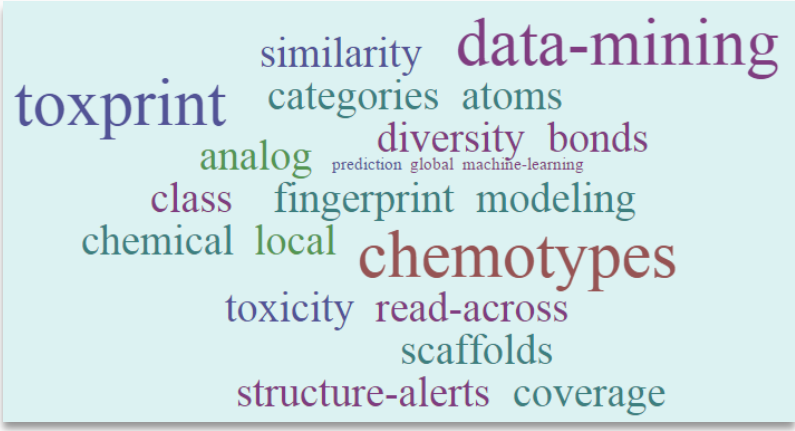
ACS Spring Meeting, San Francisco, CA, April 2-6, 2017
ENVR 932

Use of Chemotypes for Profiling and Exploring the ToxCast Chemical-Assay Landscape

Ann Richard, Chris Grulke, Grace Patlewicz, Chihae Yang, Antony Williams*

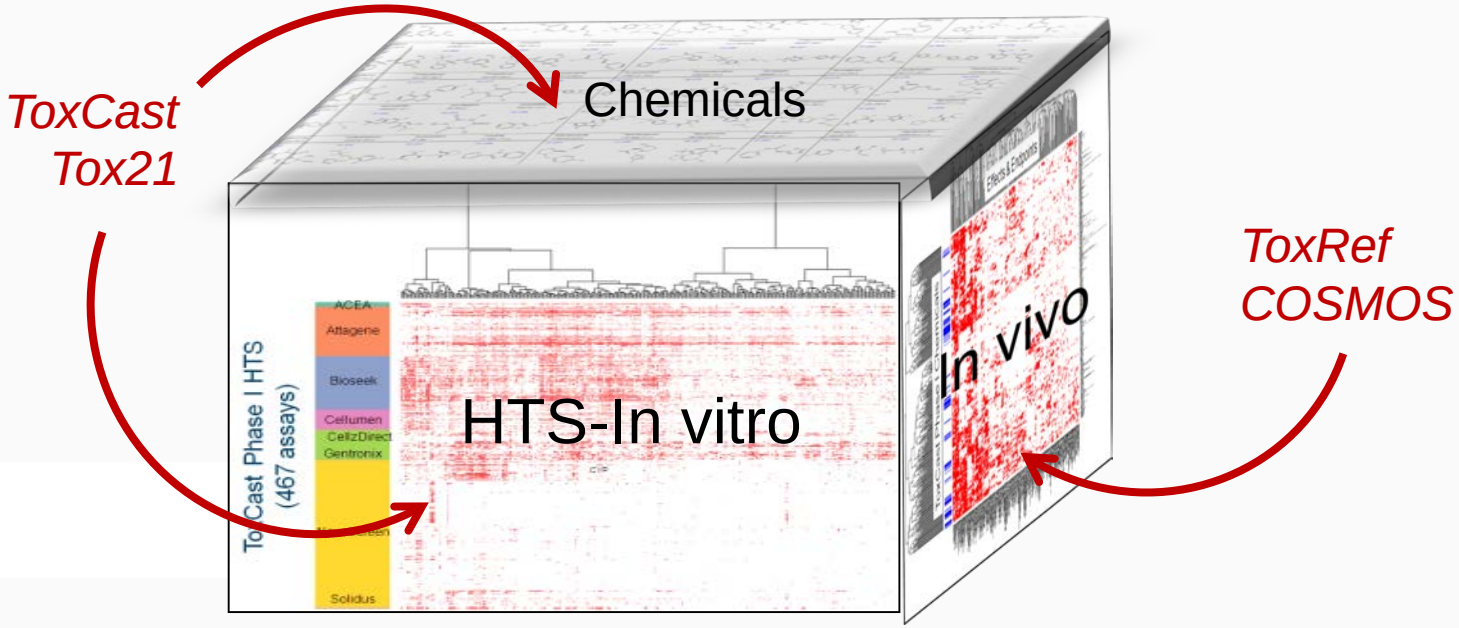
National Center for Computational Toxicology
U.S. Environmental Protection Agency, RTP, NC



This work was reviewed by the U.S. EPA and approved for presentation but does not necessarily reflect official Agency policy.

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



Toxicity Prediction Challenge:

Bringing all knowledge & data to bear on problem

EPA United States Environmental Protection Agency

Biologically-based QSAR & Cheminformatics

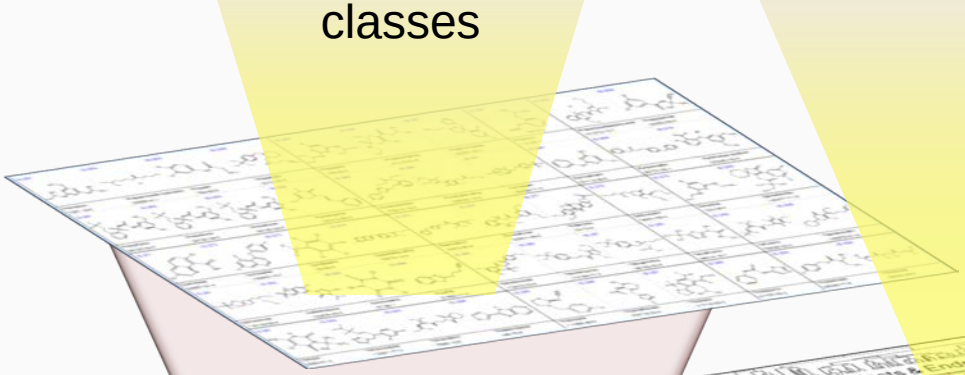
Reactivity & toxicity-informed features & classes

Aggregation

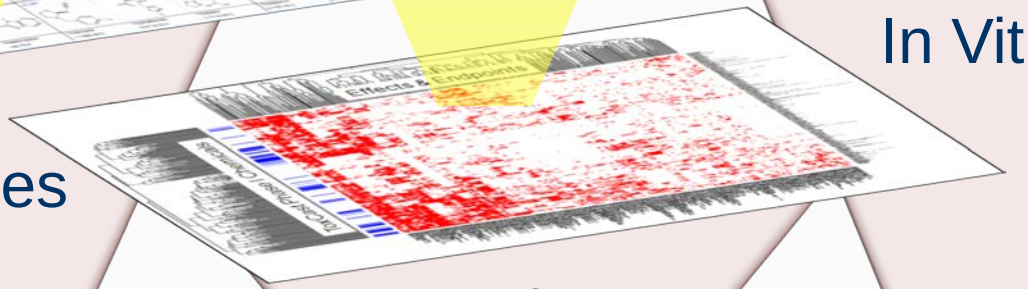
Mechanistically well-defined toxicity endpoint

Data-mining

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



Structures



In Vivo



In Vitro/HTS

Existing knowledge

Chemotyper & CSRML



www.chemotyper.org

The ChemoTyper application was developed by Molecular Networks GmbH, Erlangen, Germany (http://www.molecular-networks.com) under a contract from the U.S. FDA, Center for Food Safety and Applied Nutrition (CFSAN), Office of Food Additive Safety. The XML-based substructure (or chemotype) definition language CSRML was co-developed in collaboration with Altamira LLC, Columbus, OH, USA (http://www.altamira-llc.com).

Publicly Available Resources: Chemotyper & ToxPrint Chemotypes

EPA United States Environmental Protection Agency

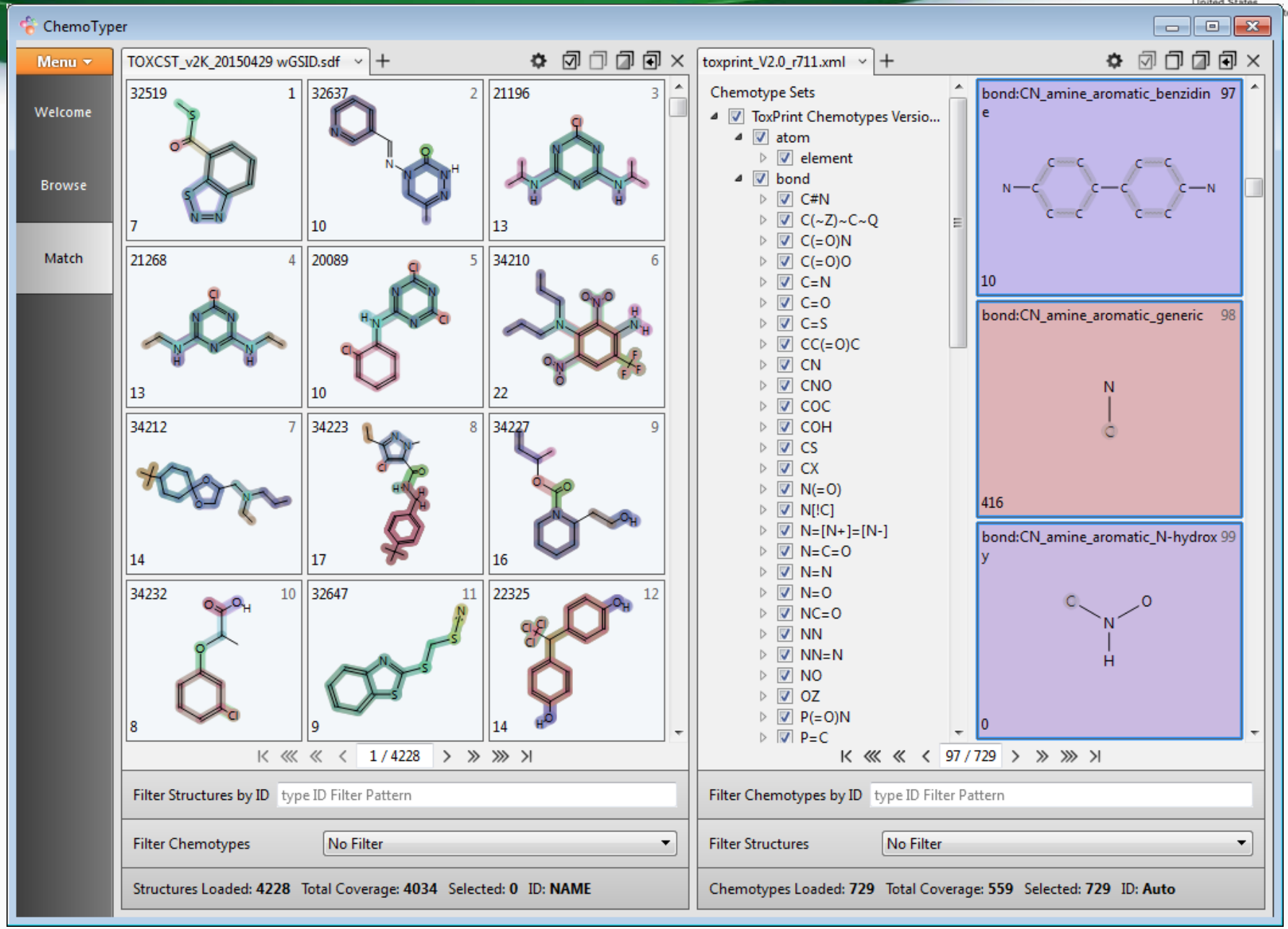
Chemotyper: http://www.chemotyper.org

- MS Windows application allows for searching and highlighting of chemical chemotypes (chemical substructures or subgraphs) in structure files.
- Developed by *Molecular Networks GmbH (MN)* under contract from FDA CFSAN to house the "ToxPrint" chemotypes

