ToxCast:
Chemical List(s), Chemical Space, and Chemotype Information

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

CompTox Communities of Practice, February 27, 2014

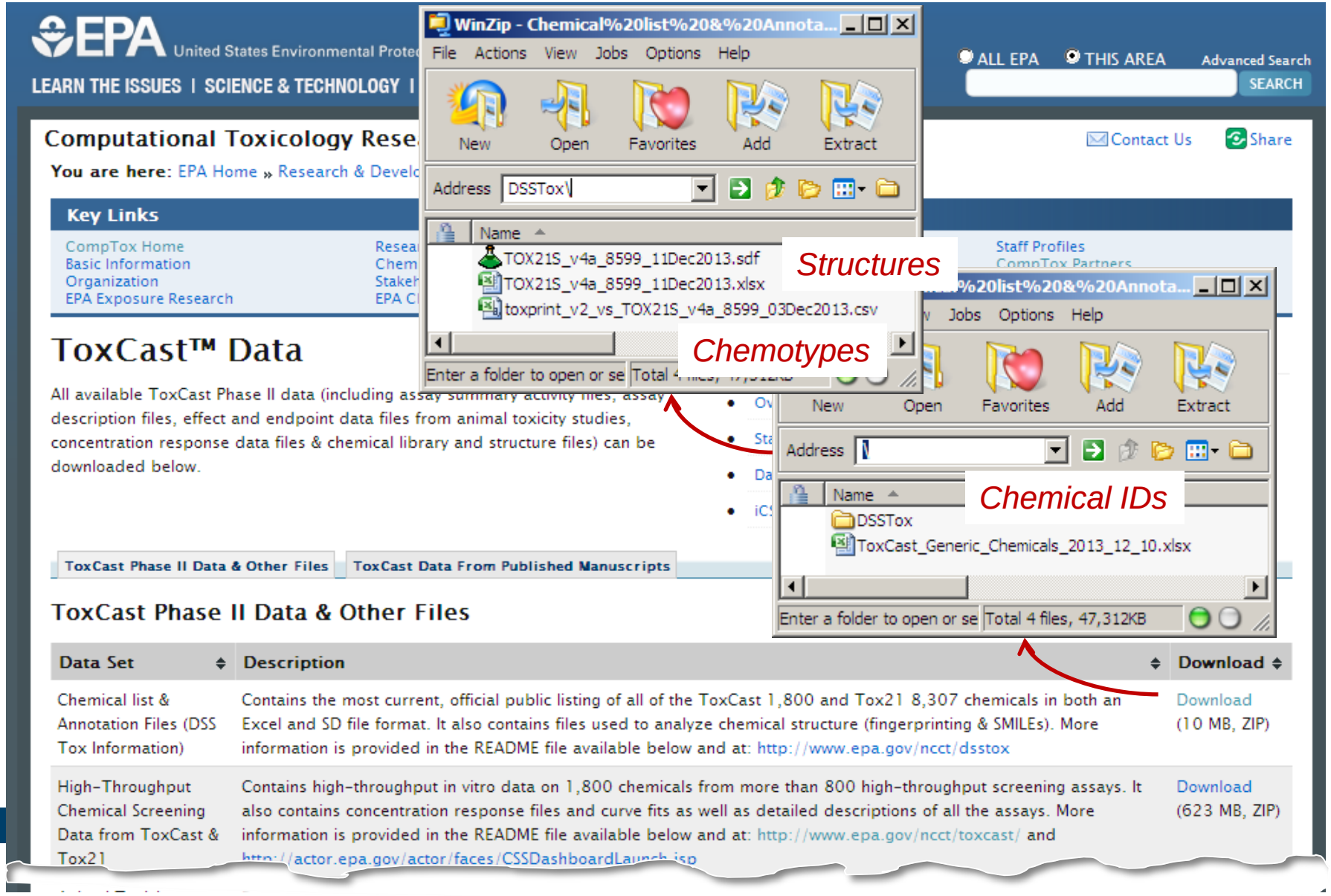
ToxCast Phase II Data Release

http://www.epa.gov/ncct/toxcast/

- ToxCast Assay Summary Activity Files
- ToxCast Assay Annotation Files
- ToxCast Chemical Library & Structure Files (DSSTox)
- ToxCast Concentration Response Data Files
- ToxRefDB Effect & Endpoint Data Files

