US EPA: 2018 Non-Targeted Analysis Collaborative Research Trial (ENTACT) Workshop



Analytical QC Challenges in EPA's Computational Toxicology Program: *ToxCast data to environmental samples*

> Ann Richard¹, Chris Grulke¹, Andrew McEachran², Ryan Lougee², Antony Williams¹ ¹EPA, National Center for Computational Toxicology, RTP, NC, USA ²ORISE Fellow, Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA

> > August 13-15, 2018 Research Triangle Park, NC

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

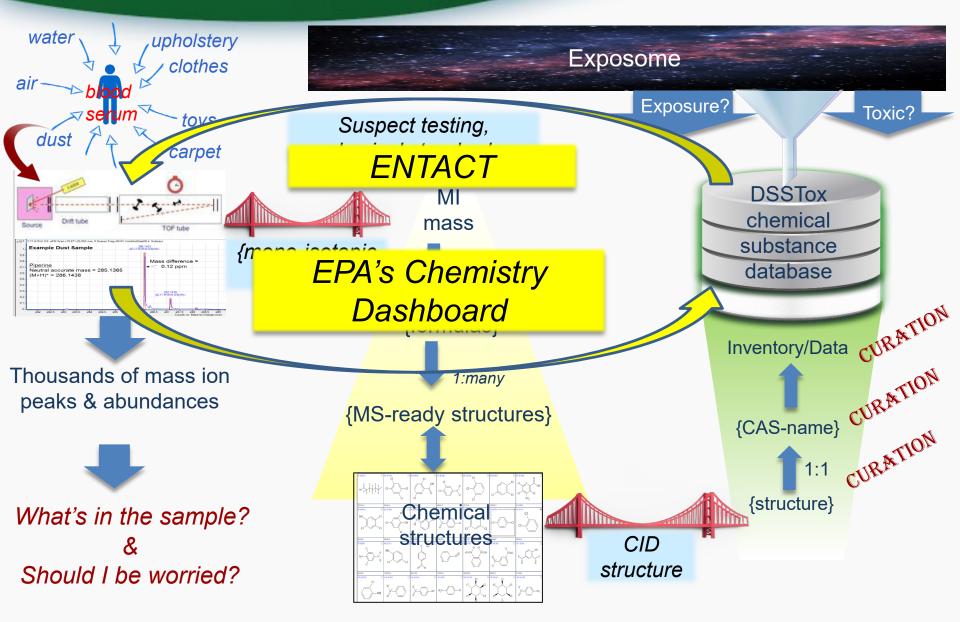
Lots of questions ...



Tuesday, August 14,2018 Judge a person by their questions rather than their answers. - Voltaire (post-enlightenment)

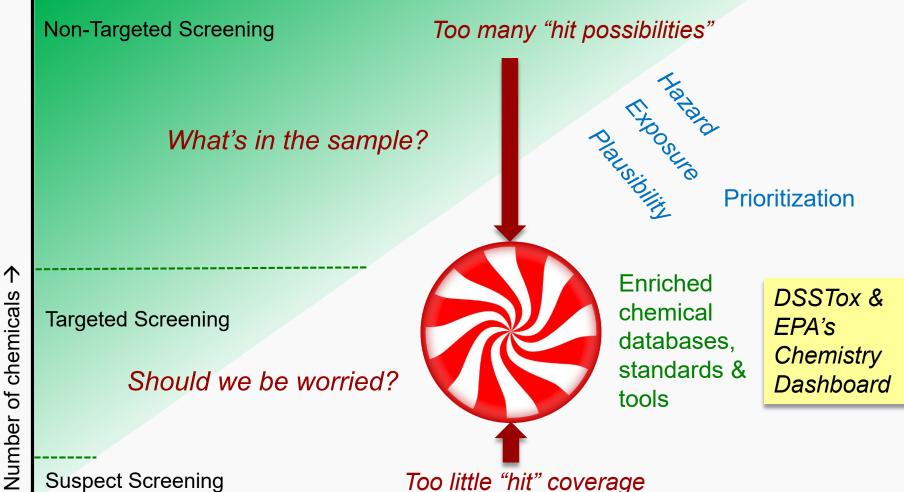
Cheminformatics view of non-targeted testing problem





Cheminformatics view of non-targeted testing problem

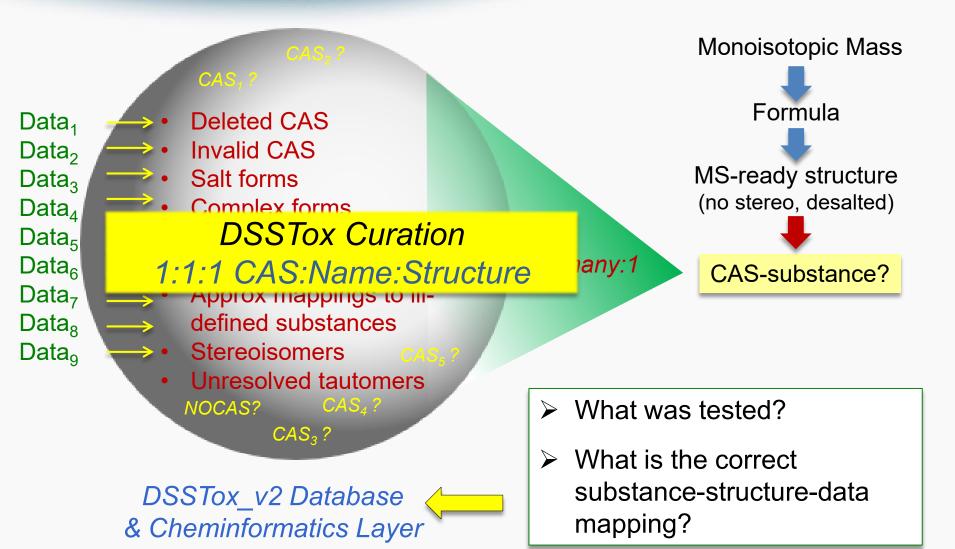




Too little "hit" coverage

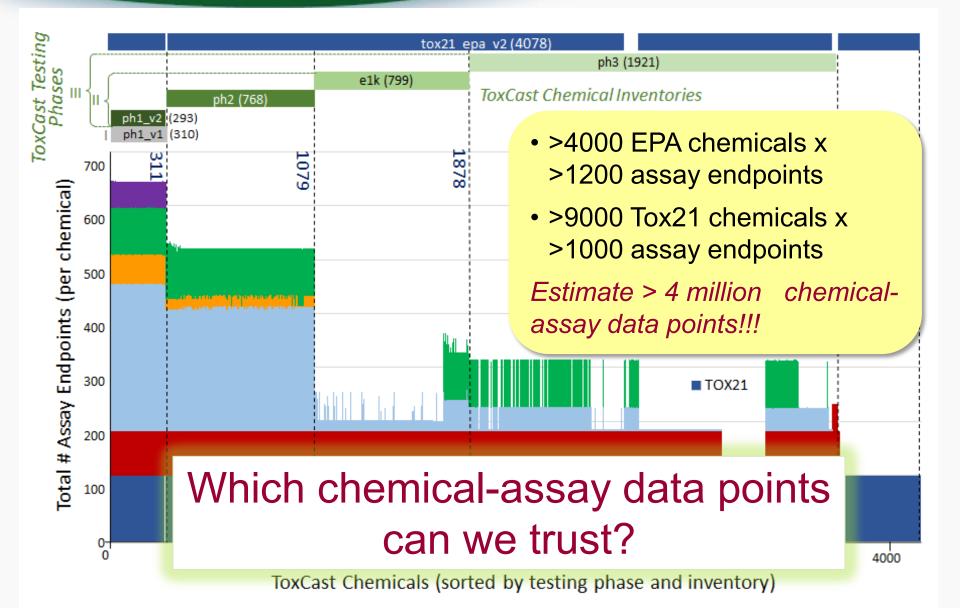
CAS-Structure "Sphere of Confusion"