ToxPrint Chemotypes: http://www.toxprint.org

- Developed by *Altamira LLC* for FDA CFSAN's CERES project
- Designed to provide coverage of EPA & FDA inventories and capture chemical features important for chemical safety assessment workflow
- Contains three subsets:
  - *generic structural fragments (729 total)*
  - *Ashby-Tennant genotoxic carcinogen alerts (Ashby, J; Tennant, RW, 1988)*
  - *cancer TTC categories (Kroes, R. et. al. 2004)*

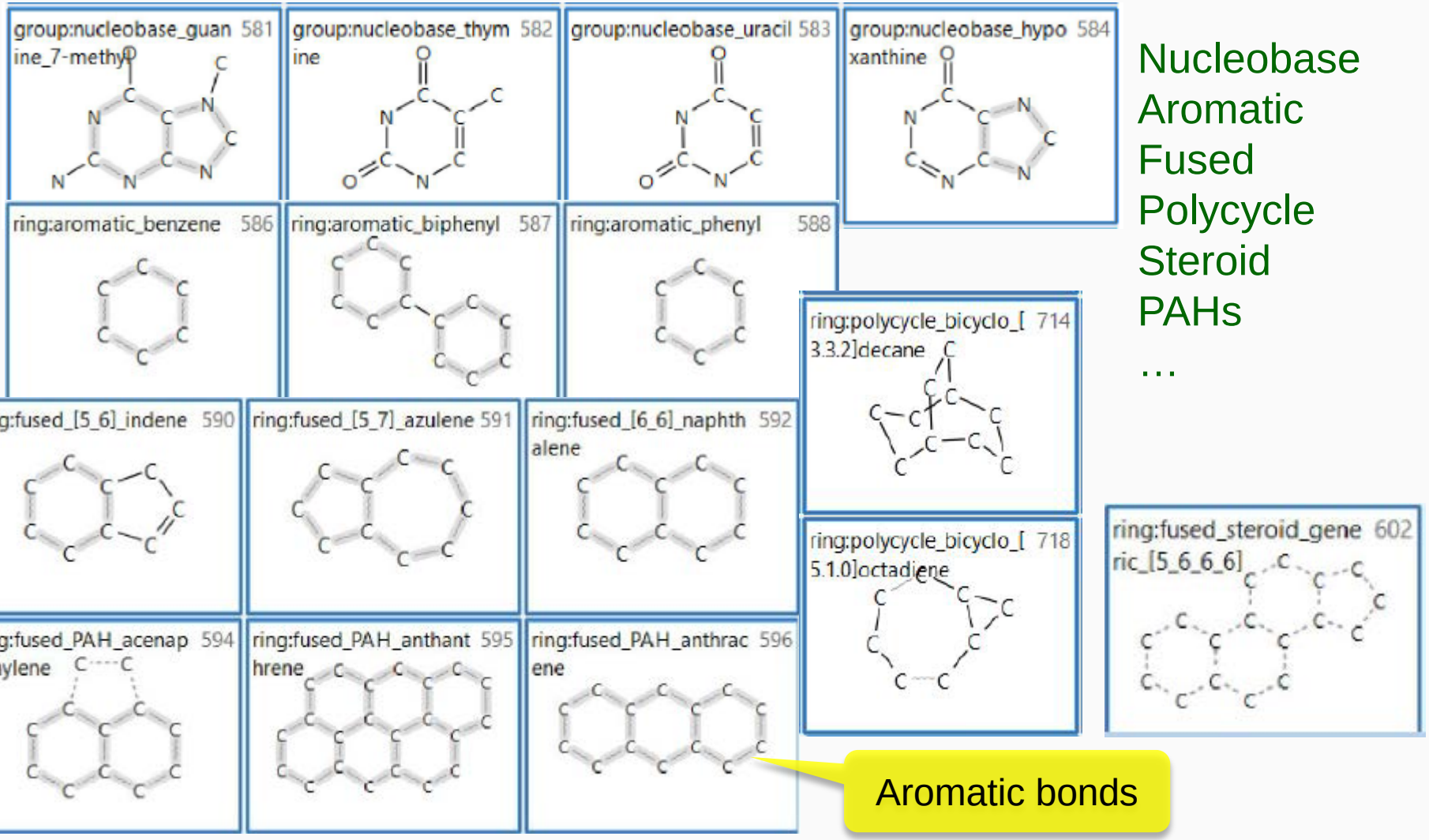
www.chemotyper.org



ToxPrints: Capturing Chemistry

EPA United States Environmental Protection Agency

Groups & Scaffolds - Defining local chemical domains



Nucleobase
Aromatic
Fused
Polycycle
Steroid
PAHs
…

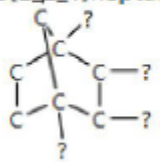
ToxPrints: Capturing Chemistry

EPA United States Environmental Protection Agency

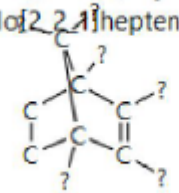
Bonding – e.g. CX_halide_alkyl-X (15 types)

bond type, order, frequency, context

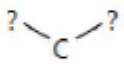
bond:CX_halide_alkyl-X 157 _bicyclo[2_2_1]heptane



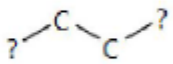
bond:CX_halide_alkyl-X 158 _bicyclo[2_2_1]heptene



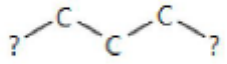
bond:CX_halide_alkyl-X 159 _dihalo_(1_1-)



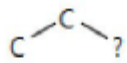
bond:CX_halide_alkyl-X 160 _dihalo_(1_2-)



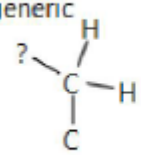
bond:CX_halide_alkyl-X 161 _dihalo_(1_3)



bond:CX_halide_alkyl-X 162 _ethyl



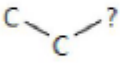
bond:CX_halide_alkyl-X 163 _ethyl_generic



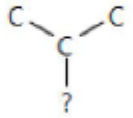
bond:CX_halide_alkyl-X 164 _generic



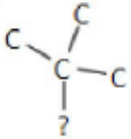
bond:CX_halide_alkyl-X 165 _primary



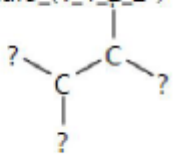
bond:CX_halide_alkyl-X 166 _secondary



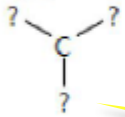
bond:CX_halide_alkyl-X 167 _tertiary



bond:CX_halide_alkyl-X 168 _tetrahalo_(1_1_2_2-)

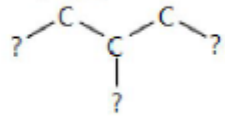


bond:CX_halide_alkyl-X 169 _trihalo_(1_1_1-)



bond:CX_halide_alkyl-X 170 _trihalo_(1_1_2-)

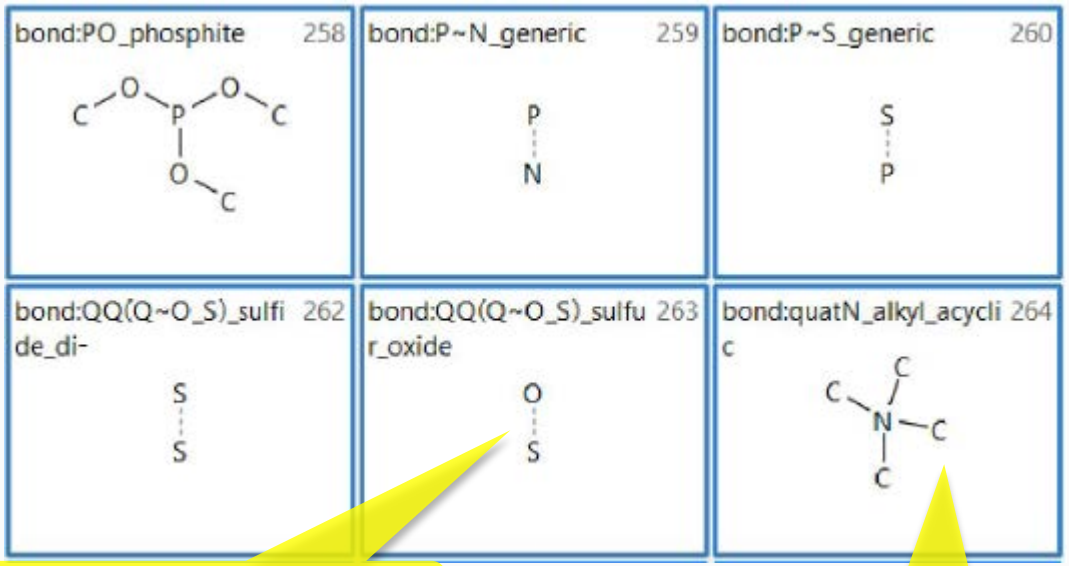
bond:CX_halide_alkyl-X 171 _trihalo_(1_2_3-)



