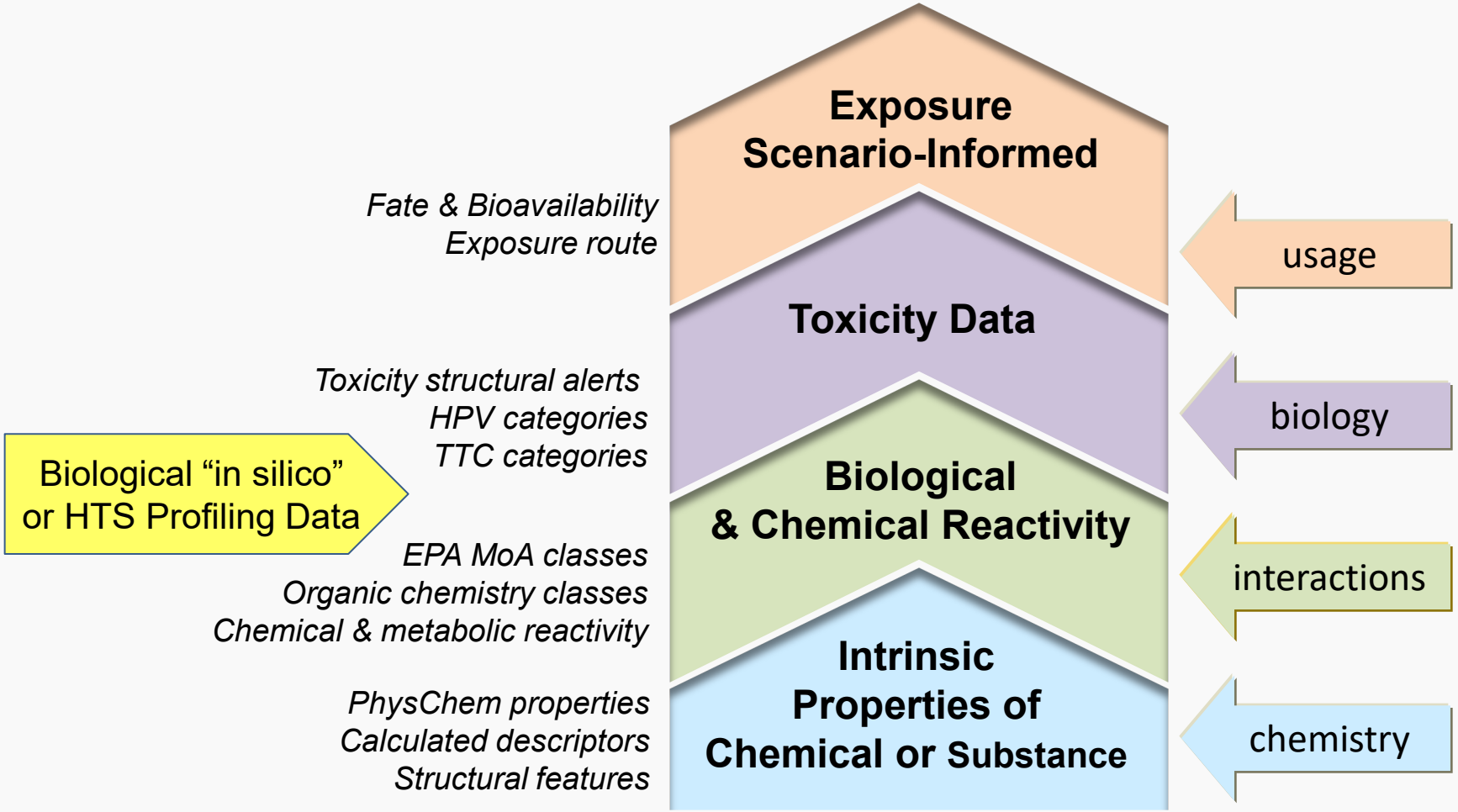
Inverting the SAR Paradigm:

Applications of a Chemotype-Enrichment Approach within EPA's Computational Toxicology Programs

Ann Richard
Computational Chemistry & Cheminformatics Branch
Chemical Characterization & Exposure Division
Office of Research & Development
U.S. Environmental Protection Agency

QSAR 2021 Virtual Meeting, June 7-9

Chemical/Substance representations for hazard & risk evaluation



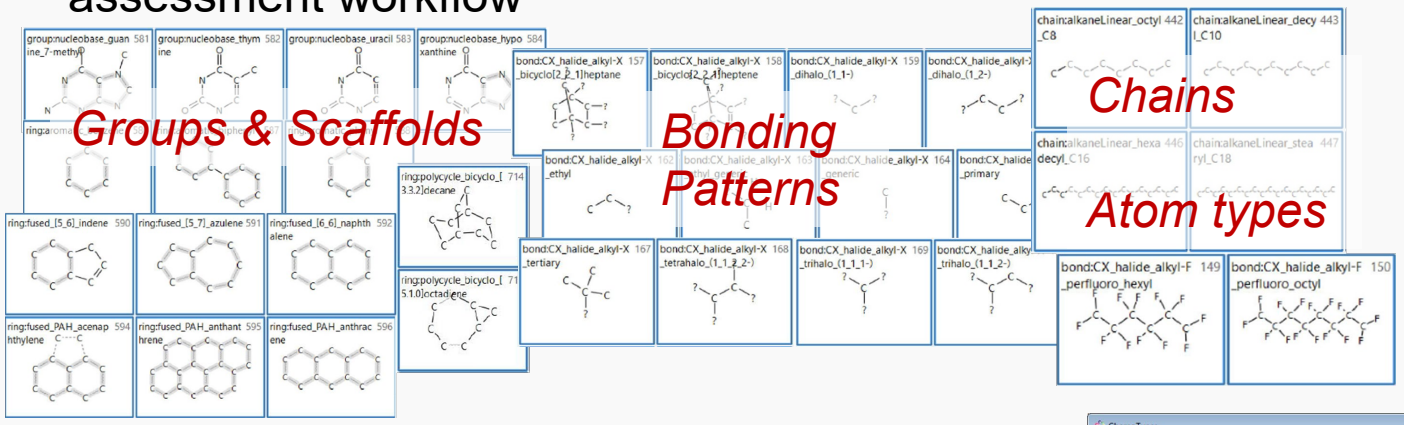
CSS workshop, April 2011

ToxPrints: A Public Set of Chemotypes

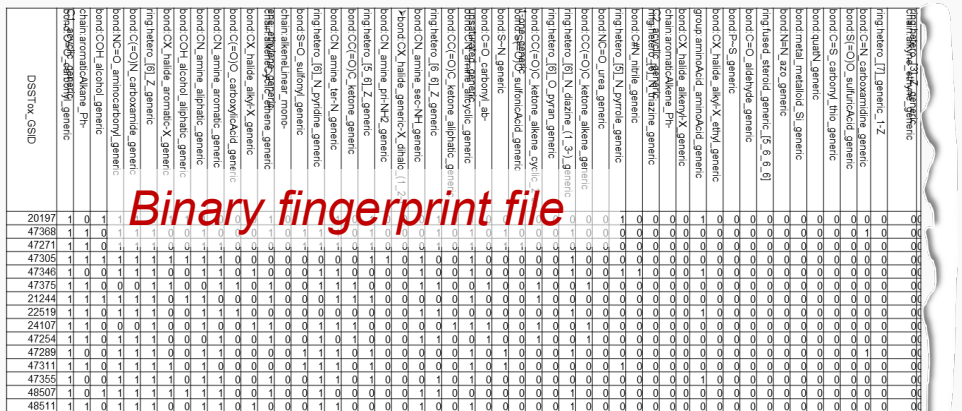
→ *Clear, reproducible means for defining & visualizing local chemistry*

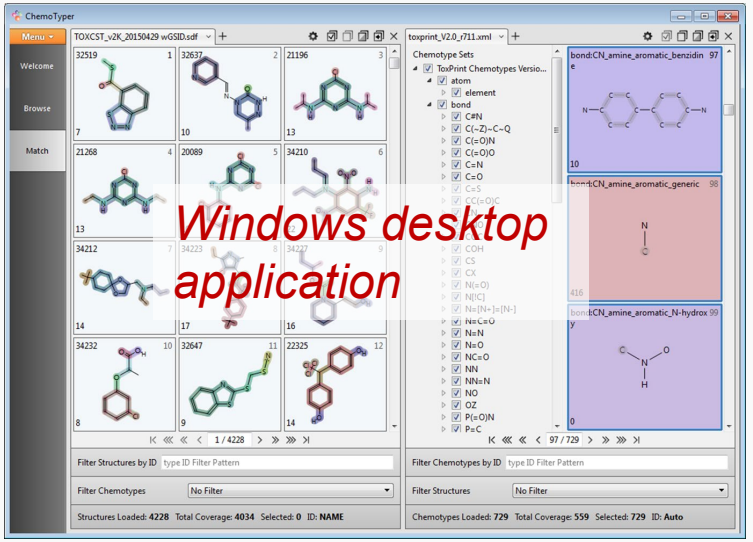
ToxPrints: http://www.toxprint.org

729 features important to EPA & FDA's "chemical exposure" landscape and safety assessment workflow



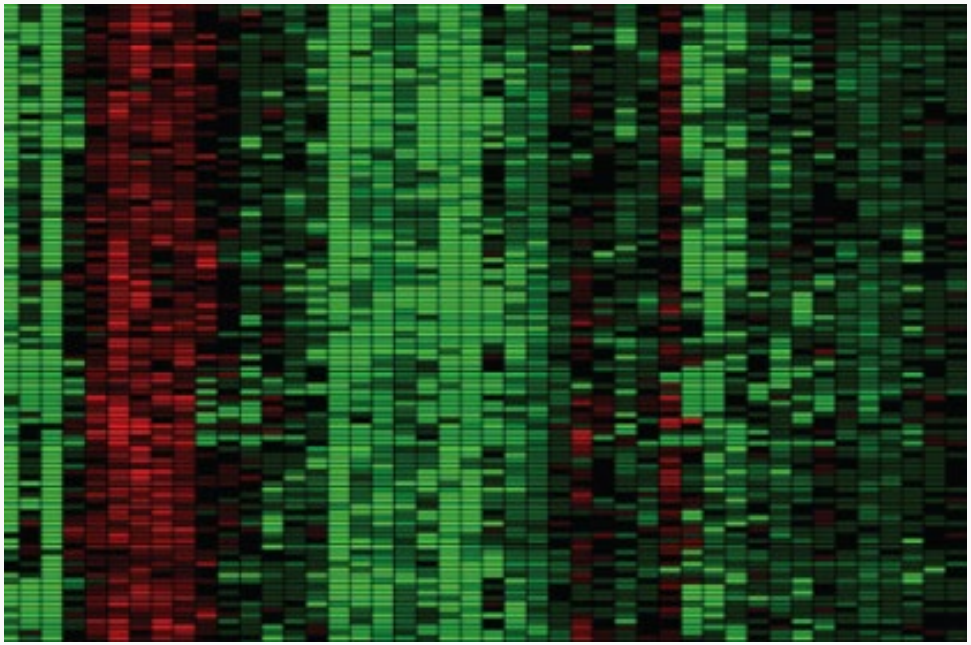
Chemotyper: http://www.chemotyper.org





Chemotype-Enrichment Approaches:

Finding chemical-activity patterns in the noise



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

42 chemicals labeled as "Flame Retardants" included in ToxCast & Tox21

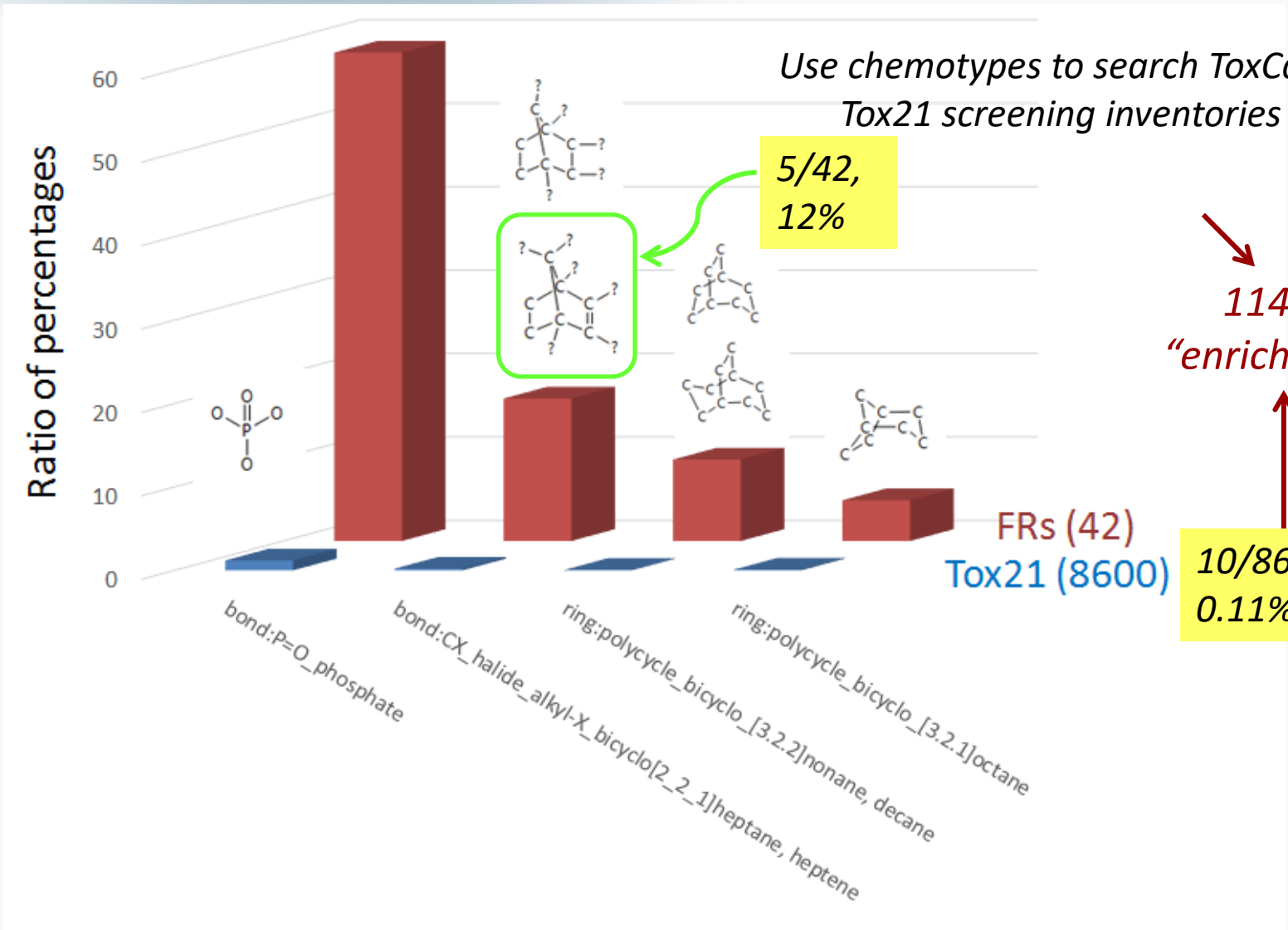
3296-90-0	2,2-Bis(bromomethyl)-1,3-propanediol
115-28-6	Chlorendic acid
2921-88-2	Chlorpyrifos
2385-85-5	Mirex
115-96-8	Tris(2-chloroethyl) phosphate
78-42-2	Tris(2-ethylhexyl) phosphate
115-86-6	Triphenyl phosphate
126-73-8	Tributyl phosphate
79-94-7	3,3',5,5'-Tetrabromobisphenol A
13674-87-8	Tris(1,3-dichloro-2-propyl)phosphate
1163-19-5	Decabromodiphenyl
19660-16-3	2,3-Dibromopropyl a
563-04-2	Tri-m-cresyl phosph
20120-33-6	Phosphonic acid, (3- dimethyl ester
868-85-9	Dimethyl hydrogen p
756-79-6	Dimethyl methylpho
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
26444-49-5	Cresyl diphenyl phosphate
1241-94-7	2-Ethylhexyl diphenyl phosphate
29761-21-5	Isodecyl diphenyl phosphate
632-79-1	Tetrabromophthalic anhydride
117-08-8	Tetrachlorophthalic anhydride
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?

