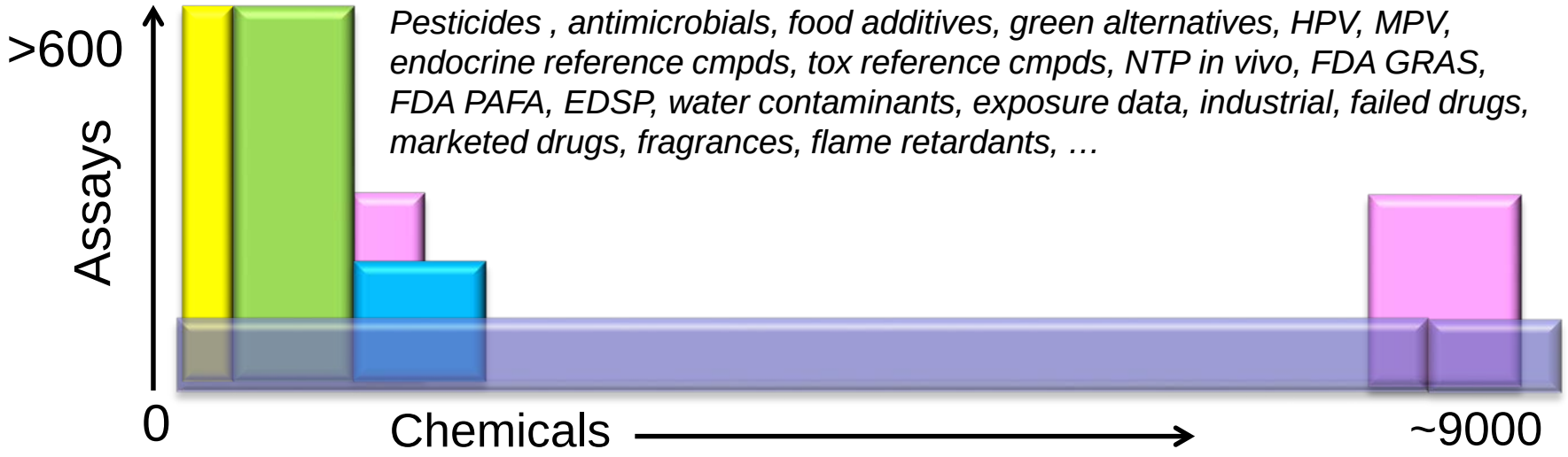
A Knowledge-Informed Chemotype Approach to Mining the ToxCast/Tox21 Chemical-Data Landscape

Ann Richard
U.S. EPA, National Center for Computational Toxicology
Office of Research and Development

9th World Congress on Alternatives and Animal Use in the Life Sciences
Prague, Czech Republic Aug 24-28, 2014

ToxCast & Tox21 Inventories: Chemicals, Data & Timelines

Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	293	~600	~700	2011	Now
ToxCast Phase II	767	~600	~700	03/2013	Now
ToxCast E1K	800	~50	~120	03/2013	Now
Tox21	~9000	~80	~150	Ongoing	Ongoing
ToxCast Phase III	~900	~300	~300	Just starting	2014-2015



Toxicity Prediction Challenge:

Bringing all knowledge & data to bear on problem

Data-mining

Biologically-based QSAR & Cheminformatics

Aggregation

Adverse Outcomes:
> Pathways
> Genes
> Assays
+ Statistical associations

Reactivity & toxicity-informed features & classes

Mechanistically well-defined toxicity endpoint

Structures

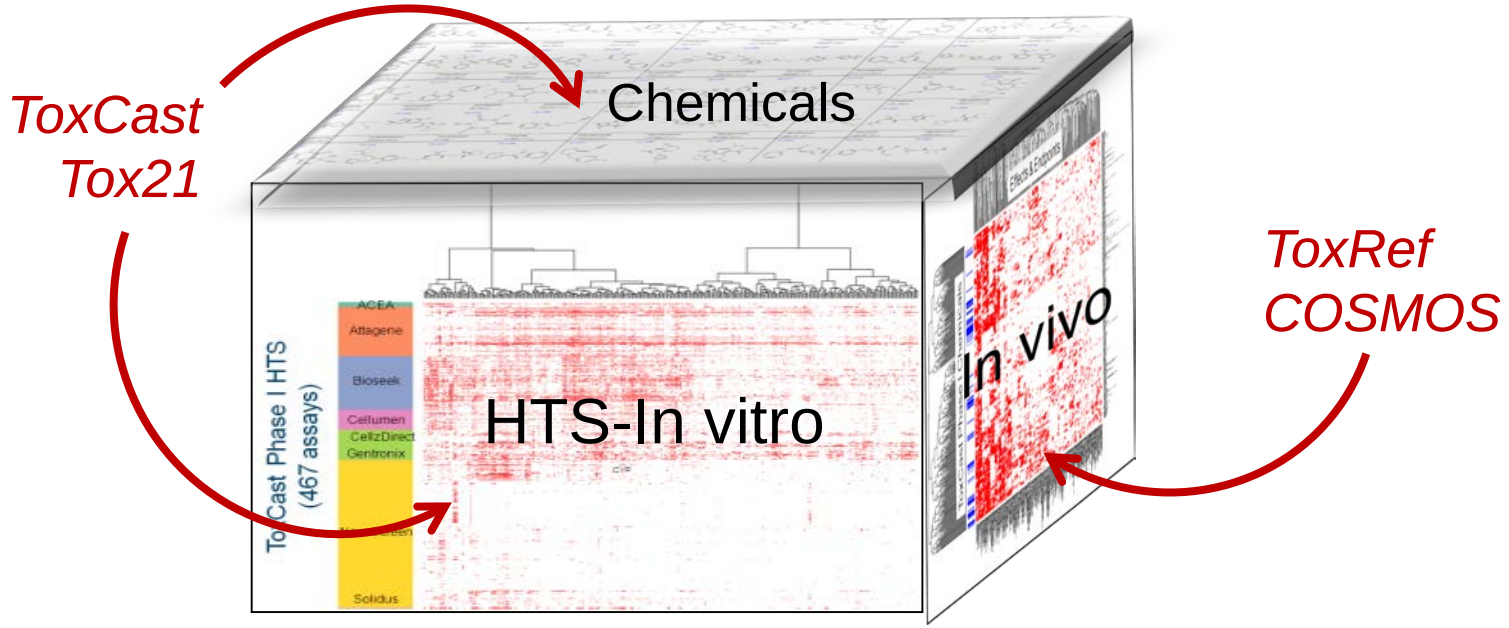
In Vivo

In Vitro/HTS

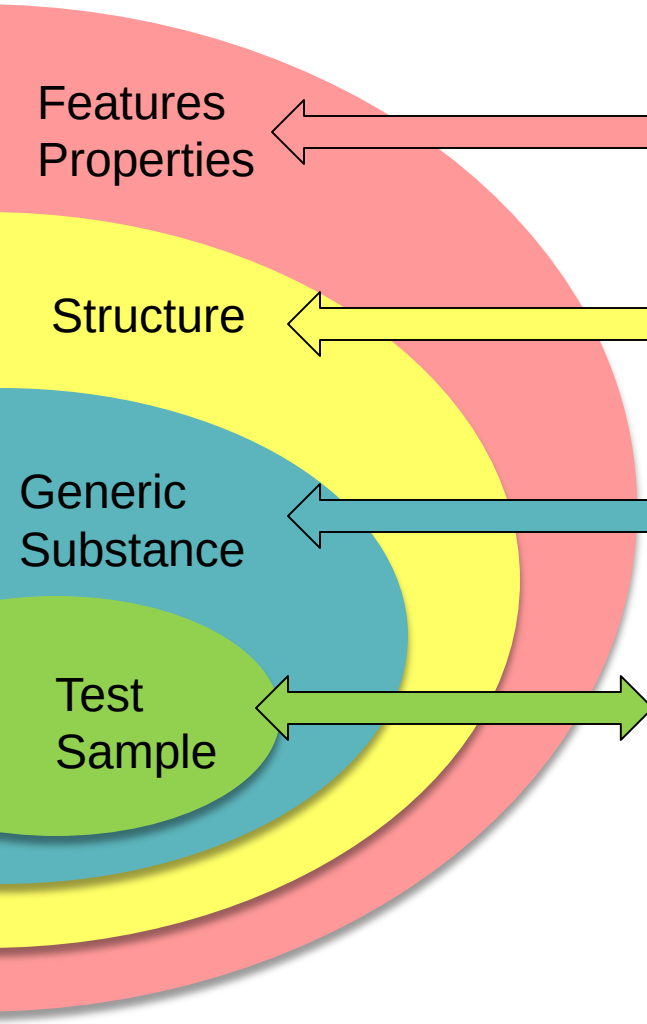
Existing knowledge

Chemical “probes” of biological activity

- Use existing knowledge & SAR to mine HTS data
- Use HTS data to inform & refine SAR models & approaches
- Use all of these data to improve ability to model toxicity

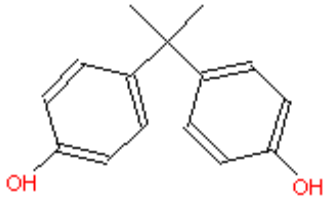


Chemical Elements to Data Integration: *Chemical representations → Uses*



Chemotypes, fingerprints, phys-chem properties, ...