Any halide: F, Cl, Br

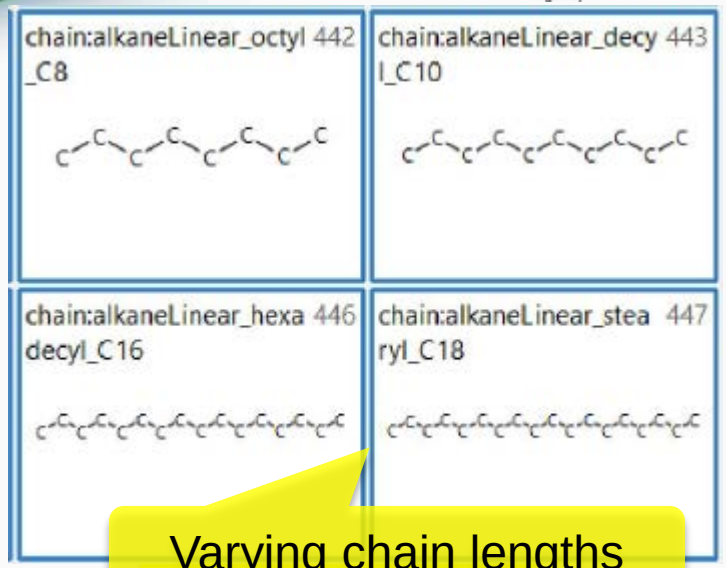
ToxPrints: Capturing Chemistry

EPA United States Environmental Protection Agency

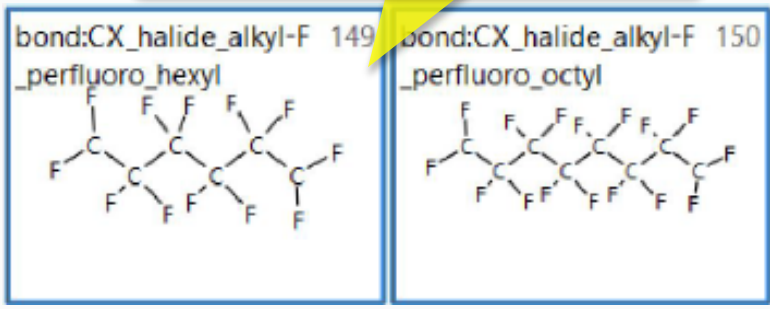


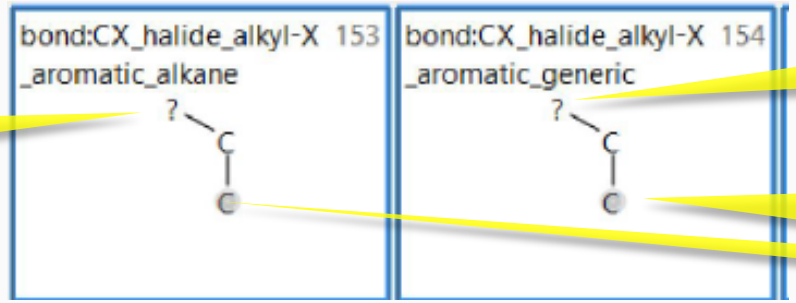
Generic bond type

Quaternary N



Varying chain lengths





Alkane [C] attachment

Generic: any atom

Aromatic [C] attachment

ToxPrint Chemotype Fingerprints

EPA United States Environmental Protection Agency

DSSTox_GSID	bond:C=O_carbonyl_generic	chain:aromaticAlkane_Ph-C1_acyclic_generic	bond:COH_alcohol_generic	bond:NC=O_aminocarbonyl_generic	bond:CX_halide_aromatic-X_generic	ring:hetero_[6]_Z_generic	bond:C(=O)N_carboxamide_generic	bond:CX_halide_alkyl-X_generic	bond:COH_alcohol_aliphatic_generic	bond:CN_amine_aliphatic_generic	bond:CN_amine_aromatic_generic	bond:C(=O)O_carboxylicAcid_generic	bond:CX_halide_alkyl-X_generic	chain:alkeneCyclic_ethene_generic	chain:alkeneLinear_mono-ene_ethylene_generic	bond:S=O_sulfonyl_generic	ring:hetero_[6]_N_pyridine_generic	bond:CC(=O)C_ketone_generic	bond:CN_amine_ter-N_generic	ring:hetero_[5_6]_Z_generic	ring:hetero_[6]_Z_generic	bond:CN_amine_pri-NH2_generic	➤bond:CX_halide_generic-X_dihalo_(1_2-)	bond:CN_amine_sec-NH_generic	bond:CN_amine_alicyclic_generic	bond:CC(=O)C_ketone_aliphatic_generic	ring:hetero_[6_6]_Z_generic	bond:C=O_carbonyl_ab-unsaturated_generic	bond:S~N_generic	bond:S(=O)O_sulfonicAcid_generic	bond:CC(=O)C_ketone_alkene_cyclic_2-en-1-one_generic	bond:CC(=O)C_ketone_alkene_generic	ring:hetero_[6]_O_pyran_generic	ring:hetero_[6]_N_diazine_(1_3-)_generic	bond:CC(=O)C_ketone_alkene_generic	bond:NC=O_urea_generic	ring:hetero_[5]_N_pyrrole_generic	bond:C#N_nitrile_generic	ring:hetero_[6]_N_triazine_generic	ring:fused_steroid_generic_[5_6_6_6]	bond:C=O_aldehyde_generic	bond:P~S_generic	bond:CX_halide_alkyl-X_ethyl_generic	group:aminoAcid_aminoAcid_generic	bond:CX_halide_alkenyl-X_generic	chain:aromaticAlkene_Ph-C2_acyclic_generic	ring:hetero_[6]_N_azo_generic	bond:metal_metalloid_Si_generic	bond:quatN_generic	bond:C=S_carbonyl_thio_generic	bond:S(=O)O_sulfuricAcid_generic	bond:C=N_carboxamidine_generic	ring:hetero_[7]_generic_1-Z	chain:alkyne_ethyne_generic	ring:hetero_[3]_Z_generic
20197	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47368	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47271	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47305	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47346	1	0	0	1	1	1	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47375	1	1	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21244	1	1	1	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22519	1	1	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24107	1	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47254	1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47289	1	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47311	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47355	1	0	0	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48507	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48511	1	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20822	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21097	1	0	1	1	1	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21233	1	1	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21777	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22588	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23322	1	0	1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23412	1	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23645	1	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25234	1	1	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34260	1	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47316	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47325	1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47339	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47347	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47351	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Similarity – in the "eye" of the beholder

Similarity, clustering results are database-dependent, no clear, extractable chemical basis for similarity provided

Chemical Similarity

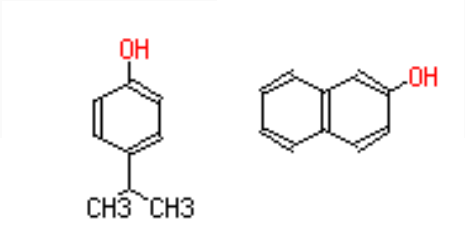
Pre-defined, standardized, invariant chemical moieties provide communicable basis for similarity assessment

Structural fingerprint

1000110101000000110
(all parts considered equal)

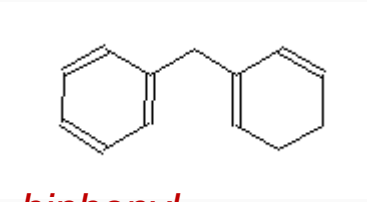
Tanimoto Coeff

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



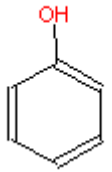
Chemotypes

Pre-defined features or scaffolds designed for toxicity database coverage and informed by SAR, public inventories, and FDA safety assessment workflows



biphenyl

phenol

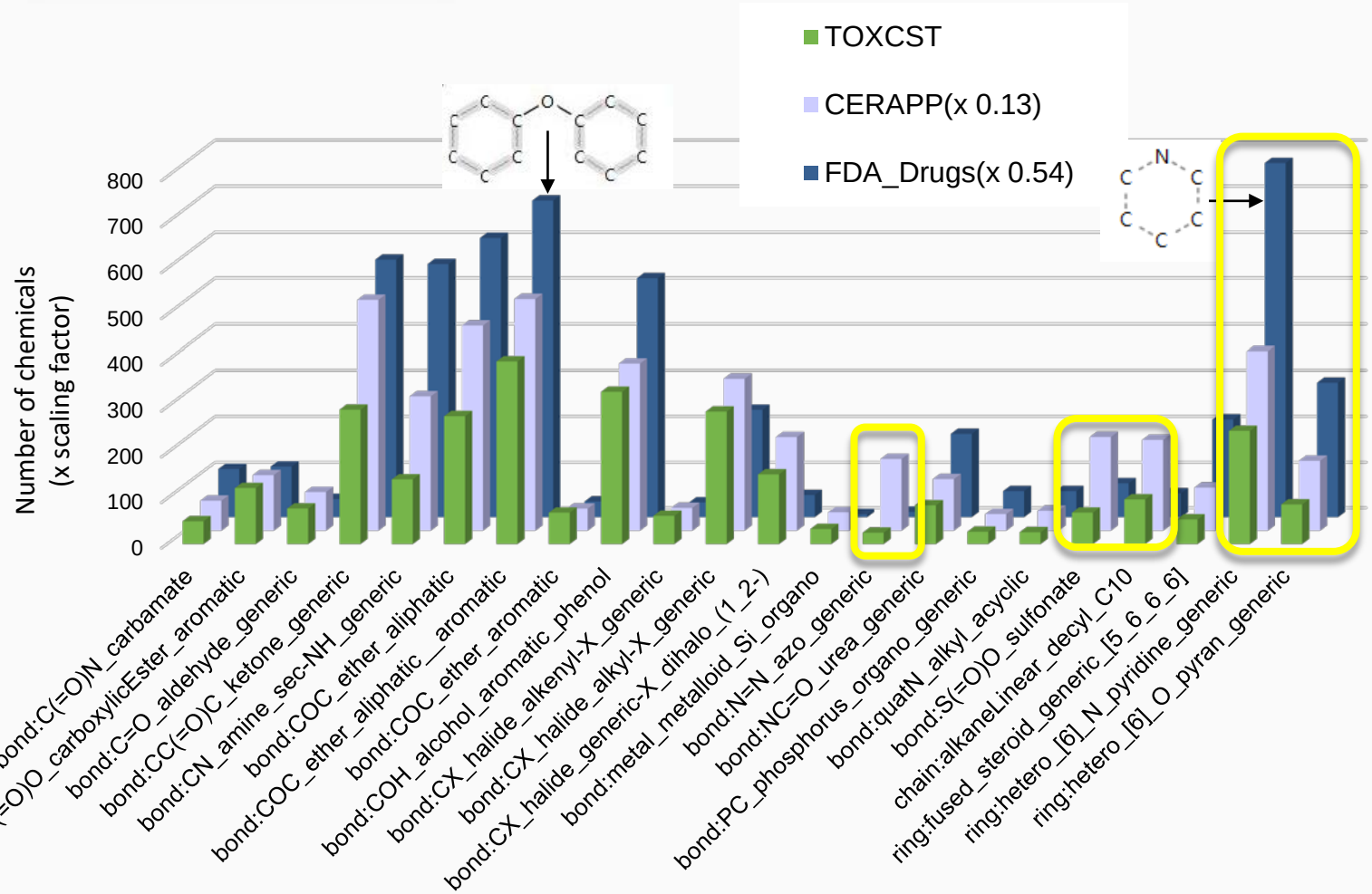


Rules for judging similarity are clearly defined, interpretable, computable, and transferable.