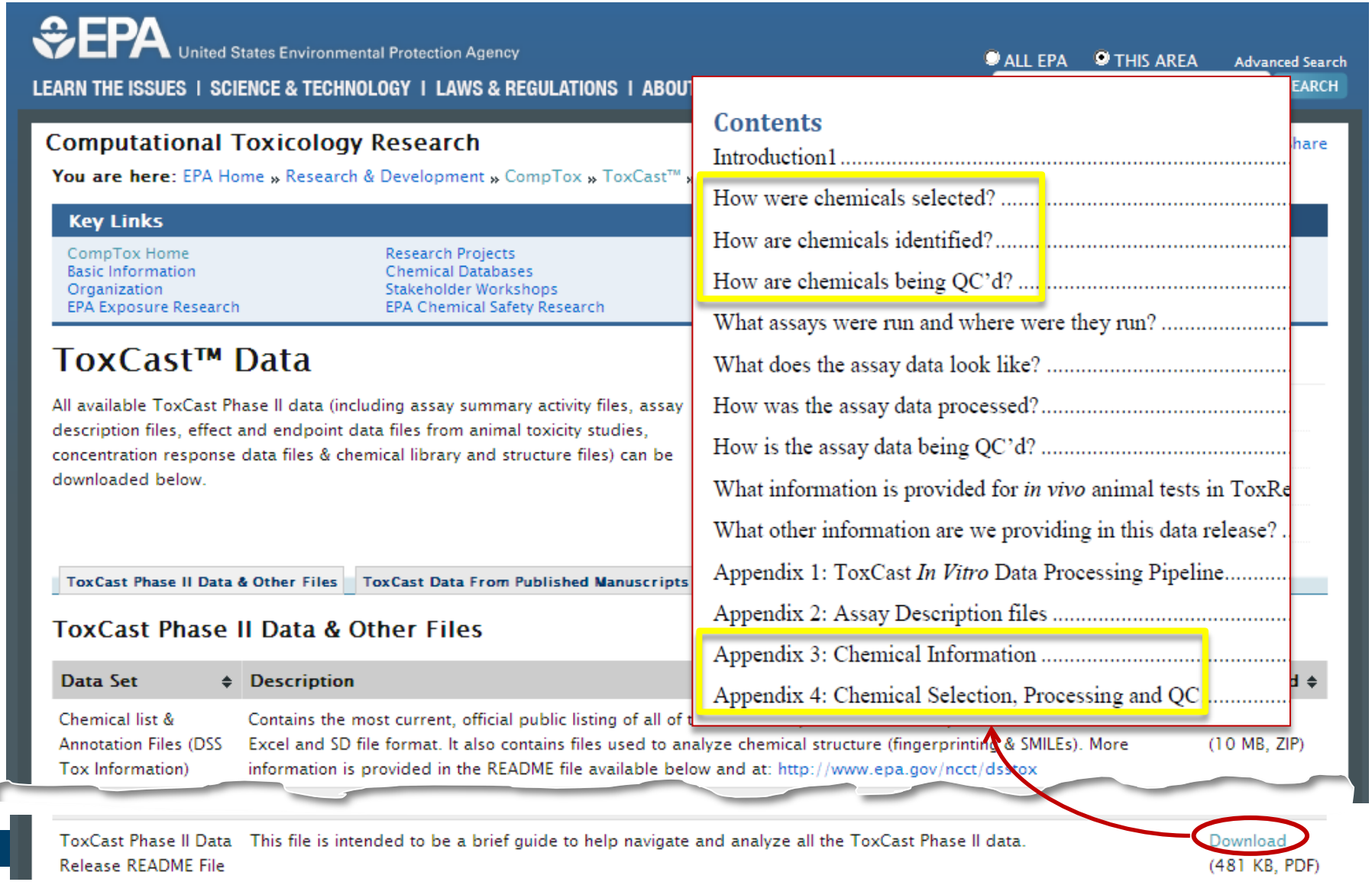
ToxCast Data: Chemical Files

http://www.epa.gov/ncct/toxcast/data.html



ToxCast Data: Chemical Files

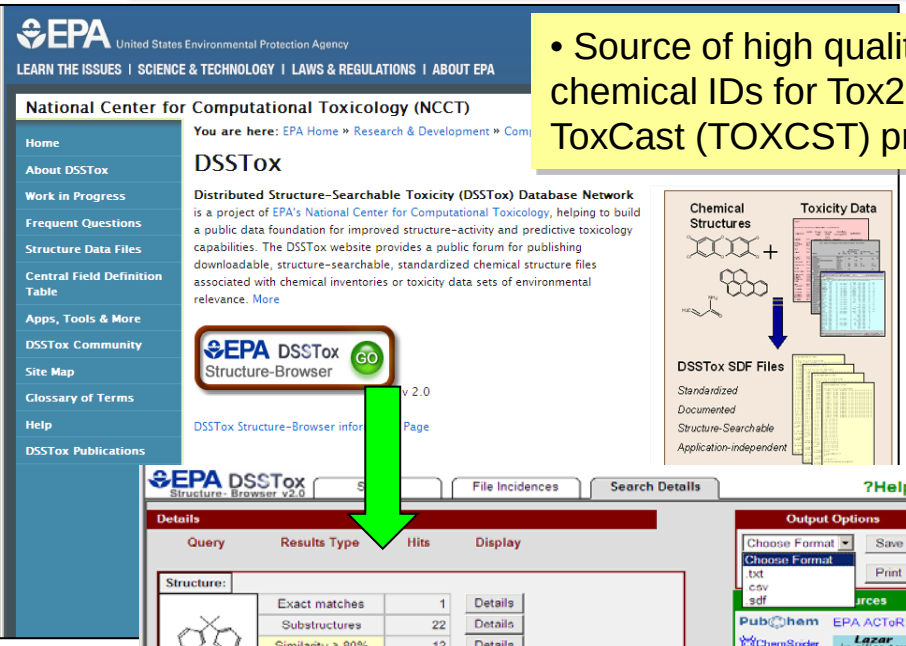
http://www.epa.gov/ncct/toxcast/data.html