Chemical analogs, Read-across, SAR modeling



SMILES
InChI

Structure searching & modeling

Chemical Name
CASRN

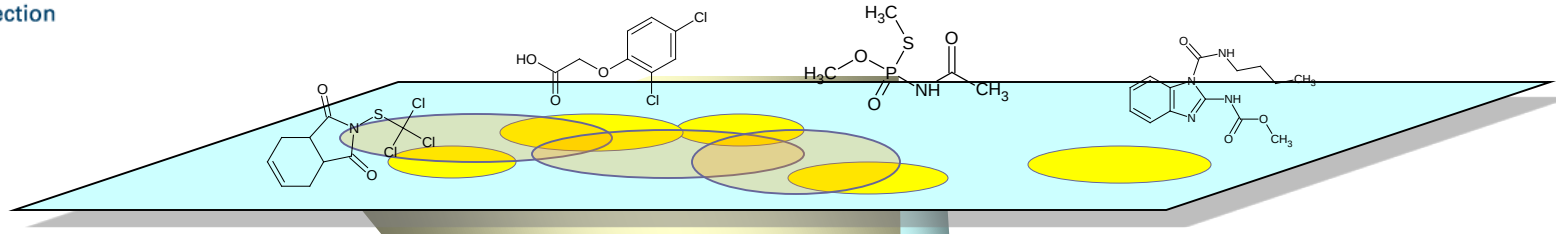
Public toxicity datasets

Supplier, Lot/Batch, physical description

Experimental Endpoint Data

Structure vs. Bioactivity Similarity

United States Environmental Protection Agency

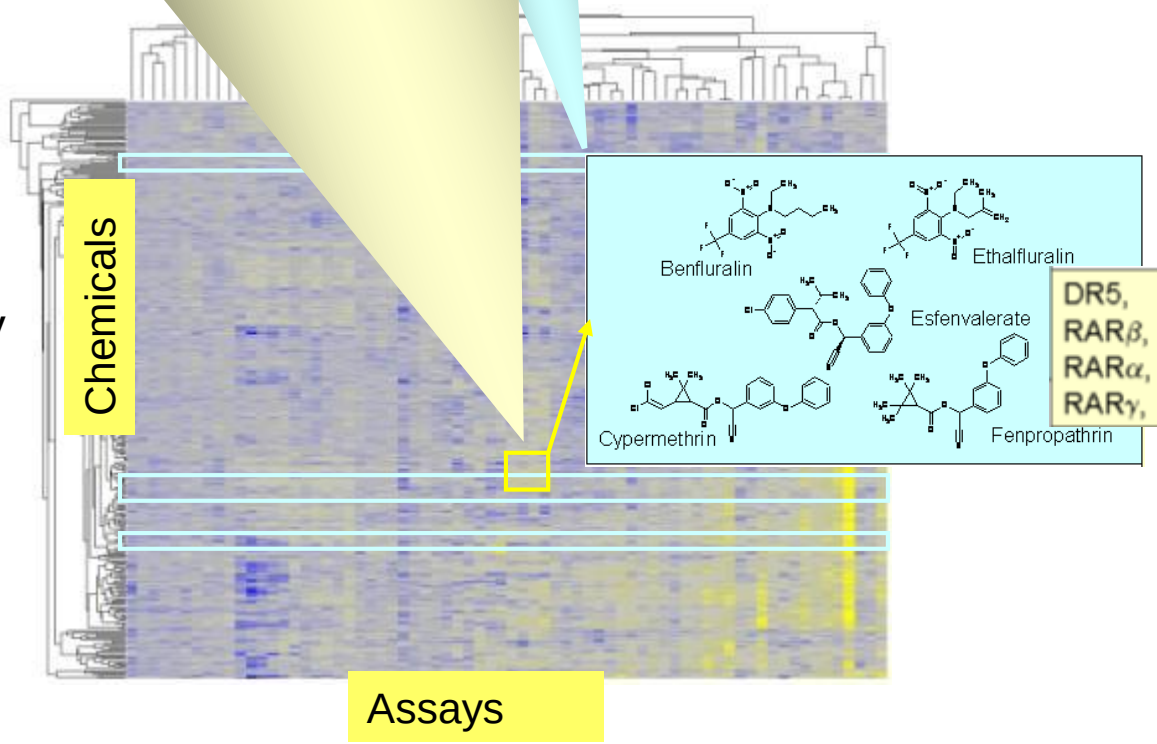


Structure similarity:

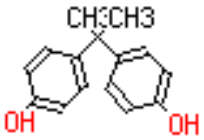
- implies biological similarity
- limited to local chemistry
- subject to “activity cliffs”

HTS bioactivity similarity:

- implies mechanistic similarity
- can link diverse local chemistries to common biological activities
- noisy data, difficult to extract clear signal



Similarity – in the "eye" of the beholder



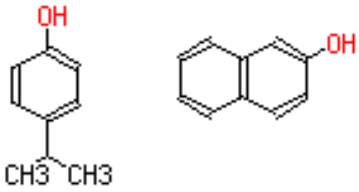
Chemical Similarity

Structural fingerprint

100011010100000110
(all parts considered equal)

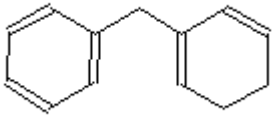
Tanimoto Coeff

1=exact match
0.9 = pretty good match
0.5 = who knows, depends



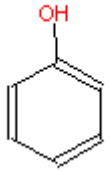
Chemotypes

Pre-defined features or scaffolds informed by SAR, public inventories, and FDA safety assessment workflows



biphenyl

phenol

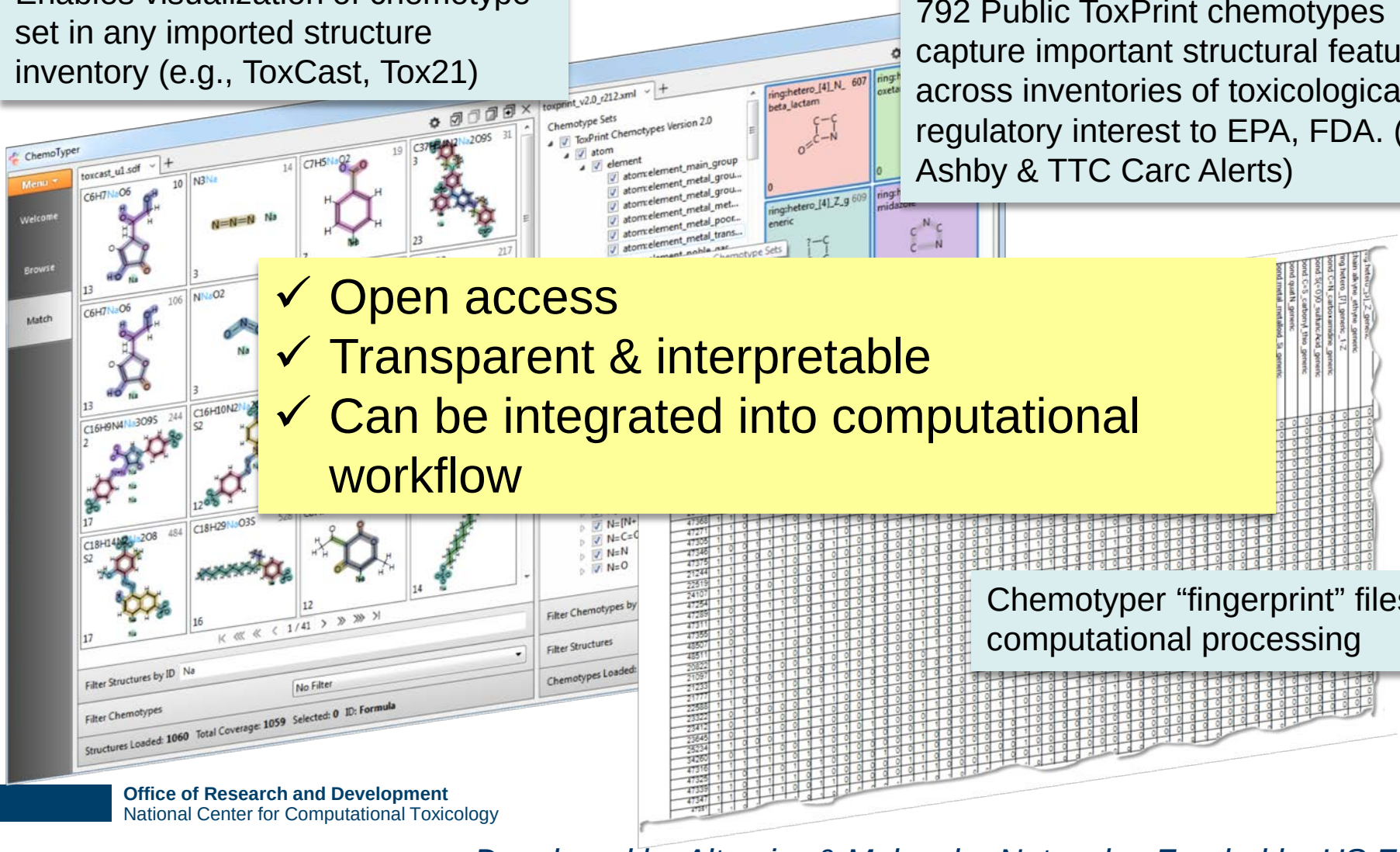


Rules for judging similarity are clearly defined, interpretable, computable, and transferable.