Chemotype (CT) Applications

- Chemical Inventory Profiling Comparisons
  - Assess coverage and diversity of inventories
  - Assess local chemistry data gaps or enrichments
- {CT-Activity} enrichments:
  - Use-categories *(e.g., Flame retardants)*
  - HTS assays *(e.g., ToxCast HTS assays)*
  - in vivo activity subsets *(e.g., Rat Carc w and w/o metabolism)*
- "Chemo-across"
  - use chemotype layer to mine data and identify HTS assay associations through {CT-Activity} enrichments

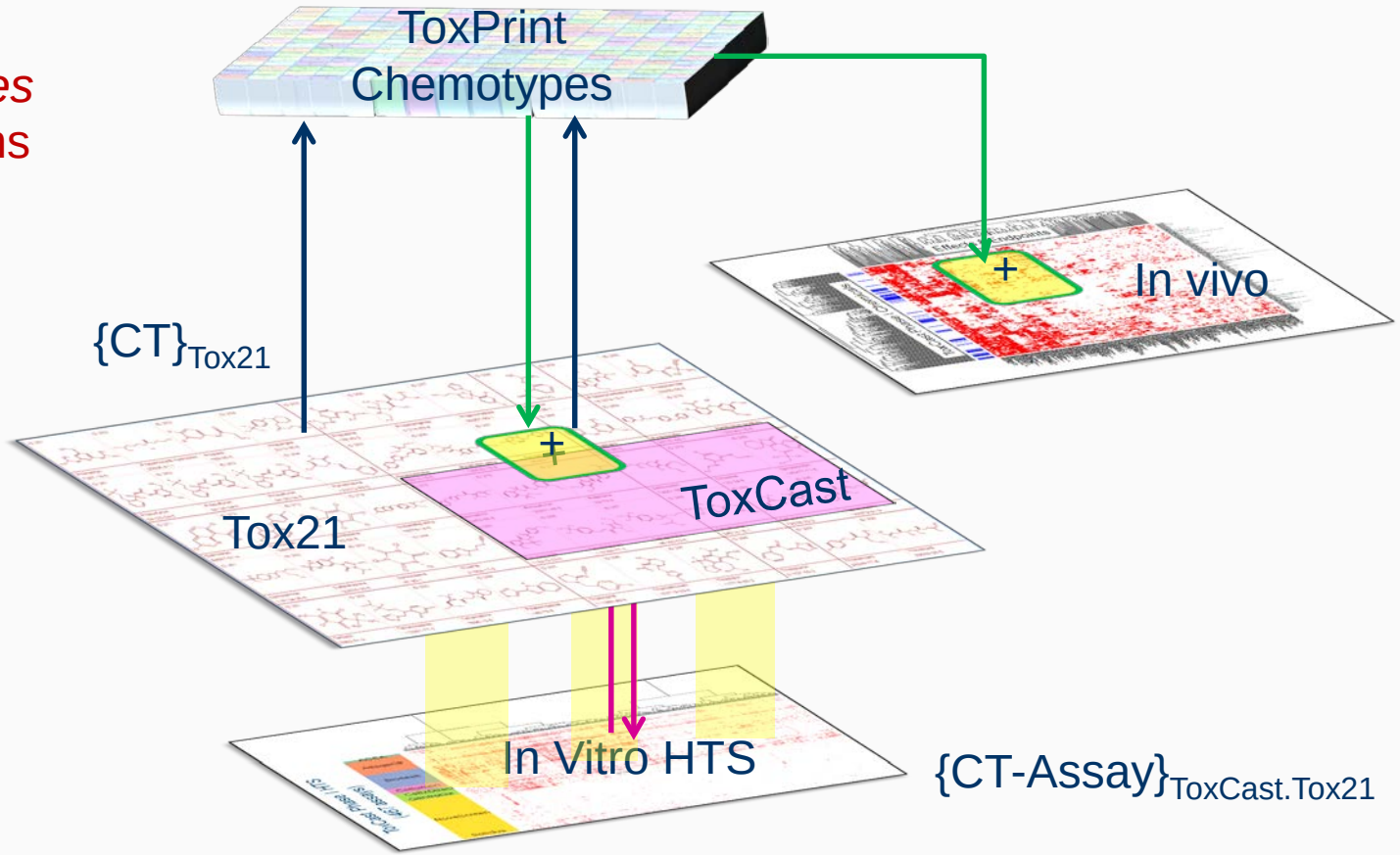
ToxPrint inventory profile comparisons (scaled)



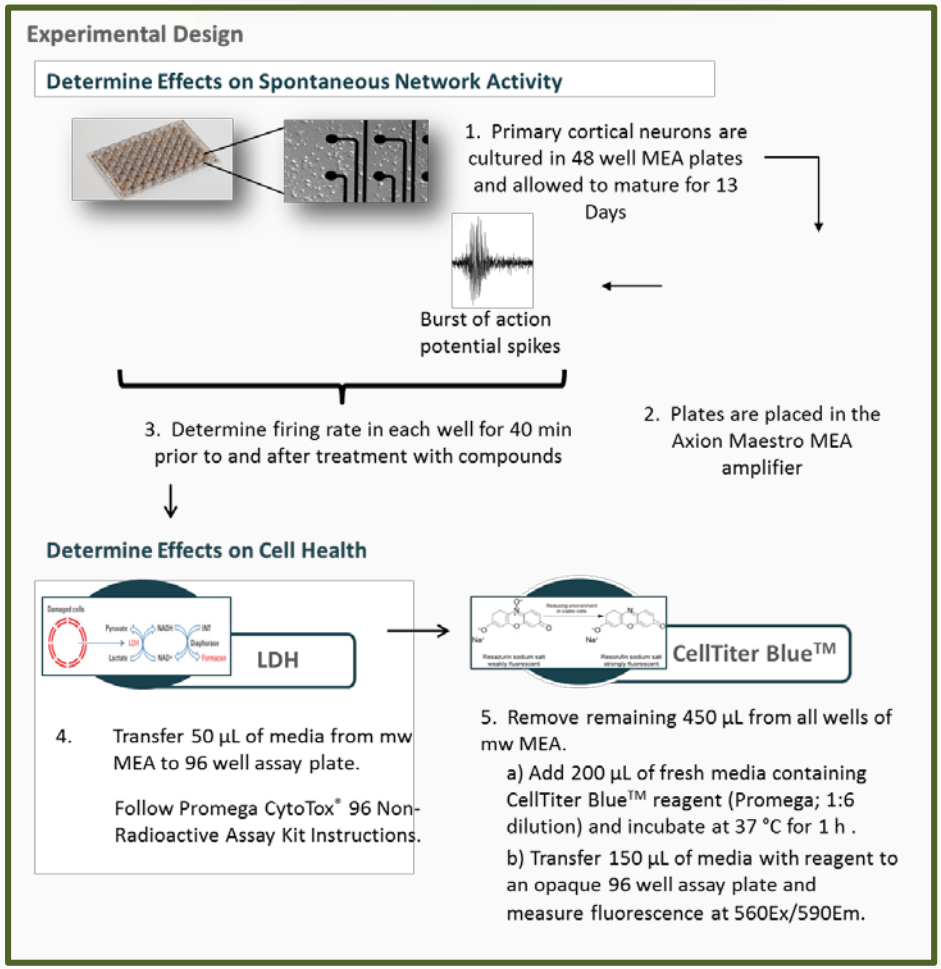
Chemotype-Activity Enrichments

Create {CT-Activity} profiles for any biological activity subset to identify CT enrichments, focus studies in local CT domains, and enhance HTS or structure-activity signal

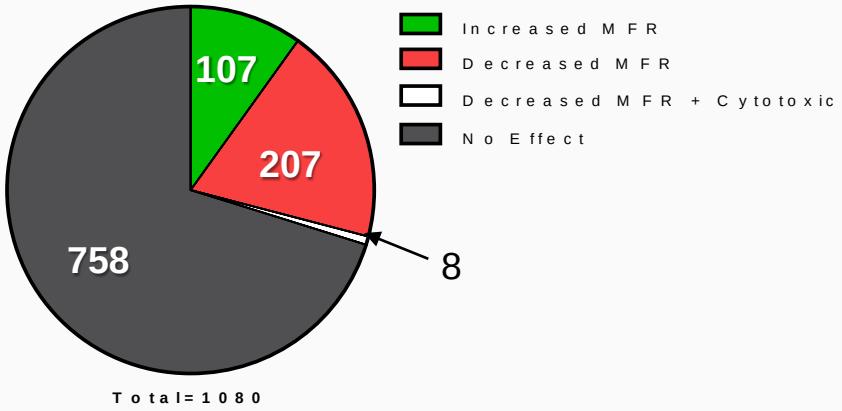
- *HTS actives*
- *Pathway actives*
- SAR predictions
- *In vivo* results



Multi-electrode array (MEA) neurotoxicity assay *(T.Shafer et al, 2017)*



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



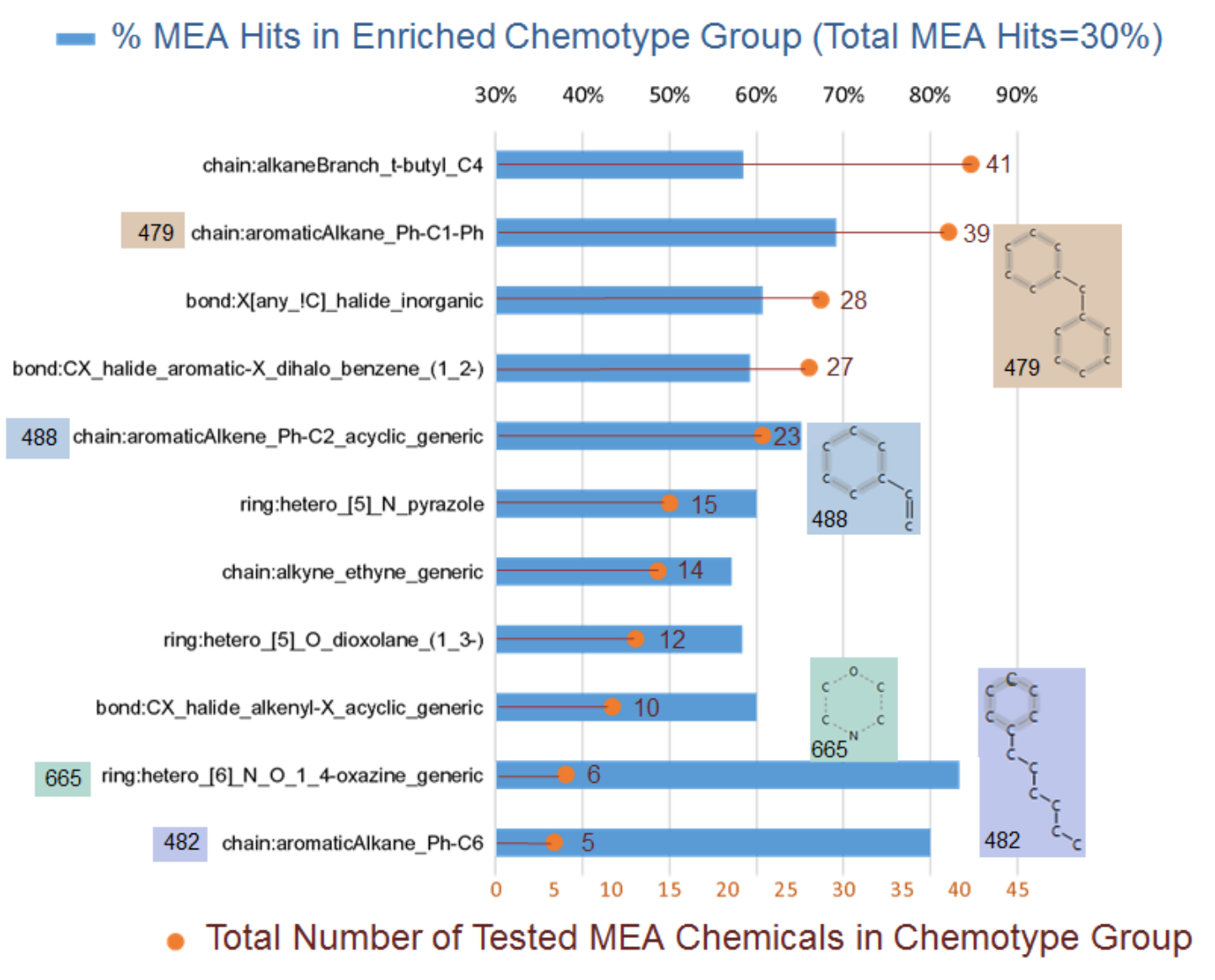
- 314 total “Actives” or Hits
- 758 “Inactives” or No-hits (8 cytotoxic)
- Total Tested = 1080 (30% hit rate)