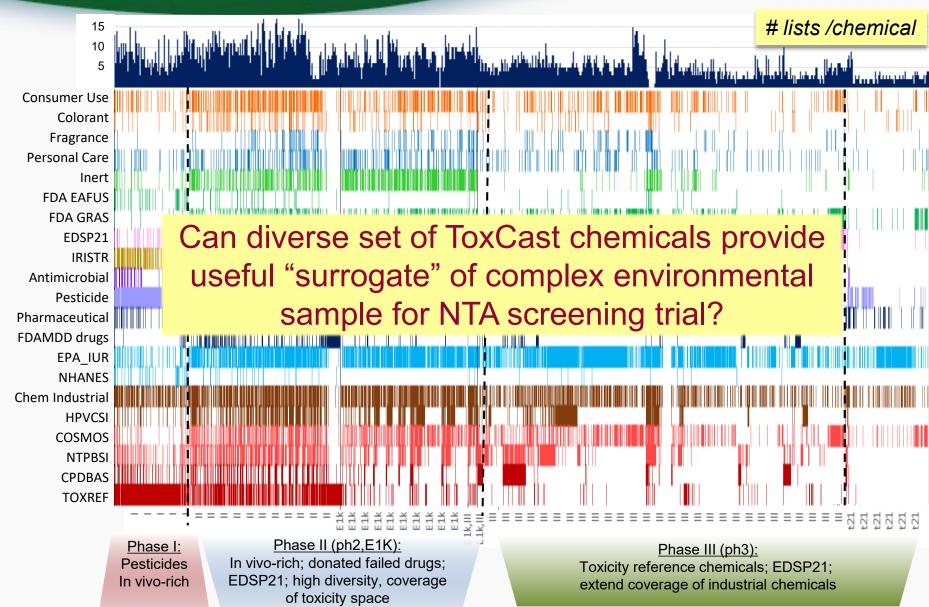
ToxCast HTS data





ToxCast Chemical Coverage: Use, Exposure, Toxicity





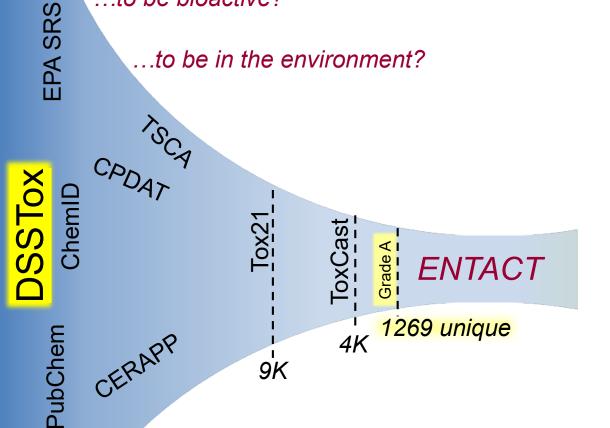
...to be hazardous?

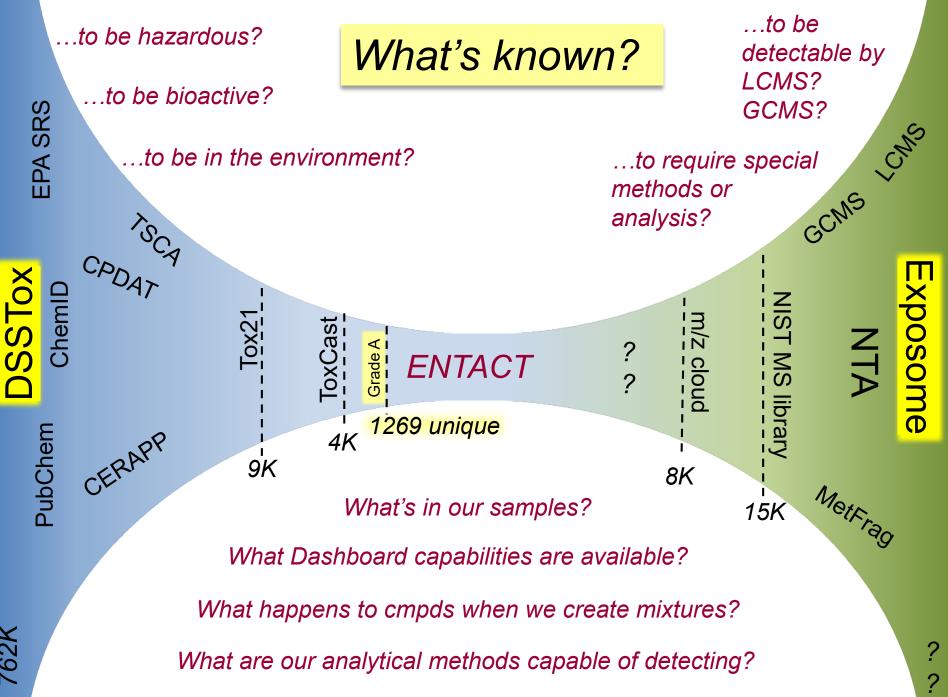
What's known?

...to be bioactive?

762K

...to be in the environment?

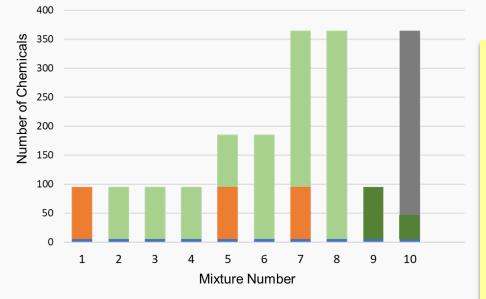




762K

ENTACT Mixture Trial











1269 unique ToxCast chemicals used to construct 10 mixtures

Chemicals in Mix 1-8 (amenable):

- *"Grade A" analytical QC results (LC/GC only)*
- single DSSTox structure
- contain no metals
- chosen to avoid identification issues
- broad range of logP values

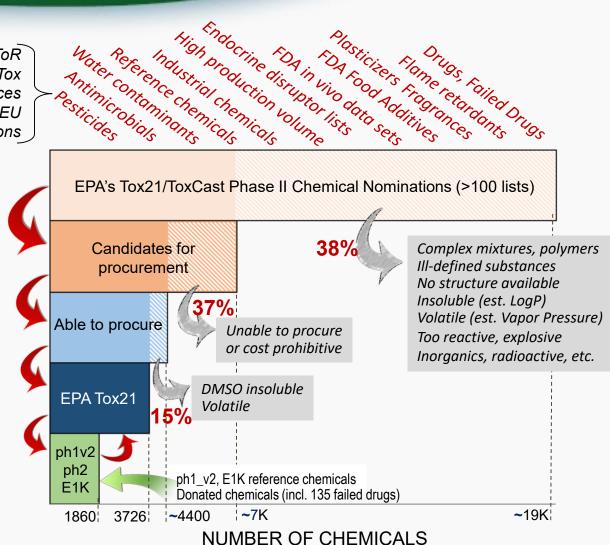
Chemicals in Mix 9,10 (challenging):

- contain isomeric & isobaric cmpds
- contain cmpds graded as <80% purity
- contain a few organometallics