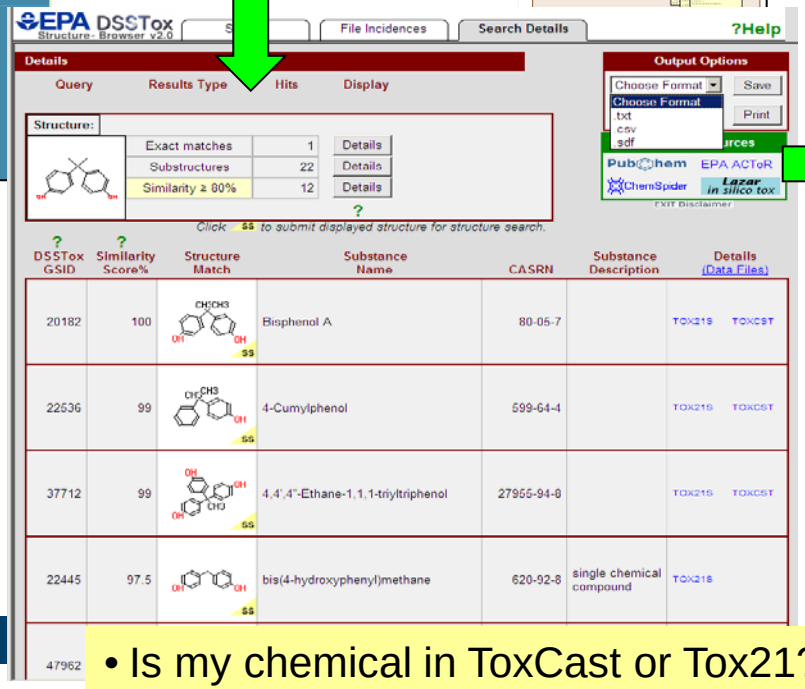


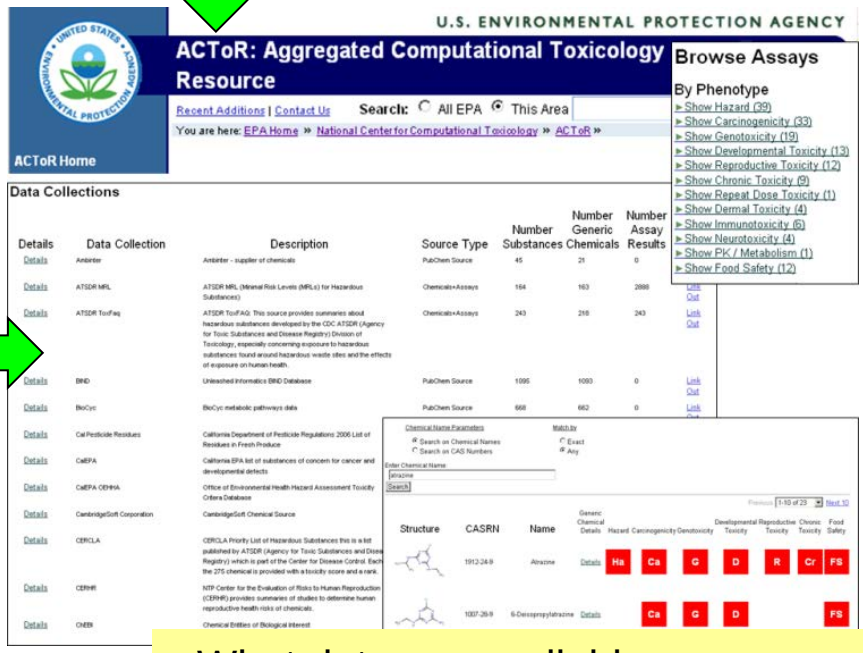
DSSTox Chemical Structure Files & Browser

http://www.epa.gov/ncct/dsstox/

• Source of high quality structure files and chemical IDs for Tox21 (TOX21S) and ToxCast (TOXCST) programs



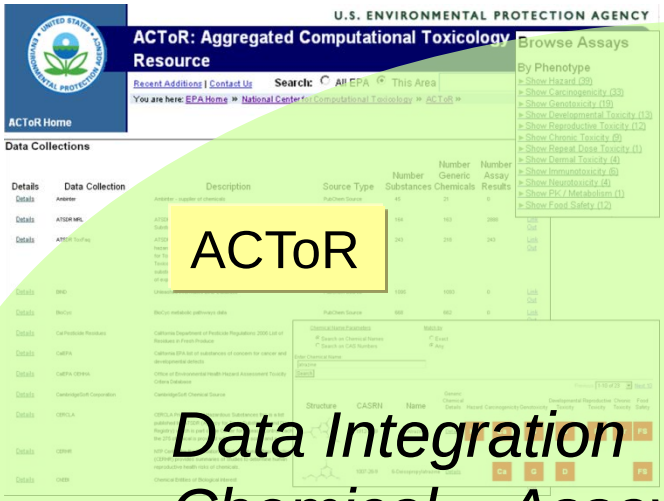




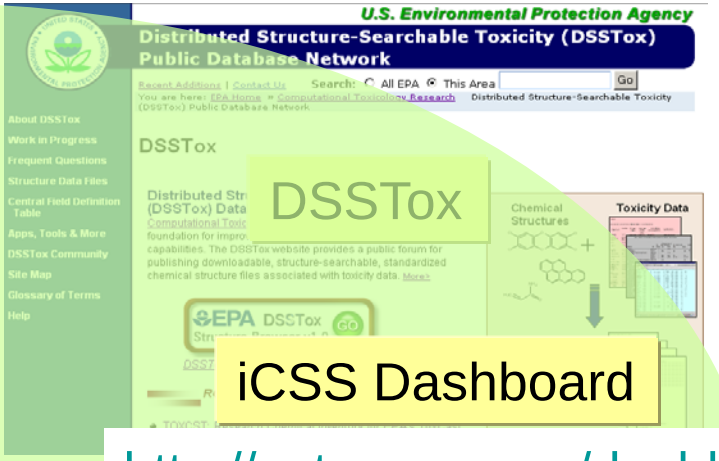
• What data are available across EPA and Internet resources for my chemical or its analogs?