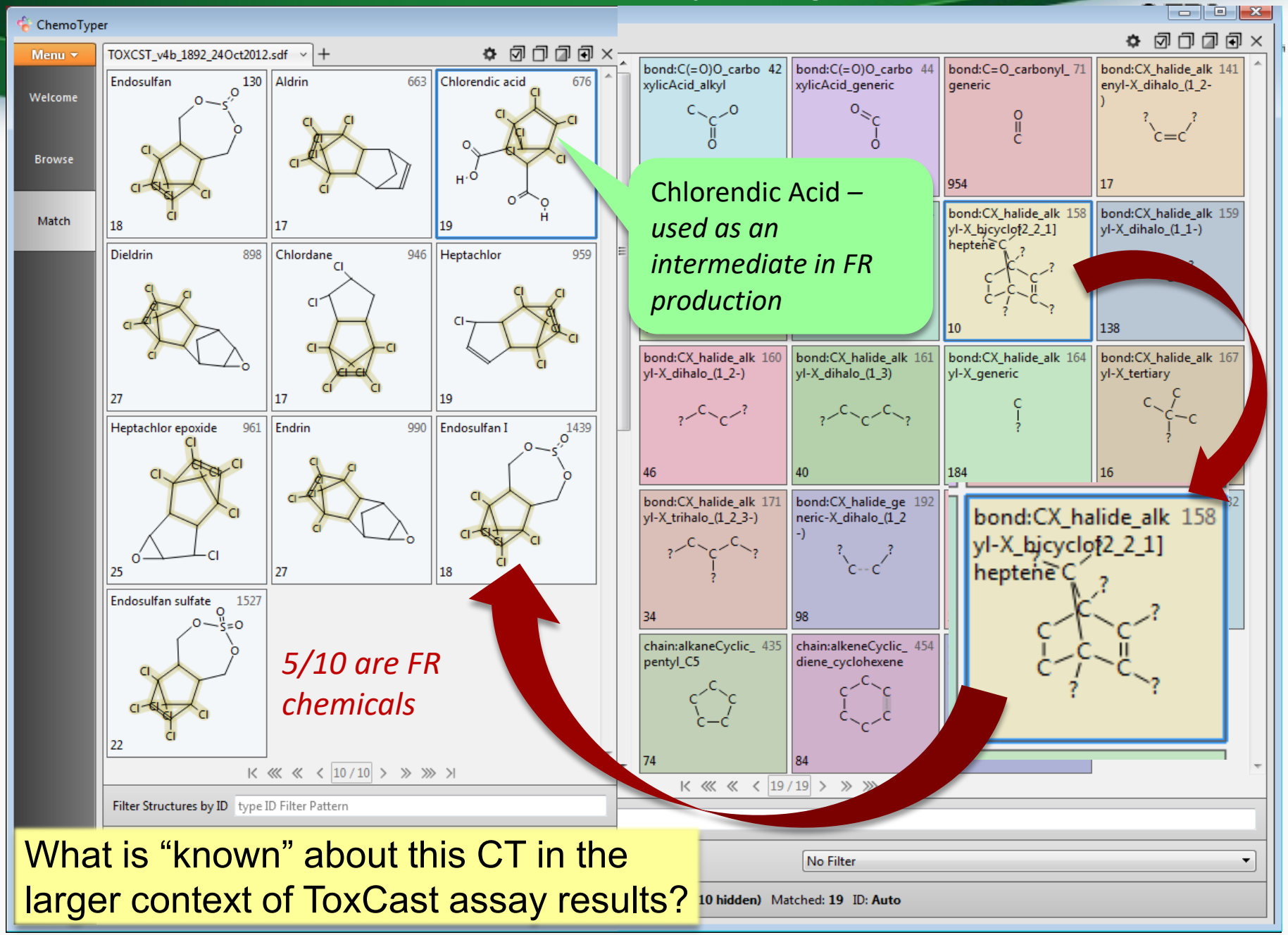
w/ Tom Knudsen (NCCT), Bhavesh Ahir, NCCT (ORISE)

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



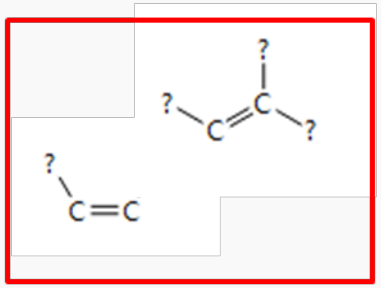
w/ Tom Knudsen and Bhavesh Ahir

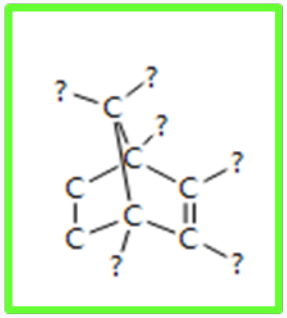
www.chemotyper.org

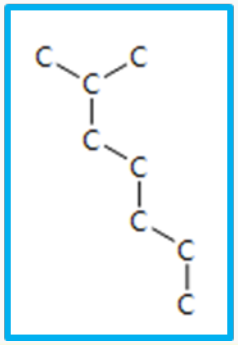


{CT-Assay}$_{ToxCast-FR}$

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





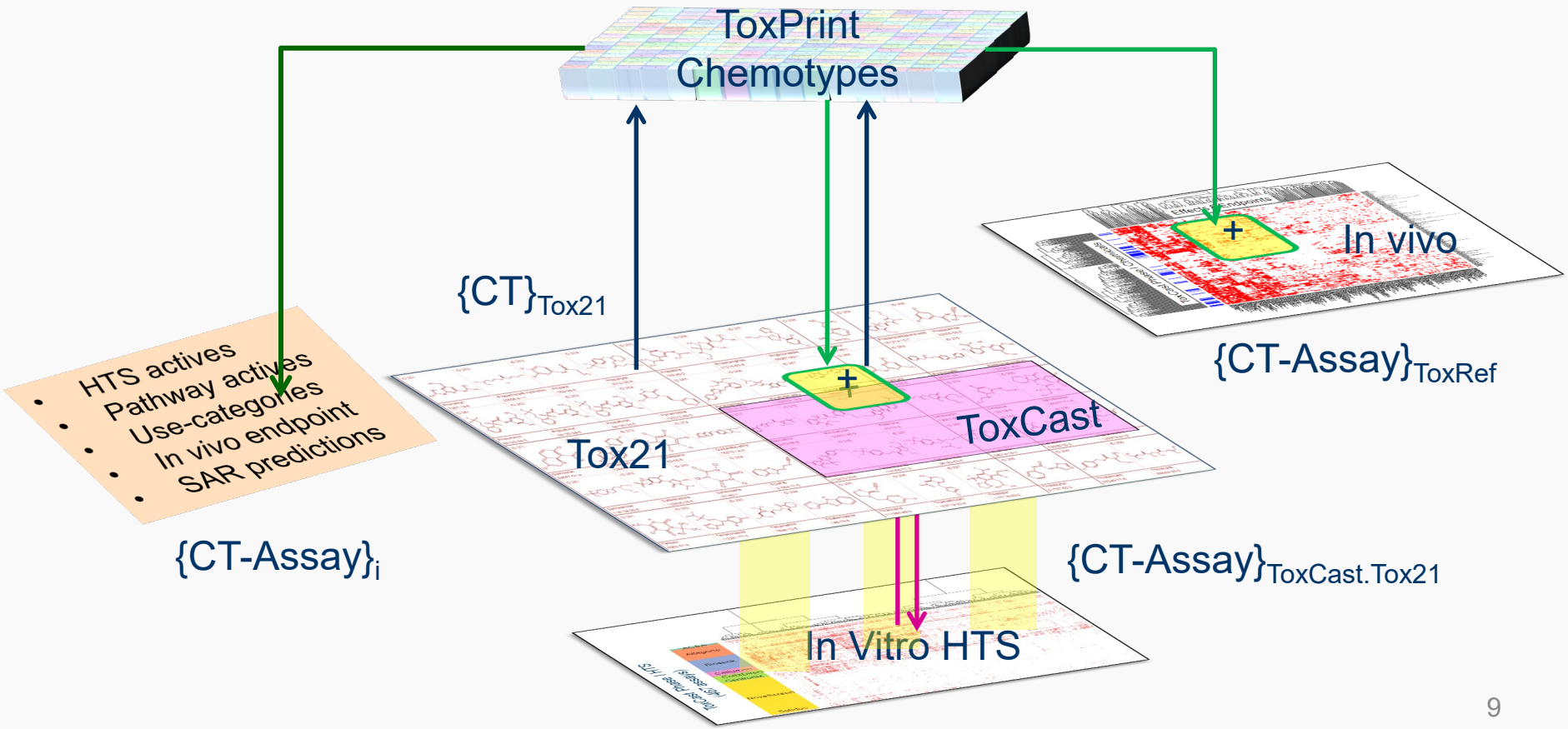


Significant {CT-Assay}$_{ToxCast-FR}$ associations potentially related to developmental outcomes

w/ Tom Knudsen and Bhavesh Ahir

Chemotype-Activity Enrichments

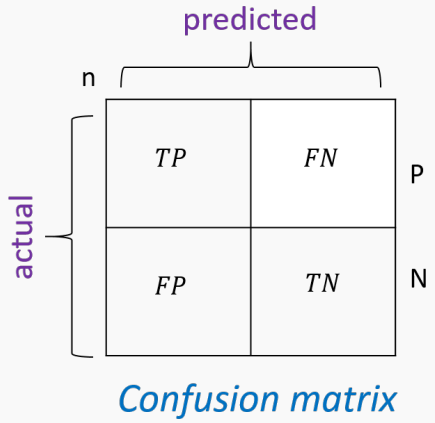
- ✓ Create {CT-Activity} enrichment profiles for any “active” subspace of a test set
- ✓ Focus studies in local CT domains
- ✓ Enhance HTS or structure-activity signal



Computing CT-Assay "Enrichments"

Simple statistical thresholds & filters of significance:

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-butyl_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**



Confusion matrix

- Odds Ratio ≥ 3, *conveys simple fractional enrichment*
- Fischer's exact p value ≤ 0.05, *compensates for size of dataset*
- T_{pos} (TP) ≥ 3, *require at least 3 chemicals with CT in Positives*

Binary "Activity" Hit Call + Fingerprint = CT Enrichment Table

DTXSID	PREFERRED_NAME	Hit_Call	bond:CX_halide_alkenyl-Cl_dichloro_(1_1-)	bond:CX_halide_alkenyl-X_acyclic_generic	bond:CX_halide_alkenyl-X_dihalo_(1_1-)	bond:CX_halide_alkenyl-X_generic	bond:CX_halide_alkyl-Cl_trichloro_(1_1_1-)	bond:CX_halide_alkyl-X_aromatic_alkane	bond:CX_halide_alkyl-X_aromatic_generic	bond:CX_halide_alkyl-X_benzyl_alkane	bond:CX_halide_alkyl-X_dihalo_(1_3)	bond:CX_halide_alkyl-X_secondary	bond:CX_halide_alkyl-X_trihalo_(1_1_1-)	bond:CX_halide_alkyl-X_trihalo_(1_1_2-)	bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O-Ph)	bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O-Ph)_ger
DTXSID7021233	FD&C Red 3	1	0	0	0											
DTXSID1041978	Flufenoxuron	0	0	0	0											
DTXSID7024160	Lactofen	1	0	0	0											
DTXSID7024241	Oxyfluorfen	1	0	0	0											
DTXSID5035957	Cyfluthrin	1	1	1	1											
DTXSID5032498	Triclosan	1	0	0	0											
DTXSID8023216	3,5,3'-Triiodot[illegible]nin[illegible]	1	0	0	0											
DTXSID0020022	5-(2-Chloro-4-([illegible]uor	0	0	0	0											
DTXSID7024112	Fomesafen	0	0	0	0											
DTXSID1034503	Cyhalofop-butyl	0	0	0												
DTXSID0032605	Methyl (RS)-2-[4-(2,4-	0	0	0												
DTXSID5048186	Tetrac	0	0	0	0											
DTXSID2045232	Tiratricol	0	0	0	0											
DTXSID7020970	Nitrofen	0	0	0	0											
DTXSID6028022	1-(Bromomethyl)-3-ph	1	0	0	0											
DTXSID7032559	λ-Cyhalothrin	1	0	1	0											
DTXSID1032648	Z-Tetrachlorvinphos	1	0	1	0											
DTXSID5032315	4,5-Dichloro-2-octyl-3	1	0	0	0											
DTXSID5032654	Triforine	1	0	0	0											
DTXSID4027527	1,2,5,6,9,10-Hexabrom	1	0	0	0											
DTXSID5021413	Tris(2,3-dibromopropy	1	0	0	0											
DTXSID8034324	Phoxim	1	0	0	0											
DTXSID3032416	Cybutryne	1	0	0	0											

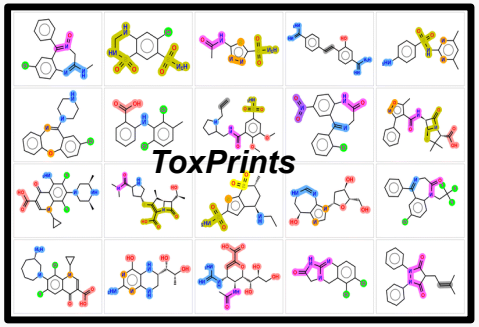
ToxPrint ID	Chemotype Label	CT-Tot	TP	FP	FN	TN	OR	P-Value
Txp-117	bond:COC_ether_aromatic	41	9	32	103	1627	4.44	7.4E-04
Txp-13	bond:C#N_nitrile_generic	52	9	43	103	1616	3.28	4.3E-03
Txp-134	bond:CS_sulfide_dialkyl	32	7	25	105	1634	4.36	3.0E-03
Txp-137	bond:CX_halide_alkenyl-Cl_dichloro_(1_1-)	7	3	4	109	1655	11.39	7.1E-03
Txp-139	bond:CX_halide_alkenyl-X_acyclic_generic	16	5	11	107	1648	7.00	2.3E-03
Txp-140	bond:CX_halide_alkenyl-X_dihalo_(1_1-)	8	3	5	109	1654	9.10	1.1E-02
Txp-142	bond:CX_halide_alkenyl-X_generic	40	9	31	103	1628	4.59	6.1E-04
Txp-146	bond:CX_halide_alkyl-Cl_trichloro_(1_1_1-)	15	4	11	108	1648	5.55	1.2E-02
Txp-153	bond:CX_halide_alkyl-X_aromatic_alkane	6	3	3	109	1656	15.19	4.3E-03
Txp-154	bond:CX_halide_alkyl-X_aromatic_generic	6	3	3	109	1656	15.19	4.3E-03
Txp-155	bond:CX_halide_alkyl-X_benzyl_alkane	41	7	34	105	1625	3.19	1.3E-02
Txp-161	bond:CX_halide_alkyl-X_dihalo_(1_3)	39	7	32	105	1627	3.39	9.5E-03
Txp-166	bond:CX_halide_alkyl-X_secondary	13	3	10	109	1649	4.54	4.4E-02
Txp-169	bond:CX_halide_alkyl-X_trihalo_(1_1_1-)	96	16	80	96	1579	3.29	2.1E-04
Txp-170	bond:CX_halide_alkyl-X_trihalo_(1_1_2-)	29	5	24	107	1635	3.18	3.2E-02
Txp-183	bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O	40	9	31	103	1628	4.59	6.1E-04
Txp-184	bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O	14	6	8	106	1651	11.68	1.1E-04
Txp-239	bond:P=O_phosphate_thioate	14	5	9	107	1650	8.57	1.2E-03
Txp-260	bond:P~S_generic	36	8	28	104	1631	4.48	1.4E-03
Txp-423	chain:alkaneBranch_t-butyl_C4	71	12	59	100	1600	3.25	1.2E-03
Txp-433	chain:alkaneCyclic_propyl_C3	25	6	19	106	1640	4.89	3.7E-03
Txp-557	group:ligand_path_4_bidentate_ethylenediamine	10	3	7	109	1652	6.50	2.1E-02
Txp-6	atom:element_metal_transistion_metal	10	3	7	109	1652	6.50	2.1E-02

Automated Chemotype-Enrichment Workflow (CTEW)

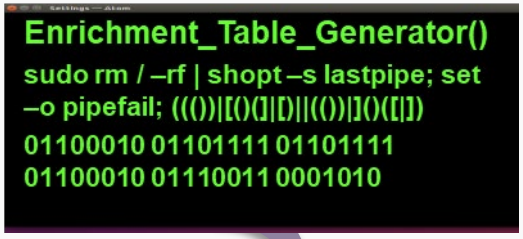
ID	Endpoint
DTXCID50675904	0
DTXCID50675924	0
DTXCID50675929	0
DTXCID50675904	1
DTXCID50675924	1
DTXCID50675929	0
DTXCID50675904	1
DTXCID50675924	0
DTXCID50675929	0
DTXCID50675904	0
DTXCID50675924	0
DTXCID50675929	0

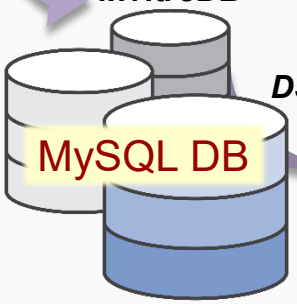
Generate binary "Hitcall Matrix"