MEA Chemotype (CT) Analysis:
1) Process Fingerprint file

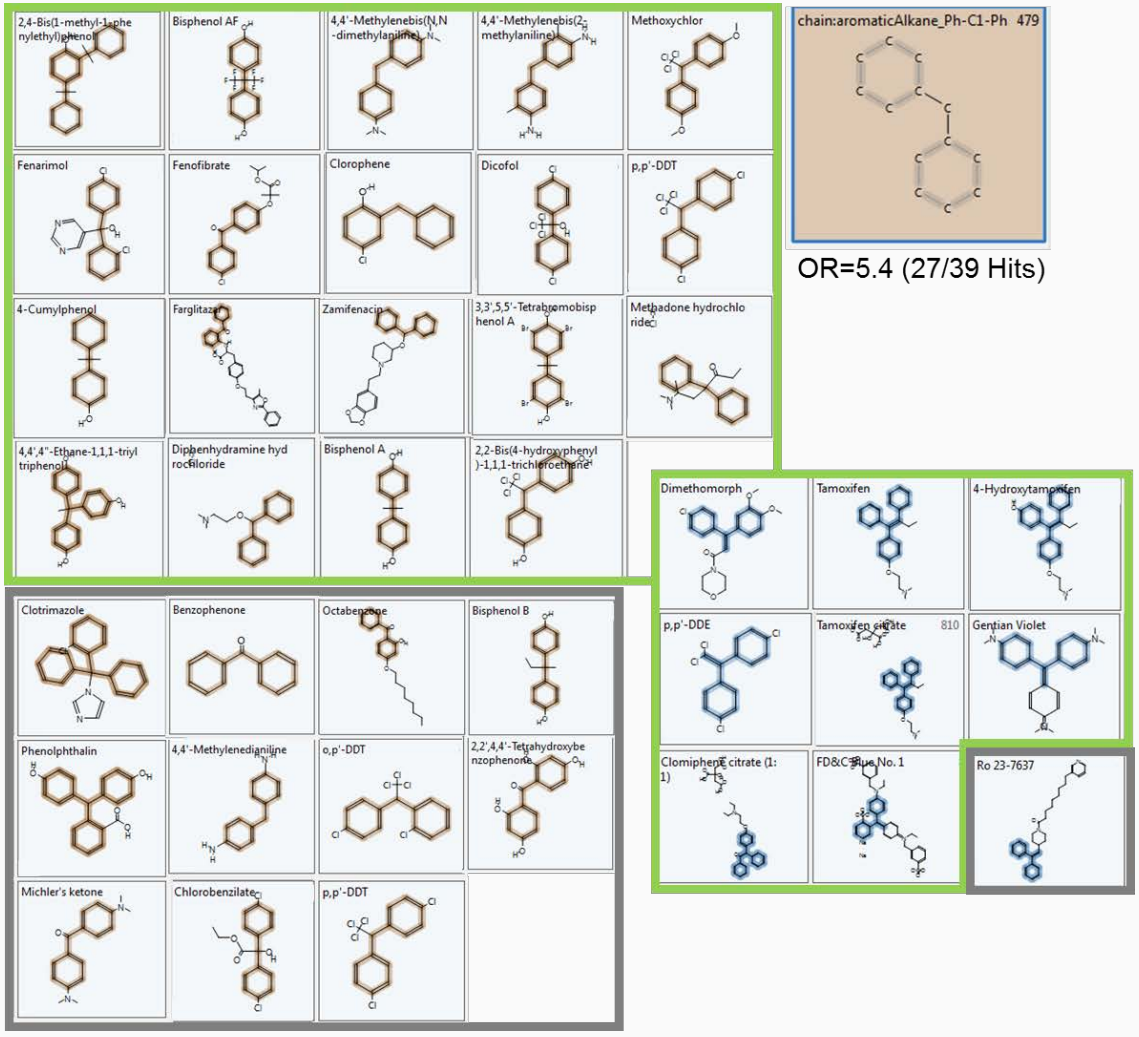
Top 11 statistically enriched ToxPrint CTs for the MEA hits vs total test set
(top 11 in 34% actives; top 5 in 26% of hits)

TP_ID[1]	ToxPrint_CT_name[2]	CT-Tot	TPos	FPos	FNeg	TNeg	OR	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
180	bond:CX_halide_aromatic-X_dihalo_benzene_(1_2-)	27	16	11	302	699	3.4	0.002
488	chain:aromaticAlkene_Ph-C2_acyclic_generic	23	15	8	303	702	4.3	6.0E-05
611	ring:hetero_[5]_N_pyrazole	15	9	6	309	704	3.4	0.02
471	chain:alkyne_ethyne_generic	14	8	6	310	704	3.0	0.04
624	ring:hetero_[5]_O_dioxolane_(1_3-)	12	7	5	311	705	3.2	0.04
139	bond:CX_halide_alkenyl-X_acyclic_generic	10	6	4	312	706	3.4	0.05
665	ring:hetero_[6]_N_O_1_4-oxazine_generic	6	5	1	313	709	11.3	0.01
482	chain:aromaticAlkane_Ph-C6	5	4	1	314	709	9.0	0.03

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



MEA Chemotype (CT) Analysis:
3) Examine enriched CT subsets

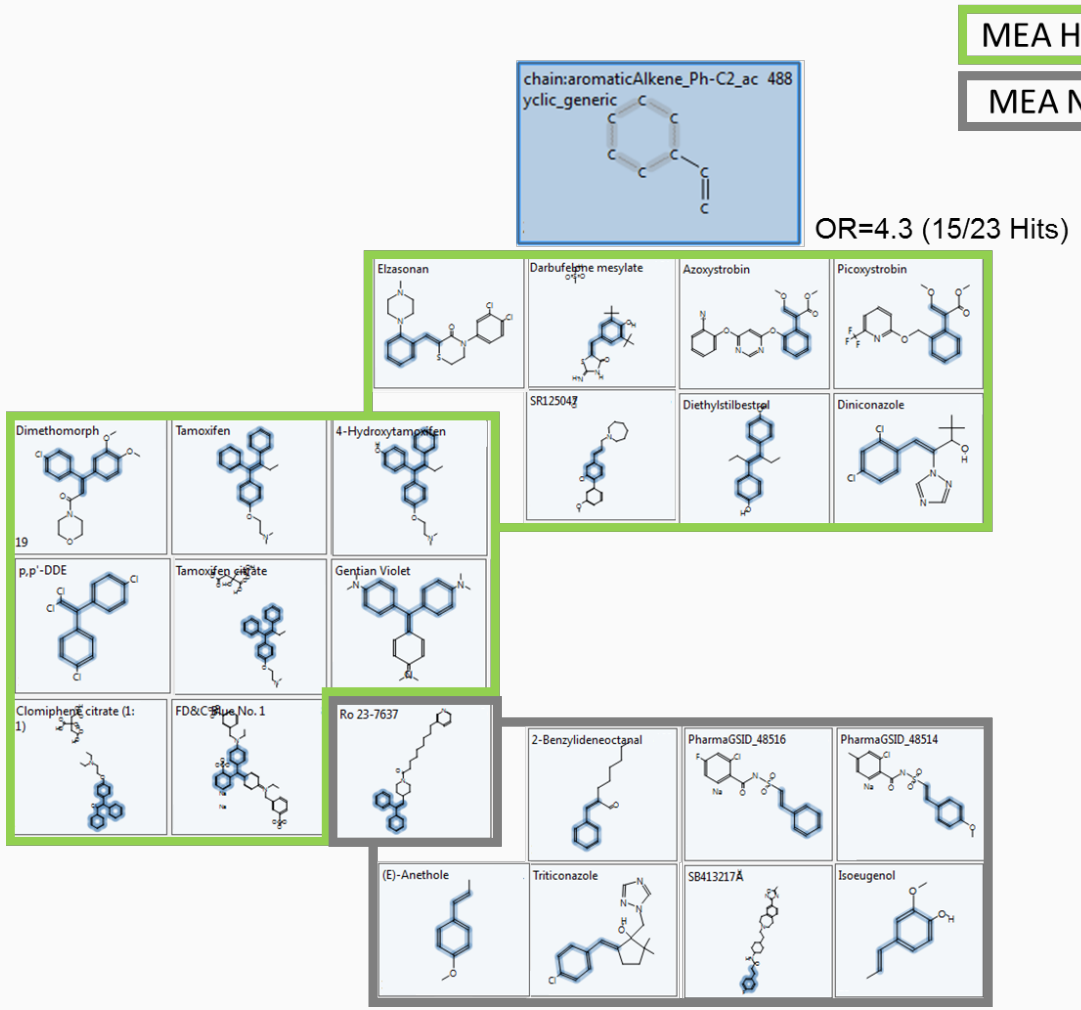


OR=5.4 (27/39 Hits)