Public Resources: Chemotyper & ToxPrint Chemotypes

Enables visualization of chemotype set in any imported structure inventory (e.g., ToxCast, Tox21)

792 Public ToxPrint chemotypes capture important structural features across inventories of toxicological & regulatory interest to EPA, FDA. (incl. Ashby & TTC Carc Alerts)



- ✓ Open access
- ✓ Transparent & interpretable
- ✓ Can be integrated into computational workflow

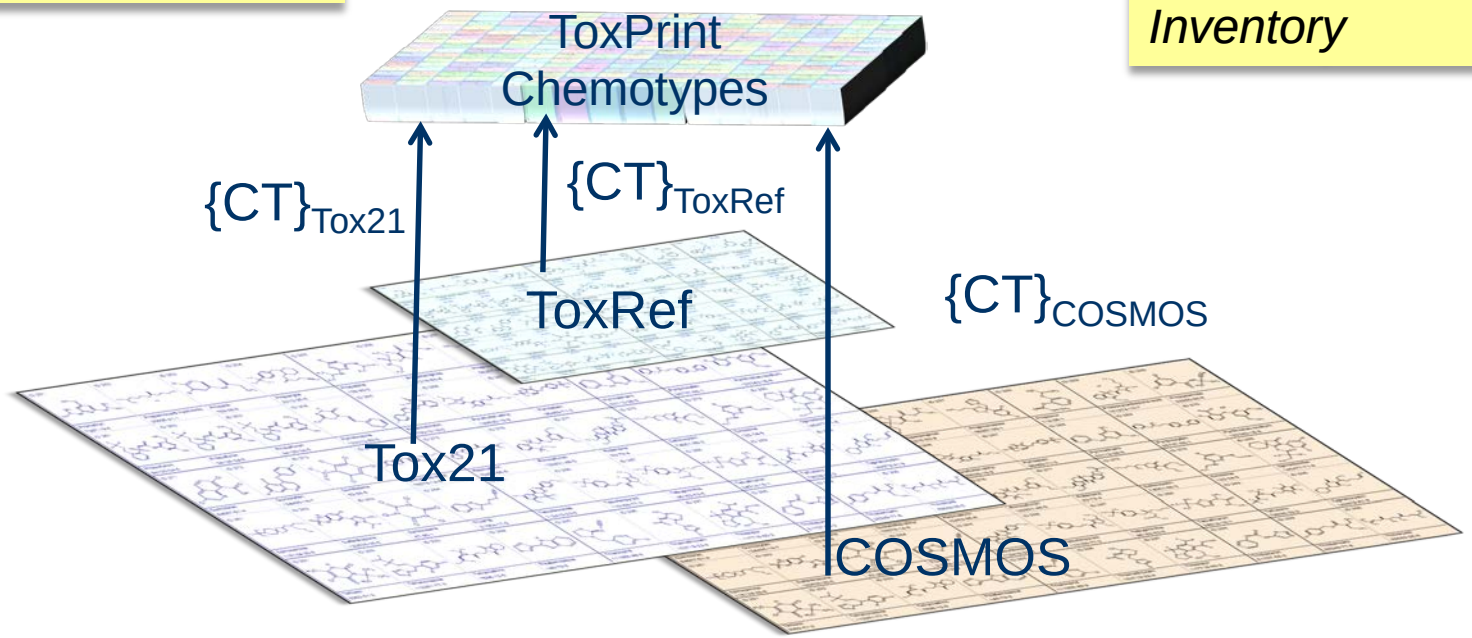
Chemotyper “fingerprint” files for computational processing

Developed by Altamira & Molecular Networks, Funded by US FDA

Chemotype (CT) inventory profiling

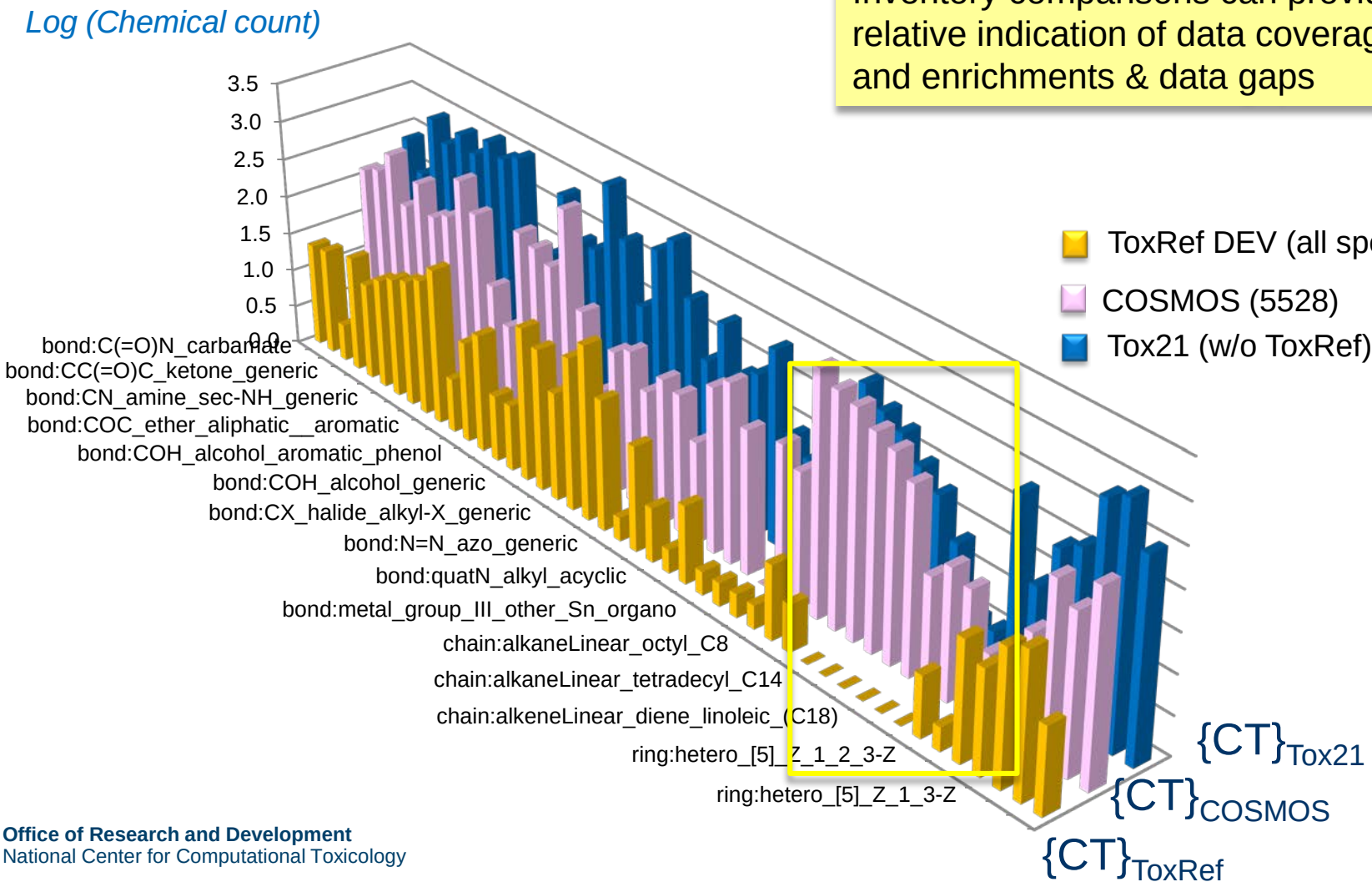
CT fingerprint provide a common structural basis for Inventory profiling & comparison