• Is my chemical in ToxCast or Tox21?
• Are "similar" chemicals being tested?

NCCT Public Data http://www.epa.gov/ncct/



ACToR

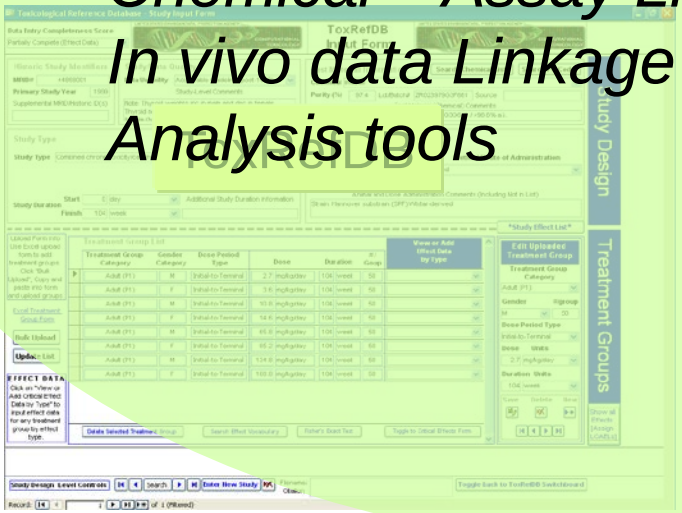


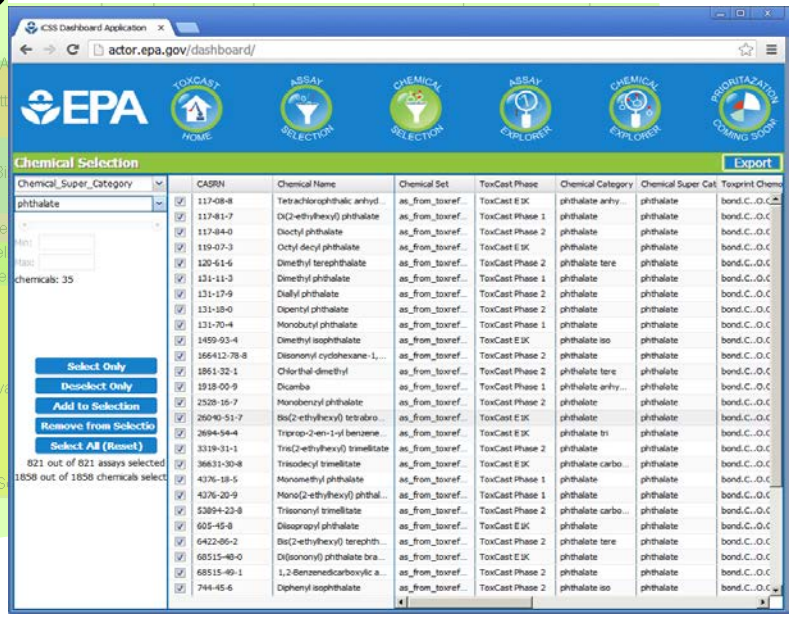
DSSTox

iCSS Dashboard

Data Integration
Chemical – Assay Linkage
In vivo data Linkage
Analysis tools

http://actor.epa.gov/dashboard/





ToxCast Inventories

293 • ToxCast Phase I (293 unique cmpds)

- *EPA pesticidal actives w/ rich in vivo data*
- *PFOAs, BPA, approx 12 metabolite/parent pairs*

1060 • ToxCast Phase II (767 unique new cmpds)

- *EPA pesticides, high interest EPA and stakeholder inventories, data rich chemicals (EDSP, OPPT, antimicrobials, inerts, green alternatives, fragrances, water …)*
- *FDA CFSAN data rich, NCTR LTKB Priority 1 drugs*
- *Toxicity reference chemicals, data-rich chemicals, NTP immunotox*
- *135 Donated pharma cmpds -- failed drugs w/ pre-clinical or clinical tox data*