Bias in ToxCast Library Build



EPA ACTOR EPA DSSTox EPA Program Offices OECD, EU Stakeholder Nominations



Chemical & Data Quality Issues



DSSTox Substance-Structure DB

ToxCast & Tox21 Sample Inventory DB

Solutions

Bottles

Supplier/Lot/Batch

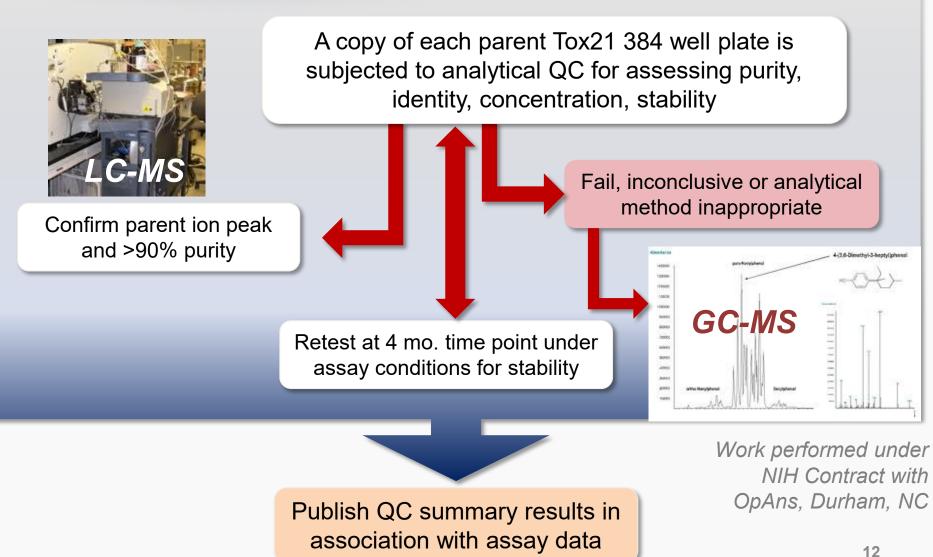
Compound Libraries

Valid Structures Accurate substance annotation CAS – Name - Structure

> Supplier-provided info COA - method/purity DMSO solubility Volatility Stability, age of sample Supplier lot/batch variability

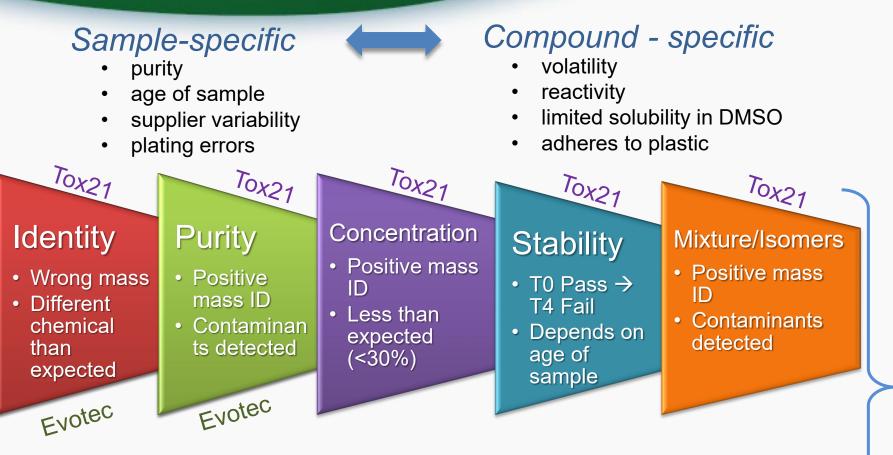
Tox21 Analytical QC





What can analytical QC data tell us?



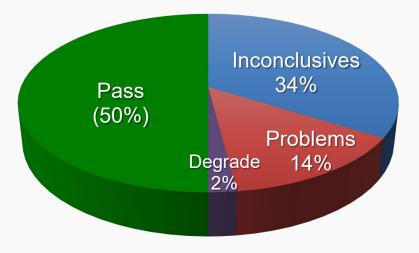


- Is activity due to cmpd or contaminant?
- Is lack of activity due to cmpd not present or present at low concentration?

Tox21 and ToxCast Chemical Library Analytical QC Results (8/2015)



Tox21_QC_Sum-GSID (8593 total)



- 50% pass purity/ID/concentration checks
- A third(34%) of library pose analytical QC challenges (LCMS and GCMS only)
- 2% degrade after 4 months under testing conditions
- 14% problems purity (<75%), ID and/or low concentration (<30% of expected [C])

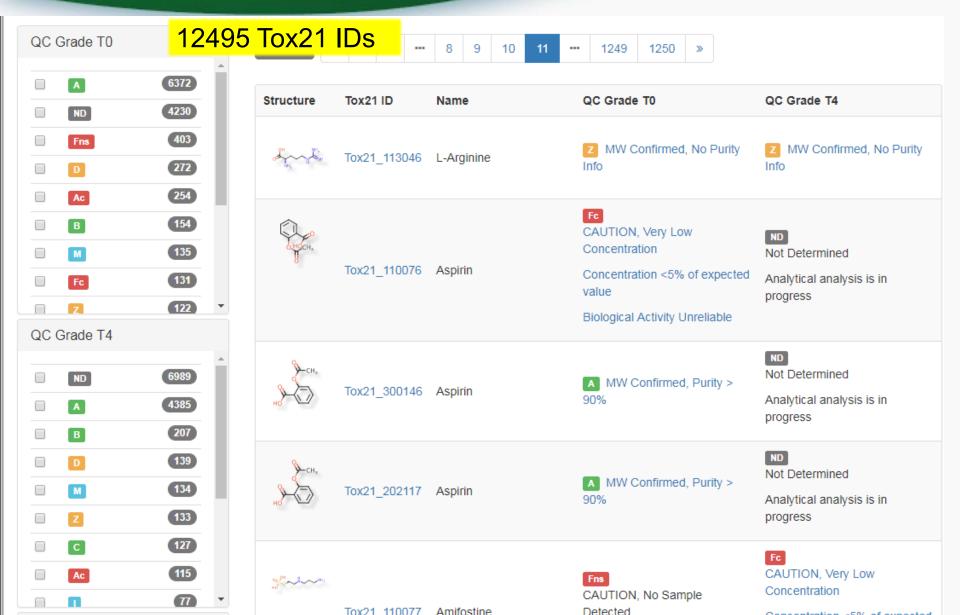
Which chemicals have QC issues? (e.g., SVOCs?)

- > Which chemicals were not analyzed? (e.g., mixtures, inorganics, etc.)
- How are HTS activity profiles linked to QC?

Tox21 Analytical QC Data



https://tripod.nih.gov/tox21/



Chemistry Dashboard



United States Environmental Protection Aaencv

Bisphenol A

United States Environmental Protection \$€P

Home Advanced Search Batch Search Lists v Predictions Downloads

Bisphenol A 80-05-7 | DTXSID7020182

Searche ld.

DETAILS

EXECUTIVE SUMMARY PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS

COMMENTS

Searched by DSSTox Substance Id.					
General	Toxicology				
(a) EPA Substance Registry Service	(a) ACTOR				
A Household Products Database	며 DrugPortal				
Chemical Entities of Biological Interest (ChEBI)	CCRIS				
PubChem					
) Chemspider	eChemPortal				
(CPCat	Gene-Tox				
🔊 DrugBank	HSDB				
hmp HMDB	ToxCast Dashboard 2				
W Wikipedia	LactMed				
Q MSDS Lookup	International Toxicity Estimates for				
ChEMBL	Risk				
Q Chemical Vendors	ATSDR Toxic Substances Portal				
CalEPA Office of Environmental Health Hazard Assessment	Superfund Chemical Data matrix INOSH IDLH Values				
W NIOSH Chemical Safety Cards	ACToR PDF Report				
b ToxPlanet	Toxics Release Inventory				
ACS Reagent Chemicals	CREST				
W Wikidata	National Air Toxics Assessment				
ChemHat: Hazards and Alternatives Toolbox					
🌞 Wolfram Alpha					
ScrubChem					
ECHA Brief Profile					

- ECHA Brief Profile ECHA Infocard
- ChemAgora Consumer Product Information Database (CPID)
- ChEBI

K Sigma-Aldrich Chemicals

NST NIST Chemistry Webbook

Publications

Copy 💌

Сору 🔻

Share 💌

Toxline

Environmental Health Perspectives

NH NIEHS National Toxicology Prog G Google Books G Google Scholar G Google Patents PPRTVWEB