{CT} profile is a property of the Inventory

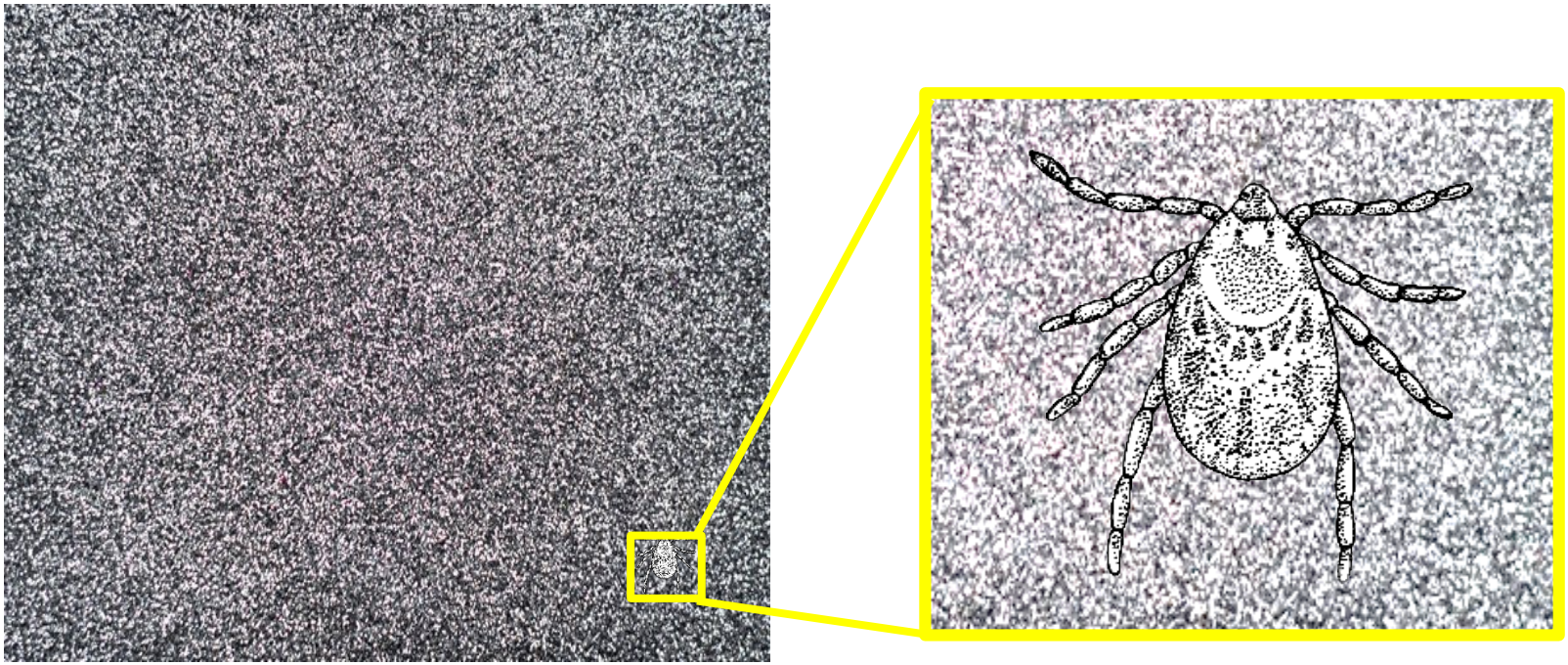


Chemotype inventory profiling

Inventory comparisons can provide relative indication of data coverage, and enrichments & data gaps



Concept of Enrichment: *Focus & Amplify to See*



Chemotype Enrichment, e.g.
Flame Retardant (FR) Use Category

- 42 FRs amenable to HTS included in ToxCast Ph3 & Tox21

$\{CT\}_{FR}$

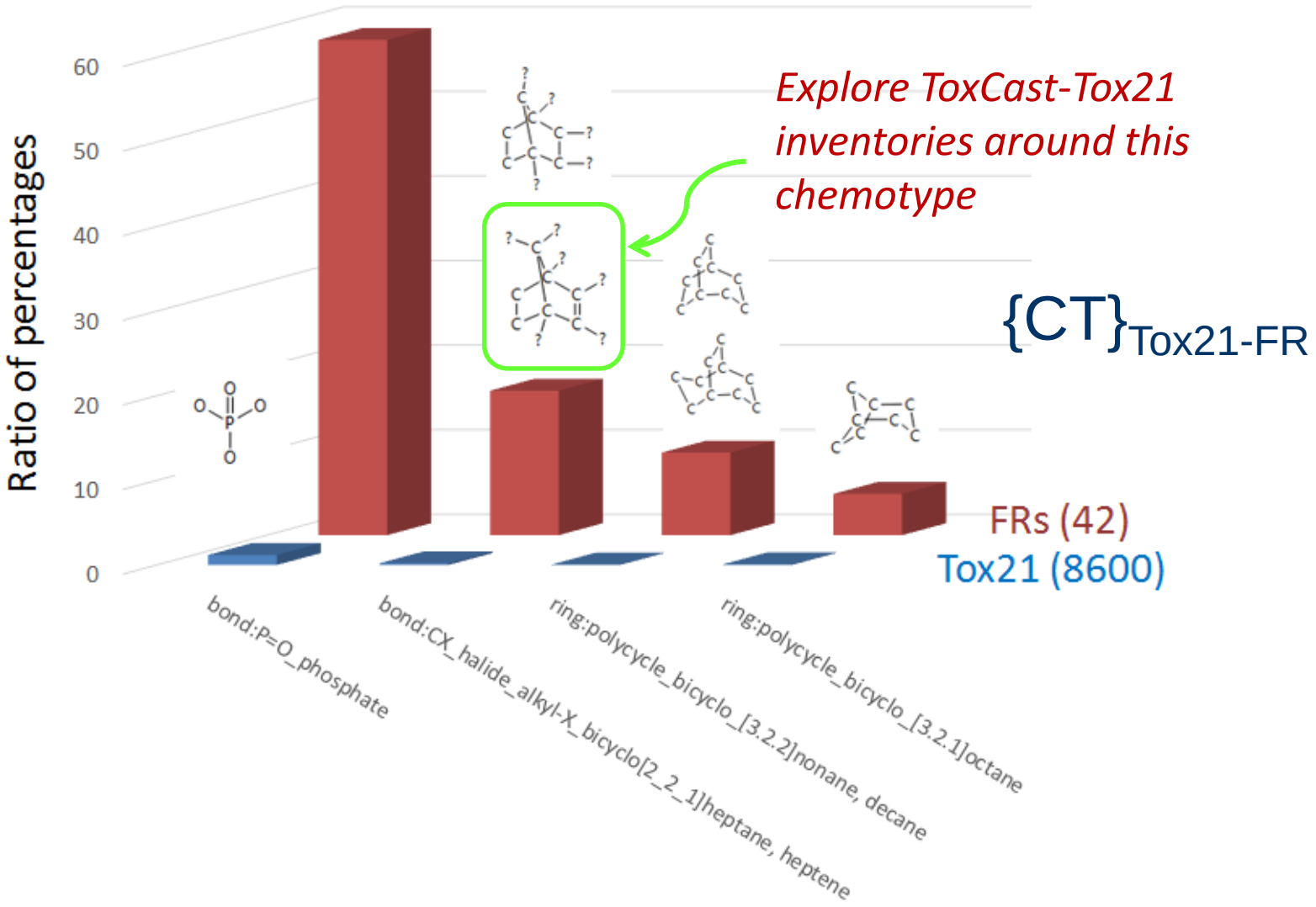
3296-90-0	2,2-Bis(bromomethyl)-1,3-propanediol
115-28-6	Chlorendic acid
2921-88-2	Chlorpyrifos
2385-85-5	Mirex
115-96-8	Tris(2-chloroethyl) phosphate
78-42-2	Tris(2-ethylhexyl) phosphate
115-86-6	Triphenyl phosphate
126-73-8	Tributyl phospha
79-94-7	3,3',5,5'-Tetrabro
13674-87-8	Tris(1,3-dichloro
1163-19-5	Decabromodiph
19660-16-3	2,3-Dibromoprop
563-04-2	Tri-m-cresyl pho
20120-33-6	Phosphonic acid dimethyl ester
868-85-9	Dimethyl hydrogen phosphite
756-79-6	Dimethyl methylphosphonate
124-64-1	Tetrakis(hydroxymethyl)phosphonium chloride
55566-30-8	Tetrakis(hydroxymethyl)phosphonium sulfate
1330-78-5	Tricresyl phosphate
512-56-1	Trimethyl phosphate

126-72-7	Tris(2,3-dibromopropyl) phosphate
78-51-3	Tris(2-butoxyethyl) phosphate
79-95-8	2,2',6,6'-Tetrachlorobisphenol A
118-79-6	2,4,6-Tribromophenol
56803-37-3	tert-Butylphenyl diphenyl phosphate
	te
	Tris(2-chloroisopropyl)phosphate
25155-23-1	TXP
598-72-1	2-Bromopropionic acid
3194-55-6	1,2,5,6,9,10-Hexabromocyclododecane
6145-73-9	Tris(2-chloropropyl) phosphate
26040-51-7	Bis(2-ethylhexyl) tetrabromophthalate
68937-41-7	Phenol, isopropylated, phosphate (3:1)
2781-11-5	Phosphonic acid, [[bis(2-hydroxyethyl)amino]methyl]-, diethyl ester
78-30-8	Tri-o-cresyl phosphate
4162-45-2	Ethanol, 2,2'-((1-methylethylidene)bis((2,6-dibromo-4,1-phenylene)oxy))bis-

