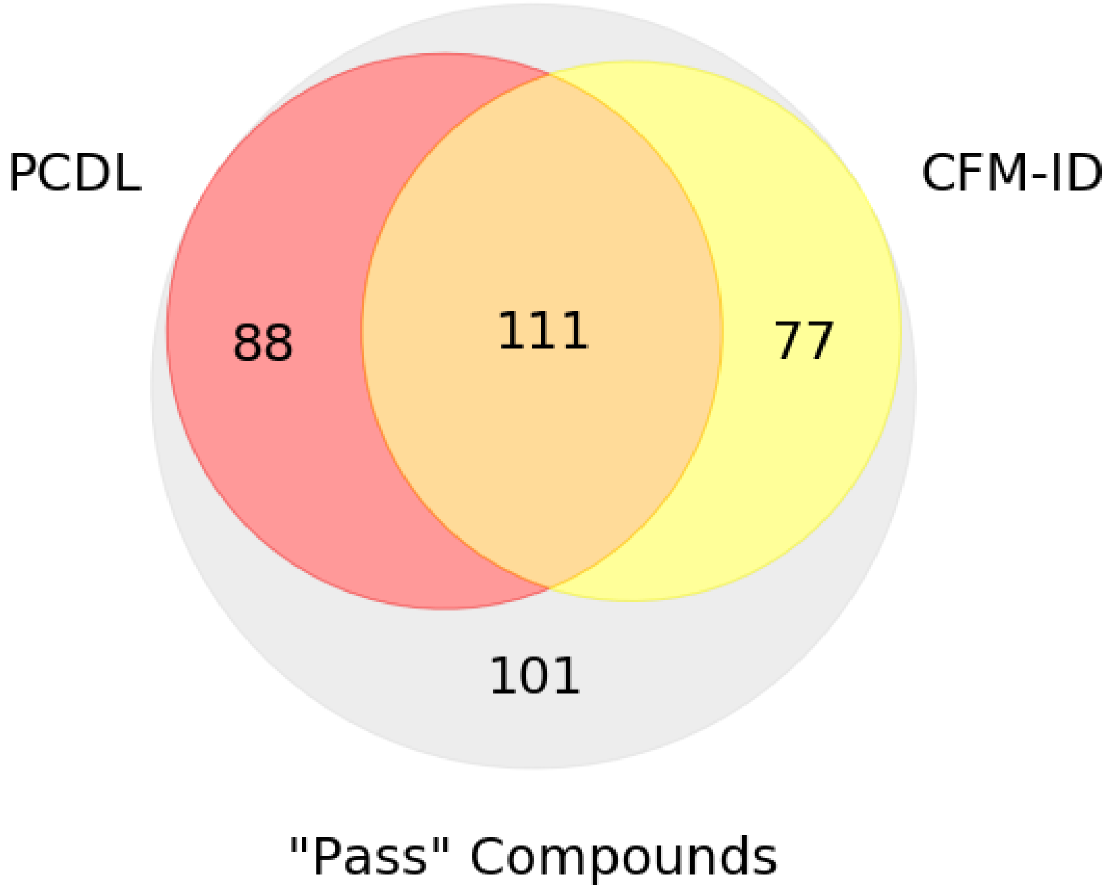
EPA Setup:



Agilent 1290 UPLC
Agilent 6530B Q-TOF with ESI source

Ulrich, Elin M., et al. *Analytical and bioanalytical chemistry* 411.4 (2019): 853-866.
Sobus, Jon R., et al. *Analytical and bioanalytical chemistry* 411.4 (2019): 835-851.

Reference vs. *in silico* Library Coverage



MS2 Library	% of "Pass" Compounds Identified
Agilent PCDL	53%
CFM-ID Top Hit	50%
PCDL and/or CFM-ID Top Hit	73%

PCDL → Agilent reference MS^2 library

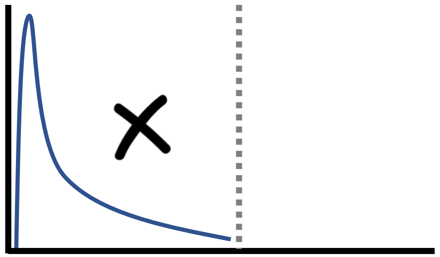
"Pass" compounds (n=377) → ENTACT chemicals observed with MS^2 data

NTA Workflows: Using CFM-ID Results as Filters

Score
Filter out candidates below score cutoff

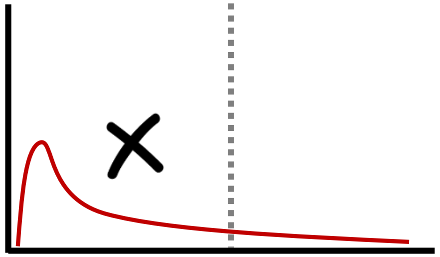
Variability in score distribution

MS2 Spectrum 1



Candidate Scores

MS2 Spectrum 2

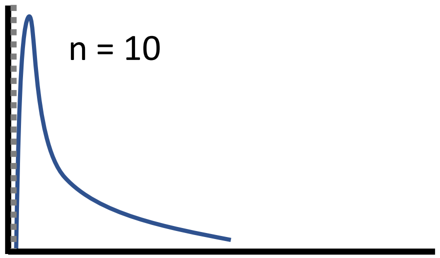


Candidate Scores

Rank
Filter out candidates above rank cutoff

Variability in number of candidate compounds

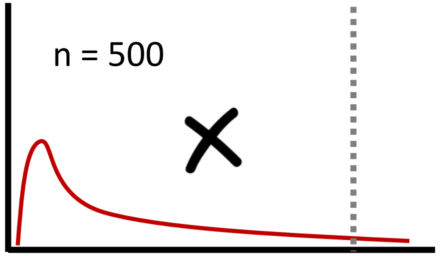
MS2 Spectrum 1



Candidate Scores

Filter by Top 20

MS2 Spectrum 2



Candidate Scores

Normalizing CFM-ID Results Values

Score Quotient
Normalize score to the highest candidate compound score

Score Percentile
Normalize rank to the number of candidate compounds

	Rank	CFM-ID Score	Maximum Score	Score Quotient	Score Percentile
Candidate Compound 1	1	0.5	0.5	1	100
Candidate Compound 2	2	0.4	0.5	0.8	80
Candidate Compound 3	3	0.39	0.5	0.78	60
Candidate Compound 4	4	0.1	0.5	0.2	40
Candidate Compound 5	5	0.05	0.5	0.1	20

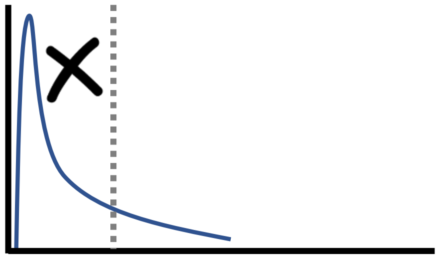
Score Quotient = Score / Maximum Score

NTA Workflows: Using CFM-ID Normalized Results as Filters

Score Quotient
Filter out candidates below score quotient cutoff

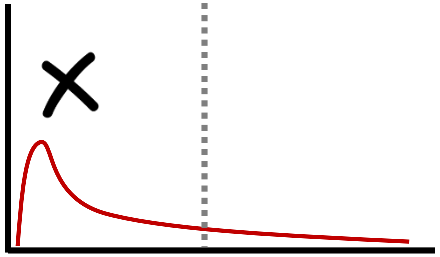
Score quotient cutoff = 0.5
Keep candidates scoring at least half of max score