Generate fingerprints



Calculate enrichment statistics





Store CTEW results in DB

Generate CT-Enrichment Table Results

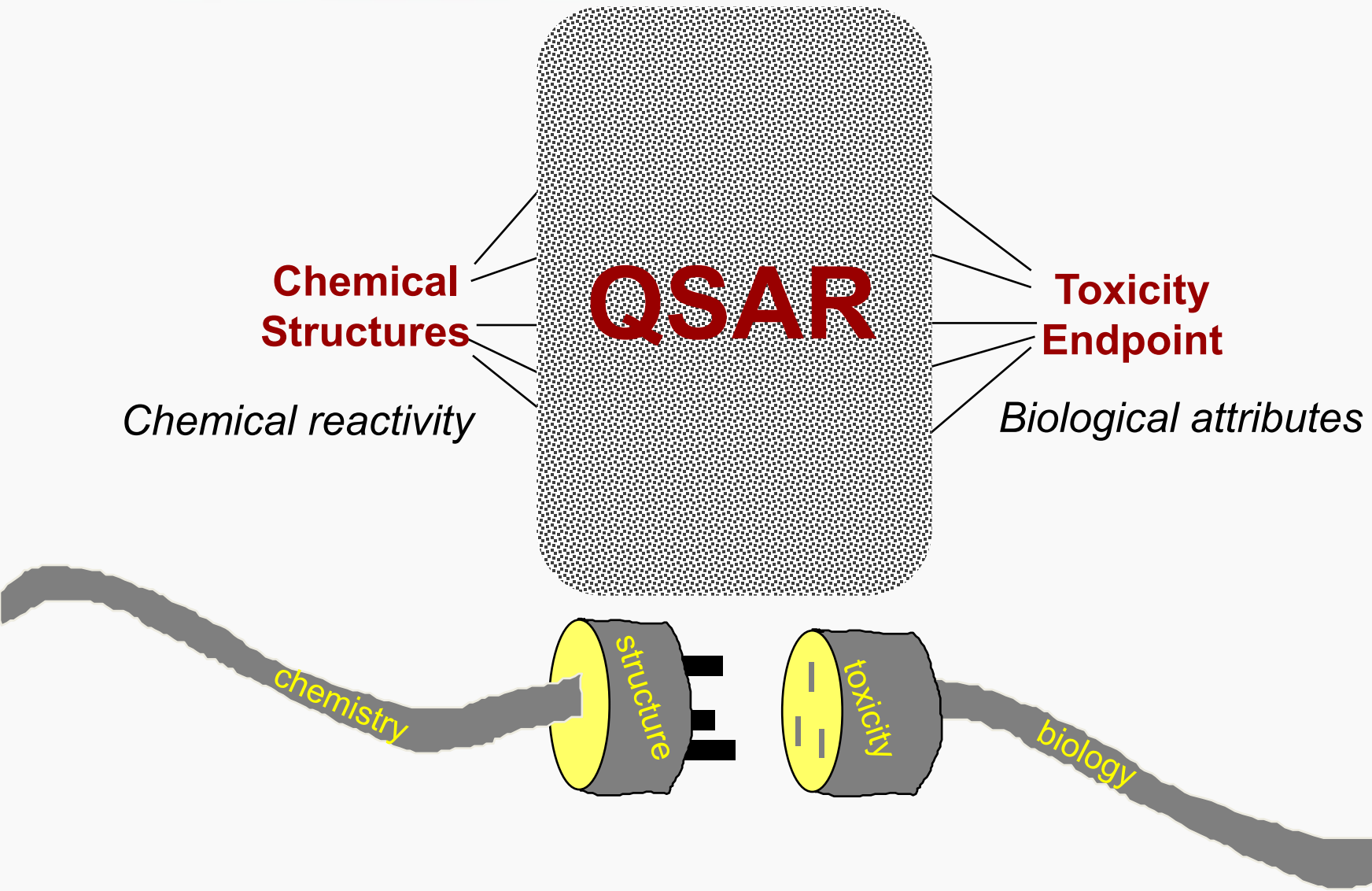
descriptors_name	label	CT-Tot	TP	FP	FN	TN	BA	OR	P-Val	Inv OR	Inv P-val
Txp-1	atom:element_main_group	2	1	1	1827	5180	0.6196303963661194	2.8352489471435547	0.4536220133304596	0.35270270705223083	0.9320695551467896
Txp-10	bond:C#N_cyano_cyanohydrin	5	1	4	1827	5177	0.46957454085350037	0.7084017395973206	0.7794185876846313	1.4116283655166626	0.6100181937217712
Txp-100	bond:CN_amine_pri-NH2_alkyl	112	25	87	1803	5094	0.4808981418609619	0.8118652701377869	0.8468714356422424	1.2317314147949219	0.21217265725135803
Txp-101	bond:CN_amine_pri-NH2_aromatic	341	84	257	1744	4924	0.4923933148384094	0.9228215217590332	0.7525069713592529	1.0836331844329834	0.28962212800979614
Txp-102	bond:CN_amine_pri-NH2_generic	459	110	349	1718	4832	0.4886806607246399	0.8864842653274536	0.8697083592414856	1.1280516386032104	0.1554924845695495
Txp-103	bond:CN_amine_sec-NH_alkyl	199	44	155	1784	5026	0.4795689284801483	0.7997395992279053	0.9174827933311462	1.2504069805145264	0.11147873103618622
Txp-104	bond:CN_amine_sec-NH_aromatic	151	36	115	1792	5066	0.4885549545288086	0.8849766850471497	0.7643690705299377	1.129973292350769	0.298209011554718
Txp-105	bond:CN_amine_sec-NH_aromatic_aliphatic	97	22	75	1806	5106	0.4827597141265869	0.82932448387146	0.8107256293296814	1.2058006525039673	0.2608259320259094
Txp-106	bond:CN_amine_sec-NH_generic	253	58	195	1770	4986	0.48362982273101807	0.8378617763519287	0.8932034373283386	1.1935142278671265	0.13699662685394287
Txp-107	bond:CN_amine_ter-N_aliphatic	449	161	288	1667	4893	0.5522294044494629	1.6408655643463135	0.00000143775969263515 44	0.6094344258308411	0.9999991655349731
Txp-108	bond:CN_amine_ter-N_aromatic	167	66	101	1762	5080	0.568841278553009	1.8839976787567139	0.00008360712672583759	0.5307862162590027	0.9999573826789856

w/ Ryan Lougee and Chris Grulke

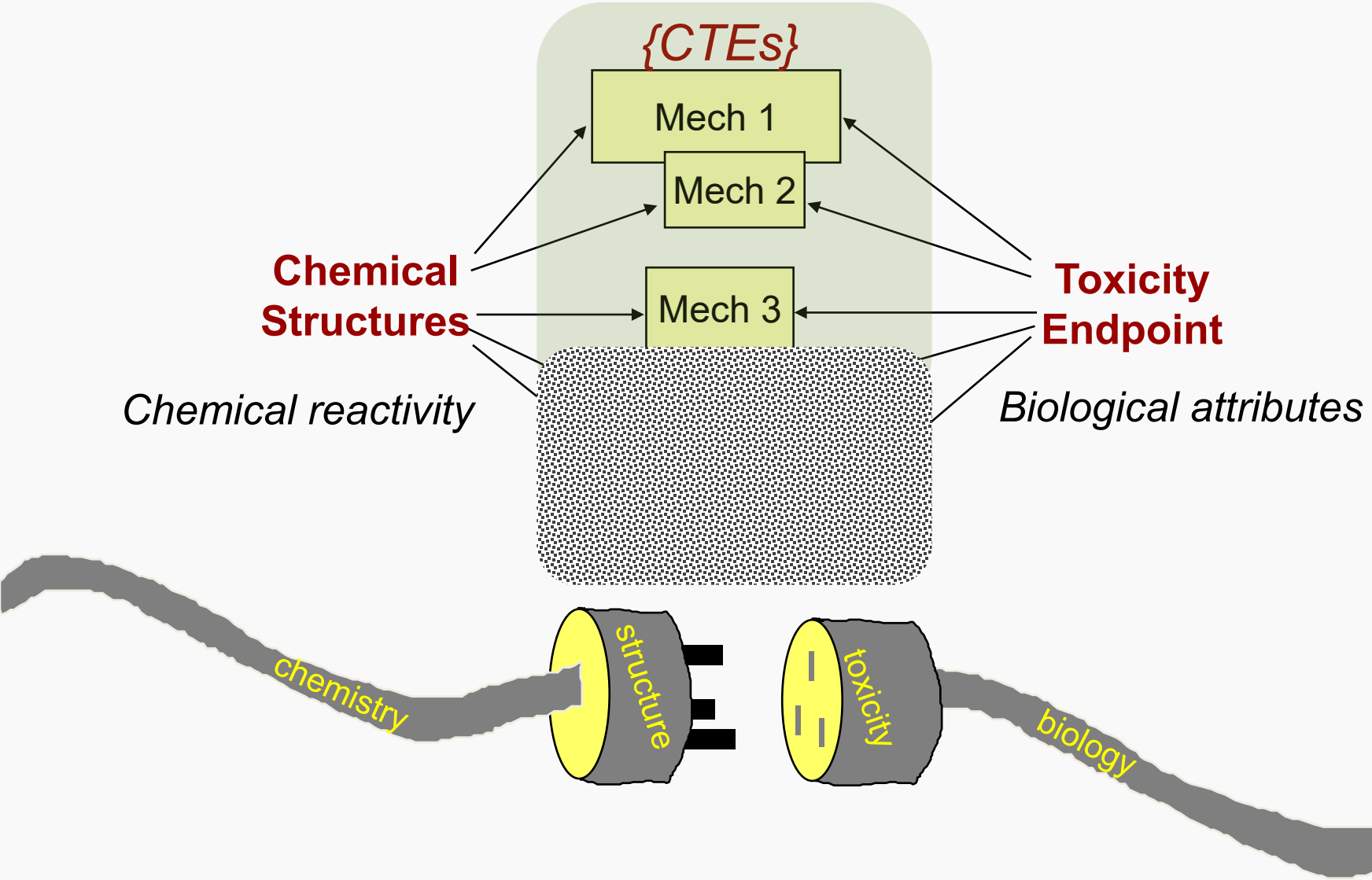
Chemotype-Enrichment vs. QSAR

It's not a contest!

QSAR vs. CTE Approach



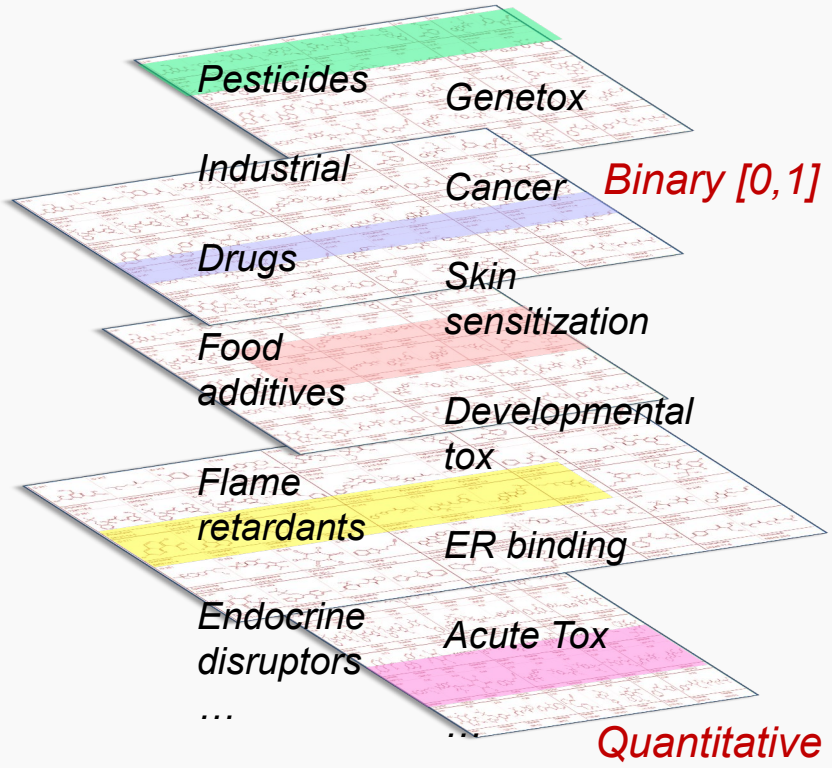
QSAR vs. CTE Approach



QSAR vs. CTE Approach

Activity = f {structure}

$\{\text{Training set}\}_i$ + $\{\text{Descriptor set}\}_i$ + $\{\text{Method}\}_i$ → $\{\text{Model}\}_i$



Physchem
Fragments
Topological
Electronic
Steric
2D
3D
…

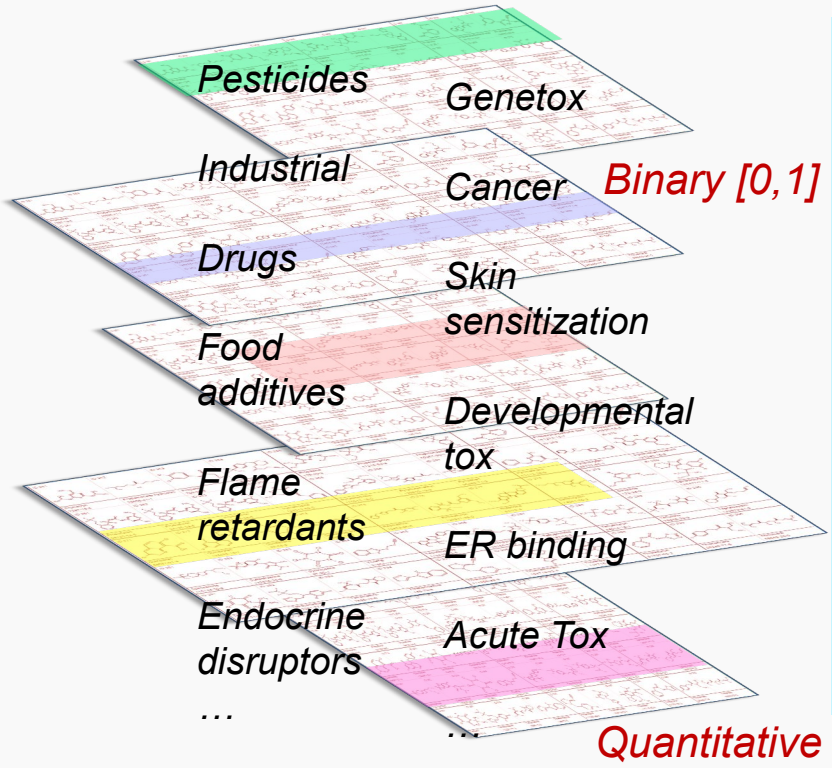
Linear regression
Multiple regression
kNN
SVM
Random Forest
Decision trees
…

Model 1
Model 2
Model 3
Model 4
Model 5
Model 6
…

QSAR vs. CTE Approach

Activity = f {structure}

{Training set}$_i$ + {Descriptor set}$_i$ + {Method}$_i$ → {CTE}$_i$



Binary [0,1]

Quantitative

{ToxPrints}

CTEW Statistical Enrichment

{CTE}$_1$

{CTE}$_2$

{CTE}$_3$

{CTE}$_4$

{CTE}$_5$

{CTE}$_6$

…

Use Chemotypes to Explore the Global ToxCast Chemical-Activity Landscape

L:\Lab\NCCT_ToxCast\CTEW-Chemotype enrichments

TXP_ENRICHMENTS_INVITRODBV2_2018.xlsx

assay_component_endpoint_name	descriptors_name	label	CT-Tot	TP	FP	FN	TN	OR	P-Val
ACEA_ER_80hr	Txp-116	bond:COC_ether_alkenyl	19	10	9	293	1425	5.403868	0.000475
ACEA_ER_80hr	Txp-124	bond:COH_alcohol_aromatic_phenol	216	76	140	227	1294	3.094525	1.03E-11
ACEA_ER_80hr	Txp-167	bond:CX_halide_alkyl-X_tertiary	15	6	9	297	1425	3.198653	0.033018
ACEA_ER_80hr	Txp-287	bond:S(=O)O_sulfuricAcid_generic	18	7	11	296	1423	3.059275	0.025425
ACEA_ER_80hr	Txp-435	chain:alkaneCyclic_pentyl_C5	69	32	37	271	1397	4.458363	1.22E-08
ACEA_ER_80hr	Txp-447	chain:alkaneLinear_stearyl_C18	3	3	0	300	1434	1E+08	0.005265
ACEA_ER_80hr	Txp-471	chain:alkyne_ethyne_generic	16	8	8	295	1426	4.833898	0.002766
ACEA_ER_80hr	Txp-479	chain:aromaticAlkane_Ph-C1-Ph	63	24	39	279	1395	3.076923	5.84E-05
ACEA_ER_80hr	Txp-482	chain:aromaticAlkane_Ph-C6							
ACEA_ER_80hr	Txp-483	chain:aromaticAlkane_Ph-C8							
ACEA_ER_80hr	Txp-560	group:ligand_path_4_polydentate							
ACEA_ER_80hr	Txp-593	ring:fused_[6_6]_tetralin							
ACEA_ER_80hr	Txp-602	ring:fused_steroid_generic_[5_6_6_6]							
ACEA_ER_80hr	Txp-611	ring:hetero_[5]_N_pyrazole							
ACEA_ER_80hr	Txp-689	ring:hetero_[6_6]_O_benzopyran	25	10	15	293	1419	3.228669	0.006246
ACEA_ER_80hr	Txp-691	ring:hetero_[6_6]_O_benzopyrone_(1_4-	10	9	1	294	1433	43.86735	1.15E-06
ACEA_ER_80hr	Txp-75	bond:CC(=O)C_ketone_aliphatic_generic	77	30	47	273	1387	3.242927	3.79E-06
ACEA_ER_80hr	Txp-76	bond:CC(=O)C_ketone_alkane_cyclic	66	30	36	273	1398	4.267399	6.45E-08
ACEA_ER_80hr	Txp-78	bond:CC(=O)C_ketone_alkane_cyclic_(C5	8	6	2	297	1432	14.46465	0.000551
ACEA_ER_80hr	Txp-80	bond:CC(=O)C_ketone_alkene_cyclic_(C6	24	14	10	289	1424	6.89827	6.73E-06
ACEA_ER_80hr	Txp-83	bond:CC(=O)C_ketone_alkene_cyclic_2-e	59	27	32	276	1402	4.286005	2.61E-07
ACEA_ER_80hr	Txp-85	bond:CC(=O)C_ketone_alkene_generic	56	27	29	276	1405	4.739505	6.64E-08
ACEA_ER_80hr	Txp-97	bond:CN_amine_aromatic_benzidine	5	3	2	300	1432	7.16	0.039927
APR_Hepat_Apop	302	bond:X[anyl_halide	169	8	161	1		6.70	038