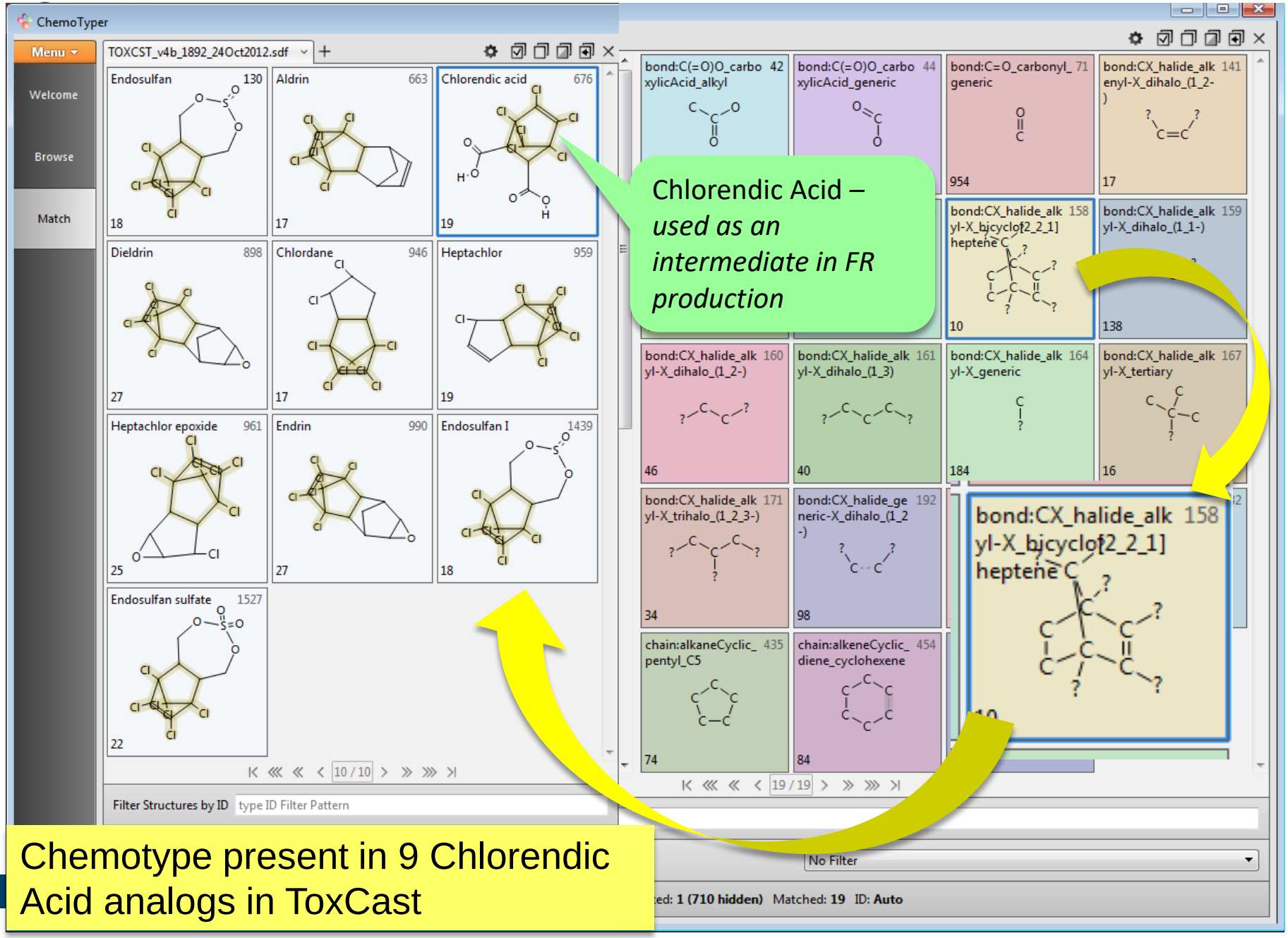
Are there enriched chemotypes within this FR subset relative to ToxCast & Tox21?

w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

Chemotype Enrichment, e.g.
Flame Retardant (FR) Use Category



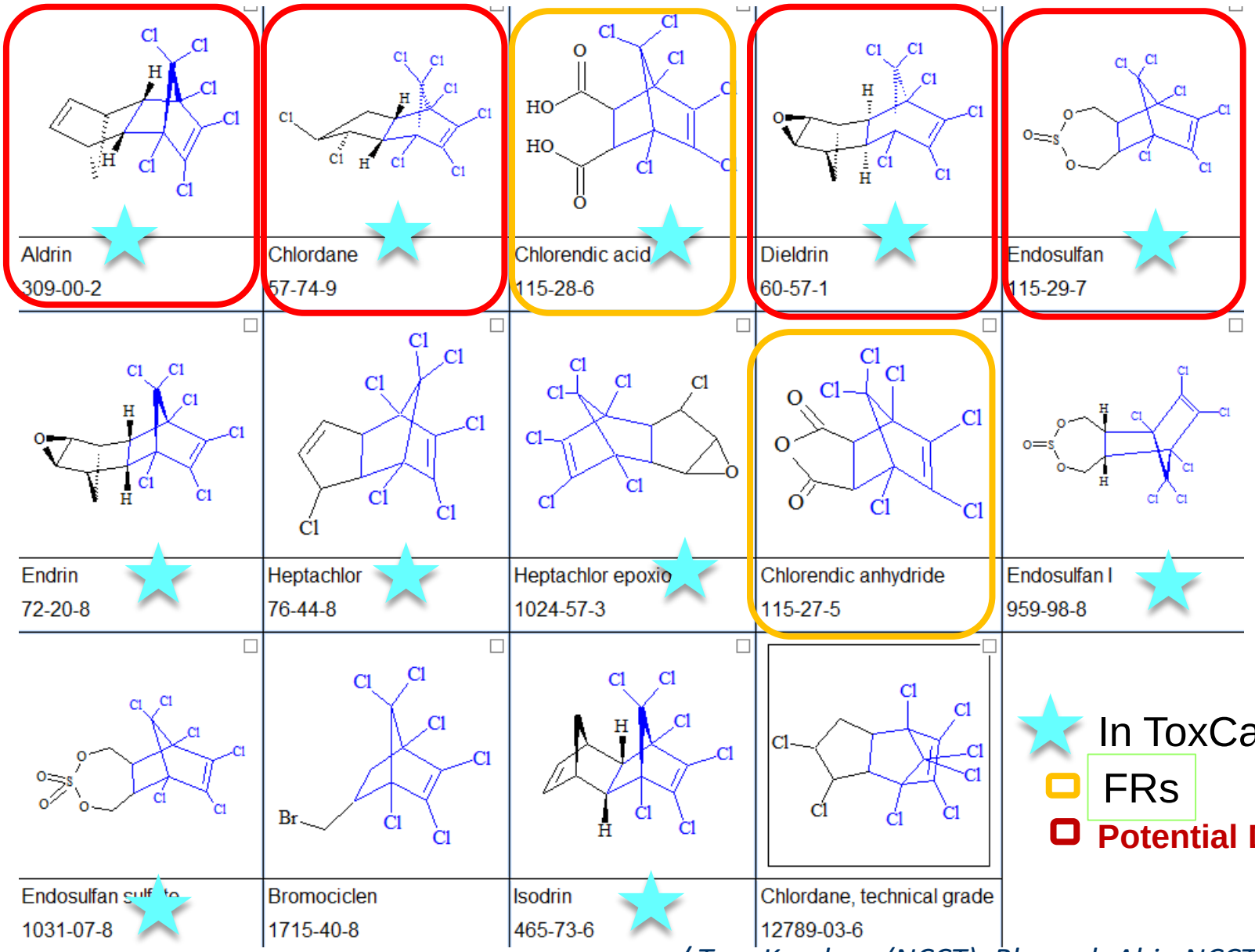
w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)



Chlorendic Acid – *used as an intermediate in FR production*

Chemotype present in 9 Chlorendic Acid analogs in ToxCast

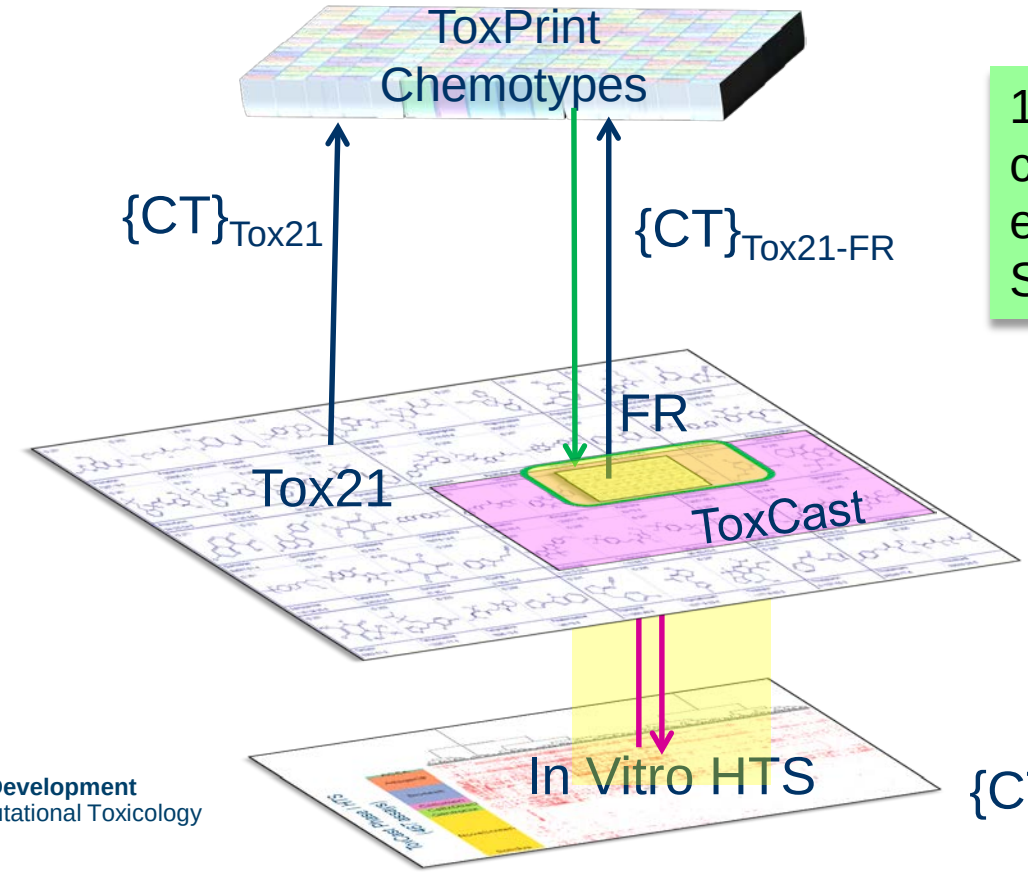
14 Tox21 Chemotype Analogs of Chlorendic Acid



w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

Chemotype-Assay Enrichments {CT-Assay}: Mining the data

Use FR-chemotypes to explore possible ToxCast assay associations relevant to this dataset