MS2 Spectrum 1



Candidate Scores

MS2 Spectrum 2

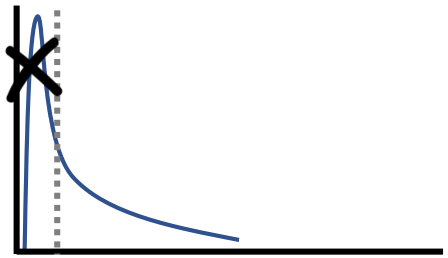


Candidate Scores

Score Percentile
Filter out candidates below percentile cutoff

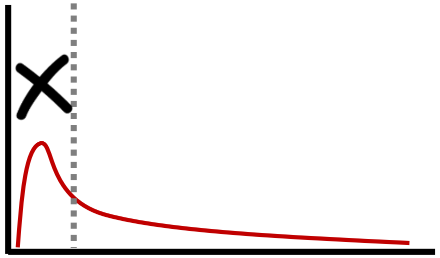
Score percentile cutoff = 0.5
Keep the top 50% of candidates

MS2 Spectrum 1



Candidate Scores

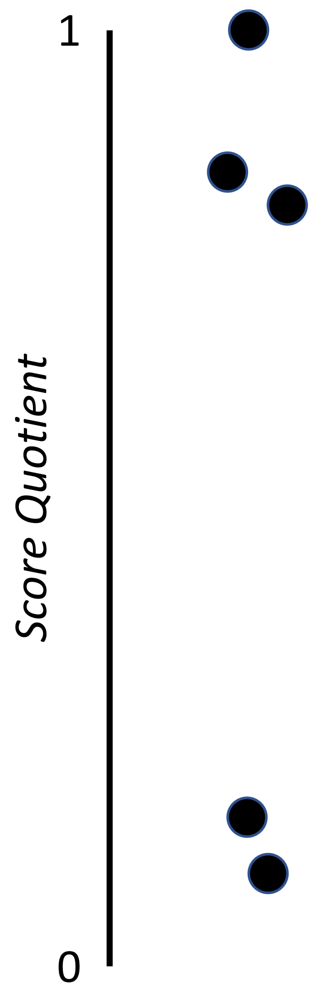
MS2 Spectrum 2



Candidate Scores

Applying Cut-off Filters to Data

	CFM-ID Score	Maximum Score	Score Quotient
Candidate Compound 1	0.5	0.5	1
Candidate Compound 2	0.4	0.5	0.8
Candidate Compound 3	0.39	0.5	0.78
Candidate Compound 4	0.1	0.5	0.2
Candidate Compound 5	0.05	0.5	0.1



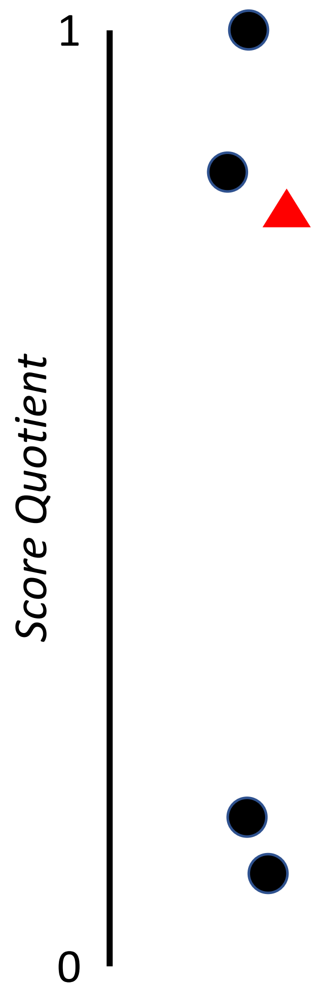
Applying Cut-off Filters to Data

	CFM-ID Score	Maximum Score	Score Quotient
Candidate Compound 1	0.5	0.5	1
Candidate Compound 2	0.4	0.5	0.8
Candidate Compound 3	0.39	0.5	0.78
Candidate Compound 4	0.1	0.5	0.2
Candidate Compound 5	0.05	0.5	0.1

▲ True Compound

● Other Candidate Compounds

True Positives	
False Negatives	
True Negatives	
False Positives	



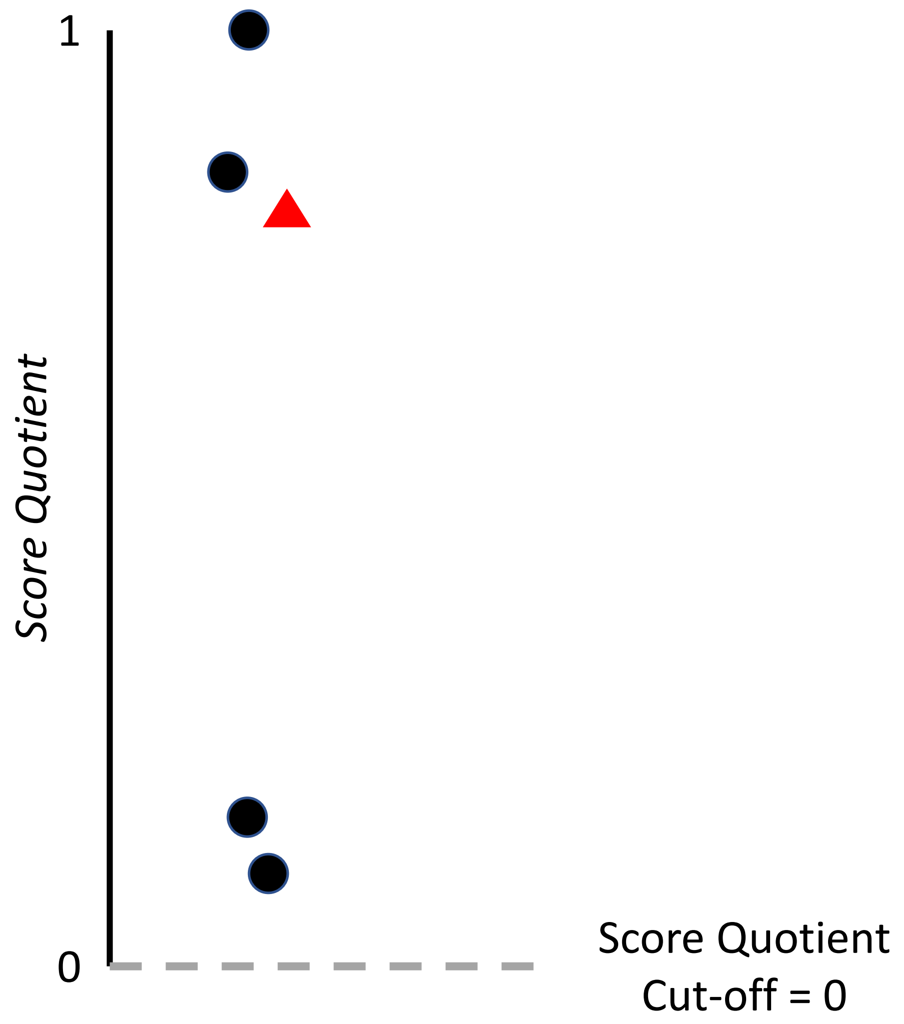
Applying Cut-off Filters to Data

	CFM-ID Score	Maximum Score	Score Quotient
Candidate Compound 1	0.5	0.5	1
Candidate Compound 2	0.4	0.5	0.8
Candidate Compound 3	0.39	0.5	0.78
Candidate Compound 4	0.1	0.5	0.2
Candidate Compound 5	0.05	0.5	0.1