NH) PubMed IRIS Assessments EPA HERO

WINSH Skin Notation Profiles

WINDSH Pocket Guide

RSC Publications BioCaddie DataMed

Springer Materials

Federal Register

Regulations.gov

Bielefeld Academic Search Engine

T CORE Literature Search

Analytical

Share 💌

PR-IDENT MEMI: National Environmental

Submit Comment

Submit Comment

Q Search all data

Prediction

2D NMR HSQC/HMBC Prediction

Carbon-13 NMR Prediction

Proton NMR Prediction

Methods Index RSC Analytical Abstracts

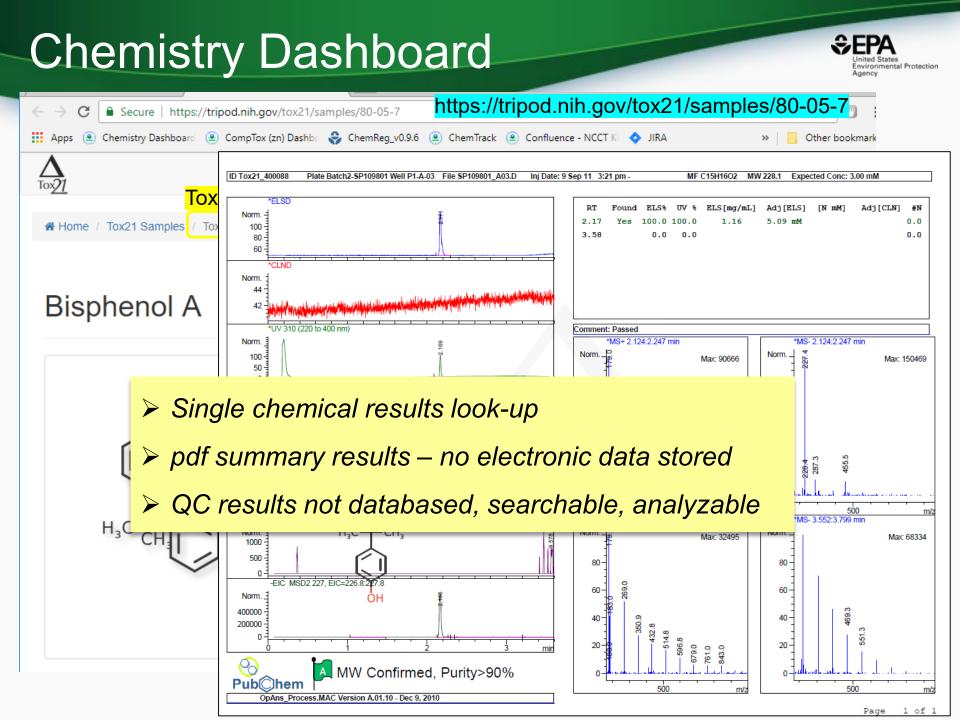
ChemRTP Predictor Tox21 analytical data

MONA: MassBank North America

mzCloud

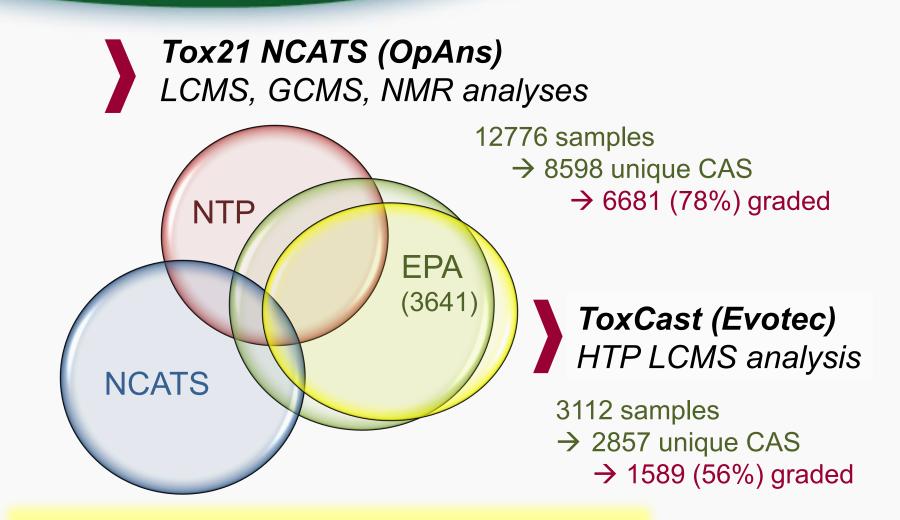
NET NIST IR Spectrum

NIST NIST MS Spectrum



Available Analytical QC Data





1724/3641 EPA_Tox21 chems graded (47%)

Evotec Analytical QC Data



gsid	20006	20014]
preferred_name	Acetamino <u>phen</u>	en Dehydroacetic acid		
casrn	103-90-2 Purity_UV214 E	SI POS MS+	96.24	100
Container TX	TX002867 Compound four	Compound found (1=yes 0=no)		0
FileName SAMPLE	01129195 State	State		Diversity fail
Date	15-Apr-15 Target MS+	Target MS+		169.0501
Exact MF	C8H9NO2 Observed MS+	Observed MS+		
Exact MW	151.06333 Purity UV214 E	Purity_UV214 ESI NEG MS-		100
Real MW(salted Average)	151.1626 Compound four	Purity_UV214 ESI NEG MS-		0
Salted MF				

How do we store, process, interpret, deliver & make use of all these data?

LCMS run in 4 detection modes for 3125 samples

Smiles

Compoundfound	1	0		
State	ОК	Diversity fail		
Target MS+	152.0711	169.0501		
Observed MS+	152.0573			
Purity_ELS ESI NEG MS-	100	52.86		
Compound found	1	0		
State	ОК	Diversity fail		
Target_MS-	150.0555	167.0345		
Observed_MS-	150.0819			
Fail 90	no	yes		
EPA_PDF_Reference File	SAM002476551_100.00perc.pdf	SAM002696017_0.00perc.pdf		
Comments				

Tox21 Grades: Confidence Levels

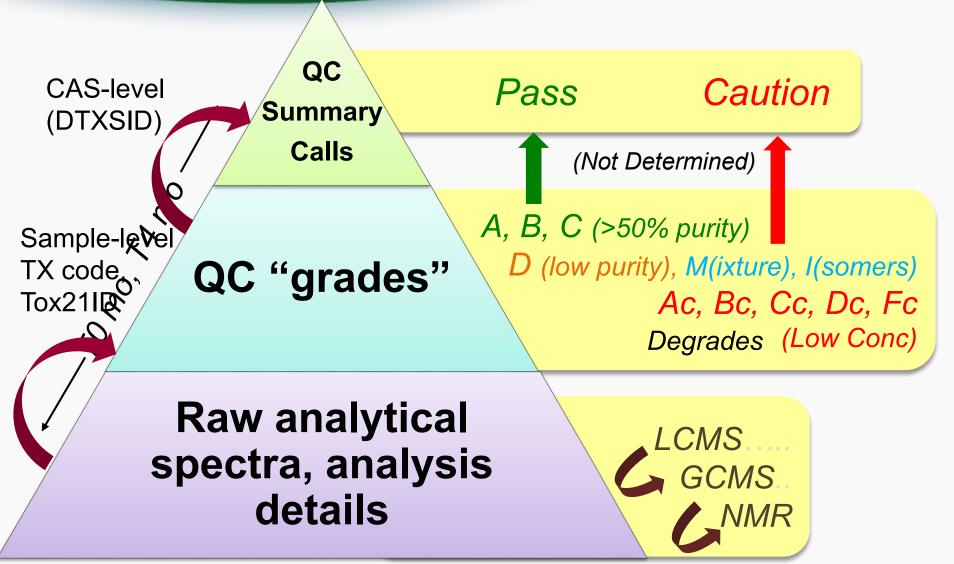


Grades	%Purity	Conc		QC-Detailed grade	QC-	Detailed grade	
А	>90		P				
В	75-90		osit	Pass	ecr		
С	50-75%		ositive		ea	Pass	
Z	?			Pass_NoPurityInfo	Decreasing		. <u>-</u>
T0-T4 fail			MW	Pass_degrades			s C
D	<50%		ide	Caution_LowPurity	confidence		confidenc 「S results
Ac	90		nt		ide		fid es
Bc	80	5-30% of expec	ification	Caution_LowConc	nc		S r
Сс	75		ltio		e in	Caution	
М	?	Mixture	n	Caution_Mixture		Caution	
Fc	?	<5% of expecte	d	Caution_VeryLowConc	sample		easir Cast
F	wrong ID			Caution_WrongID	ple		Incre: ToxC:
Fns	no sample			Caution_NotDetected			μĒΡ
ND	Not determined			Tox21_ND		Tox21_ND	