MEA Hits

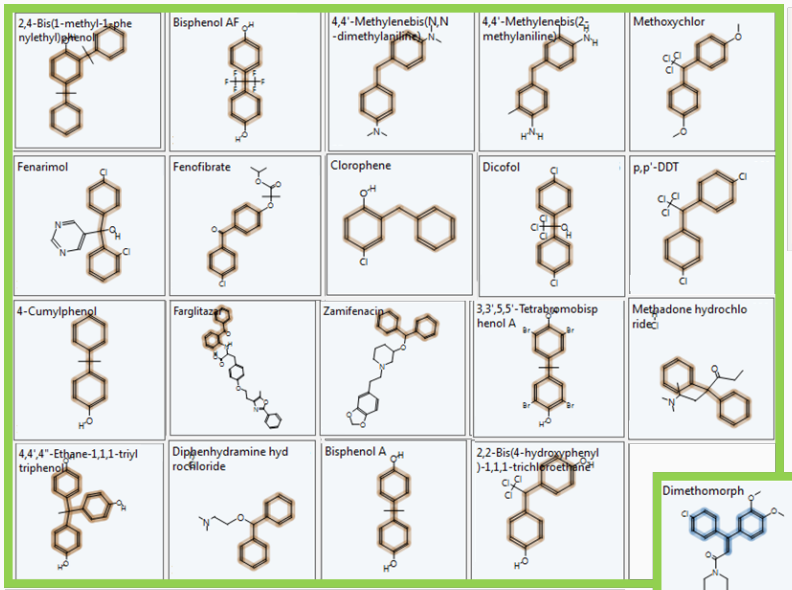
MEA Non-Hits

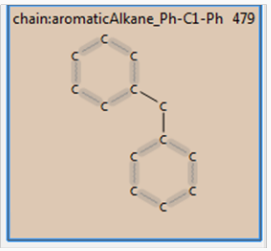
MEA Chemotype (CT) Analysis:
3) Examine enriched CT subsets



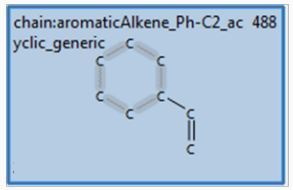
MEA Chemotype (CT) Analysis:
4) Examine overlapping CT subsets

EPA United States Environmental Protection Agency





OR=5.4 (27/39 Hits)



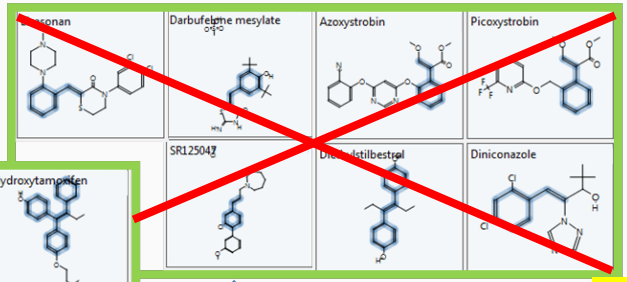
OR=4.3 (15/23 Hits)

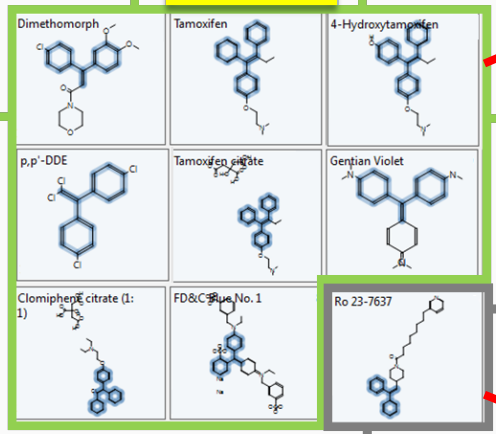
MEA Hits

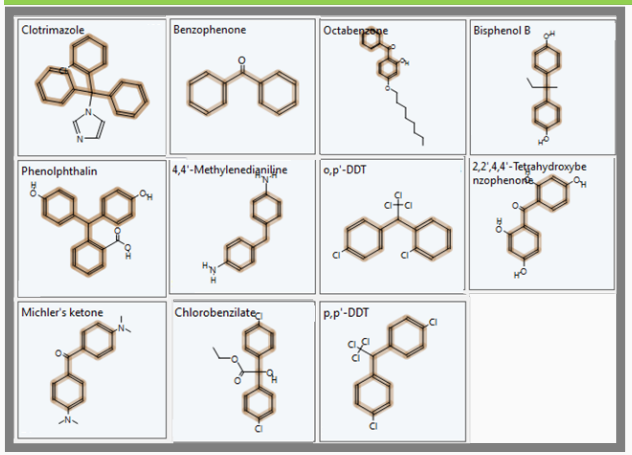
MEA Non-Hits

OR=18.3

479 & 488

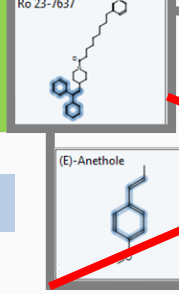






488 minus 479

OR=1.9



479 minus 488

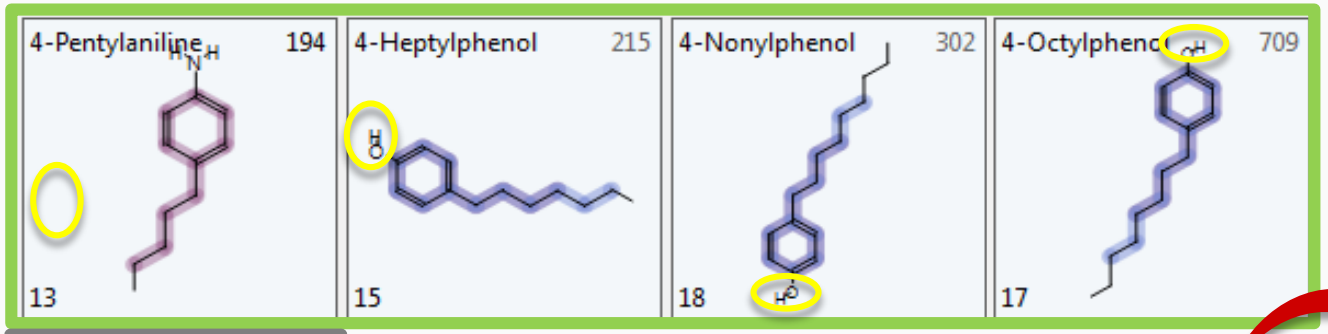
OR=4.0

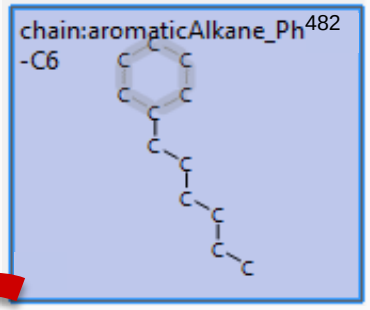
New Rules:
CT488
CT488 & CT479
~~CT479~~

MEA Chemotype (CT) Analysis:
5) Modify rules for CT subsets

MEA Hits

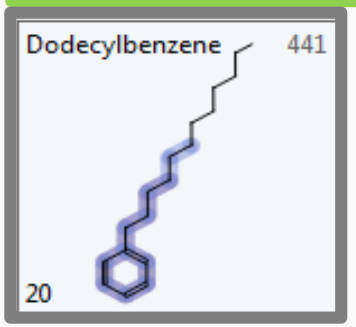
MEA Non-Hits





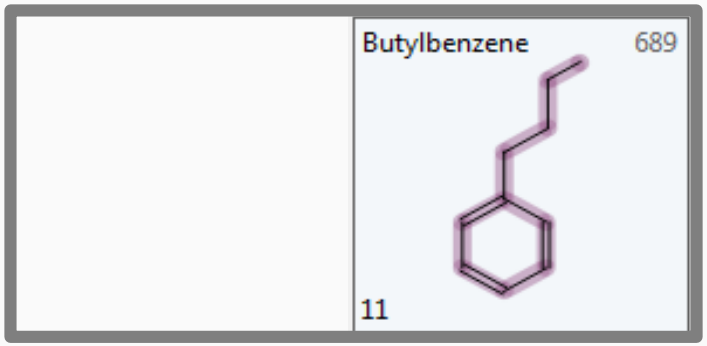
OR=9 (4/5 Hits)

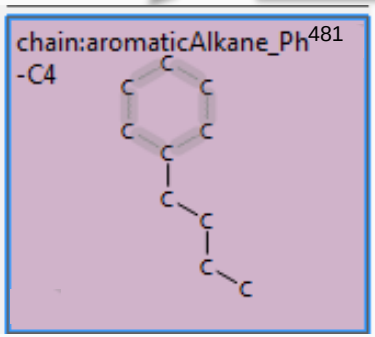
Hypothesis 1:
Hits require [phenol]
(or [para-OH])



New Rule:
CT481 + [phenol]

OR= ∞ (7/7 Hits)



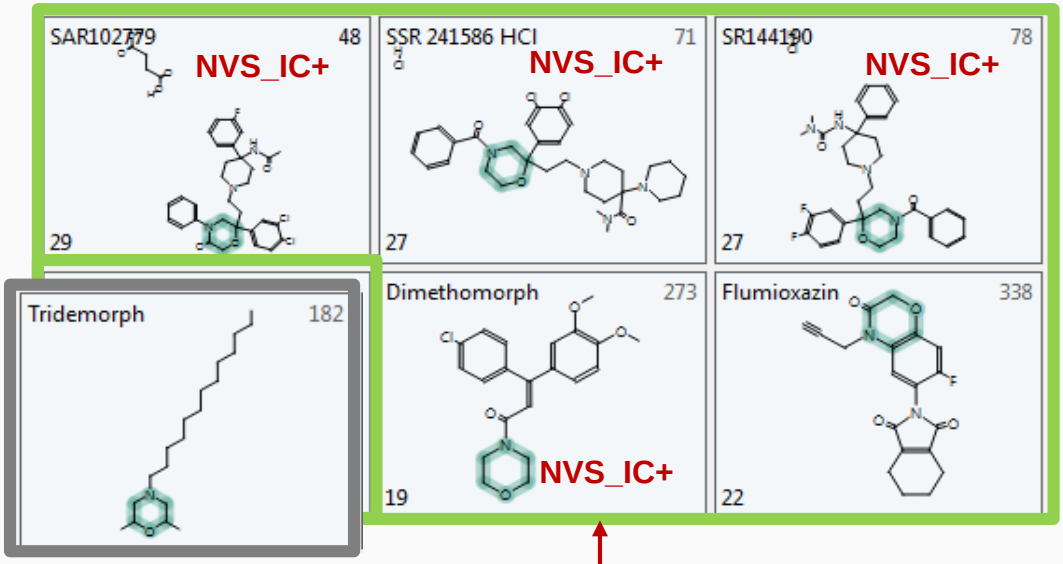


OR=3 (4/7 Hits)

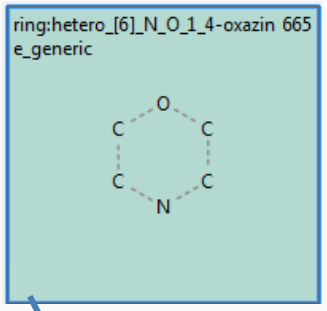
MEA Chemotype (CT) Analysis:
6) Use CTs to build biological linkages

MEA Hits

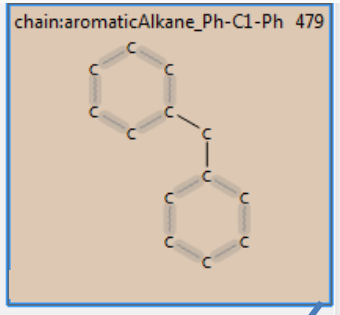
MEA Non-Hits



OR=11.3 (5/6 Hits)