True Compound

Other Candidate Compounds

True Positives	1
False Negatives	0
True Negatives	0
False Positives	4



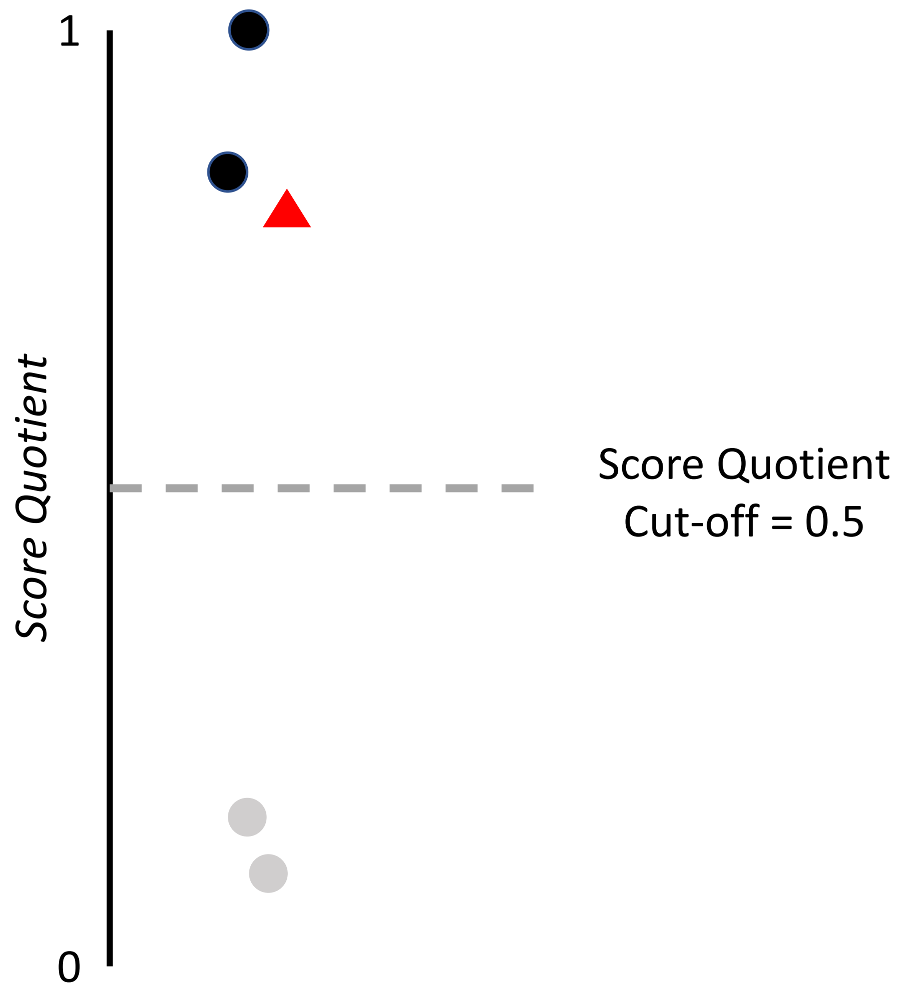
Applying Cut-off Filters to Data

	CFM-ID Score	Maximum Score	Score Quotient
Candidate Compound 1	0.5	0.5	1
Candidate Compound 2	0.4	0.5	0.8
Candidate Compound 3	0.39	0.5	0.78
Candidate Compound 4	0.1	0.5	0.2
Candidate Compound 5	0.05	0.5	0.1

▲ True Compound

● Other Candidate Compounds

True Positives	1
False Negatives	0
True Negatives	2
False Positives	2



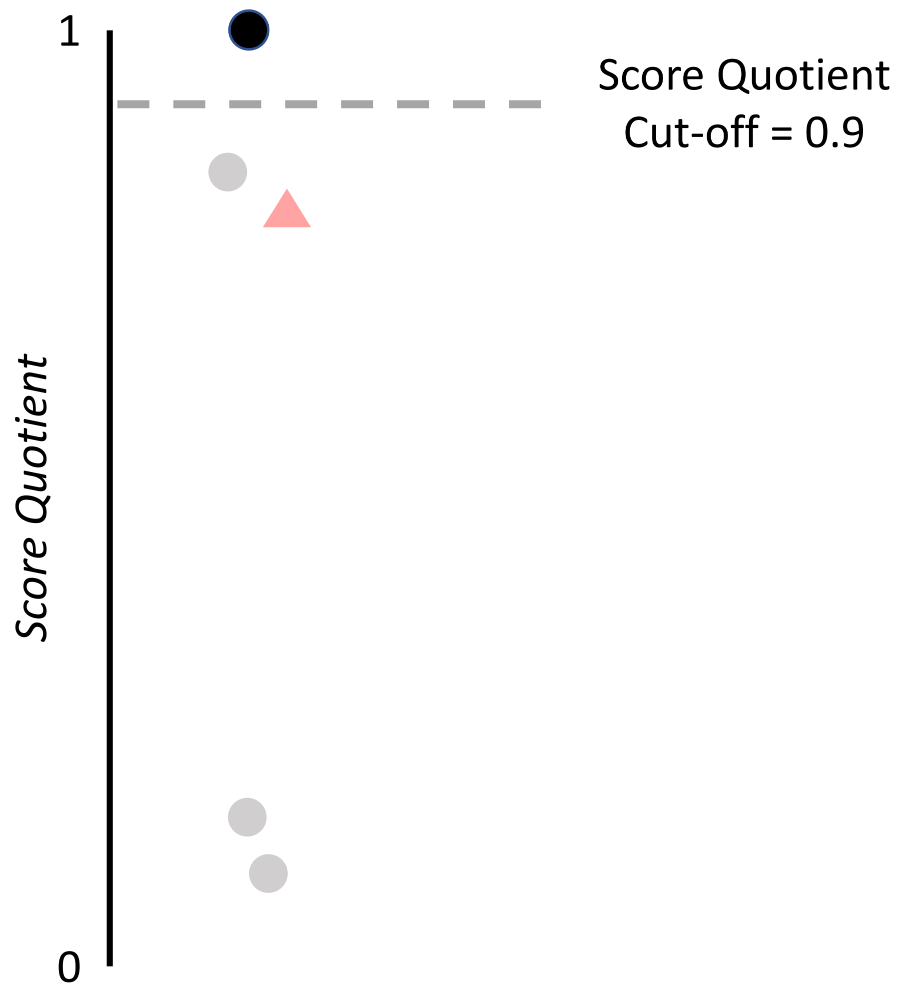
Applying Cut-off Filters to Data

	CFM-ID Score	Maximum Score	Score Quotient
Candidate Compound 1	0.5	0.5	1
Candidate Compound 2	0.4	0.5	0.8
Candidate Compound 3	0.39	0.5	0.78
Candidate Compound 4	0.1	0.5	0.2
Candidate Compound 5	0.05	0.5	0.1