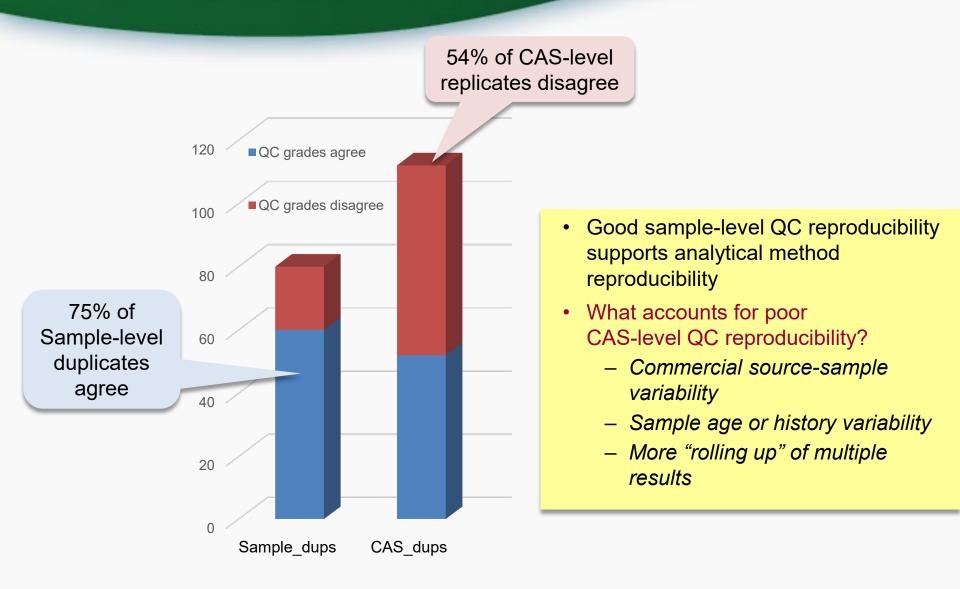
What level of reporting conveys "useful" guidance & level of confidence in QC results along with assay results?

Processing Tox21 QC Data





Evotec QC Results: Reproducibility

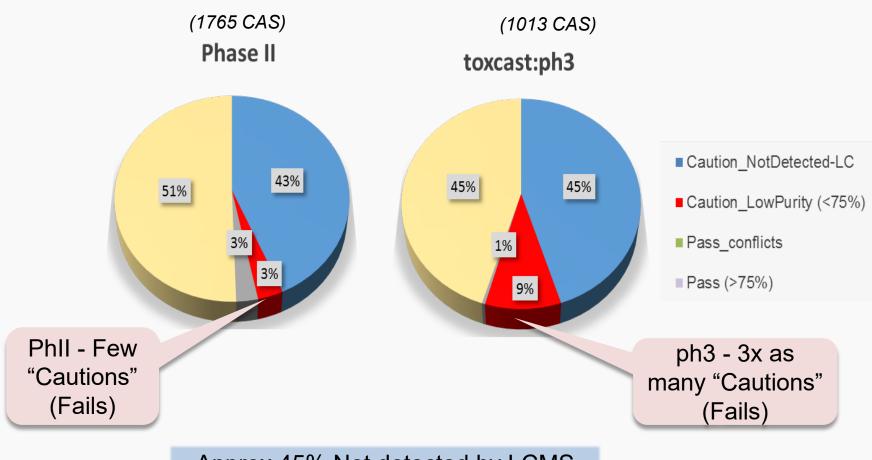


nvironmental Protection

Agency

Evotec QC (LCMS) Results





Approx 45% Not detected by LCMS

Use-cases benefiting from access to detailed QC data



- 1) ExpoCast projects attempting to detect ToxCast/Tox21 chemicals in consumer products or environmental samples
 - knowledge of appropriate analytical method
- 2) Use of ToxCast data to support safety assessments of individual chemicals with questionable QC (e.g., 2,4,6-Triazole)
- 3) Evaluate ToxCast HTS models (ER, Vascular) for sensitivity to results for chemicals with low QC *(e.g., TNP-40)*
- Non-targeted screening projects of environmental & biological media (NERL)
 - Reference spectra & knowledge of successful method for detecting chemical to guide analyses
- 5) EPA chemical library management
 - Prioritize reprocurements, solubilizations, platings
 - Propose QC analysis schedule for all/portions of library

Ongoing/future work with QC Data (w/ Tony Williams & Andrew McEachran)



Process & load all "raw" QC data into database to support structurebased modeling of analytical methods:

> Ionization mode, chromatography/retention time, detector used

- Goals of analysis:
 - Explain "Not Determined" and "Failed" QC labels

Provide guidance for use of QC data to

- improve our science moving forward!
 - *Reducing number of "Failed" QC labels and uncertainty in QC grades*

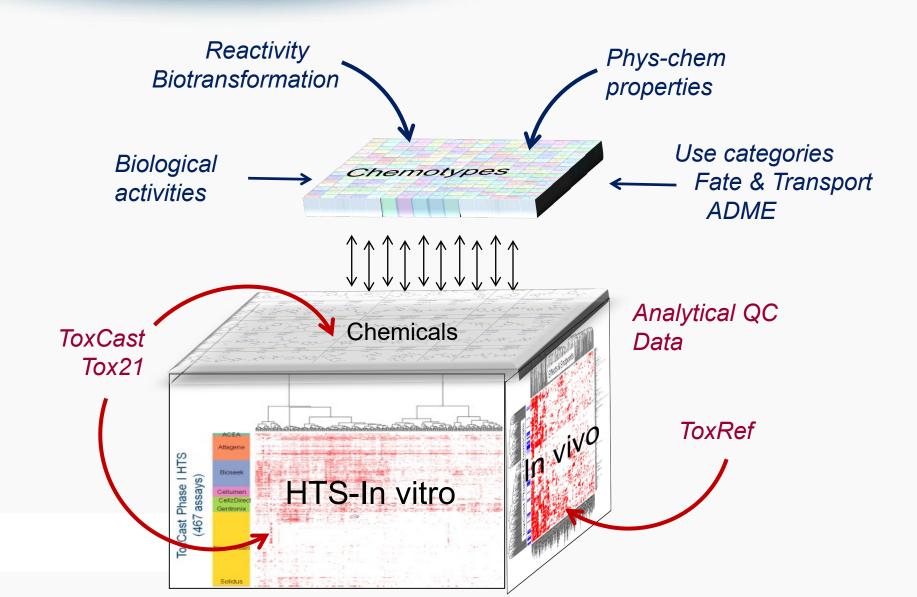
EPA Chemistry Dashboard

- Provide access to all available analytical QC results from CAS level summary grades to sample-level results
- Serve up predicted & actual MS spectra for >700K DSSTox chemicals for use in wide ranging environmental research projects

interest

Building a chemical "feature" knowledge-base



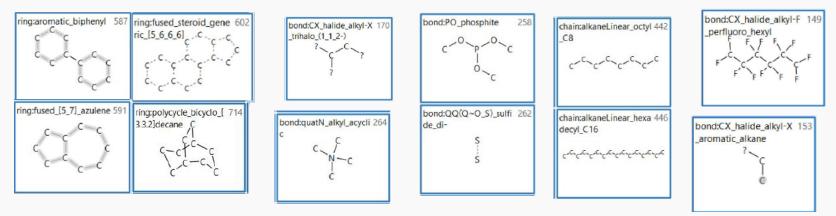


ToxPrint Chemotypes



- 1) Publicly available Chemotyper application <u>http://www.chemotyper.org</u> & ToxPrint CSRML file <u>http://www.toxprint.org</u> for computing ToxPrint "fingerprints", as implemented within EPA's Chemistry Dashboard using Molecular Networks CORINA command-line code (*Yang et al., 2016*)
- 2) ToxPrints include well-defined chemical features relevant to environmental datasets & toxicity (e.g., structure-alerts), and are computational, visualizable, and interpretable