OR=5.4 (27/39 Hits)



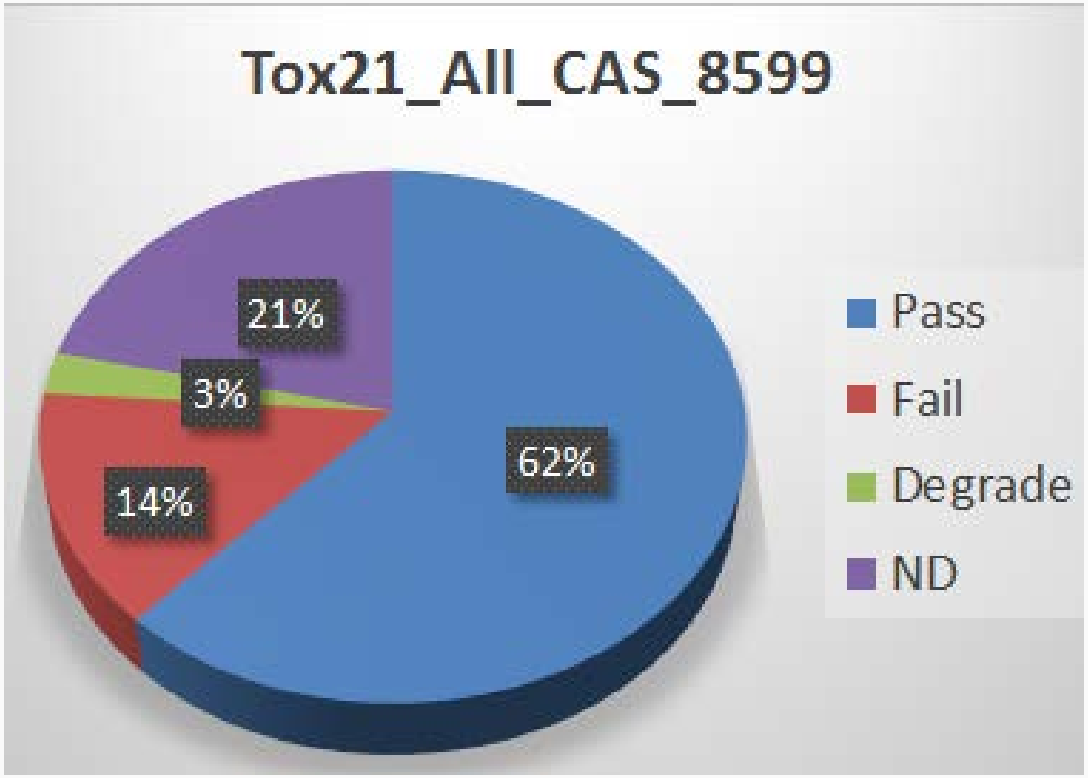
ToxCast assays enriched with hits in CT665 (OR > 4)

- 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

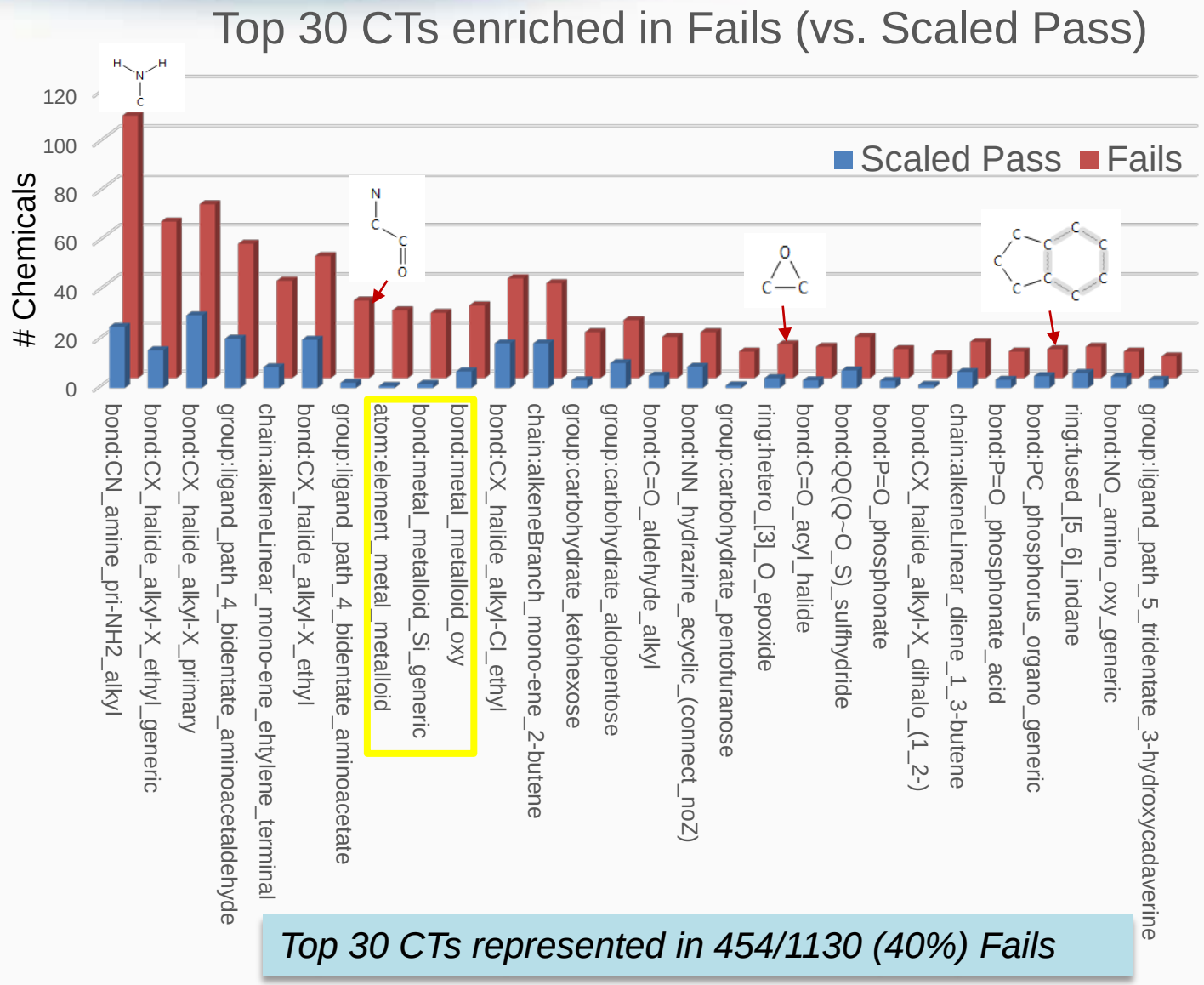
16/27 MEA hits in CT479 also hits in NVS_IC assays

Summarized QC Results



Pass = C (75%) or greater
Fail = D, F, Ac, Bc, Cc
Degrade = T0-Pass, T4-Fail
ND = Not Determined

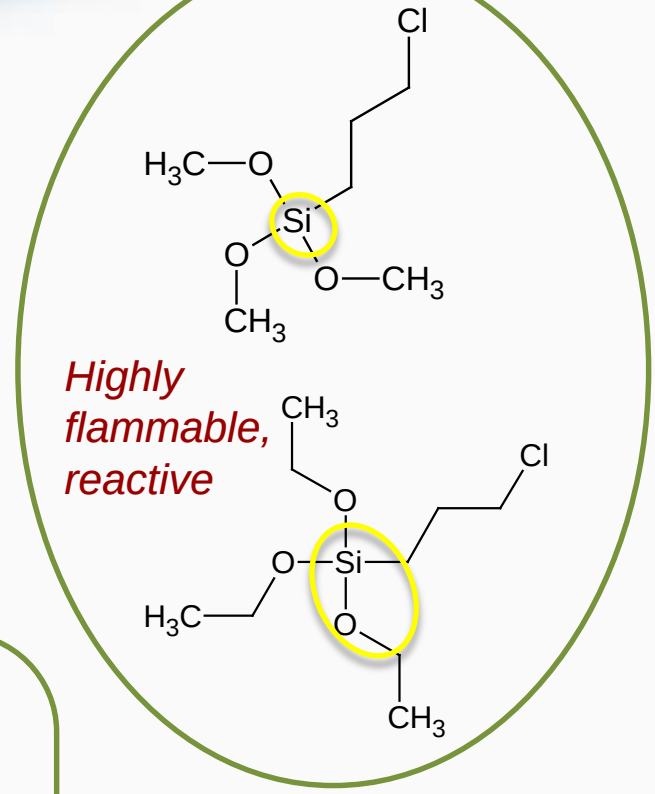
ToxPrints enriched in Fails

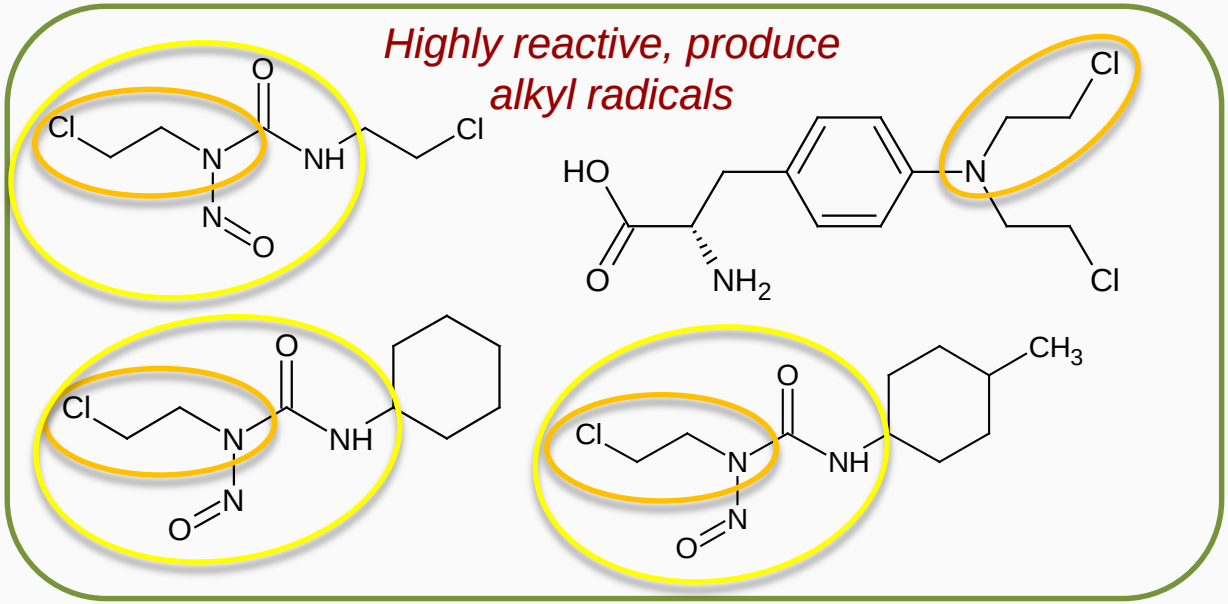


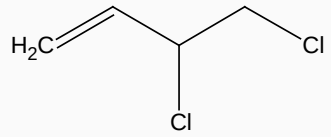
Top 30 CTs represented in 454/1130 (40%) Fails