- 13757 assay(aeid)–CTE pairs
- 773/1440 unique assays have enriched CTs
- 424/729 unique ToxPrints are enriched

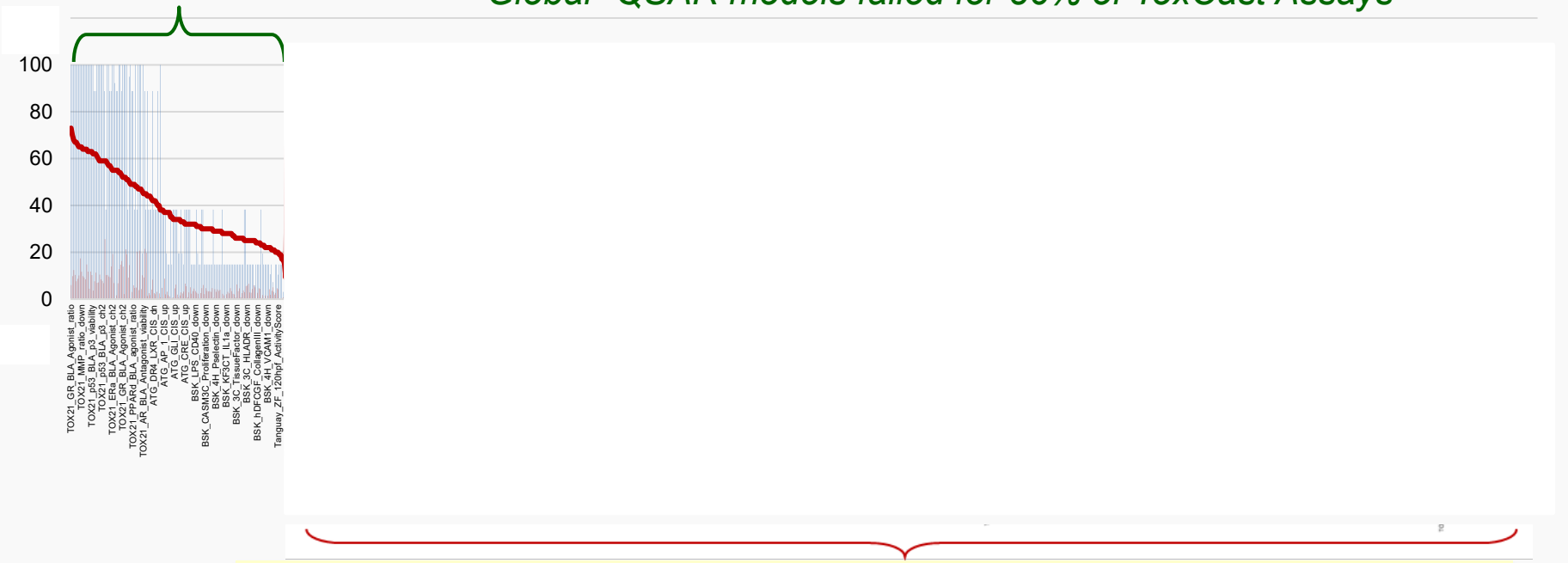
Assays x Enriched CT count

Enriched CT-count across 1032 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* for 167 ToxCast Assays
Avg % Median BA = 0.67

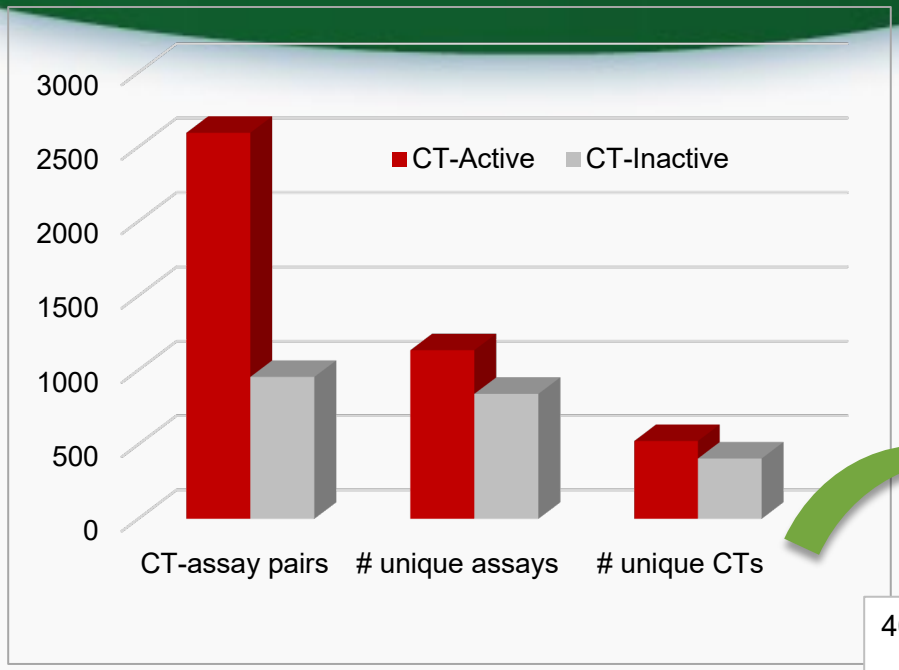
"Global" QSAR models failed for 86% of ToxCast Assays



Significant CT-enrichments across ToxCast assay space considered unmodelable by traditional "global" QSAR methods

w/ Jeremy Fitzpatrick and Grace Patlewicz

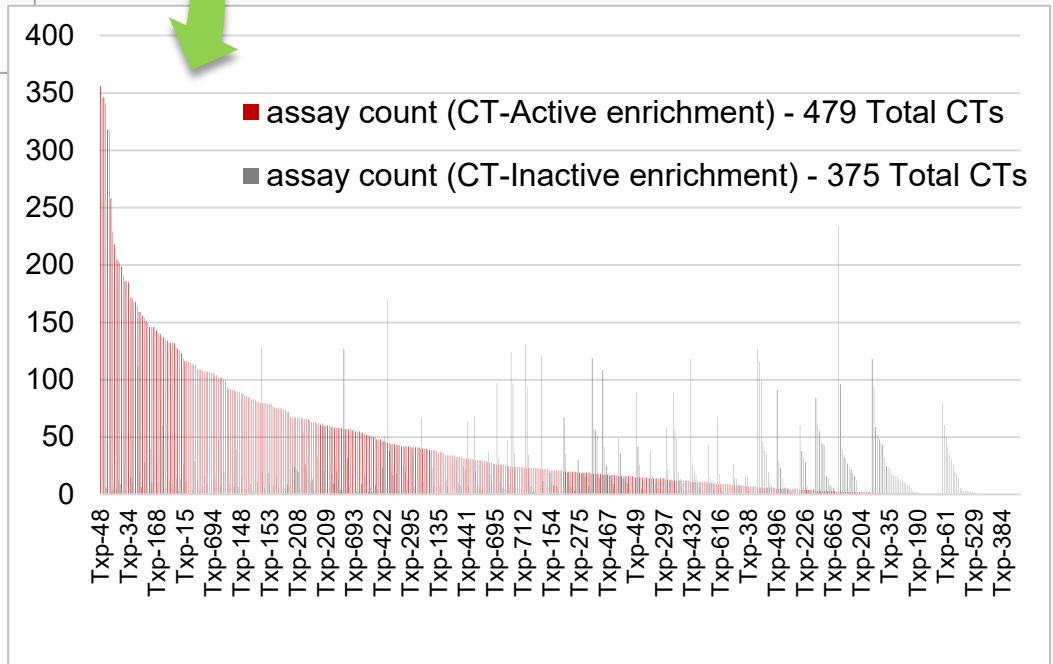
What about the “Inactives”?



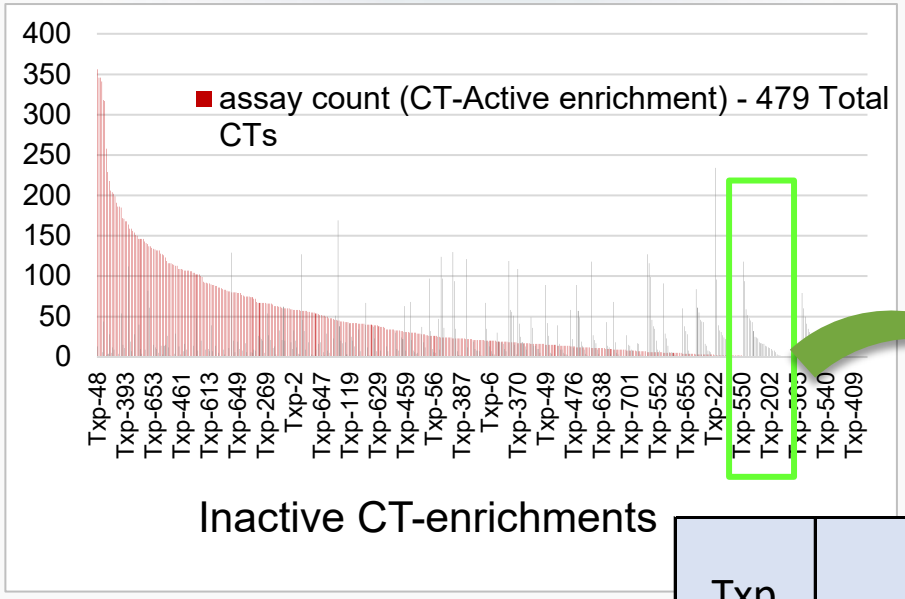
More “signal” in Actives, but significant amount of signal in Inactives

Inactive CT-enrichments

- Span ToxCast assay space
- More frequent in assays with fewer Active CT-enrichments



Top 10 enriched CT-Inactives (skewed from actives)

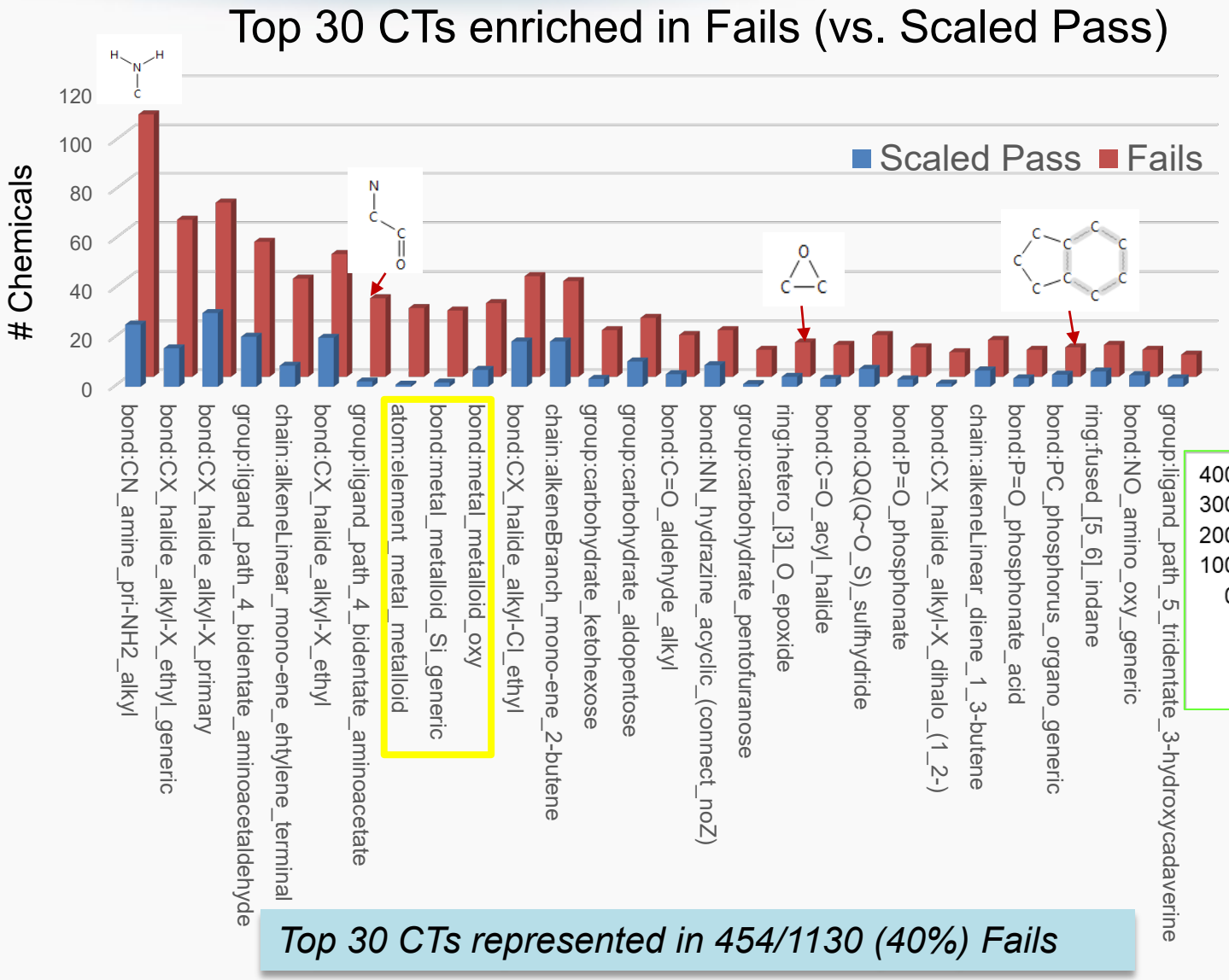


CT-Inactive >75 assays, CT-active <10 assays

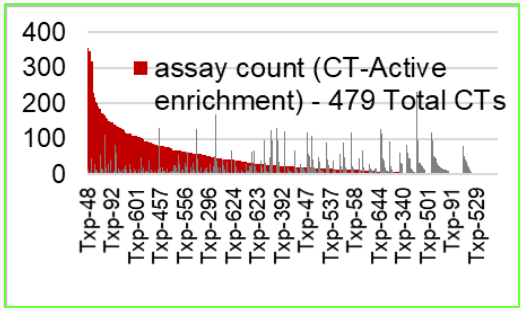
- *True inactivity?*
- *Assay artifacts?*
- *QC failure?*

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

ToxPrints enriched in Fails



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



Promiscuous ToxPrints

Enriched in actives of >100 ToxCast Assays

ct_id	AssayCount /TP
chain.aromaticAlkane_Ar.C.Ar	345
chain.aromaticAlkane_Ph.C1.Ph	351
chain.aromaticAlkane_Ph.C1_cyclic	111
chain.aromaticAlkane_Ph.C4	137
chain.aromaticAlkane_Ph.C6	132
chain.aromaticAlkene_Ph.C2	166

ct_id	AssayCount /TP
atom.element_metal_poor_metal	271
bond.metal_group_III_other_generic	270
atom.element_metal_transistion_metal	220
bond.metal_group_III_other_Sn_generic	183
bond.metal_group_III_other_Sn_organo	183
bond.metal_group_III_other_Sn_halide	107

ct_id	AssayCount /TP
bond.CX_halide_alkenyl.X_dihalo_.1_2..	165
bond.CX_halide_alkyl.Cl_dichloro_.1_1..	107
bond.CX_halide_alkyl.F_perfluoro_octyl	118
bond.CX_halide_alkyl.X_bicyclo.2_2_1.heptene	110
bond.CX_halide_alkyl.X_dihalo_.1_3.	103
bond.CX_halide_alkyl.X_tertiary	104