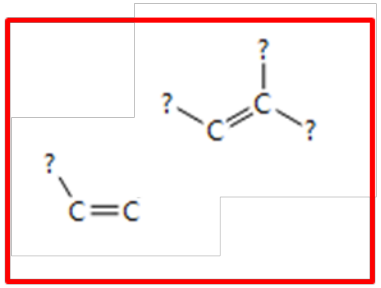
1. Identify all Tox21 chemicals containing enriched CTs for SAR follow-up

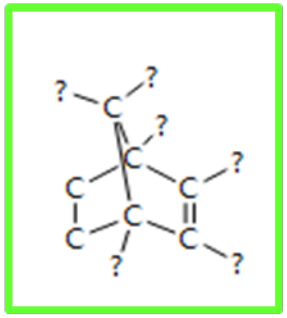
2. Identify ToxCast CT-Assay enrichments

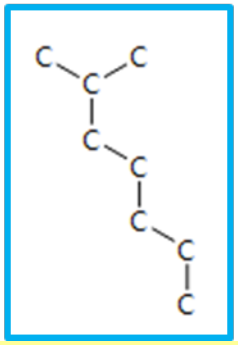
{CT-Assay}ToxCast-FR

Preliminary Results: {CT-Assay}$_{ToxCast\text{-}FR}$

Assay	FR Chemotype (CT)	Odds Ratio	# CT TP	# CT total	# assay positive	Total cmpds
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_2-)	6	8	46	71	1857
ATG_RARa_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_2-)	5	8	46	77	1857
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_dihalo_(1_3)	8	9	40	71	1857
ATG_RARg_TRANS	bond:CX_halide_alkyl-X_trihalo_(1_2_3-)	9	8	34	71	1857
ATG_RARa_TRANS	bond:CX_halide_alkyl-X_trihalo_(1_2_3-)	5	6	34	77	1857
ATG_RXRb_TRANS	chain:alkaneBranch_isooctyl_hexyl_2-methyl	6	8	17	261	1857
Tox21_TR_LUC_GH3_Antagonist	bond:CX_halide_alkyl-X_bicyclo[2_2_1]heptene	18	6	10	151	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkenyl-X_dihalo_(1_2-)	10	12	17	385	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkyl-X_bicyclo[2_2_1]heptene	35	9	10	385	1858
Tox21_MitochondrialToxicity_ratio	bond:CX_halide_alkyl-X_tertiary	8	10	15	385	1858







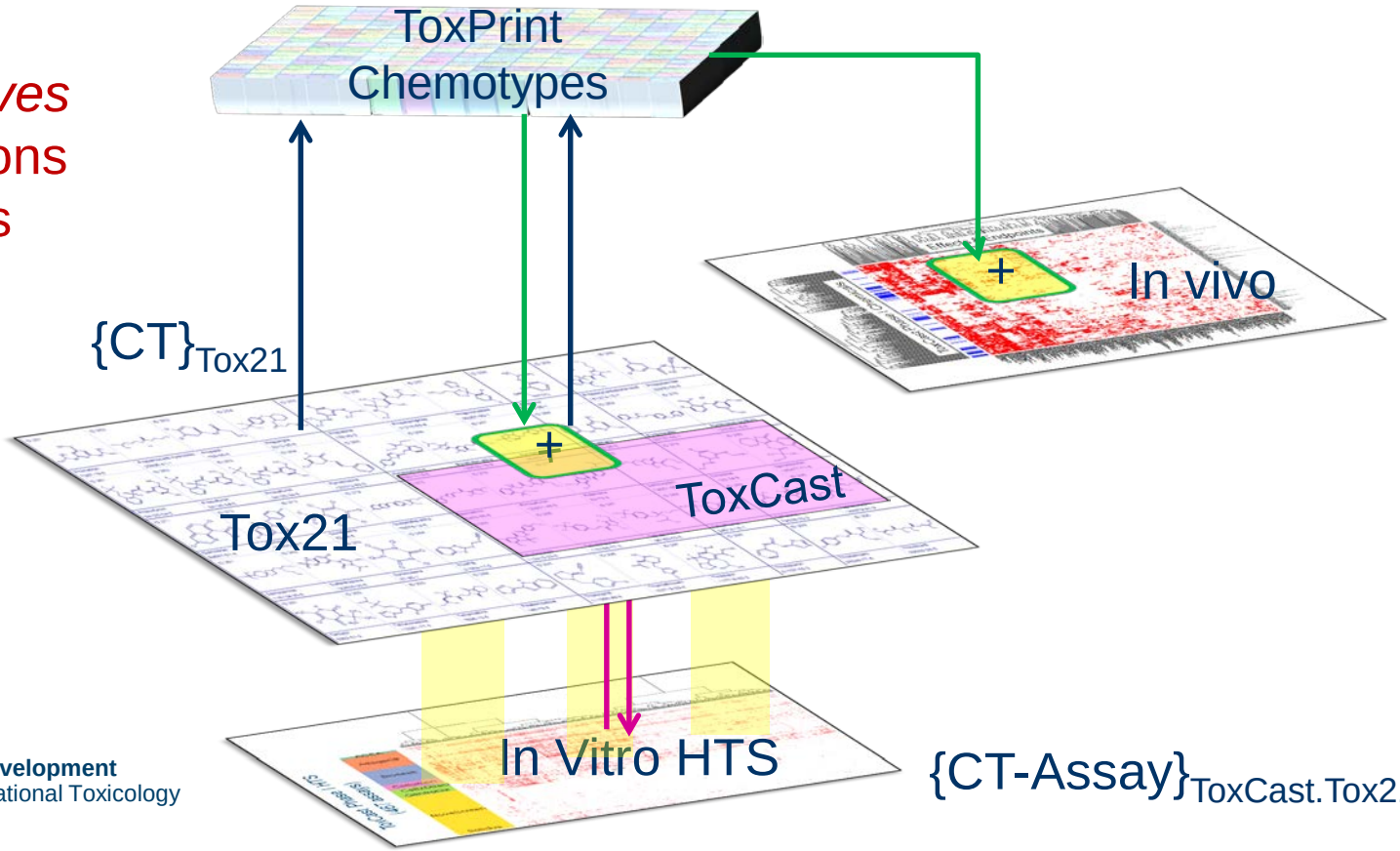
Significant {CT-Assay}$_{ToxCast\text{-}FR}$ associations potentially related to developmental outcomes

w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

Chemotype-Activity Enrichments

Create {CT-Activity} profiles for any biological activity subset to focus investigations into local CT domains and enhance HTS or structure-activity signal

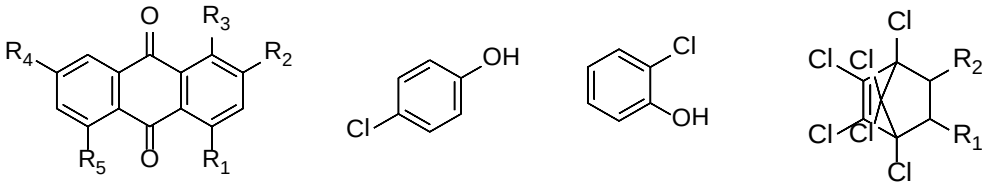
- *HTS actives*
- *Pathway actives*
- SAR predictions
- *In vivo* results



Tox21 Mitochondrial Membrane Potential (MMP) Disruption Assay Study

OBJECTIVES: To identify individual chemicals and general structural features associated with the decrease of mitochondrial membrane potential (MMP) in HepG2 cells.

- 913/8300 MMP “actives”
- Precompute 651 structural similarity clusters for Tox21 Inventory using ChemAxon proprietary fingerprints
  - ➢ 76 of 651 clusters significantly enriched for actives
  - ➢ representative features extracted from active clusters by visual inspection



(http://www.ncbi.nlm.nih.gov/pcsubstance; search term “tox21”).