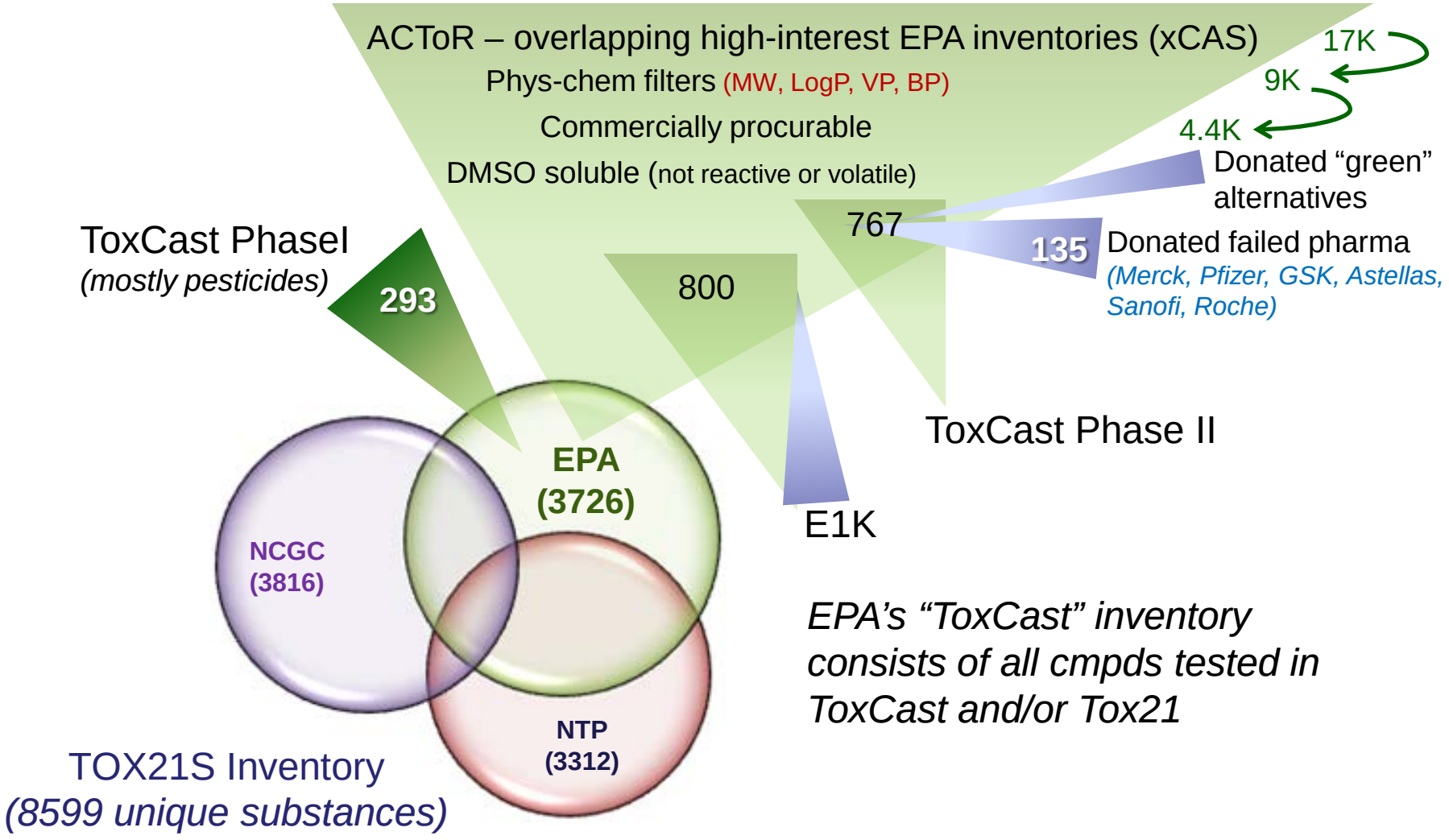
1860 • ToxCast E1K (800 unique new cmpds)

- *Endocrine active reference cmpds, SAR predicted ER-active/inactives, EDSP cmpds*

3727 • EPA's Tox21 library (3727 unique cmpds out of current 8599 total)

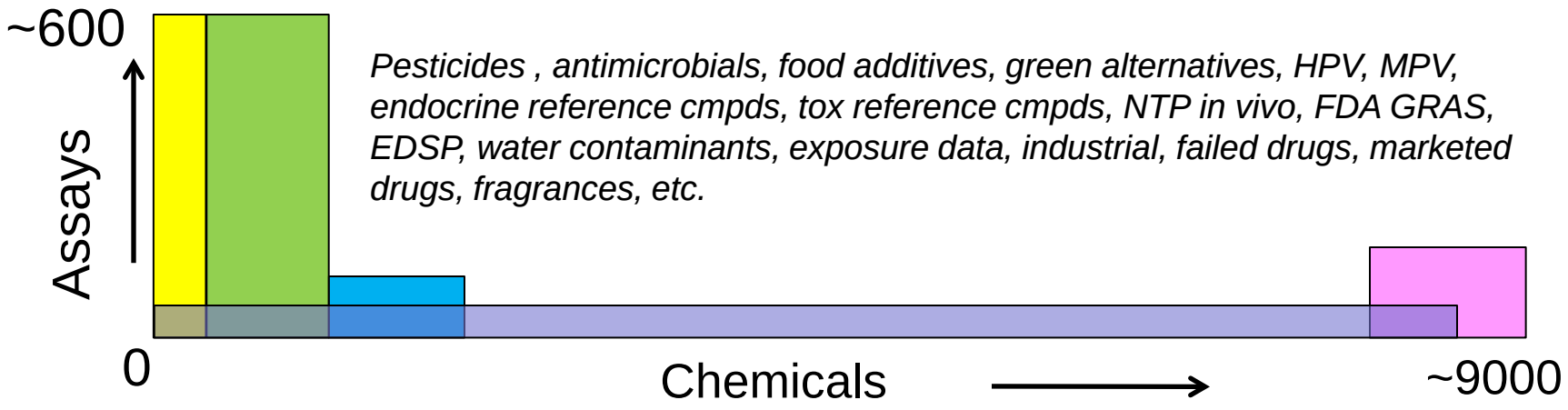
- *Complete on-hand EPA sample library used to build ToxCast inventories*