True Compound

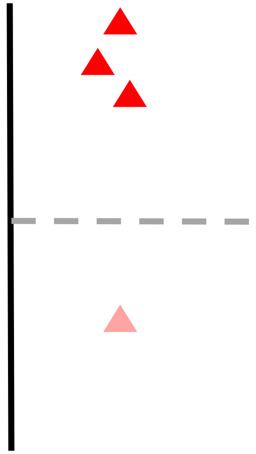
Other Candidate Compounds

True Positives	0
False Negatives	1
True Negatives	3
False Positives	1



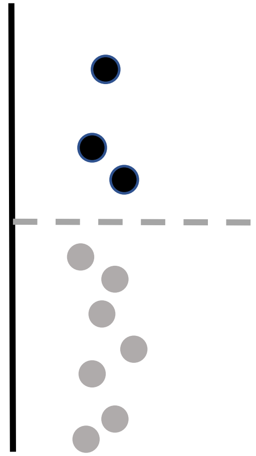
Balancing Cut-offs

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP + FN}$$



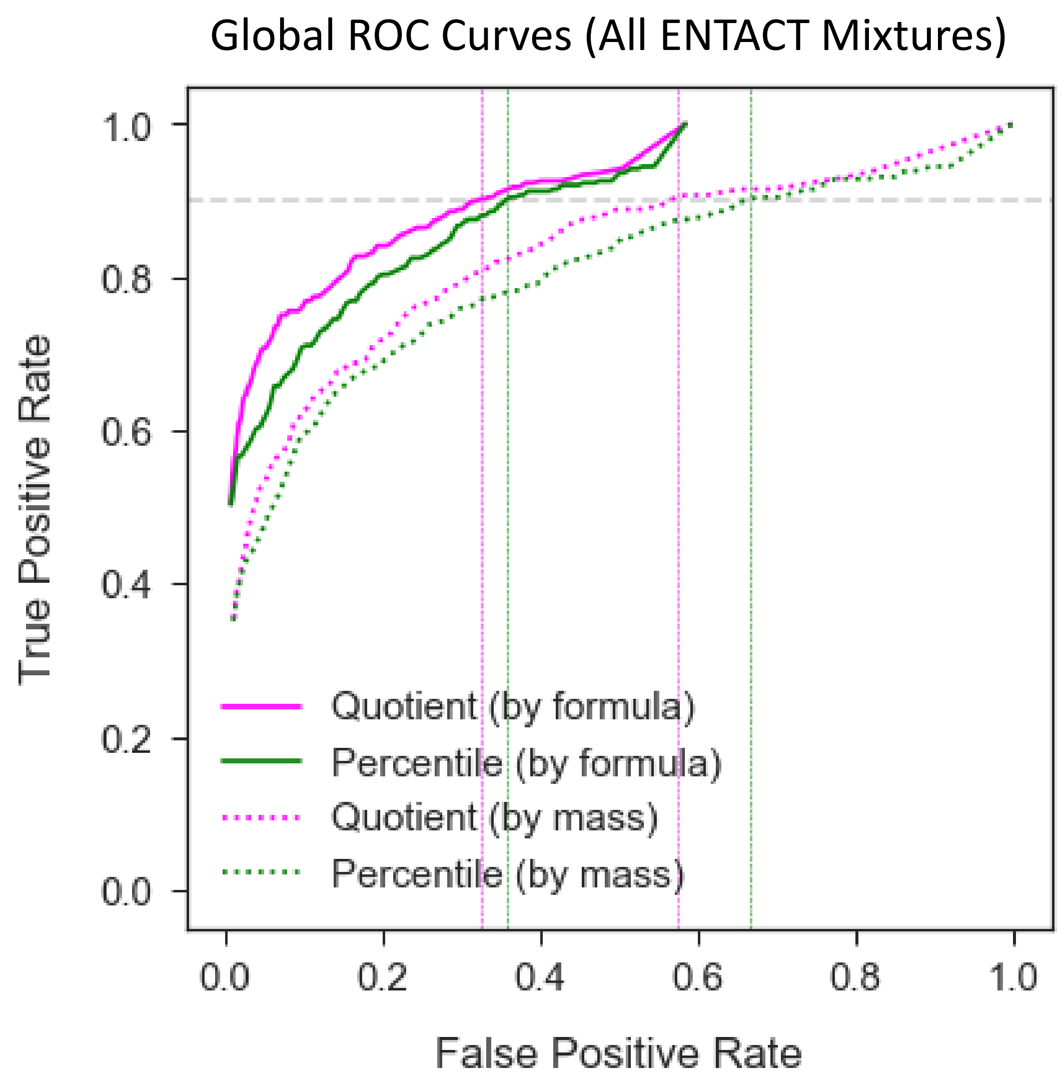
How many of the true compounds are we keeping?