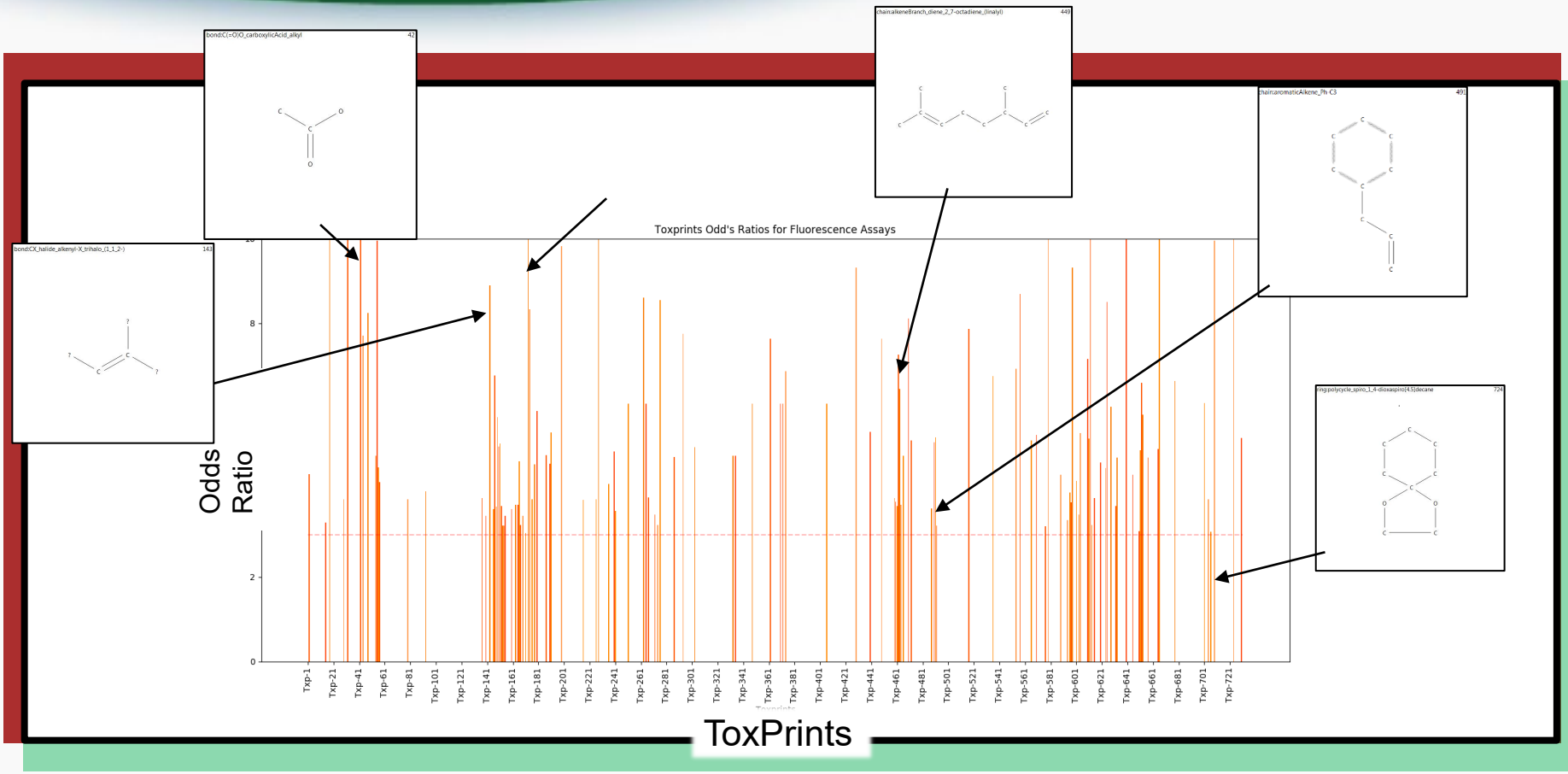
ct_id	AssayCount /TP
chain.alkaneLinear_decyl_C10	159
chain.alkaneLinear_tetradecyl_C14	159
chain.alkaneLinear_dodedyl_C12	149
chain.alkaneLinear_hexadecyl_C16	127
chain.alkaneLinear_octyl_C8	126

ct_id	AssayCount /TP
bond.CN_amine_ter.N_aliphatic	133
bond.CN_amine_ter.N_generic	133
bond.CN_amine_alicyclic_generic	115
bond.CN_amine_aliphatic_generic	111

- *True actives?*
- *Assay artifacts?*
- *Cytotoxic?*

w/ Ryan Lougee

CT Enrichments in Fluorescence Assays



CT's enriched (much more likely to be active) in "Fluorescence Assays" than in non-fluorescence detection assays

w/ Ryan Lougee

CTEs for Cytotoxicity

categories_table_invitrodb_v2

aeid	assay_component_endpoint_name	assay_format_type	organism	intended_target_type	intended_target_type_sub	intended_target_family	intended_target_family_sub	assay_desig
1	ACEA_T47D_80hr_Negative	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	growth repc
5	APR_HepG2_CellLoss_1h_dn	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
25	APR_HepG2_CellLoss_24h_dn	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
45	APR_HepG2_CellLoss_72h_dn	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
157	BSK_3C_Proliferation_down	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
159	BSK_hDFCGF_Proliferation_down	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
17		cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
19		cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
22		cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
22		cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
251	BSK_hDFCGF_Proliferation_down	cell-based	human	cellular	cellular	cell cycle	cytotoxicity	viability rep
253	BSK_hDFCGF							
269	BSK_KF3CT_							
291	BSK_LPS_SRI							
314	BSK_SAg_PB							
315	BSK_SAg_Pr							
317	BSK_SAg_SR							
763	TOX21_AR_B							
787	TOX21_ERa_							
795	TOX21_GR_B							
799	TOX21_MMP							

41 assays labeled as target_family_category "cytotoxicity"

assay_component_endpoint_name	desc_nam	s	label	CT-Tot	TP	FP	FN	TN	OR	P-Val
ACEA_ER_80hr	Txp-116		bond:COC_ether_alkenyl	19	10	9	293	1425	5.403868	0.000475
ACEA_ER_80hr	Txp-124		bond:COH_alcohol_aromatic_phenol	216	76	140	227	1294	3.094525	1.03E-11
ACEA_ER_80hr	Txp-167		bond:CX_halide_alkyl-X_te					1425	3.198653	0.033018
ACEA_ER_80hr	Txp-287		bond:S(=O)O_sulfuricAcid_					1423	3.059275	0.025425
ACEA_ER_80hr	Txp-435		chain:[illegible]aneCyc[illegible]ntyl					1397	4.458363	1.22E-08
ACEA_ER_80hr	Txp-447		chain:alka[illegible]					1434	1E+08	0.005265
ACEA_ER_80hr	Txp-471		chain:alkyne_ethyne_[illegible]ne					1426	4.833898	0.002766
ACEA_ER_80hr	Txp-479		chain:aromaticAlkane_Ph-C1-Ph	63	24	39		1395	3.076923	5.84E-05
ACEA_ER_80hr	Txp-482		chain:aromaticAlkane_Ph-C6	7	4	3	2[illegible]	[illegible]1	6.381271	0.020479
ACEA_ER_80hr	Txp-483		chain:aromaticAlkane_Ph-C8	4	3	1	300		14.33	0.01[illegible]26
ACEA_ER_80hr	Txp-560		group:ligand_path_4_polydentate	3	3	0	300	1434	1E+08	0.005265
ACEA_ER_80hr	Txp-593		ring:fused_[6_6]_tetralin	15	7	8	296	1426	4.215372	0.008257
A[illegible]										9
A[illegible]										8
ACEA_ER_80hr	Txp-689		ring:hetero_[6_6]_O_benzopyran	25	10	15	293	1419	3.228669	0.006246
ACEA_ER_80hr	Txp-691		ring:hetero_[6_6]_O_benzopyrone_(1_4-	10	9	1	294	1433	43.86735	1.15E-06
ACEA_ER_80hr	Txp-75		bond:CC(=O)C_ketone_aliphatic_generic	77	30	47	273	1387	3.242927	3.79E-06
ACEA_ER_80hr	Txp-76		bond:CC(=O)C_ketone_alkane_cyclic	66	30	36	273	1398	4.267399	6.45E-08
ACEA_ER_80hr	Txp-78		bond:CC(=O)C_ketone_alkane_cyclic_(C5	8	6	2	297	1432	14.46465	0.000551
ACEA_ER_80hr	Txp-80		bond:CC(=O)C_ketone_alkene_cyclic_(C6	24	14	10	289	1424	6.89827	6.73E-06
ACEA_ER_80hr	Txp-83		bond:CC(=O)C_ketone_alkene_cyclic_2-e	59	27	32	276	1402	4.286005	2.61E-07
ACEA_ER_80hr	Txp-85		bond:CC(=O)C_ketone_alkene_generic	56	27	29	276	1405	4.739505	6.64E-08
ACEA_ER_80hr	Txp-97		bond:CN_amine_aromatic_benzidine	5	3	2	300	1432	7.16	0.039927

TXP_ENRICHMENTS_INVITRODBV2_2018_filtered.xlsx

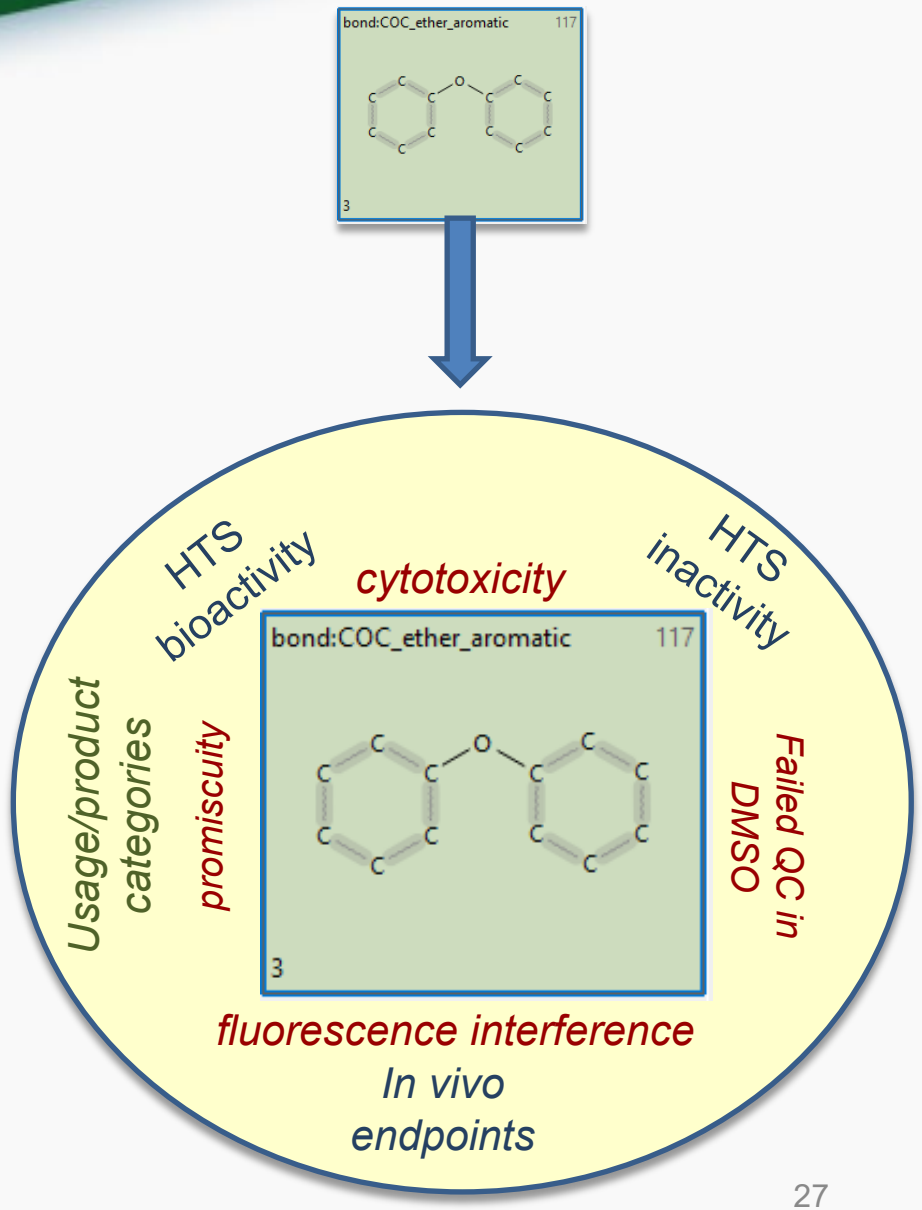
179 unique CTs enriched in the 41 assays

CTs enriched in >10 "cytotoxicity" assays	#cytotox assays
chain:aromaticAlkane_Ph-C1-Ph	36
chain:aromaticAlkane_Ar-C-Ar	35
bond:metal_group_III_other_Sn_generic	31
bond:metal_group_III_other_Sn_organo	31
atom:element_metal_poor_metal	27
bond:CX_halide_alkenyl-X_dihalo_(1_2-)	27
bond:quatN_alkyl_acyclic	27
bond:metal_group_III_other_generic	27
chain:alkaneLinear_tetradecyl_C14	27
chain:alkaneLinear_dodedyl_C12	24
chain:alkaneLinear_hexadecyl_C16	21
bond:C(=O)N_carbamate_dithio	20
chain:alkaneLinear_decyl_C10	20
atom:element_metal_transistion_metal	19
chain:aromaticAlkene_Ph-C2_acyclic_generic	19
chain:alkaneLinear_octyl_C8	18
bond:quatP_phosphonium	17
bond:metal_transition_Hg_generic	17
bond:metal_transition_Hg_organo	17
b[illegible]	17
b[illegible]	16
b[illegible]	16
chain:[illegible]kaneLinear_hexyl_C6	16
bo[illegible]halide_alkyl-X_bicyclo[2_2_1]hepte	15
bon[illegible]etal_group_III_other_Sn_halide	15
b[illegible]:CX_halide_alkyl-Cl_dichloro_(1_1-)	13
bond:quatN_generic	13
bond:CS_sulfide_di-	12
chain:aromaticAlkane_Ph-C4	12
chain:aromaticAlkene_Ph-C2	12
ring:hetero_[5]_N_pyrazole	12
bond:C=N_imine_C(connect_H_gt_0)	11
bond:CX_halide_alkenyl-X_generic	11
bond:CX_halide_alkyl-X_tertiary	11
bond:quatN_trimethyl_alkyl_acyclic	11
bond:metal_transition_Hg_halide	11
chain:alkaneLinear_stearyl_C18	11
chain:aromaticAlkane_Ph-C6	11

38 CTs enriched in >10 "cytotoxicity" assays

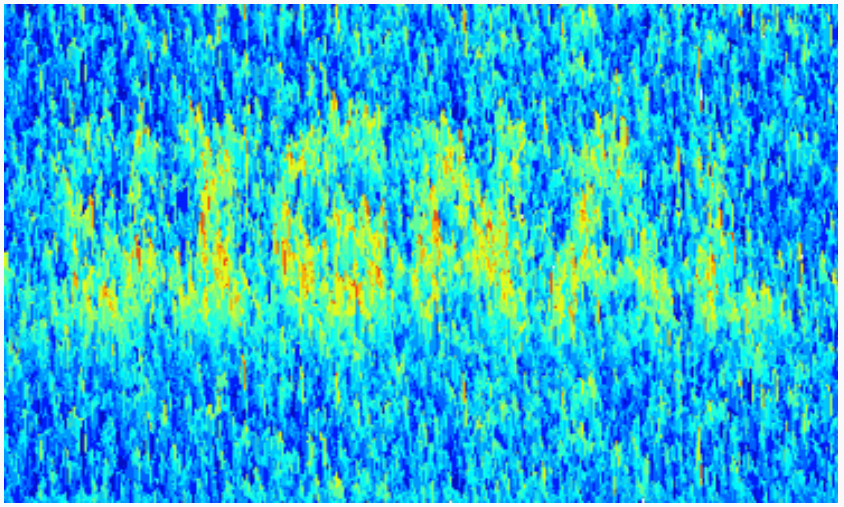
ToxPrints with ~~Attitude~~… Attributes

- Individual ToxPrints found to be significantly associated with:
  - HTS bioactivity
  - HTS inactivity
  - failed QC in DMSO
  - promiscuity
  - fluorescence interference
  - cytotoxicity
  - *in vivo* endpoints
  - Usage/product categories
  - …
- Thus, individual ToxPrints evolve to become “knowledge-informed” features, absorbing inferences from diverse activity datasets

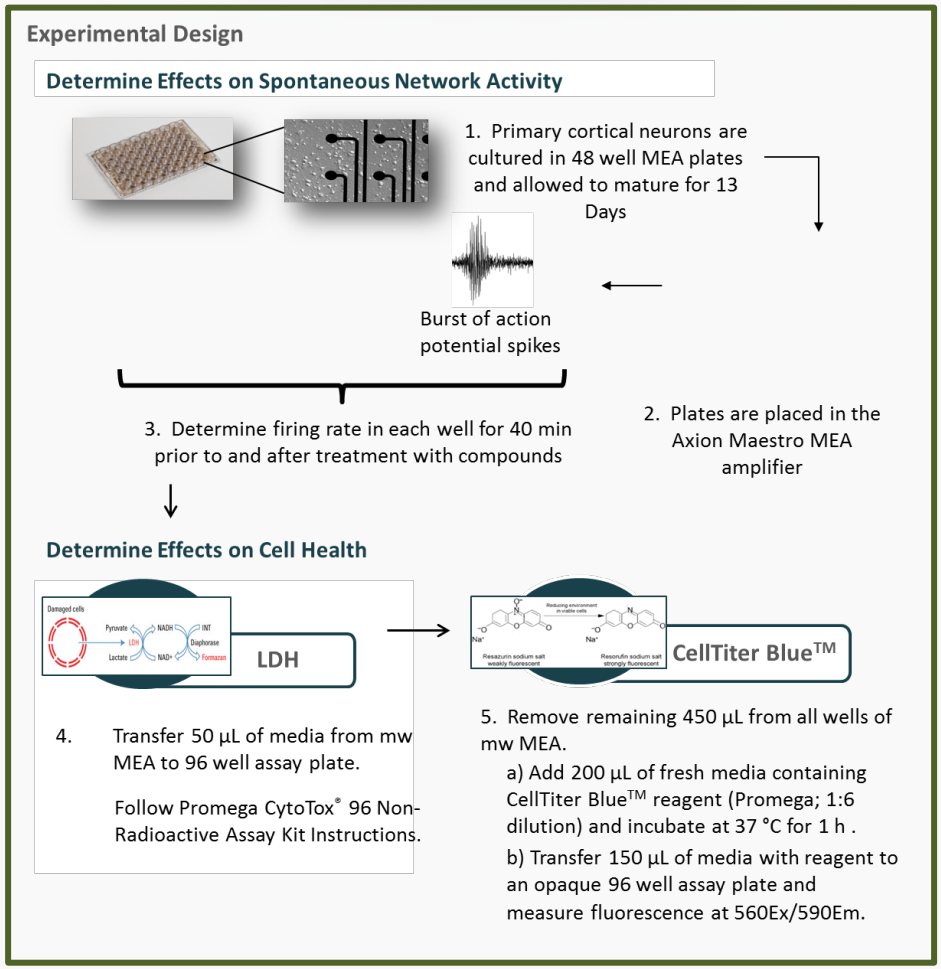


Applications of the CTE Approach

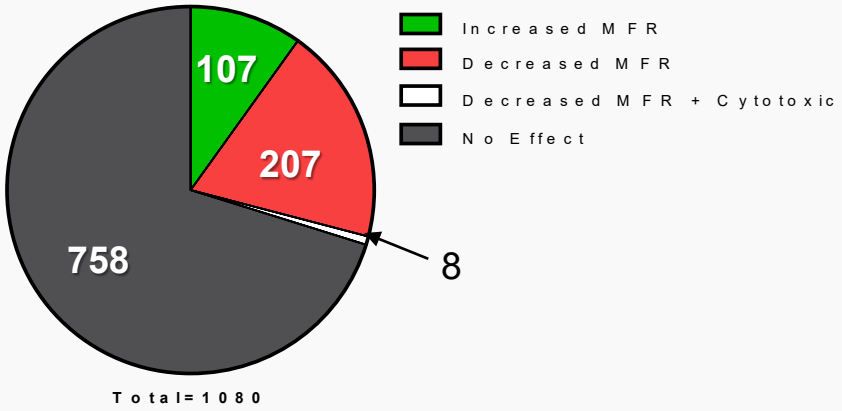
Finding Signal in the Noise



Multi-electrode array (MEA) neurotoxicity assay *(T. Shafer)*



- Screened 1080 ToxCast Phase II chemicals, incl. ph1_v2, ph2
- Activity measured as Mean Firing Rate (MFR) above or below threshold