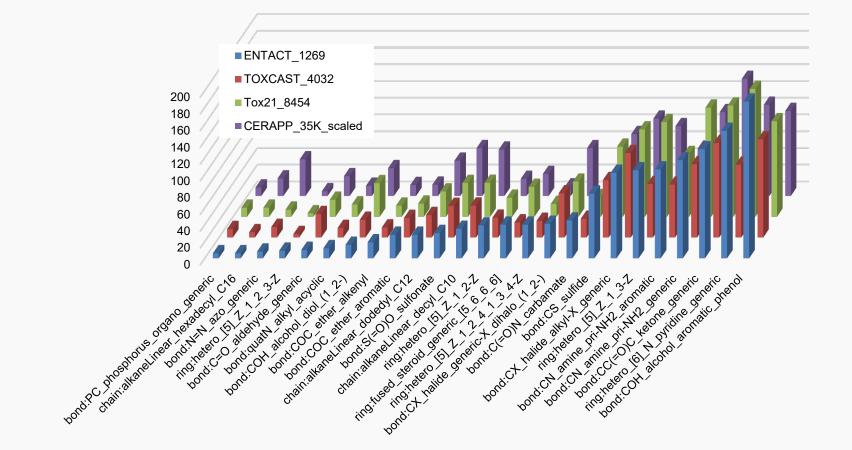
729 chemical features, capturing diverse chemistry: scaffolds, functional groups, chains, rings, bonding patterns, atom-types



 \rightarrow structure-based means for defining regions of local chemistry.

ToxPrint inventory profile comparisons

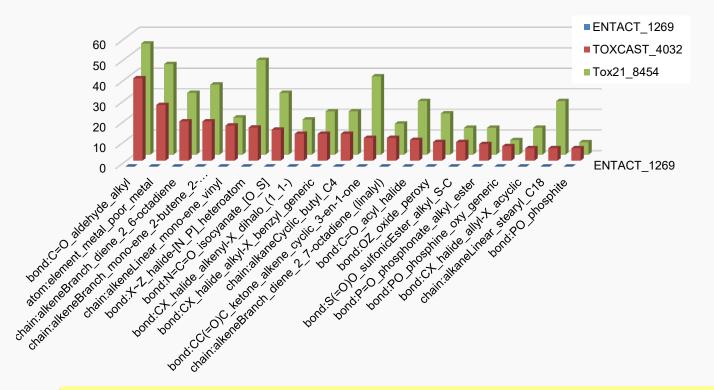




ENTACT chemical set has similar feature profile to larger ToxCast and DSSTox datasets, but ...

ToxPrint inventory profile comparisons

Top 20 ToxPrints missing in ENTACT Mixtures



Features in ToxCast & Tox21 missing in ENTACT either due to:

- ND (not detected) by LC/GC method
- Less than "Grade A" QC score
- Metal-containing

Environmental Protection

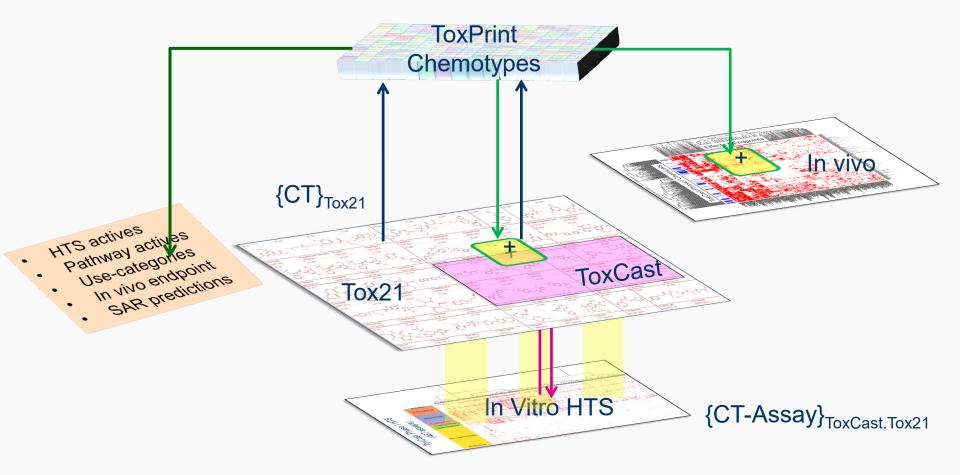
Agency

Chemotype-"Activity" Enrichments

Create {CT-Activity} enrichment profiles for binary "activity" datasets

invironmental Protection

Agency

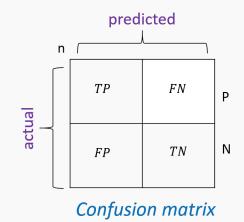


Computing CT-Assay "Enrichments"



Set statistical thresholds & filters for significance to support data-mining objectives:

TP_ID	ToxPrint_CT_name ²	CT _{Tot}	T _{pos}	F_{pos}	F _{neg}	T _{neg}	Odd's Ratio	Fischer's pval
423	chain:alkaneBranch_t-butyI_C4	41	24	17	294	693	3.3	2.0E-04
479	chain:aromaticAlkane_Ph-C1-Ph	39	27	12	291	698	5.4	6.5E-07
303	bond:X[any_!C]_halide_inorganic	28	17	11	301	699	3.6	9.0E-04



TestSet = # Pos + # Neg = # chems tested

CT _{Tot} = total # chems in TestSet w/ CT (Pos or Neg)
TP (T _{pos}) = # Pos in TestSet w/ CT
FP (F _{pos}) = # Neg in TestSet w/ CT
FN (F _{neg}) = # Pos in TestSet w/o CT
TN $(T_{pos}) = #$ Neg in TestSet w/o CT

 \blacktriangleright Odds Ratio \geq 3, conveys simple fractional enrichment

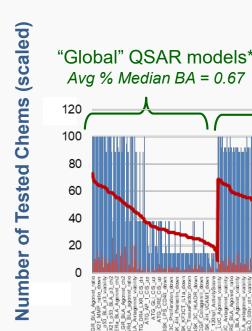
Fischer's exact p value \leq 0.05, *takes into account size of dataset*

 \succ T_{pos} (TP) \geq 3, require at least 3 chemicals with CT in Positives

Finding chemical signal in HTS assay "noise"

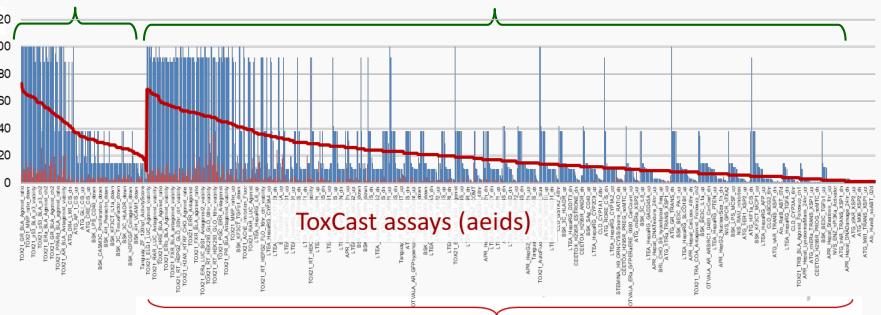


Enriched CT-count across 1032 ToxCast assays



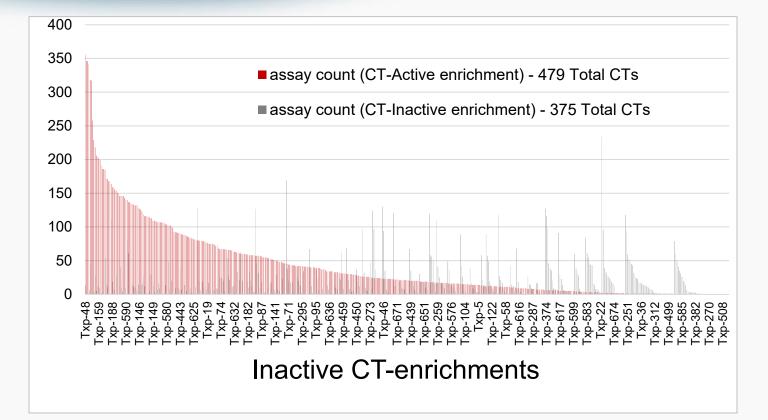
*Random Forest models based on ToxPrint CT descriptors, validated using independent Test Set & Y-randomization, with Training (100A, 100I) & Test (25A,25I) Set minimums (J. Fitzpatrick)

"Global" QSAR models failed for 86% of ToxCast Assays



Significant CT-enrichments detected across ToxCast assay space

What about the "Inactives"?