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN}$$

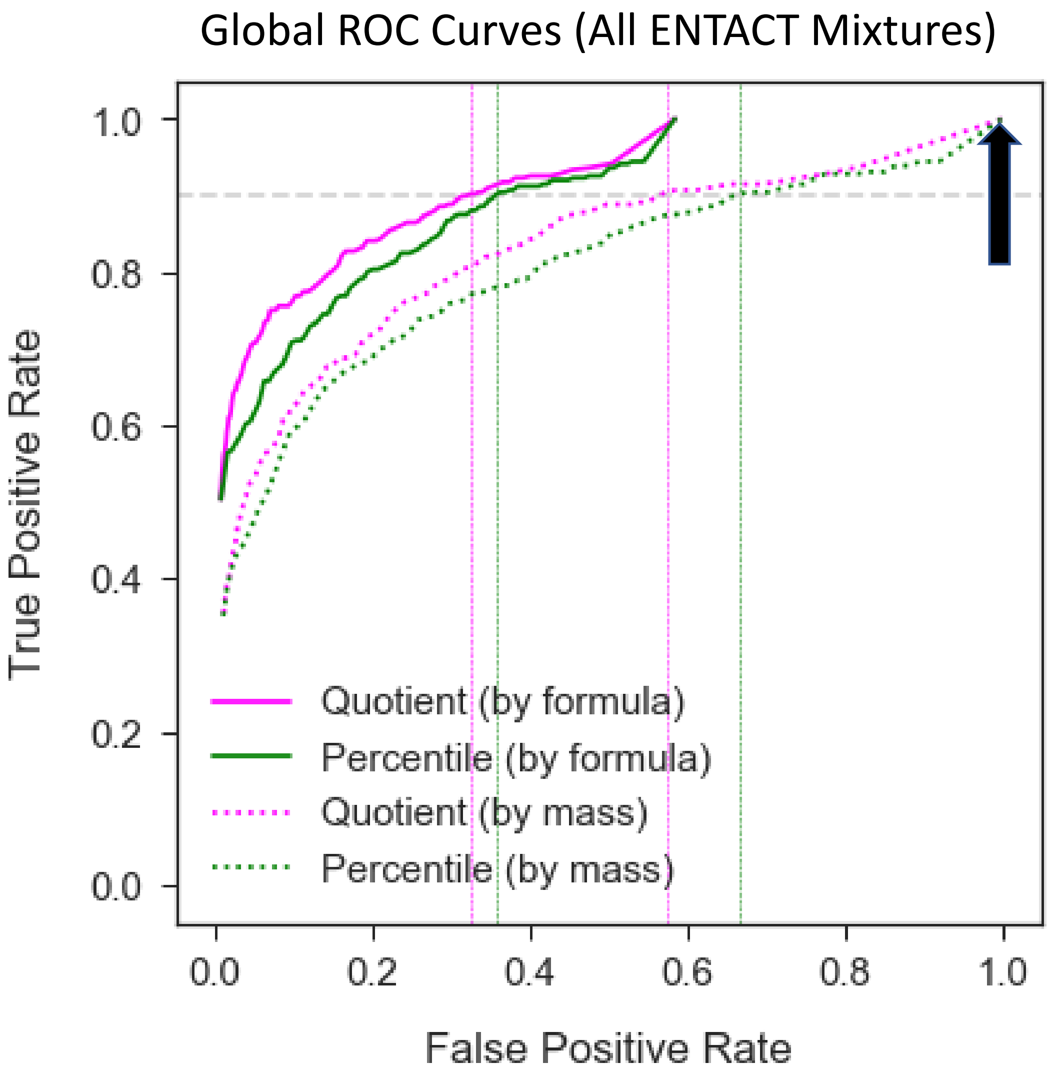


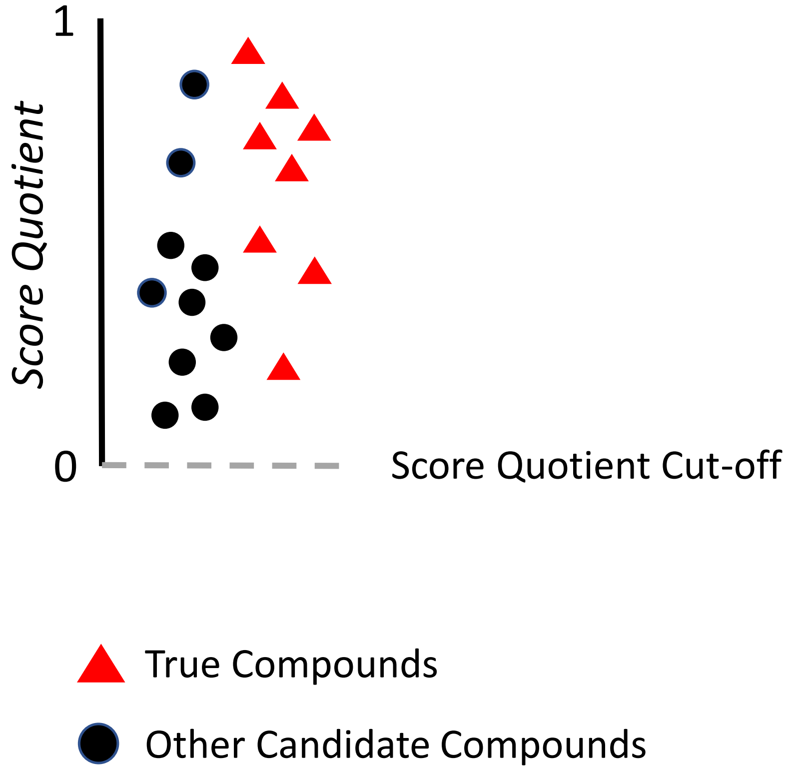
How much of the junk are we getting rid of?