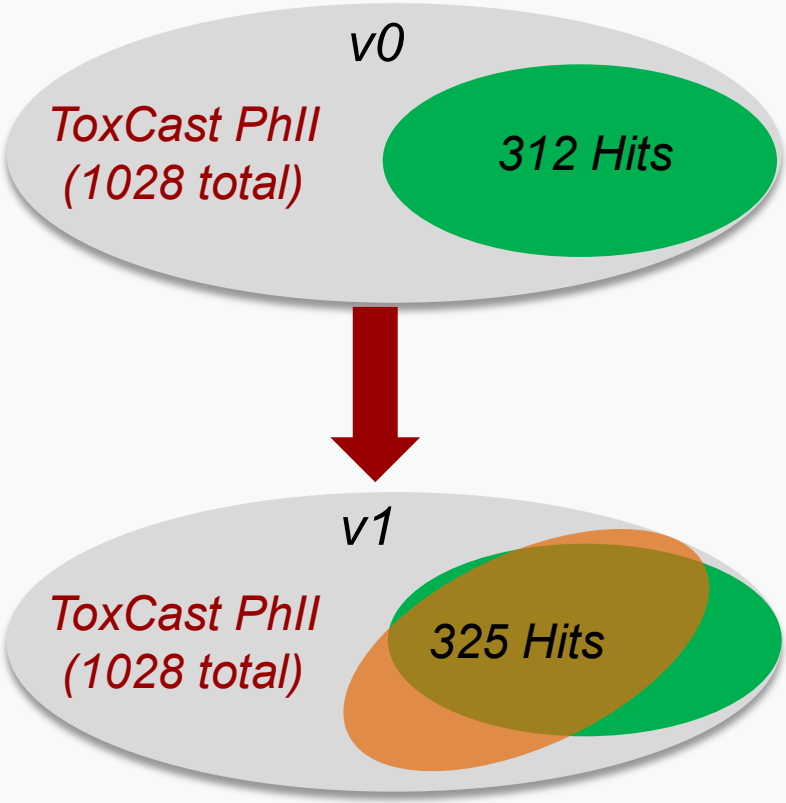


- 314 total "Actives" or Hits
- 758 "Inactives" or No-hits (8 cytotoxic)
- Total Tested = 1080 (30% hit rate)

Strickland et al., Arch Toxicol, 2018, 487-500

Optimizing MEA Assay Hit Calls

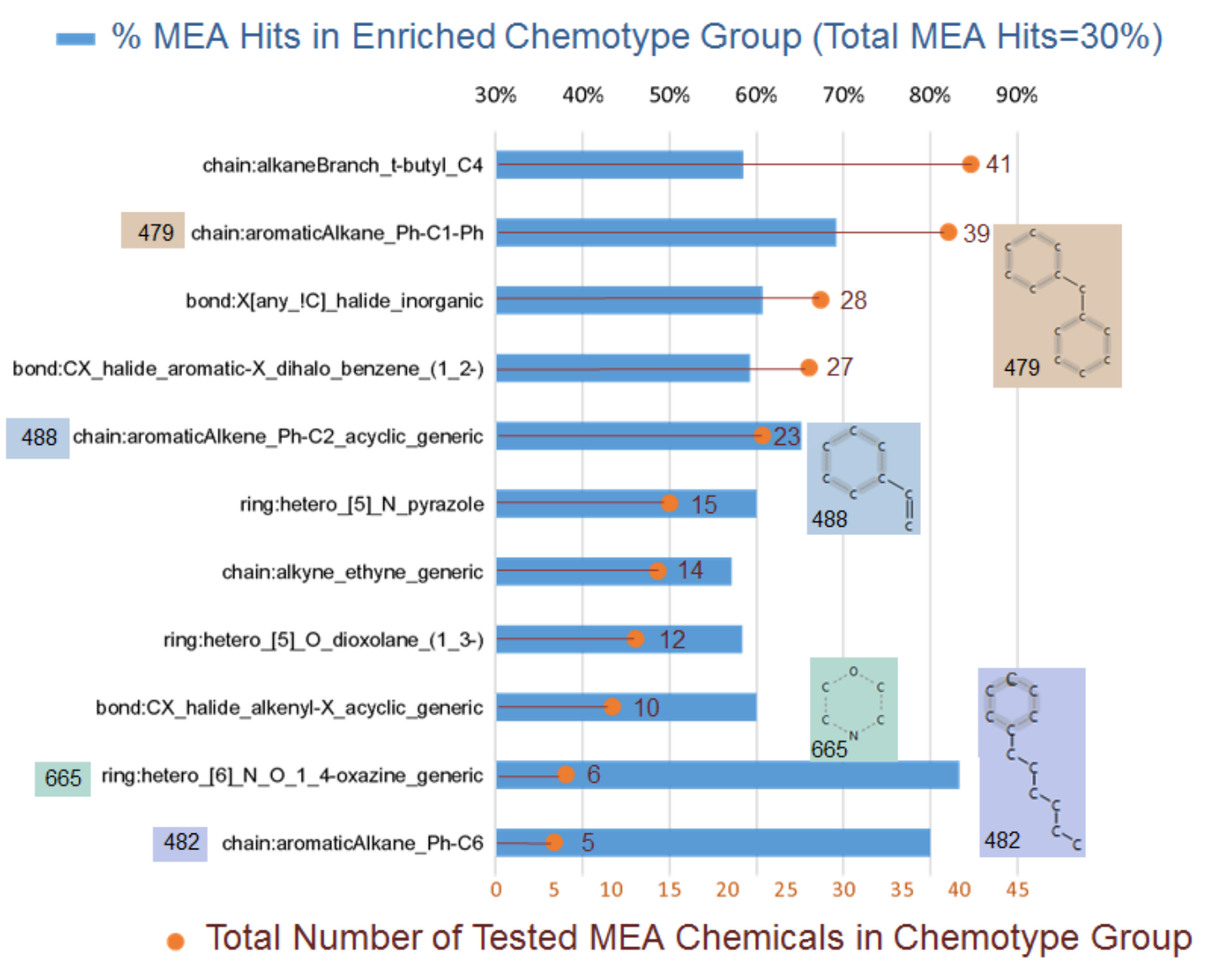
Multi-electrode array (MEA)
InvitroDB Hits



MEA	#Hits	#CTE
v0	312	18

with Tim Shafer & Matt Martin

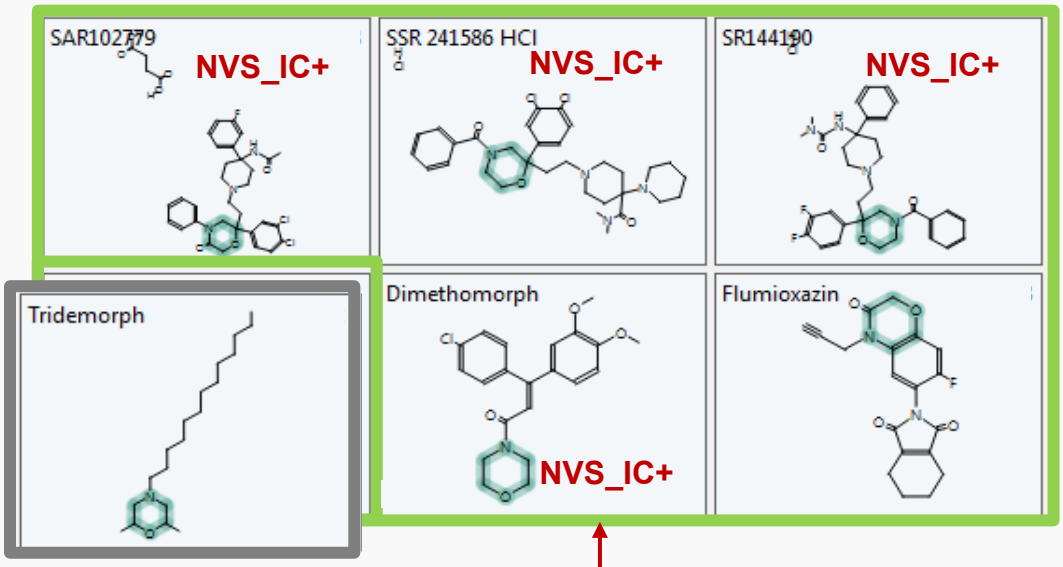
MEA Chemotype (CT) Analysis:
Visualize & convey CT enrichment



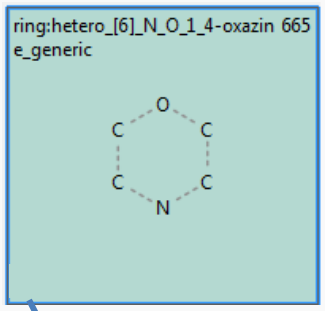
MEA Chemotype (CT) Analysis:
Use CTs to build biological linkages

MEA Hits

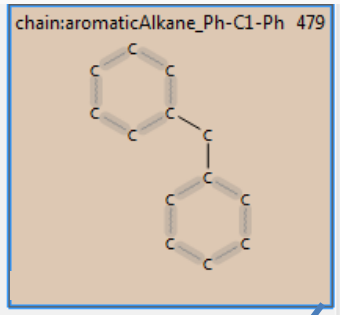
MEA Non-Hits



OR=11.3 (5/6 Hits)



OR=5.4 (27/39 Hits)



ToxCast assays enriched with hits in both CTs

- NVS_GPCR_hOpiate_mu
- NVS_GPCR_rOpiate_NonSelective
- NVS_GPCR_rOpiate_NonSelectiveNa
- NVS_IC_rCaBTZCHL
- NVS_IC_rCaDHPRCh_L
- NVS_IC_rNaCh_site2

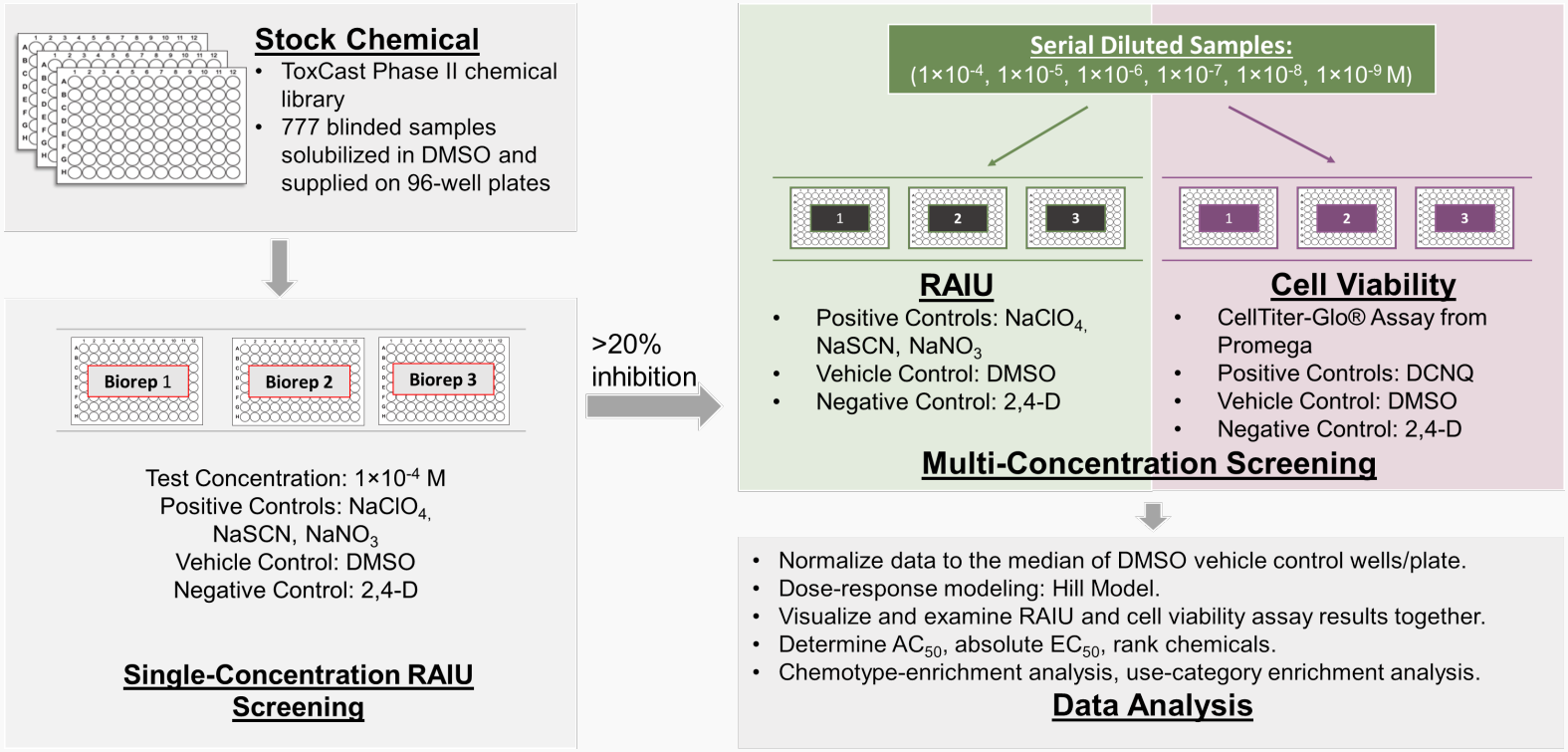
4/5 MEA hits with CT665 also MEA hits in Ion Channel (IC) assays

16/27 MEA hits in CT479 also hits in NVS_IC assays

Strickland et al., Arch Toxicol, 2018, 487-500

NIS Inhibition Assay

High-Throughput Screening and Chemotype-Enrichment Analysis of ToxCast Phase II Chemicals Evaluated for Human Sodium-Iodide Symporter (NIS) Inhibition