EPA ToxCast Inventories



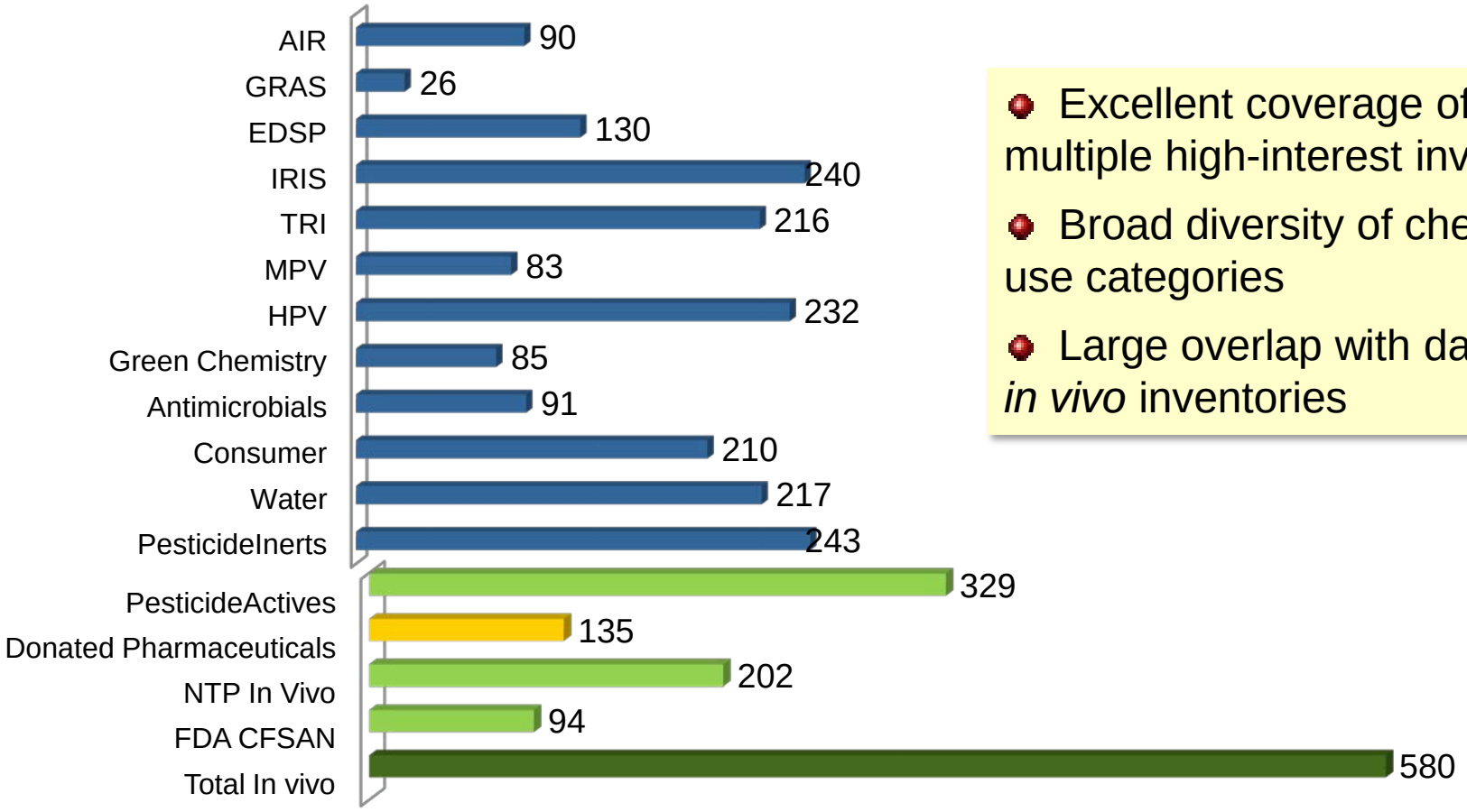
ToxCast & Tox21: Chemicals, Data and Release Timelines

Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	293	~600	~700	2011	Now
ToxCast Phase II	767	~600	~700	03/2013	Now
ToxCast Phase III	~900	~100	~100	Just starting	2014
E1K (endocrine)	800	~50	~120	03/2013	Now
Tox21	8,599	~25	~50	Ongoing	Ongoing



Pesticides , antimicrobials, food additives, green alternatives, HPV, MPV, endocrine reference cmpds, tox reference cmpds, NTP in vivo, FDA GRAS, EDSP, water contaminants, exposure data, industrial, failed drugs, marketed drugs, fragrances, etc.