Profiling of the Tox21 Chemical Collection for Mitochondrial Function I. Compounds that Decrease Mitochondrial Membrane Potential
Attene-Ramos MS, Huang R, Michael S, Witt KL, Richard A, Tice R, Simeonov A, Austin CP, Xia M, (*Submitted to EHP)*

Tox21 Mitochondrial Membrane Potential (MMP) Disruption Assay

United States Environmental Protection Agency

1. Compute $\{CT\}_{Tox21\text{-}MMP}$

	OR
chain:aromaticAlkane_Ph-C6	∞
bond:CX_halide_alkenyl-X_dihalo_(1_2-)	19.7
ring:hetero_[6_6]_O_benzopyrone_(1_4-)	19.0
ring:fused_PAH_phenanthrene	18.4
bond:metal_group_III_other_Sn_organo	17.5
bond:P=O_phosphate_thioate	14.6
bond:metal_group_III_other_Sn_generic	13.1
bond:COH_alcohol_aromatic_phenol	10.0

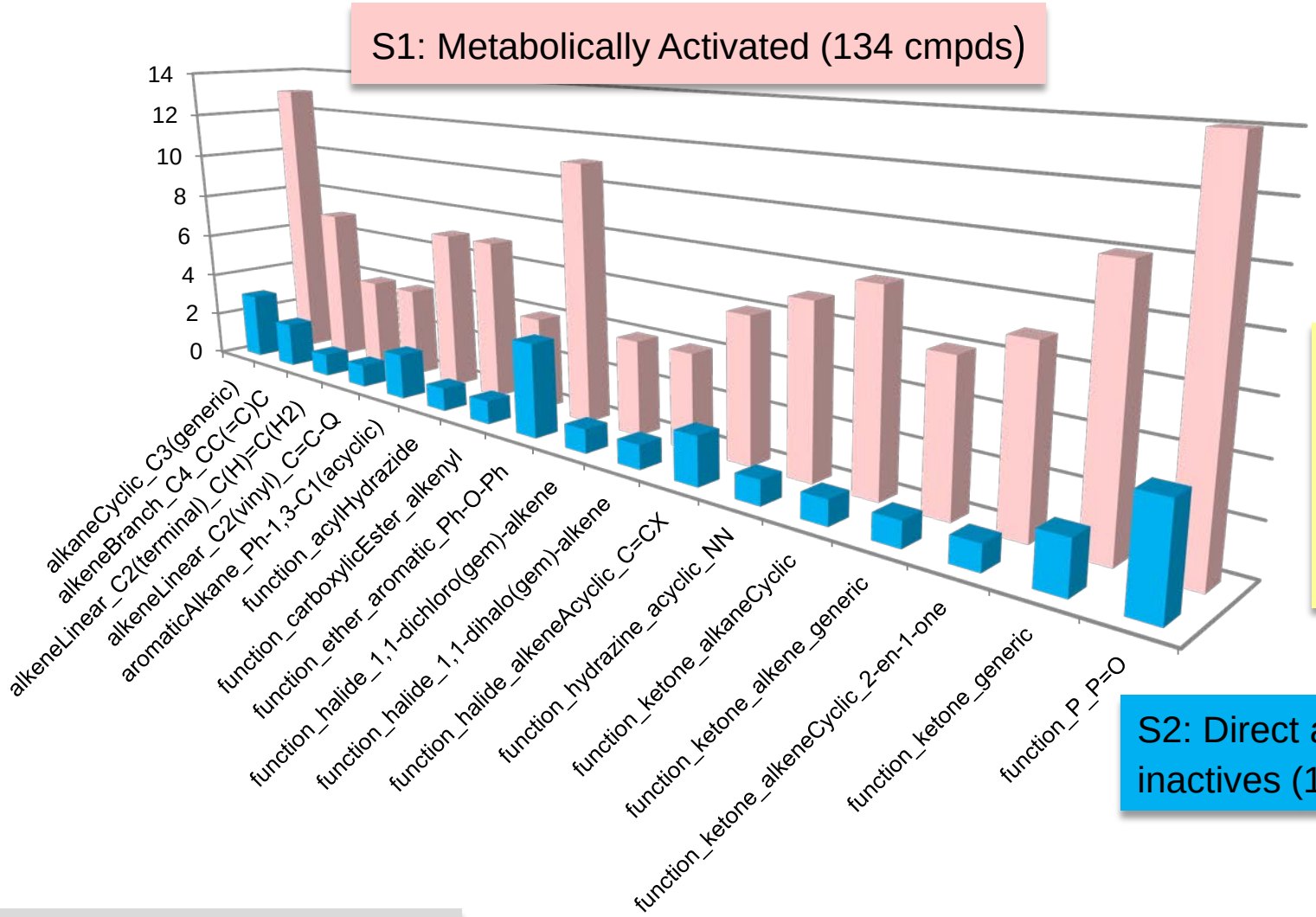
Most significantly enriched CTs in MMP active subset

Which Tox21 & ToxCast assays are enriched w/ positives in these {CT} subsets?

2. Compute $\{CT\text{-}Assay\}_{ToxCast\text{-}MMP}$

ToxCast/Tox21 Assays	{CT}MMP	OR	# TP (CT & Assay)	# CT Positive	# Assay positive	Total # Cmpds
APR_CellLoss_72h_dn	bond:C(=O)N_carbamate_dithio	∞	8	8	305	958
ATG_PXRE_CIS	bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O-Ph)_generic	∞	13	13	708	1857
ATG_NRF2_ARE_CIS	bond:metal_group_III_other_Sn_generic	∞	8	8	635	1857
NVS_GPCR_hORL1	bond:C(=O)N_carbamate_dithio	94.6	7	9	45	1067
Tox21_ERa_BLA_Agonist_ratio	ring:hetero_[6_6]_O_benzopyrone_(1_4-)	67.7	9	10	227	1858
ACEA_T47D_80hr_Positive	ring:hetero_[6_6]_O_benzopyrone_(1_4-)	61.7	9	10	240	1815
NVS_ENZ_rCOMT	bond:COH_alcohol_aromatic_phenol	50.4	11	115	13	1059
Tox21_PPARg_BLA_Agonist_ratio	bond:metal_group_III_other_Sn_generic	50.0	6	8	111	1858
OT_ER_ERaERb_1440	chain:aromaticAlkane_Ph-C6	43.7	9	11	181	1846
Tox21_AhR	bond:P=O_phosphate_thioate	43.5	12	14	237	1858
Tox21_MitochondrialToxicity_viability	bond:metal_group_III_other_Sn_generic	43.4	6	8	126	1858
APR_p53Act_24h_up	bond:C(=O)N_carbamate_dithio	36.8	7	8	160	958
Tox21_MitochondrialToxicity_ratio	ring:hetero_[6_6]_O_benzopyrone_(1_4-)	35.4	9	10	385	1858
OT_ER_ERaERb_0480	chain:aromaticAlkane_Ph-C6	35.4	9	11	218	1853
Tox21_AhR	ring:hetero_[6_6]_O_benzopyrone_(1_4-)	28.4	8	10	237	1858

Chemotype Activity Profiling: e.g. Modeling *in vivo* activity subsets



Significantly different {CT} profiles in S1 vs S2 subsets

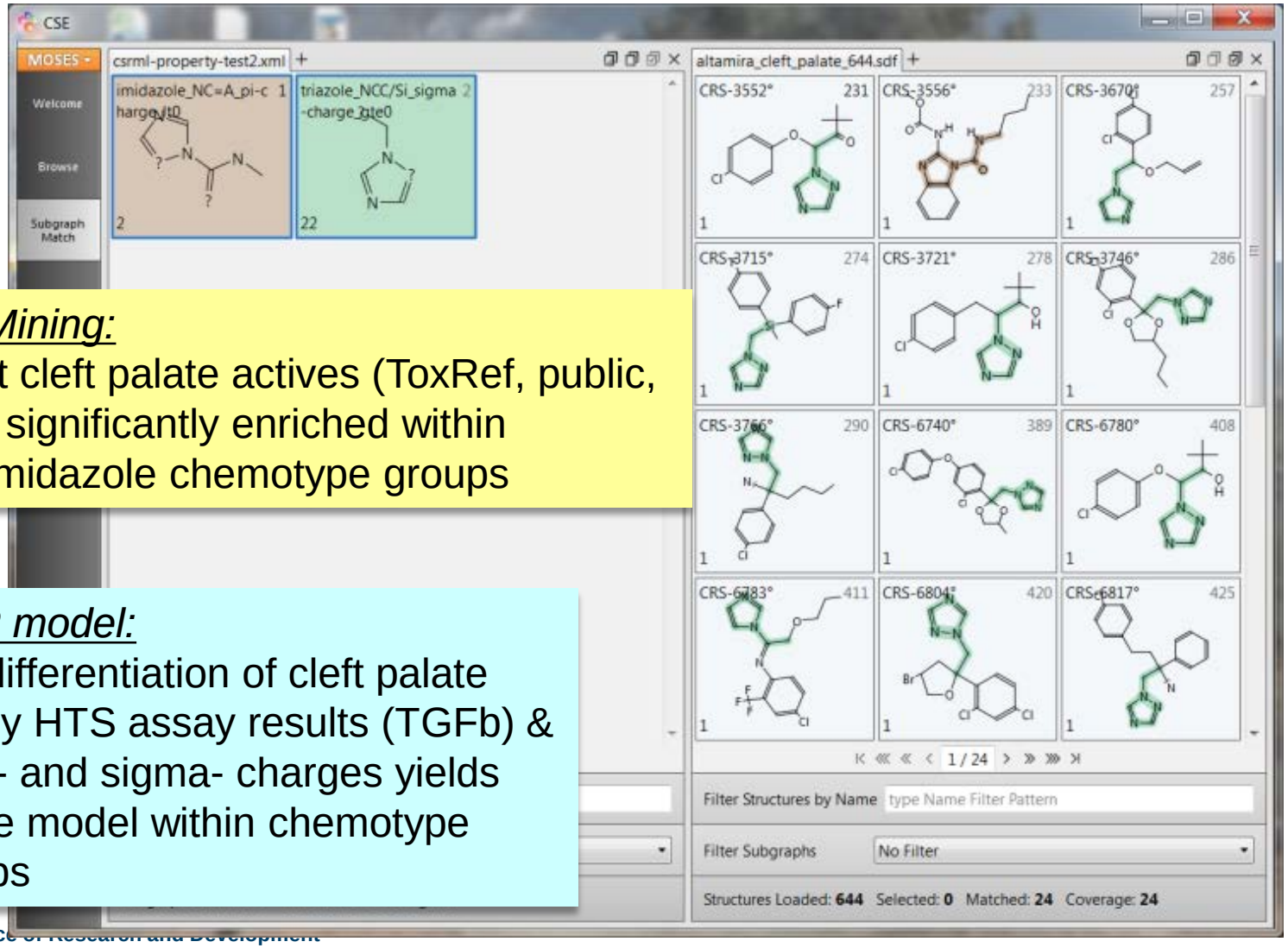
Use S1 {CT} set to predict chemical space likely to require metabolic activation for Rat Carcinogenicity