1028 total ToxCast Phase II chemicals screened

Wang et al., Environ. Int, 2019, 126:377-386

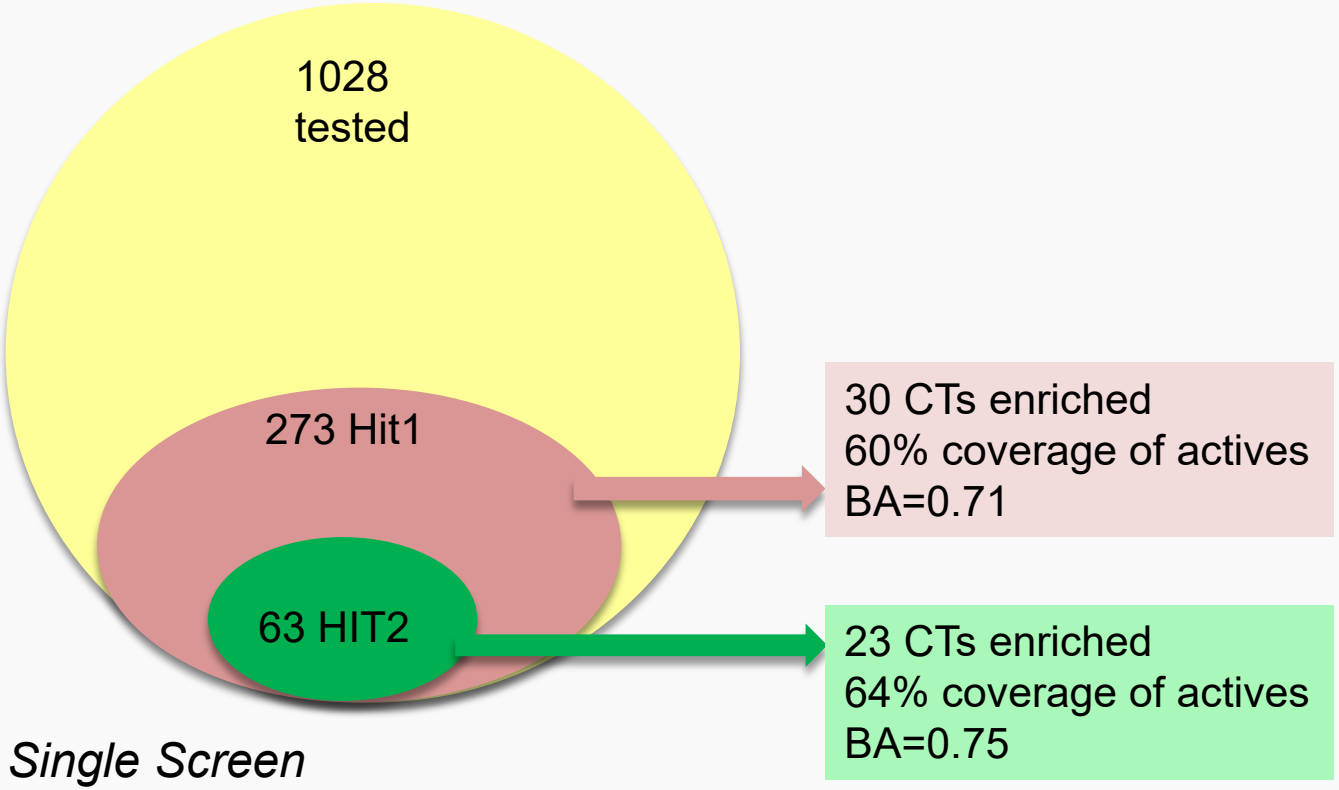
NIS Inhibition Assay

Single screen vs. Multiscreen results

1028 ToxCast Phase II chemicals screened

→ 27% positive at 20% threshold in single screen …**Hit1**

→ 6% positive in multiscreen, cell viability filtered results …**Hit2**

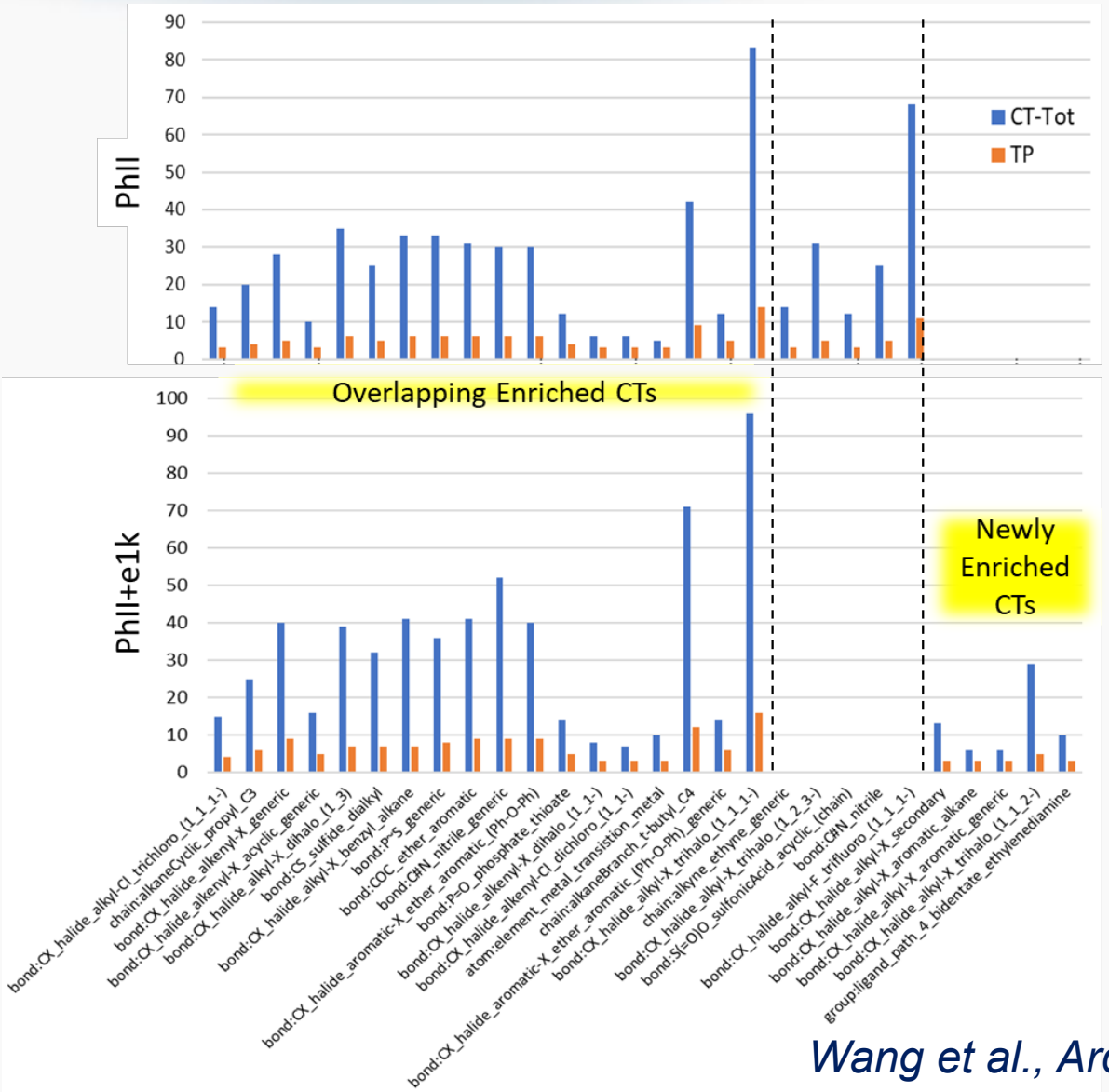


HIT1: Single Screen

HIT2: Multiscreen + cell viability filters

Wang et al., Environ. Int, 2019, 126:377-386

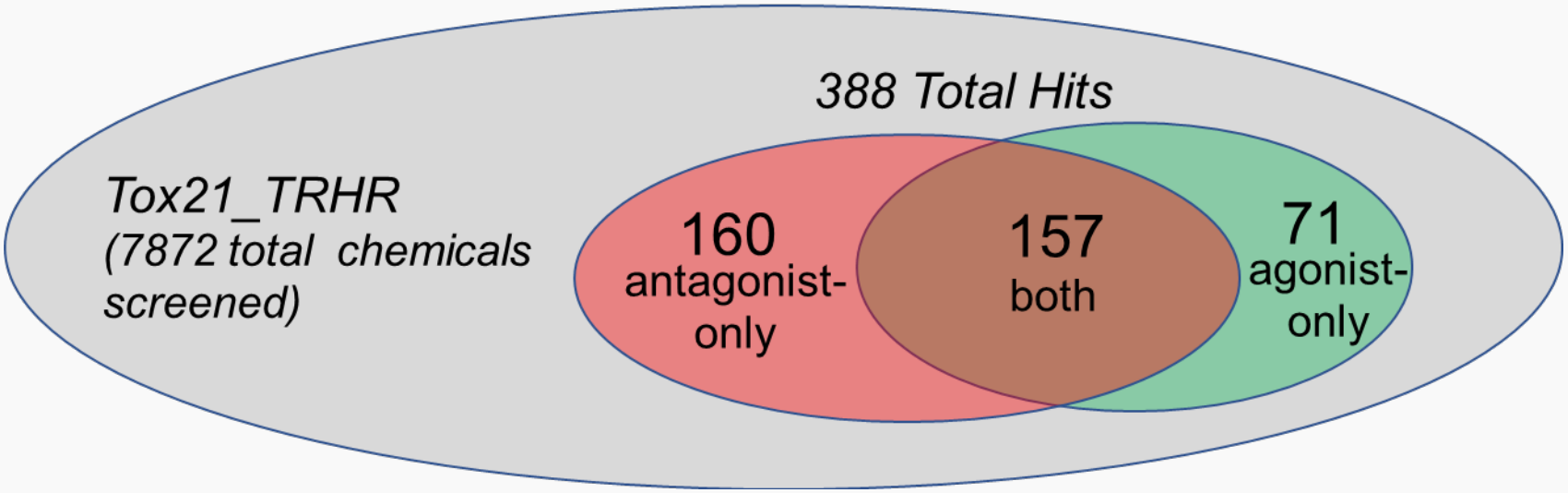
CT Enrichments (NIS Hit2) profiles from PhII to PhII+e1k



- *Majority of PhII CTEs validated for Hit2 Actives*
- *Small number of CTEs disappeared*
- *Equal number were added with e1k expanded inventory*

Wang et al., Arch Toxicol, 2021, 95:1723-37

Thyrotropin-Release Hormone Receptor Assay
Tox21_TRHR



Goal: Prioritize subset of actives (true hits) and inactives (potential false negatives) for follow-up testing using:

- *domain knowledge*
- *chemotype enrichments*
- *in silico computational chemistry models*

M. Shobair et al., Poster #252

Thyrotropin-Release Hormone Receptor Assay
Tox21_TRHR

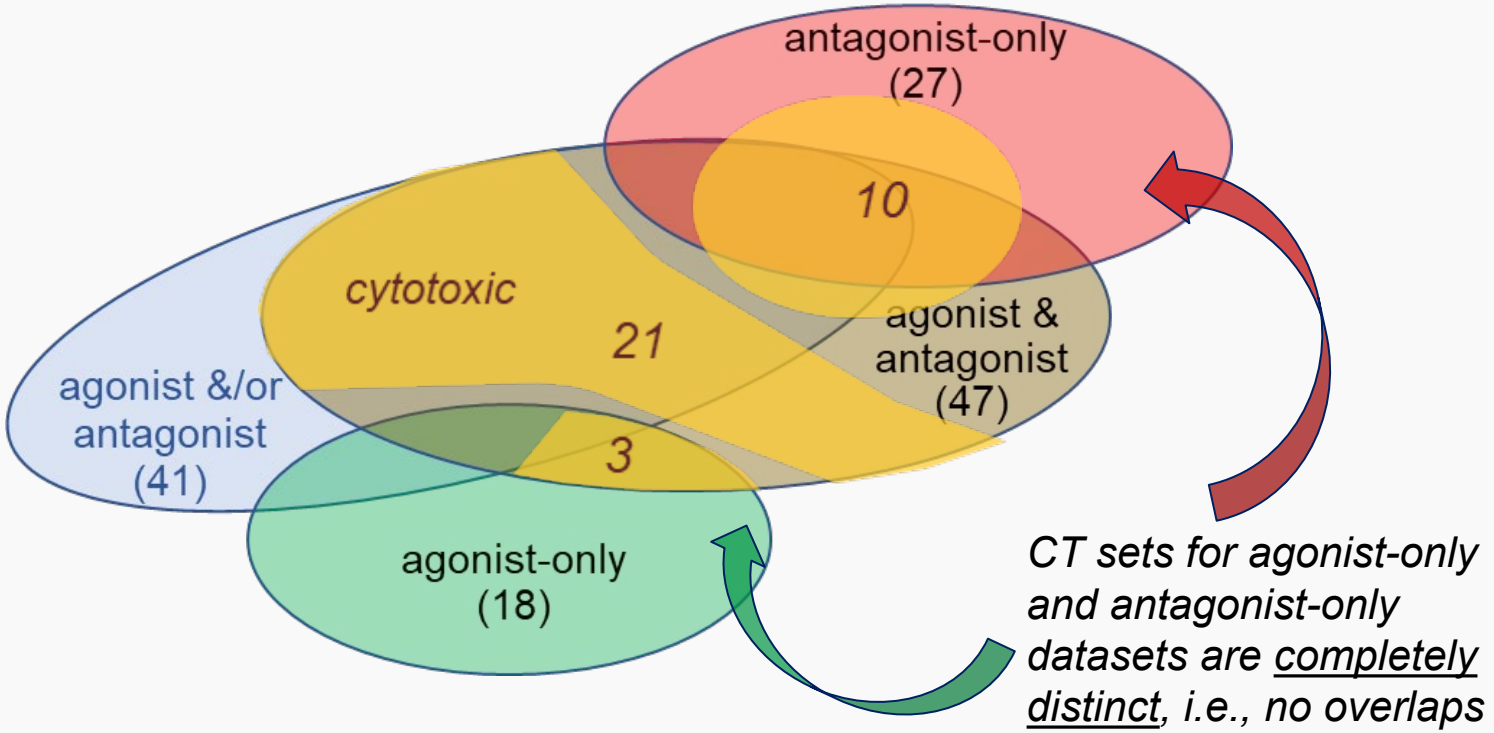
Unique chemotypes	Combined	Agonist_only	Antagonist_only	Agonist_and_antagonist
bond:CN_amine_aliphatic_generic		bond:CN_amine_aliphatic_generic		
bond:CN_amine_alkyl_ethanolamine		bond:CN_amine_alkyl_ethanolamine		
bond:CN_amine_ter-N_aliphatic		bond:CN_amine_ter-N_aliphatic		
bond:CN_amine_ter-N_aromatic		bond:CN_amine_ter-N_aromatic		
bond:CN_amine_ter-N_generic		bond:CN_amine_ter-N_generic		
bond:CS_sulfide_dialkyl		bond:CS_sulfide_dialkyl		
bond:quatN_alkyl_acyclic	bond:quatN_alkyl_acyclic	bond:quatN_alkyl_acyclic		bond:quatN_alkyl_acyclic
bond:quatN_generic		bond:quatN_generic		bond:quatN_generic
bond:quatN_trimethyl_alkyl_acyclic		bond:quatN_trimethyl_alkyl_acyclic		bond:quatN_trimethyl_alkyl_acyclic
bond:X[any_!C]_halide_inorganic	bond:quatN_trimethyl_alkyl_acyclic	bond:X[any_!C]_halide_inorganic		
group:ligand_path_4_bidentate_aminoethanol		group:ligand_path_4_bidentate_aminoethanol		
group:ligand_path_5_bidentate_propandiami	group:ligand_path_5_bidentate_pro	group:ligand_path_5_bidentate_propandiamine		
ring:hetero_[5]_N_pyrrole		ring:hetero_[5]_N_pyrrole		
ring:hetero_[6]_O_pyran_generic		ring:hetero_[6]_O_pyran_generic		ring:hetero_[6]_O_pyran_generic
ring:hetero_[6_6]_O_benzopyran	ring:hetero_[6_6]_O_benzopyran	ring:hetero_[6_6]_O_benzopyran		ring:hetero_[6_6]_O_benzopyran
ring:hetero_[6_6]_Z_generic		ring:hetero_[6_6]_Z_generic		
ring:hetero_[6_6_6]_N_S_phenothiazine	ring:hetero_[6_6_6]_N_S_phenothia	ring:hetero_[6_6_6]_N_S_phenothiazine		ring:hetero_[6_6_6]_N_S_phenothia
ring:hetero_[6_6_6]_O_benzopyran_dibenzo	ring:hetero_[6_6_6]_O_benzopyran	ring:hetero_[6_6_6]_O_benzopyran_dibenzo[b_e]		ring:hetero_[6_6_6]_O_benzopyran_
atom:element_metal_poor_metal	atom:element_metal_poor_metal		atom:element_metal_poor_metal	atom:element_metal_poor_metal
bond:C(=O)N_carbamate_dithio	bond:C(=O)N_carbamate_dithio		bond:C(=O)N_carbamate_dithio	
bond:C=N_imine_C(connect_H_gt_0)	bond:C=N_imine_N(connect_noZ)		bond:C=N_imine_C(connect_H_gt_0)	
bond:C=N_imine_N(connect_noZ)			bond:C=N_imine_N(connect_noZ)	
bond:C=S_carbonyl_thio_generic			bond:C=S_carbonyl_thio_generic	
bond:CC(=O)C_quinone_1_4-benzo	bond:CC(=O)C_quinone_1_4-benzo		bond:CC(=O)C_quinone_1_4-benzo	
bond:CC(=O)C_quinone_1_4-naphtho	bond:CC(=O)C_quinone_1_4-naphtho		bond:CC(=O)C_quinone_1_4-naphtho	
bond:COH_alcohol_aromatic			bond:COH_alcohol_aromatic	
bond:COH_alcohol_aromatic_phenol			bond:COH_alcohol_aromatic_phenol	
bond:CS_sulfide_di-			bond:CS_sulfide	
bond:CS_sulfide			bond:CS_sulfide_di-	
bond:CX_halide_alkenyl-X_dihalo_(1_2-)	bond:CX_halide_alkenyl-X_dihalo_(1_2-)		bond:CX_halide_alkenyl-X_dihalo	bond:CX_halide_alkenyl-X_dihalo_(1
bond:CX_halide_alkenyl-X_generic			bond:CX_halide_alkenyl-X_generic	
bond:CX_halide_alkyl-X_benzyl_alkane			bond:CX_halide_alkyl-X_benzyl_alkane	
bond:CX_halide_alkyl-X_trihalo_(1_1_1-)			bond:CX_halide_alkyl-X_trihalo_(1_1_1-)	
bond:CX_halide_aromatic-X_dihalo_benzene_(1_3-)			bond:CX_halide_aromatic-X_dihalo_benzene_(1_3-)	
bond:CX_halide_aromatic-X_halo_phenol_o	bond:CX_halide_aromatic-X_halo_phenol_ortho		bond:CX_halide_aromatic-X_halo	bond:CX_halide_aromatic-X_halo_p
bond:CX_halide_aromatic-X_halo_phenol_p	bond:CX_halide_aromatic-X_halo_phenol_para		bond:CX_halide_aromatic-X_halo	bond:CX_halide_aromatic-X_halo_p
bond:CX_halide_aromatic-X_trihalo_benzen	bond:CX_halide_aromatic-X_trihalo_benzene_(1_2_3-)		bond:CX_halide_aromatic-X_triha	bond:CX_halide_aromatic-X_trihalo_
bond:CX_halide_generic-X_dihalo_(1_2-)			bond:CX_halide_generic-X_dihalo_(1_2-)	
bond:metal_group_III_other_generic	bond:metal_group_III_other_generic		bond:metal_group_III_other_gener	
chain:aromaticAlkane_Ar-C-Ar	chain:aromaticAlkane_Ar-C-Ar		chain:aromaticAlkane_Ar-C-Ar	
chain:aromaticAlkane_Ph-C1-Ph	chain:aromaticAlkane_Ph-C1-Ph		chain:aromaticAlkane_Ph-C1-Ph	
chain:aromaticAlkene_Ph-C2_acyclic_generic			chain:aromaticAlkene_Ph-C2_acy	
ring:hetero_[5]_N_triazole_(1_2_4-)			ring:hetero_[5]_N_triazole_(1_2_4-	
ring:hetero_[5]_N_triazole_(1_3_4-)			ring:hetero_[5]_N_triazole_(1_3_4-	
ring:hetero_[5]_Z_1_2_4_1_3_4-Z			ring:hetero_[5]_Z_1_2_4_1_3_4-Z	
bond:CX_halide_alkyl-Cl_dichloro_(1_1-)				
bond:CX_halide_alkyl-X_bicyclo[2_2_1]hepte	bond:CX_halide_alkyl-X_bicyclo[2_2_1]heptene			
bond:CX_halide_alkyl-X_tertiary				
bond:metal_group_III_other_generic_oxy	bond:metal_group_III_other_generic_oxy			
bond:metal_group_III_other_Sn_generic	bond:metal_group_III_other_Sn_generic			
bond:metal_group_III_other_Sn_halide	bond:metal_group_III_other_Sn_halide			