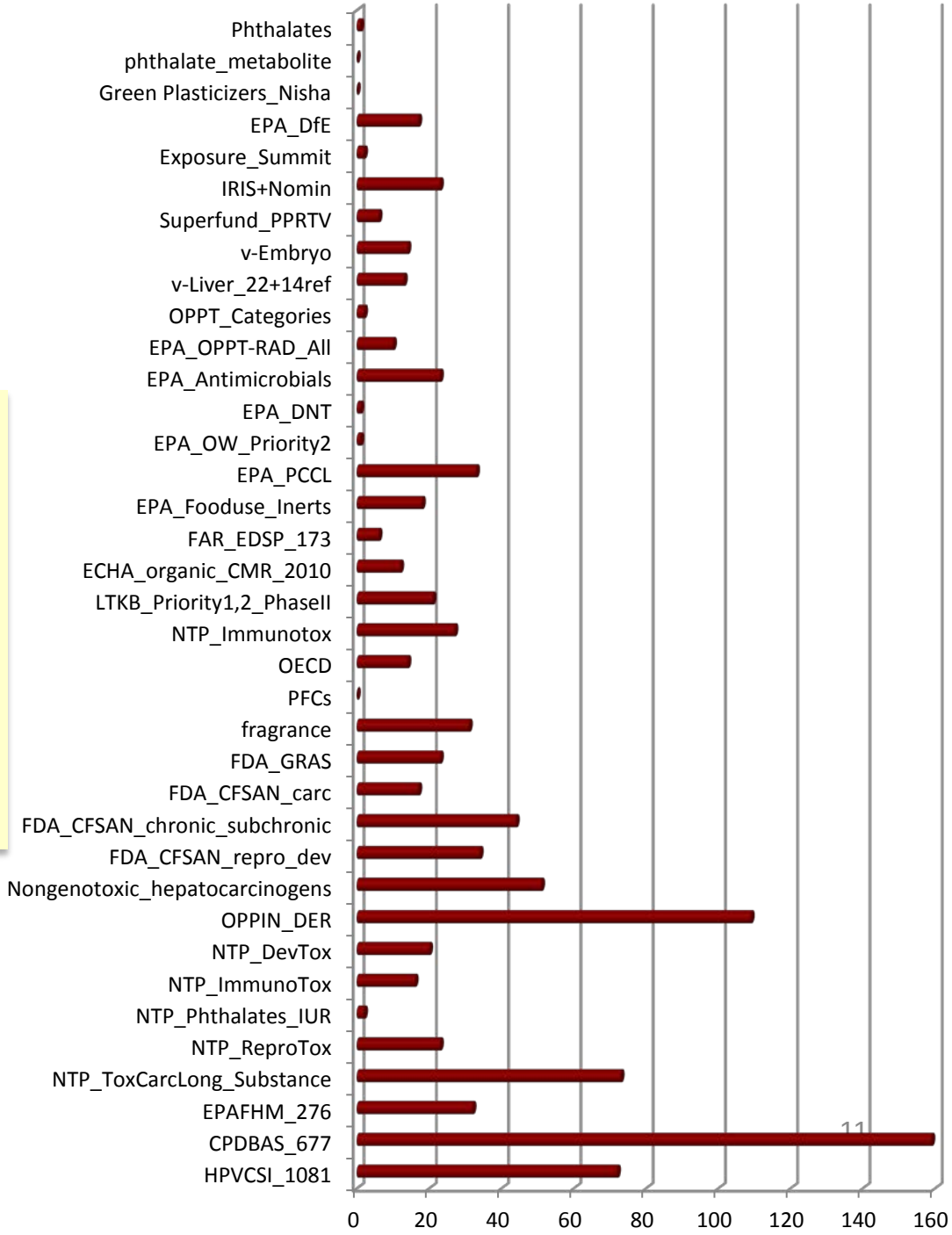
ToxCast PhI&PhII 1060: # Compounds per Inventory