Fail Chems with Mulitple Fail CTs

DSSTox_Substance_Id	GSID x top 30 CT Fails	Chemical Name
DTXSID4027490	7	(3-Chloropropyl)trimethoxysilane
DTXSID5029267	7	(3-Chloropropyl)(triethoxy)silane
DTXSID6020804	7	Melphalan
DTXSID8022113	6	3,4-Dichloro-1-butene
DTXSID8022743	6	Carmustine
DTXSID2023222	6	Lomustine
DTXSID8031603	6	Semustine







Reacts w/H_2O to form HCl

CT Use Case #1:
Compare 2 “inventories”

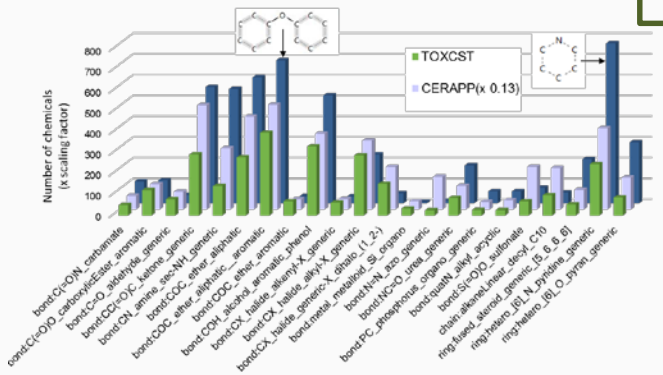
Chemical Dashboard Query

Compare DSSTox Inventories in CT space

Select DSSTox “Lists” from pull down

Enter 1 or more lists of CAS, DTXSIDs, or SMILES

Display CT-subset overlap graph for inventory “view”

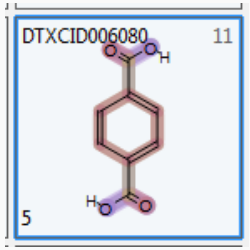


- *Mouse-over display of CT*
- *Highlight greatest differential CTs (or limit view)*
- *View chemicals within CT-inventory subspace*

CT Use Case #2: Profile new chemical

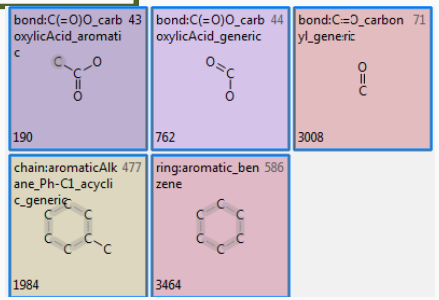
Chemical Dashboard Query

Create CT profile of new chemical- C1



Display C1 with C1-CTs superimposed

Select C1-CTs for building “analog” search



Which data sets or endpts are of interest: e.g., ToxCast, Exposure, InVivo, Carc, SkinSenstization?

CAUTION: CT associated with cytotoxicity, promiscuity

Tox21

ToxCast

CT-KB

Tox21

ToxCast

CAUTION: CT associated with degradation, failed QC, etc.

List of assays, use categories or endpts, significantly enriched with C1-CTs

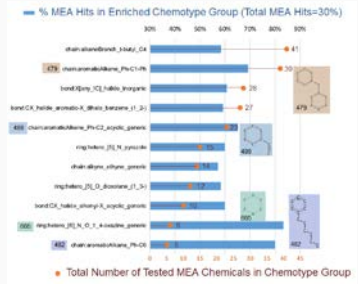
SAR within CT-limited assay-chemical space

CT Use Case #3:
CT-assay enrichment study

EPA United States Environmental Protection Agency

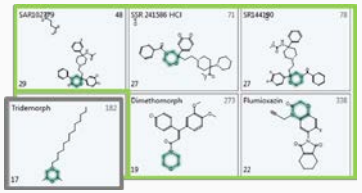
Top 11 statistically enriched ToxPrint CTs for the MEA hits vs total test set

TP_ID	ToxPrint_CT_name	CT-Tot	TPos	FPos	FNeg	TNeg	OR	pval
423	chain:alkaneBranch_1-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:Xany_!C!_halide_inorganic	28	17	11	301	699	3.6	9.0E-04
180	bond:CX_halide_aromatic-X_dihalo_benzene_(1,2-)	27	16	11	302	699	3.4	0.002
488	chain:aromaticAlkane_Ph-C2_acyclic_generic	23	15	8	303	702	4.3	6.0E-05
611	ring:hetero_[5]_N_pyrazole	15	9	6	309	704	3.4	0.02
471	chain:alkyne_ethyne_generic	14	8	6	310	704	3.0	0.04
624	ring:hetero_[5]_O_dioxolane_(1,3-)	12	7	5	311	705	3.2	0.04
139	bond:CX_halide_alkenyl-X_acyclic_generic	10	6	4	312	706	3.4	0.05
665	ring:hetero_[6]_N_O_1_4-oxazine_generic	6	5	1	313	709	11.3	0.01
482	chain:aromaticAlkane_Ph-C6	5	4	1	314	709	9.0	0.03



Chemical Dashboard Query

CT enrichment of "assay" endpt



Display top CT enrichments w/statistics



Select CTx to view enrichment details

What other ToxCast assays have CTx enrichment in hits?

CT-KB

CAUTION: CT associated with cytotoxicity, promiscuity

Tox21

ToxCast

Tox21

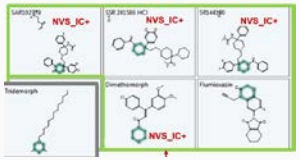
ToxCast

CAUTION: CT associated with degradation, failed QC, etc.

NVS_GPCR_hOpiate_mu
NVS_GPCR_rOpiate_NonSelective
NVS_GPCR_rOpiate_NonSelectiveNa
NVS_IC_rCaBTZCHL
NVS_IC_rCaDHPRCh_L
NVS_IC_rNaCh_site2

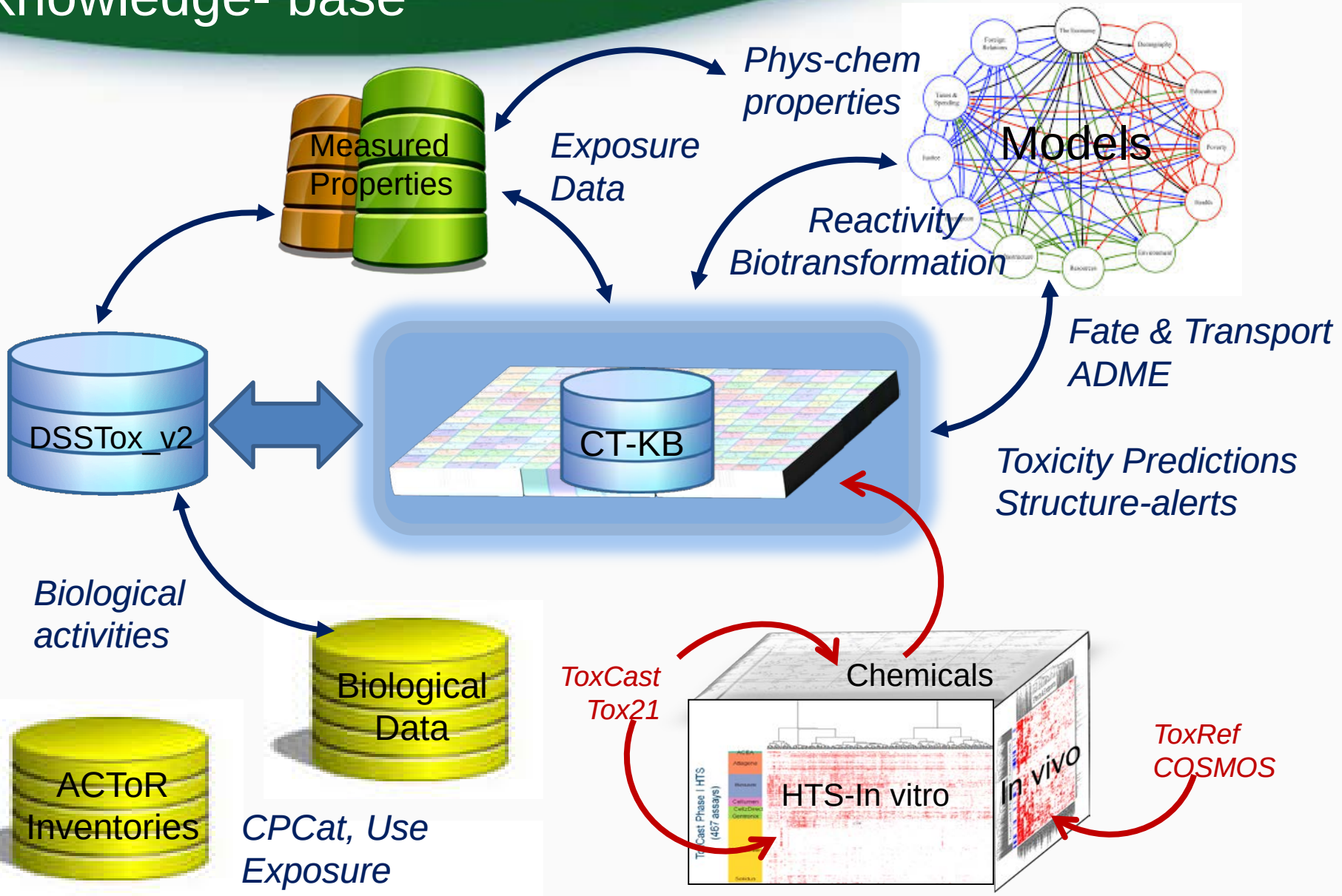
List of assays (use categories or endpts) significantly enriched with C1-CTs

Select assays with biological relevance, assess overlap of hits across endpts



→ SAR study

Building a public chemotype "knowledge- base"



Acknowledgements:

- EPA NCCT Chemistry Team
  - Chris Grulke (DSSTox, Chemotypes)
  - Indira Thillainadarajah (DSSTox)
  - Tony Williams (DSSTox, Chemistry Dashboard)
  - Grace Patlewicz (Read-across, Corina)
- MEA ToxCast study
  - Tim Shafer, Jenna Strickland, Matt Martin, Kristin Connors
- External Collaborators – ToxPrint Chemotypes
  - Molecular Networks: Chihae Yang, Aleksey Tarkhov, Christof Schwab
  - Altamira: Jim Rathman
  - US FDA: Kirk Arvidson, Patra Volarath