Quotient Vs. Percentile Cutoffs

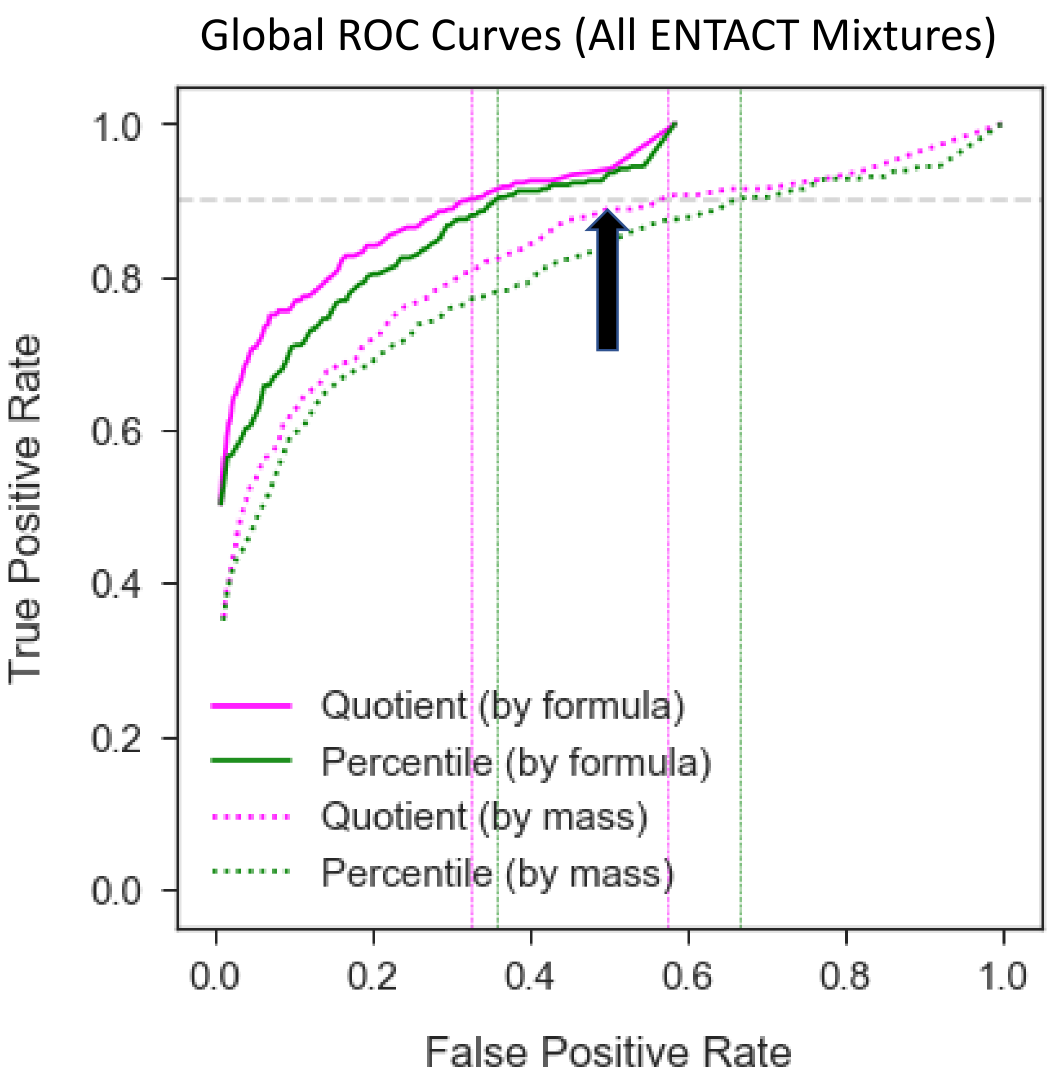


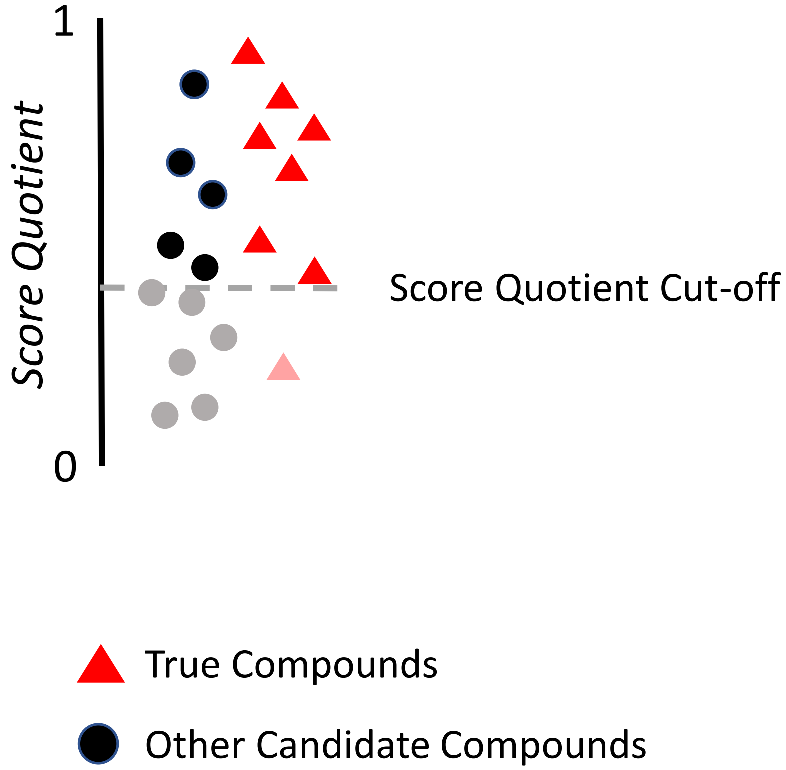
Quotient Vs. Percentile Cutoffs



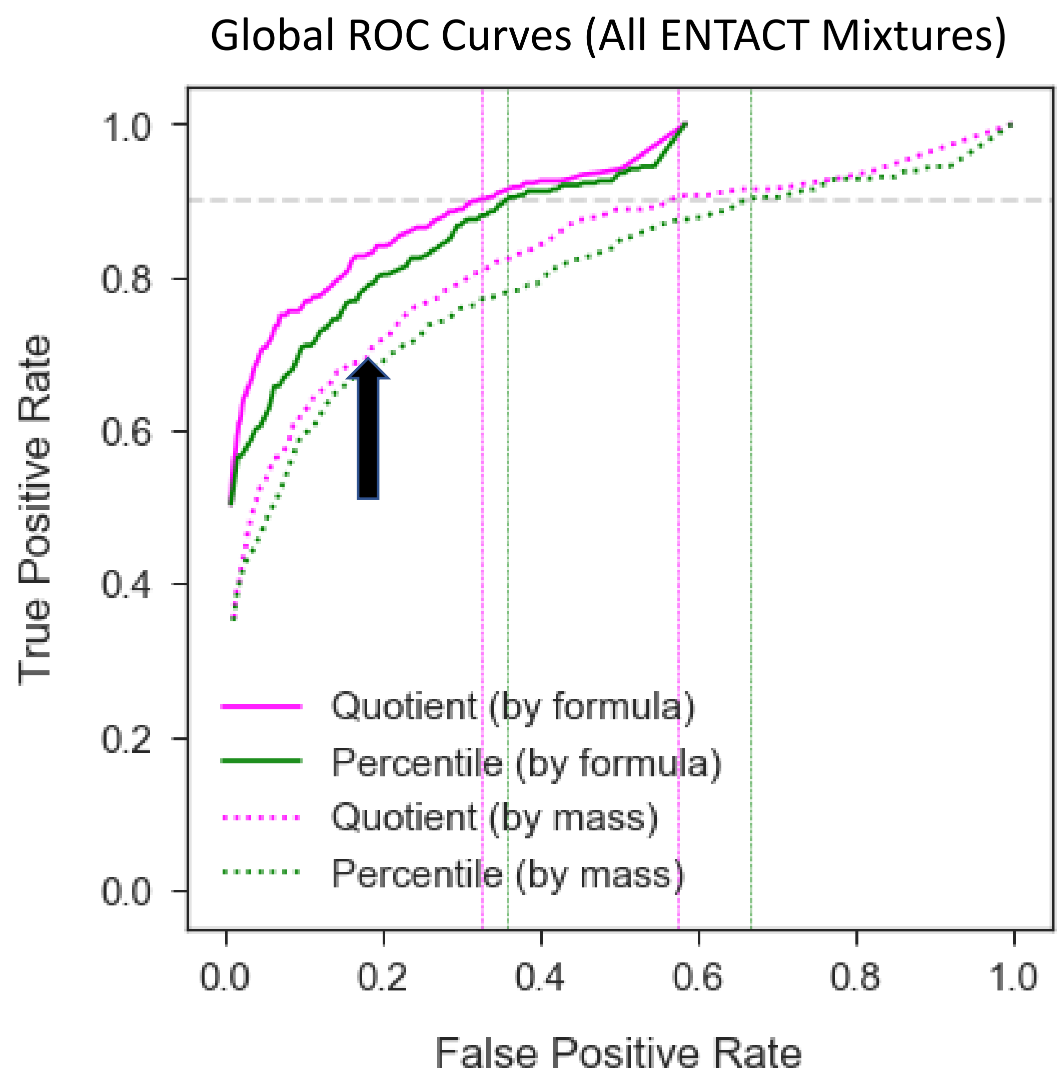


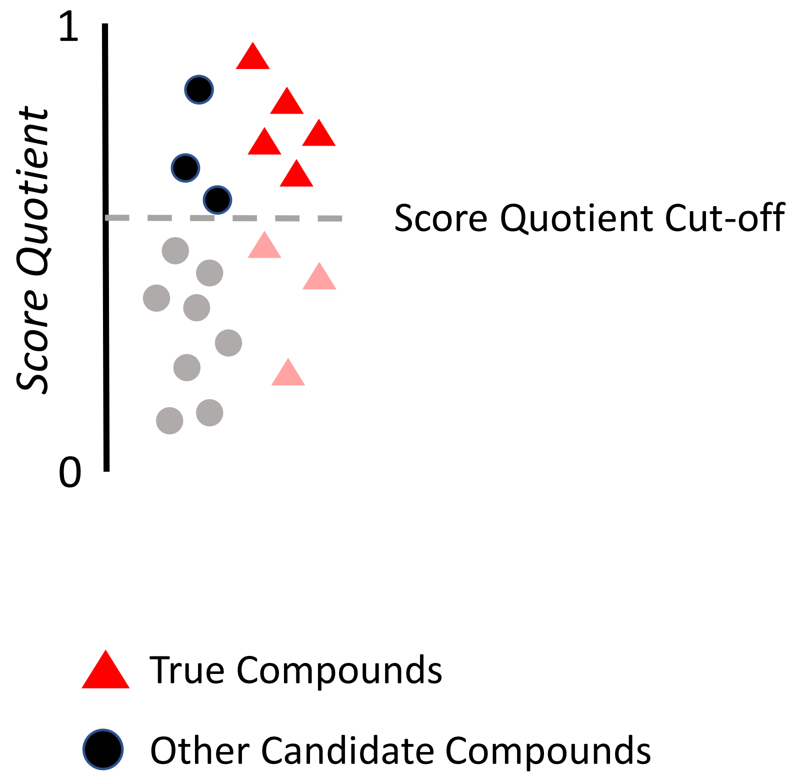
Quotient Vs. Percentile Cutoffs



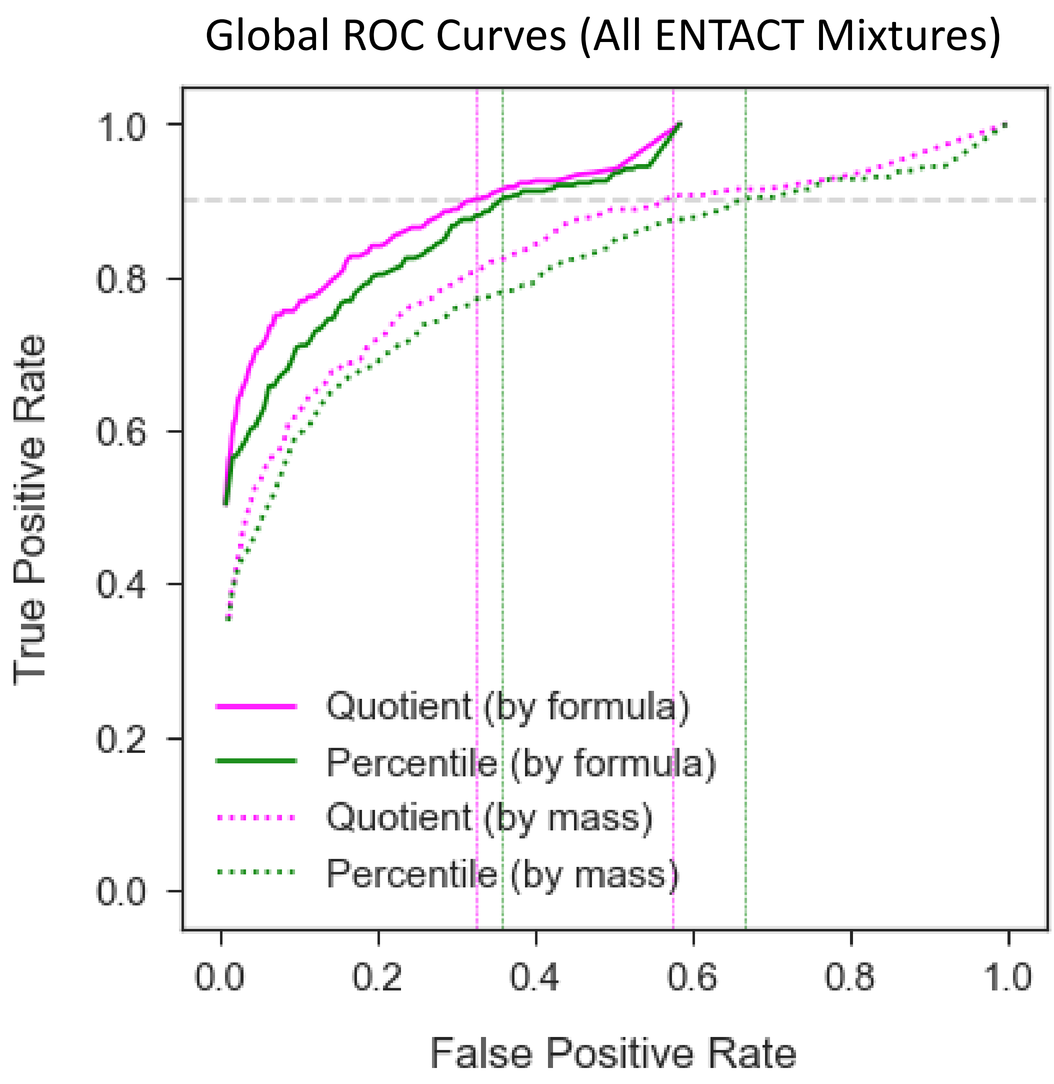


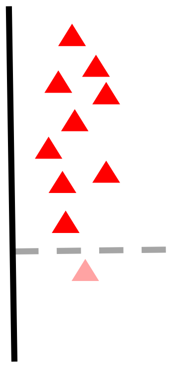
Quotient Vs. Percentile Cutoffs





Quotient Vs. Percentile Cutoffs



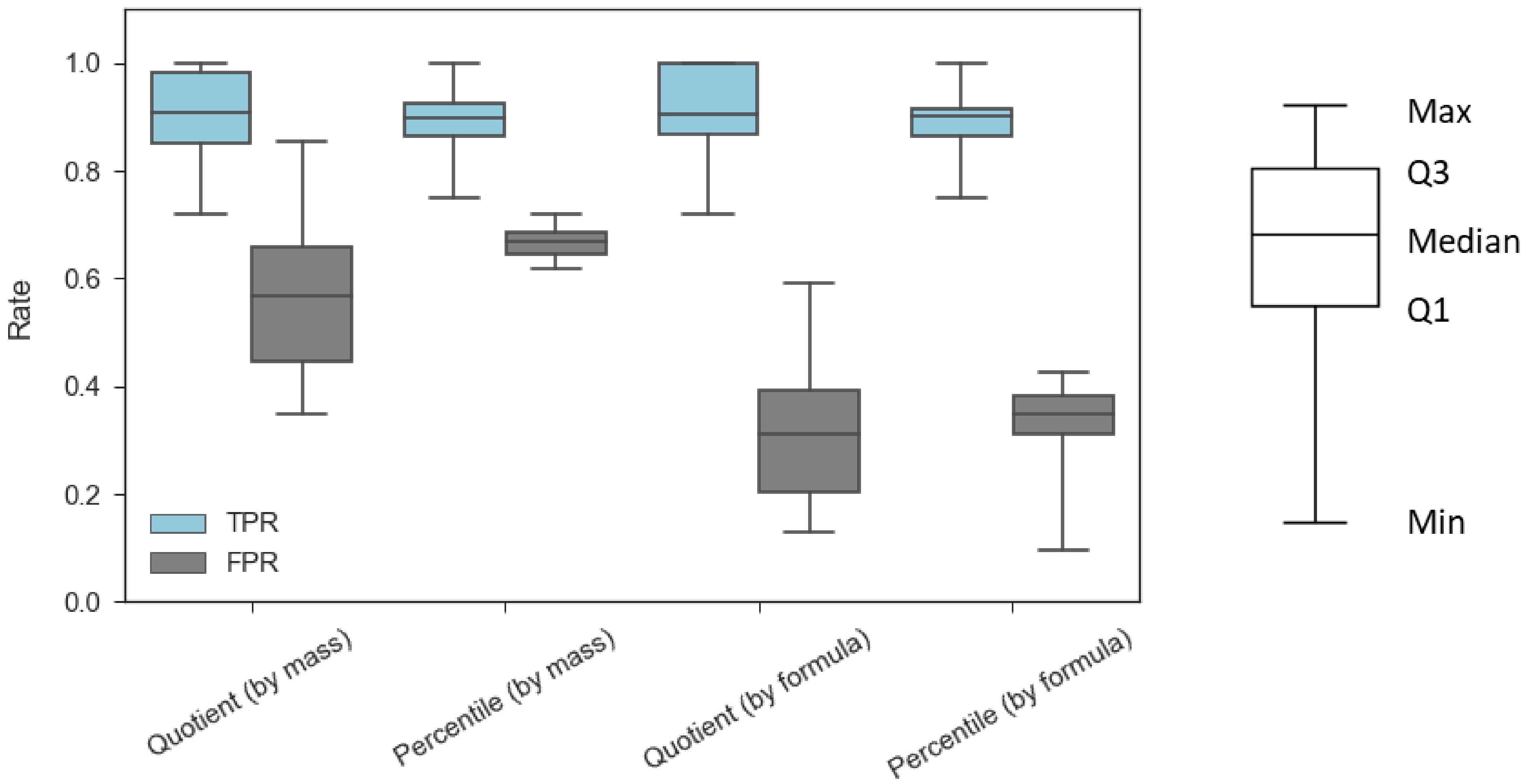


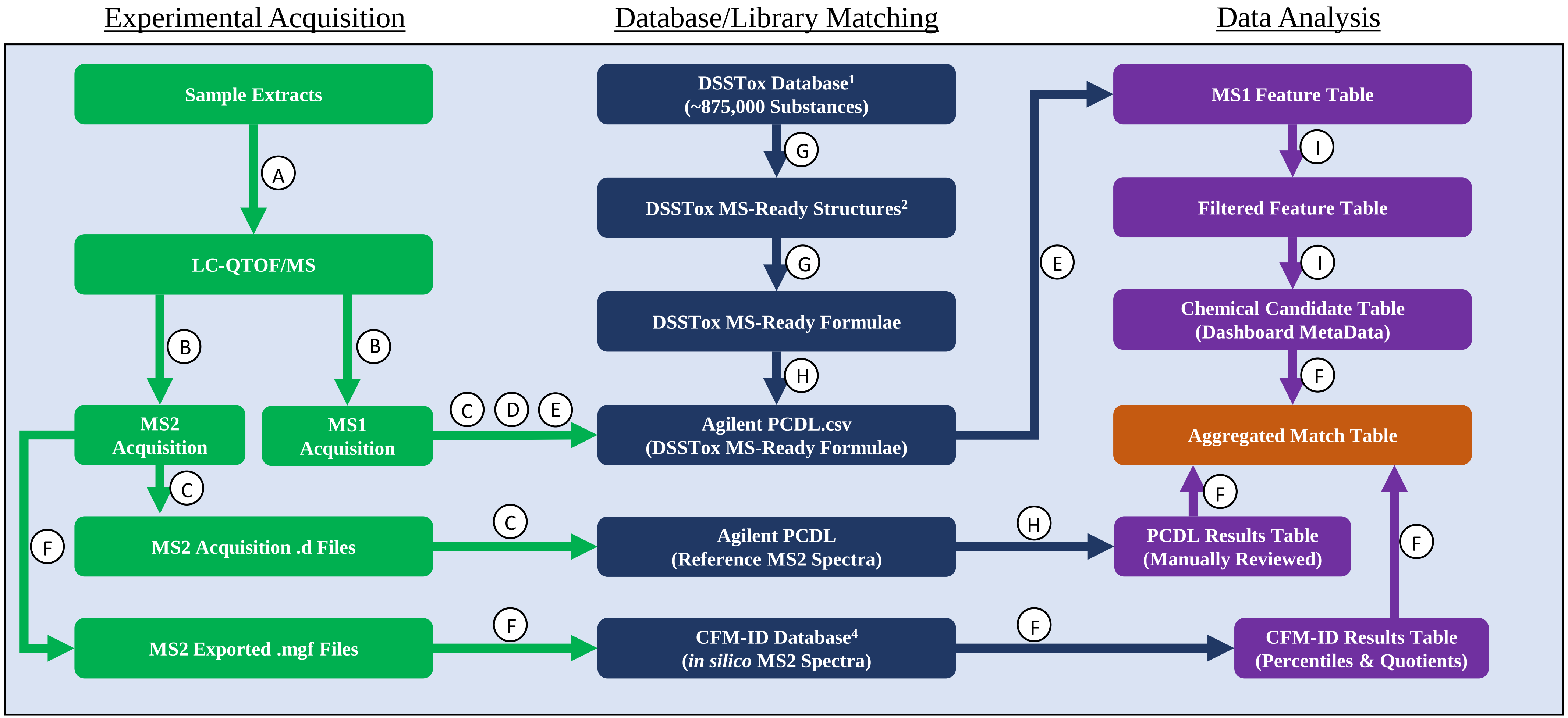
Cut-off Values for Global TPR = 0.9

	Cut-off value
Quotient (by formula)	0.18
Percentile (by formula)	38
Quotient (by mass)	0.13
Percentile (by mass)	32

Apply to individual ENTACT mixtures

CFM-ID Cut-off Filtering: Individual ENTACT Mixtures





Software & Tools

A) Excel Macro (naming & randomization)
B) Agilent MassHunter Data Acquisition
C) Agilent MassHunter Qualitative Analysis
D) Agilent Profinder (peak picking & alignment)
E) Agilent Mass Profiler Professional (formula matching)