Generate CT enrichments for the 4 activity subsets relative to the larger Tox21 test set

M. Shobair et al., Poster # 252

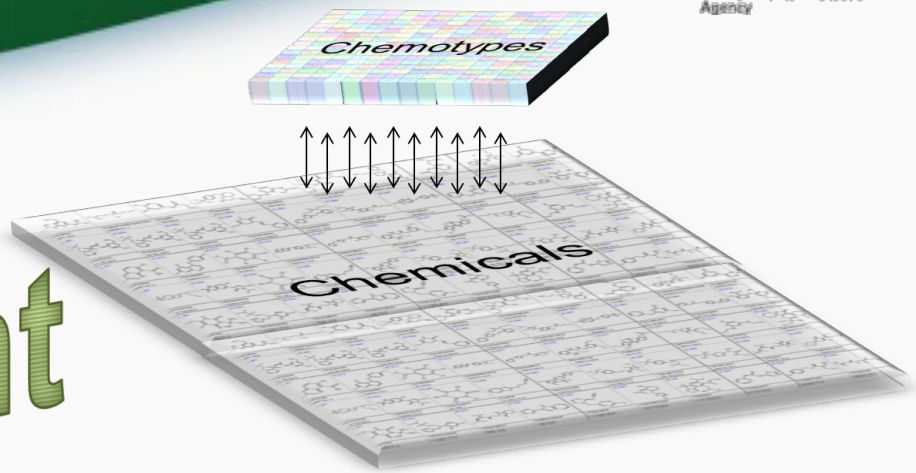
Thyrotropin-Release Hormone Receptor Assay
Tox21_TRHR

(#) of Enriched CTs per TRHR Activity Subset



CTs included in the "cytotoxic" regions are significantly enriched in >10/42 Tox21 "cytotoxicity"-labeled assays (invitrodb_v2)

M. Shobair et al., Poster # 252


Chemotypes
Chemicals

Enrichment


NO TAXATION
WITHOUT
REPRESENTATION

ToxPrint enrichments only possible with inclusion of NCATS partner library

Tox21 Library Profile

NCATS (Drugs)

vs.

EPA+NTP (Environmental)

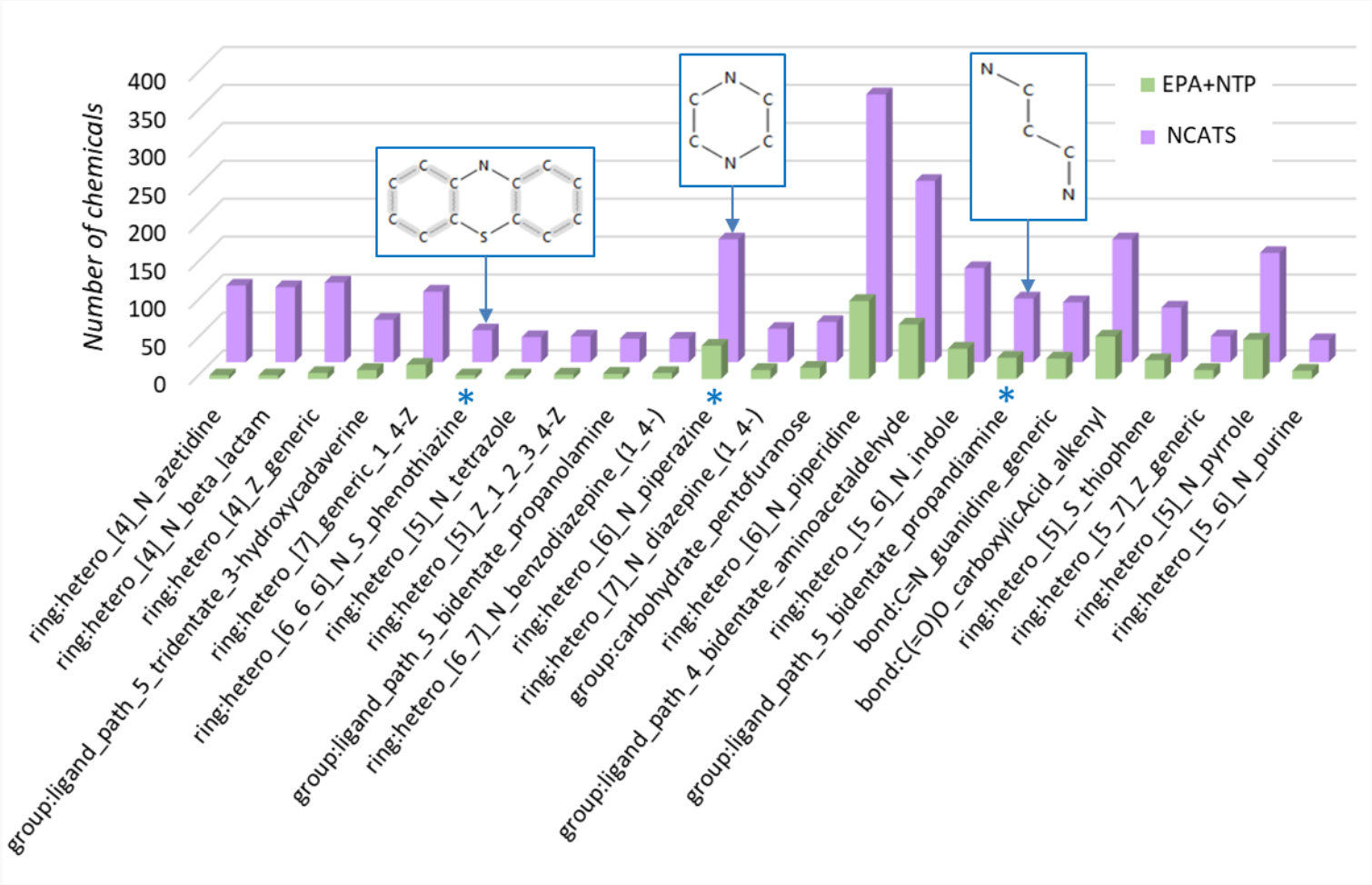
ToxPrint enrichments only possible with inclusion of EPA+NTP partner libraries

Richard et al., CRT, 2021, 34:189-216

CT Profile & Assay "Enrichments"

ToxPrint enrichments only possible with inclusion of NCATS partner library

Each ToxPrint below is enriched in 1 or more Tox21 assays

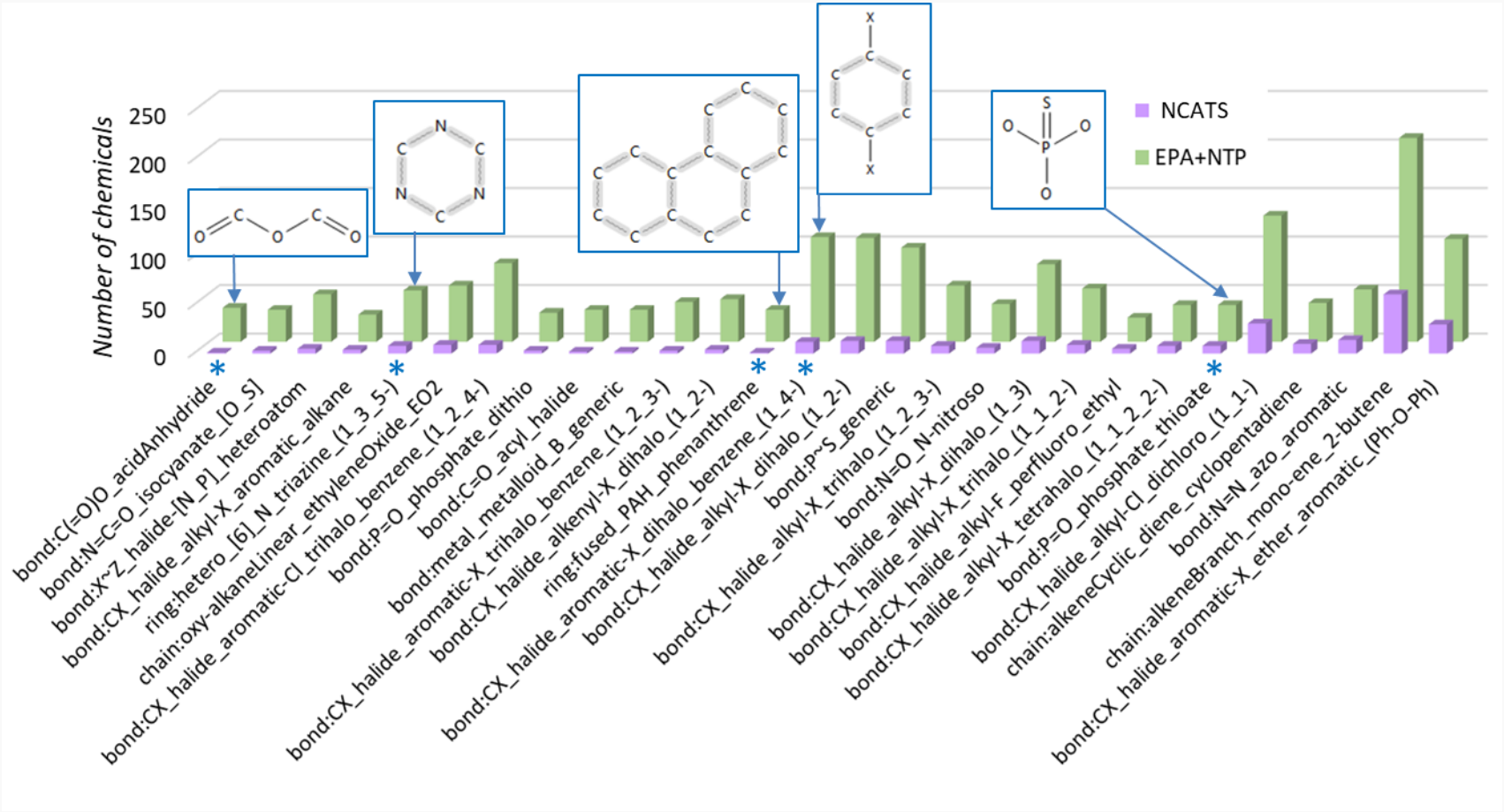


Richard et al., CRT, 2021, 34:189-216

CT Profile & Assay "Enrichments"

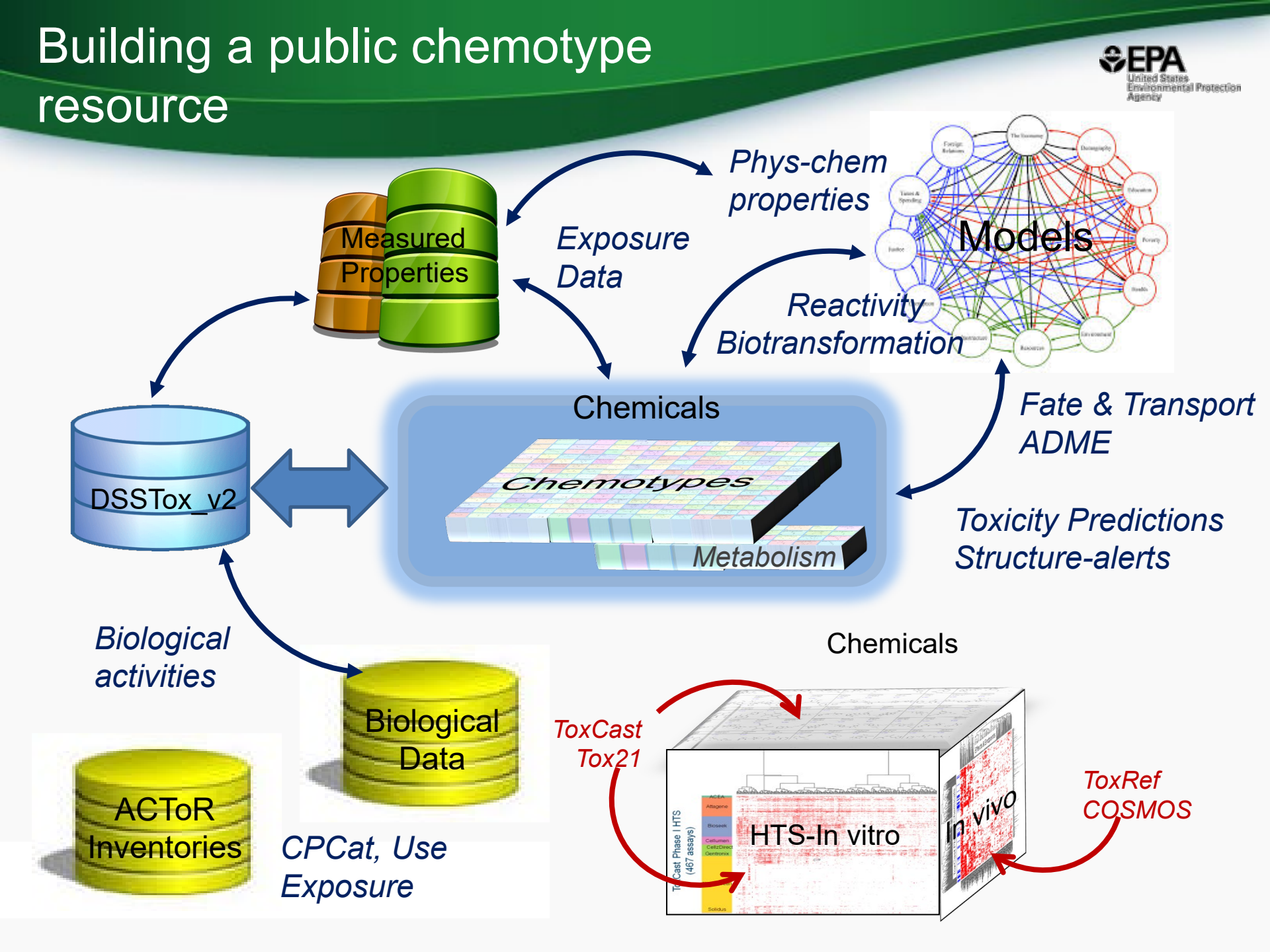
ToxPrint enrichments only possible with inclusion of EPA+NTP partner libraries

Each ToxPrint below is enriched in 1 or more Tox21 assays



Richard et al., CRT, 2021, 34:189-216

Building a public chemotype resource



Acknowledgements:

- EPA NCCT Chemistry Team
  - Ryan Lougee (CTEW workflow coding)
  - Chris Grulke (DSSTox, Chemotypes)
  - Antony Williams (DSSTox, Chemistry Dashboard)
  - Grace Patlewicz & Jeremy Fitzpatrick
- ToxCast/Tox21 case studies
  - Mahmoud Shobair, Katie-Paul Friedman,
  - Tom Knudsen, Bhavesh Ahir
  - Tim Shafer, Jenna Strickland, Matt Martin
  - Jun Wang, Susan Laws, Tammy Stoker
- ToxPrint Chemotypes
  - Chihae Yang, Jim Rathman & colleagues
    - Molecular Networks & Altamira


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
dziękuję
bedankt
спасибо
go raibh maith agat
mochchakkeram
dakujem
arigatō
grazie
sukriya
kop khun krap
lerima kasih
감사합니다
ευχαριστώ
dank je
hvala
mauruuru
sagolun
takk
tapadh leat
рахмат
Баярлалаа
merci