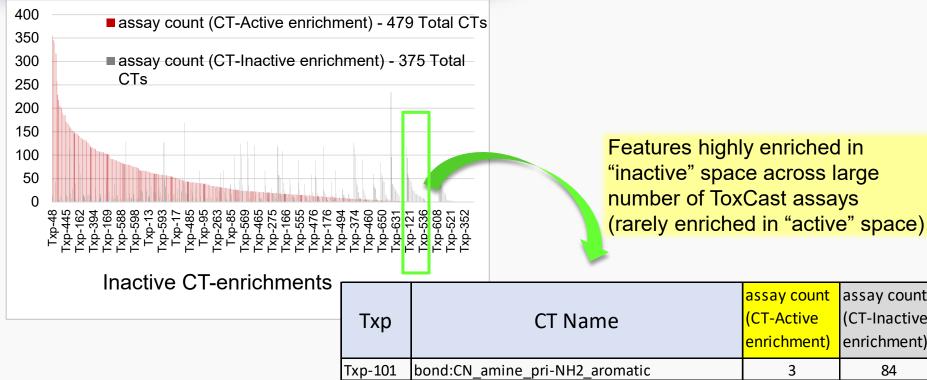
vironmental Protection

Are CT-enrichments in "negative" assay space due to:

- True biological inactivity?
- Assay artifacts?
- QC failure?

Top 10 enriched CT-Inactives (skewed from actives)





- True biological inactivity?
- Assay artifacts?
- QC failure?

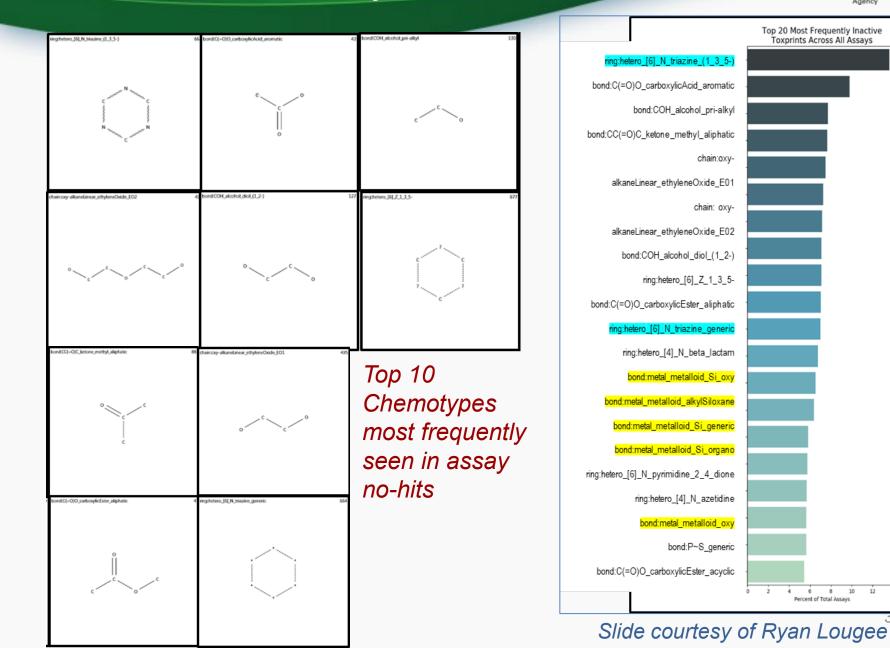
Тхр	CT Name	assay count (CT-Active enrichment)	assay count (CT-Inactive enrichment)
Тхр-101	bond:CN_amine_pri-NH2_aromatic	3	84
Тхр-145	bond:CX_halide_alkyl-Cl_ethyl	0	79
Тхр-260	bond:P~S_generic	1	94
Тхр-362	bond:metal_metalloid_oxy	2	96
Тхр-372	bond:metal_metalloid_Si_organo	6	99
Тхр-374	bond:metal_metalloid_Si _oxy	6	116
Тхр-496	chain:oxy-alkaneLinear_ethyleneOxide_EO1(O)	5	91
Тхр-497	chain:oxy-alkaneLinear_ethyleneOxide_EO2	6	127
Тхр-607	ring:hetero_[4]_N_beta_lactam	1	118
Тхр-663	ring:hetero_[6]_N_triazine_(1_3_5-)	2	234