F) Python Script
G) CompTox Chemicals Dashboard[3]
H) Excel
I) EPA NTA WebApp

General Examples

Sobus et al. https://link.springer.com/article/10.1007%2Fs00216-018-1526-4
Newton et al. https://www.sciencedirect.com/science/article/pii/S026974911732691X?via%3Dihub
Hedgespeth et al. https://www.sciencedirect.com/science/article/pii/S004896971933298X?via%3Dihub

Specific References

[1]Grulke et al. https://www.sciencedirect.com/science/article/pii/S2468111319300234
[2]McEachran et al. https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0299-2
[3]Williams et al. https://jcheminf.biomedcentral.com/articles/10.1186/s13321-017-0247-6
[4]McEachran et al. https://www.nature.com/articles/s41597-019-0145-z

Take-away Messages

- EPA/ORD NTA activities:
  - Focused on applications
    - qualitative (to date)→ semi-quantitative (soon)
    - must support HT exposure prediction & risk evaluation
  - R&D required to support applications
    - Experimental + cheminformatic + computational efforts = Viable NTA program
  - Growing capacity with new instrumentation
    - Requires flexible workflows
      - Work smarter, not harder
      - Don't reinvent the wheel
      - Build once, use many (A. Williams)

United States Environmental Protection Agency

Contributing Researchers

This work was supported, in part, by ORD's Pathfinder Innovation Program (PIP) and an ORD EMVL award

EPA ORD
Hussein Al-Ghoul*
Alex Chao*
Jarod Grossman*
Kristin Isaacs
Sarah Laughlin*
Charles Lowe
James McCord
Jeff Minucci
Seth Newton
Katherine Phillips
Tom Purucker
Randolph Singh*
Mark Strynar
Elin Ulrich

* = ORISE/ORAU

EPA ORD (cont.)
Chris Grulke
Kamel Mansouri*
Andrew McEachran*
Ann Richard
John Wambaugh
Antony Williams

Agilent
Jarod Grossman
Andrew McEachran

GDIT
Ilya Balabin
Tom Transue
Tommy Cathey

Questions?

sobus.jon@epa.gov

The views expressed in this presentation are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.