S2: Direct acting & inactives (157 cmpds)

**Altamira beta version of ToxPrint Chemotypes*

ToxCast Phase I (291 total) Rat Carcinogenicity Study using ToxRefDB & Meteor:Derek workflow, *Volarath et al.*

Chemotype Activity Profiling: e.g. Data mining & QSAR models



1. Data Mining:

Tox21 rat cleft palate actives (ToxRef, public, CERES) significantly enriched within triazole/imidazole chemotype groups

2. QSAR model:

Further differentiation of cleft palate actives by HTS assay results (TGFb) & partial pi- and sigma- charges yields predictive model within chemotype subgroups

C Yang et al., Altamira

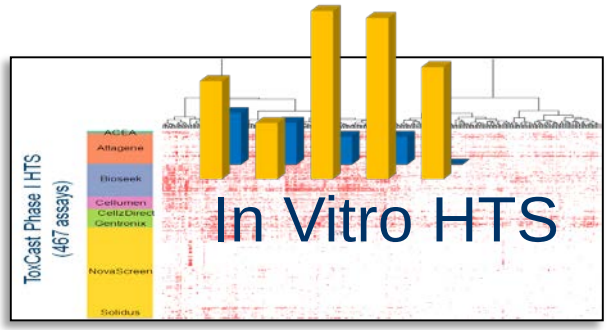
Building a public chemotype {CT} "knowledge- base"

3. Compute relative {CT} enrichments for chemical sets of interest

- *Use categories (exposure), e.g. Flame Retardants*
- *Activity subsets*
- *Metabolically activated toxicants*
- *Assay-assay associations to inform pathways*

2. Precompute CT- Assay enrichments for data inventories of interest

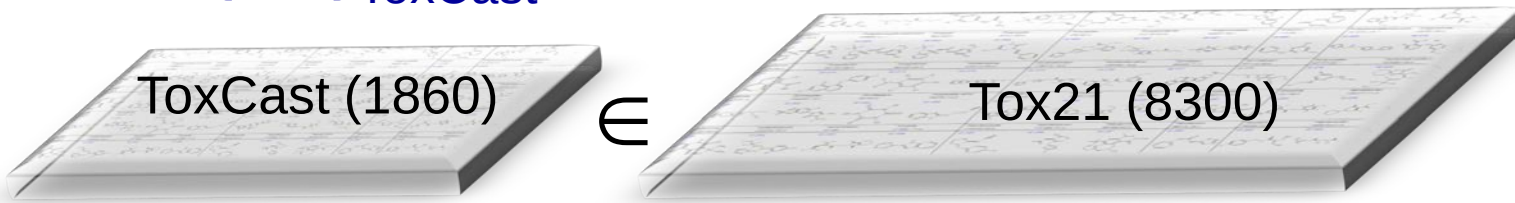
{CT-Assay}$_{ToxCast,Tox21}$



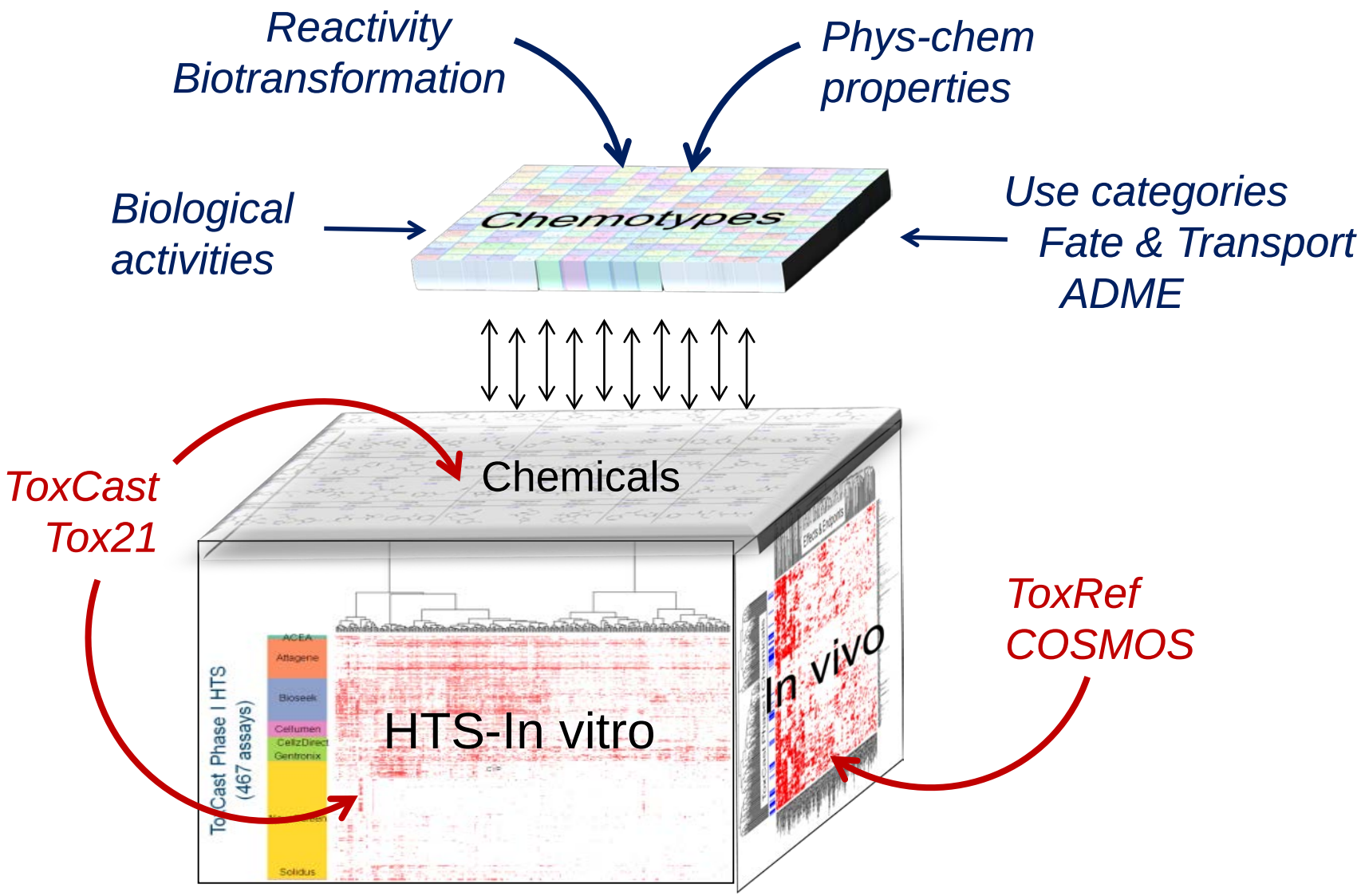
1. Precompute CT profiles for chemical inventories of interest

{CT}$_{ToxCast}$

{CT}$_{Tox21}$



Building a public chemotype "knowledge- base"



Acknowledgements:

- EPA NCCT ToxCast Team (K. Crofton, R. Thomas et al)
  - Chris Grulke (DSSTox, Chemotypes)
  - Richard Judson, Keith Houck (ToxCast)
  - Matt Martin (ToxRefDB)
- Tox21 Collaborative Team – MMP study
  - Ruili Huang, Menghang Xia - NCATS
- Flame Retardants
  - Bhavesh Ahir, Tom Knudsen, Ray Tice (NTP), Nicole Kleinstrauer
- External Collaborators – ToxPrint Chemotypes
  - Molecular Networks: Chihae Yang, Aleksey Tarkhov, Christof Schwab
  - Altamira: Jim Rathman
  - U.S. FDA: Kirk Arvidson, Patra Volarath

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