Inverse Promiscuity



Top 20 Most Frequently Inactive

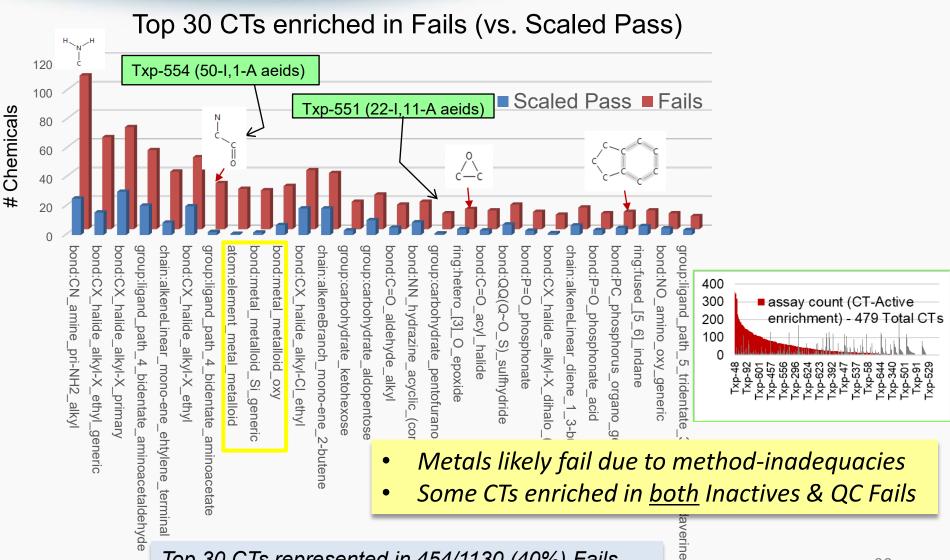
Toxprints Across All Assays



10 12

6 8 Percent of Total Assavs

ToxPrints enriched in Tox21 **QC** Fails



Top 30 CTs represented in 454/1130 (40%) Fails

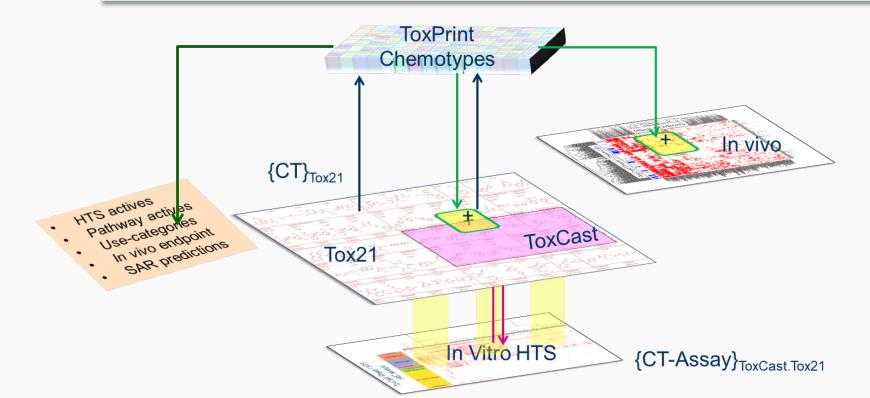
Invironmental Protection Agency

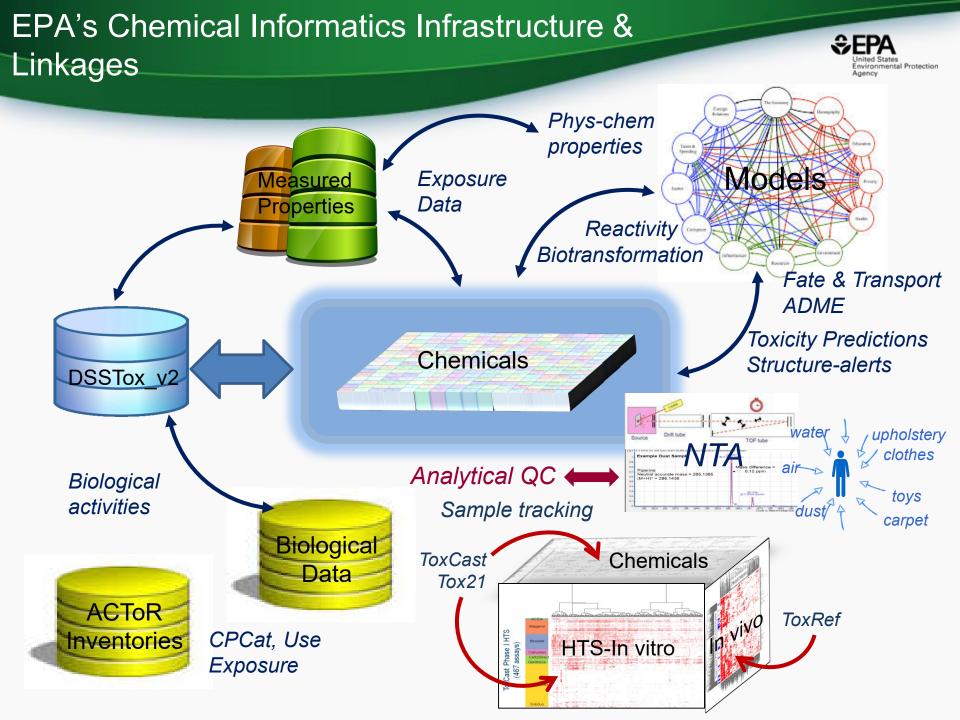
Chemotype-"Activity" Enrichments



Create {CT-Activity} enrichment profiles for any binary "activity" dataset

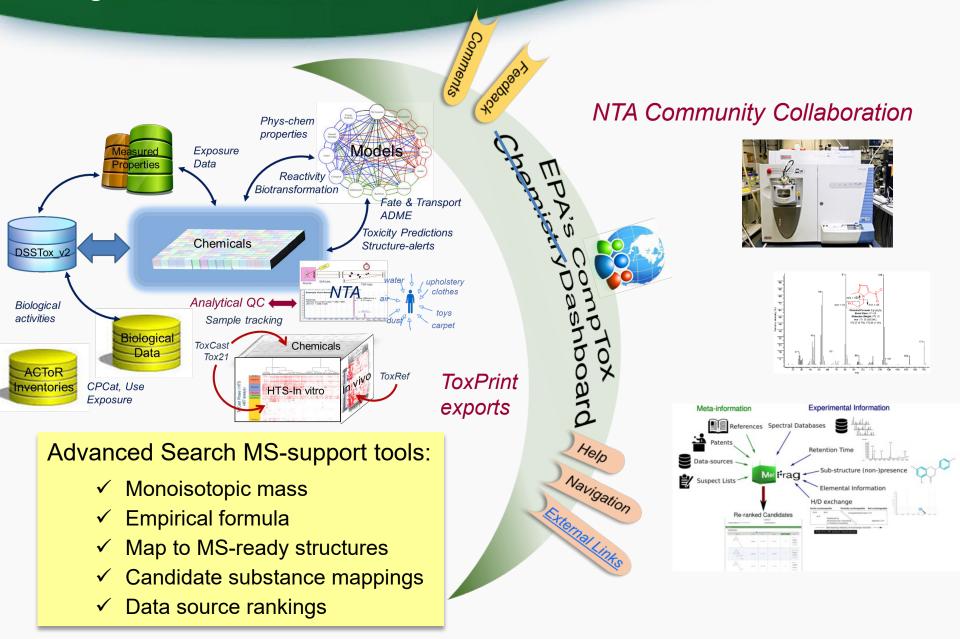
- Bioassay activity (active, inactive)
- Assay type artifacts (e.g., fluorescence detection method)
- QC "Fail" (yes, no)
- Sample degrades over time (yes, no)
- LCMS method suitability (yes, no)
- GCMS method suitability (yes, no)





EPA's Chemical Informatics Infrastructure & Linkages





Chemical "Universe" problem



Biodegradation products Metabolome Protein & DNA fragments Virtual screening libraries Combinatorial chemistry Polymer fragments Polyorganic acids Adducts, surfactants

The other 98%

Scientific literature Toxicology studies Environment/Industry Commercially available

Exposure?

Toxic?

- ✓ Where should DSSTox expand chemically?
- What part of the universe should we store in DSSTox?
- How can the ToxCast library be shared for greatest gain?
- What cheminformatics "plumbing" would be most useful to this community?

DSSTox_v2 CAS - structures CAS - no structures Structures only

MS-ready structures Category-enumerated structures

Chemical "Universe" problem



EPAHFR - EPA Chemicals associated with hydraulic fracturing

Search EPAHFR Chemicals

Q

Substring search

List Details

1640

Description: Chemicals used in hydraulic fracturing fluids and/or identified in produced water from 2005-2013, corresponding to chemicals listed in Appendix H of EPA's Hydraulic Fracking Drinking Water Assessment Final Report (Dec 2016). Citation: U.S. EPA, Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-16/236F, 2016. https://www.epa.gov/hfstudy

*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (as of Jan 2018), both were included in the current EPAHFR chemical listing for completeness; additionally, 34 previously unmapped chemicals in Table H-5 are now registered in DSSTox (all but 2 assigned CASRN) and, thus, have been added to the current EPAHFR listing.

Number of Chemicals: 1640

Endoc	ine Disruption Screening Program (EDSP) Chemicals	Univ	/erse of
	Search EDSPUOC Chemicals	Q	
9411	Substring search		
List Details			
EDSP Universe of Chemic	*	stance in th	ne CompTox Dashboard are

What do we hope to learn?

What are various analytical methods capable of detecting? How consistent are results across labs? How does creation of ToxCast mixture affect results?

• Generate mass spectra for ToxCast samples by various instrumentation/approaches

1269 unique

Improve guidance for choosing appropriate analytical methods

ENTACT

• Provide information on impurities, sample problems

Grade A

4K

Leverage use of ToxCast library & capabilities of CompTox
 Dashboard for advancing NTA workflows

Exposome

NTA

MetFrag

NIST

MS library

15K

m/z cloud

8K

?

Chis

EPA SRS

SSTox

ChemID

PubChem

CERAPP

Tox21

9K

Lots of questions, few clear answers ... YET!



Tuesday, August 14,2018 Judge ENTACTby the posed & lessons learned questions rather than by getting all of what we think may be the right - Annonymous

Acknowledgements:

 EPA NCCT Chemistry Team & DSSTox Curators
 Chris Grulke, Antony Williams, Ryan Lougee, Andrew McEachran, Grace Patlewicz, Indira Thillainadarajah, Saku Sivasupramaniam, Brian Meyer, Jeremy Fitzpatrick

- EPA NCCT ToxCast & ChemTrack Teams Rusty Thomas, Keith Houck, Richard Judson, and many others Kathy Coutros, Jeremy Dunn, Jeff Edwards
- EPA NERL ENTACT Collaborators Jon Sobus, Elin Ulrich, Mark Strynar
- Tox21 Analytical Team (NCATS, NTP)
 William Leister, Ruili Huang, Christopher LeClaire, Suramya Waidyanathas
- ToxPrint Chemotypes Molecular Networks: Chihae Yang, Aleksey Tarkhov, Christof Schwab Altamira: Jim Rathman

This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.