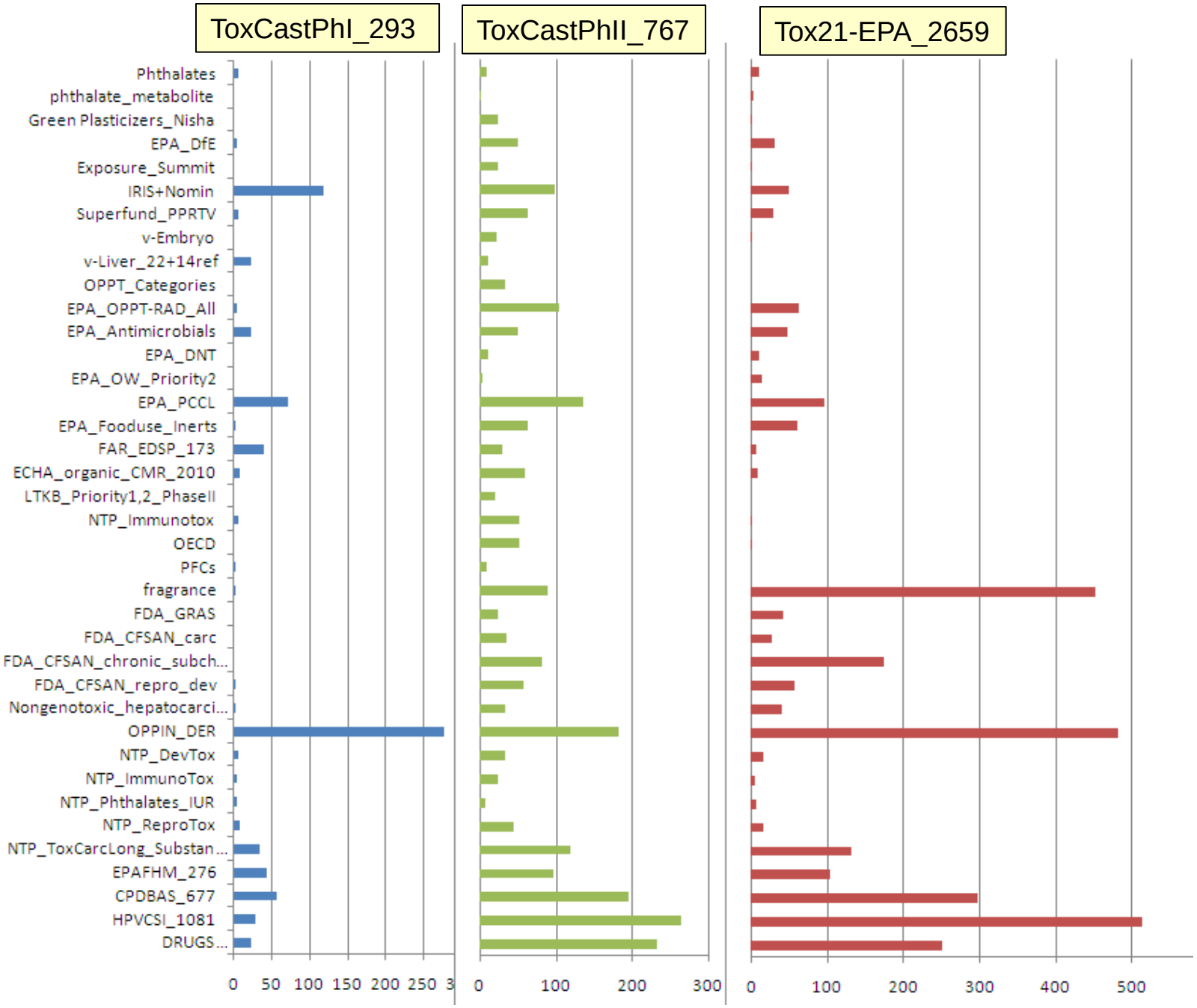


- Excellent coverage of multiple high-interest inventories
- Broad diversity of chemical-use categories
- Large overlap with data-rich *in vivo* inventories

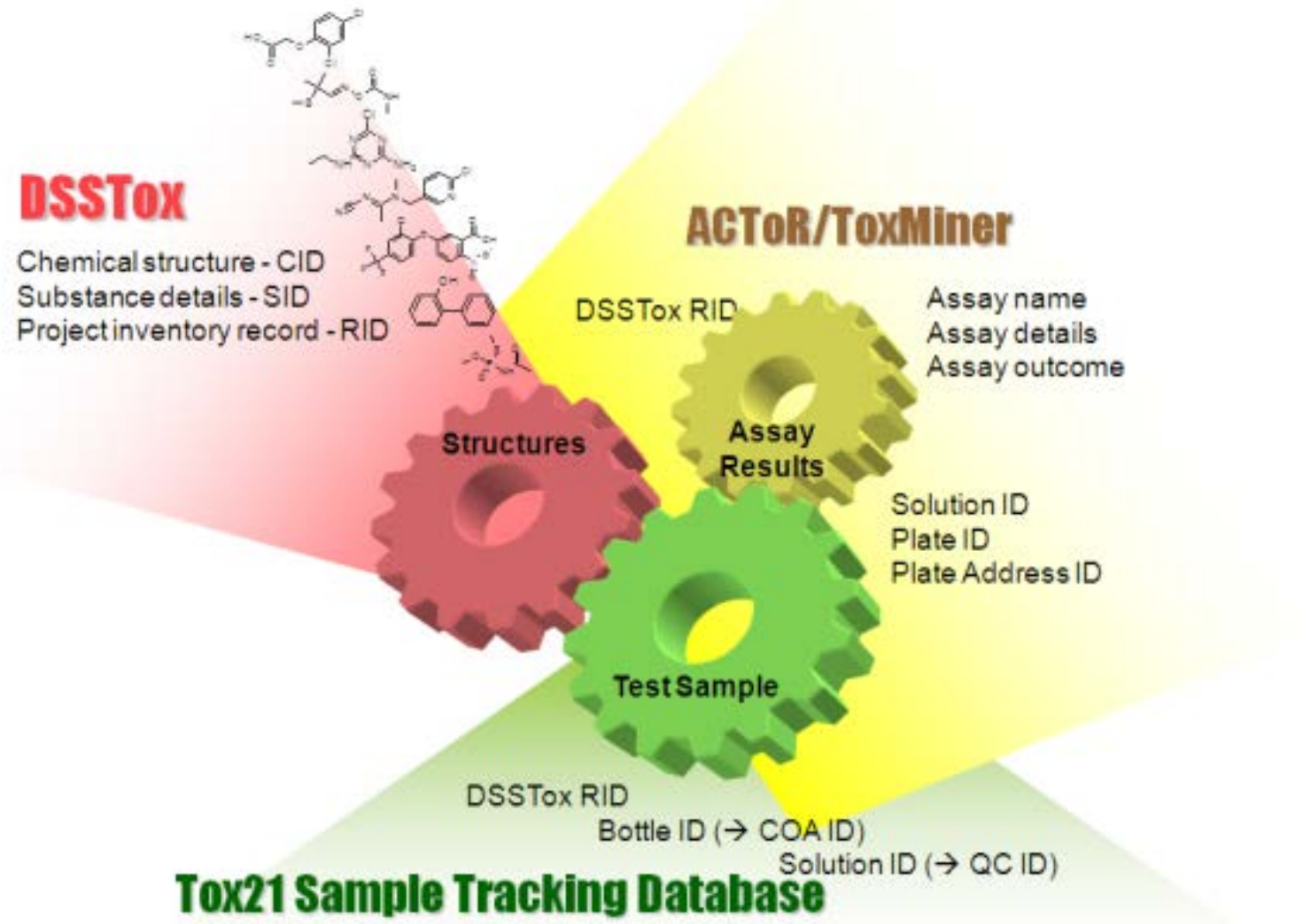
ToxCast Phase II Pharmaceuticals: *Multipurposing*

- 275 PhII cmpds classified as drugs based on presence on FDAMDD, NCGC, NCTR LKB, Donated pharma lists
- 150 drugs appear on additional lists (i.e., multipurposing)
- Caffeine appears on 18 lists

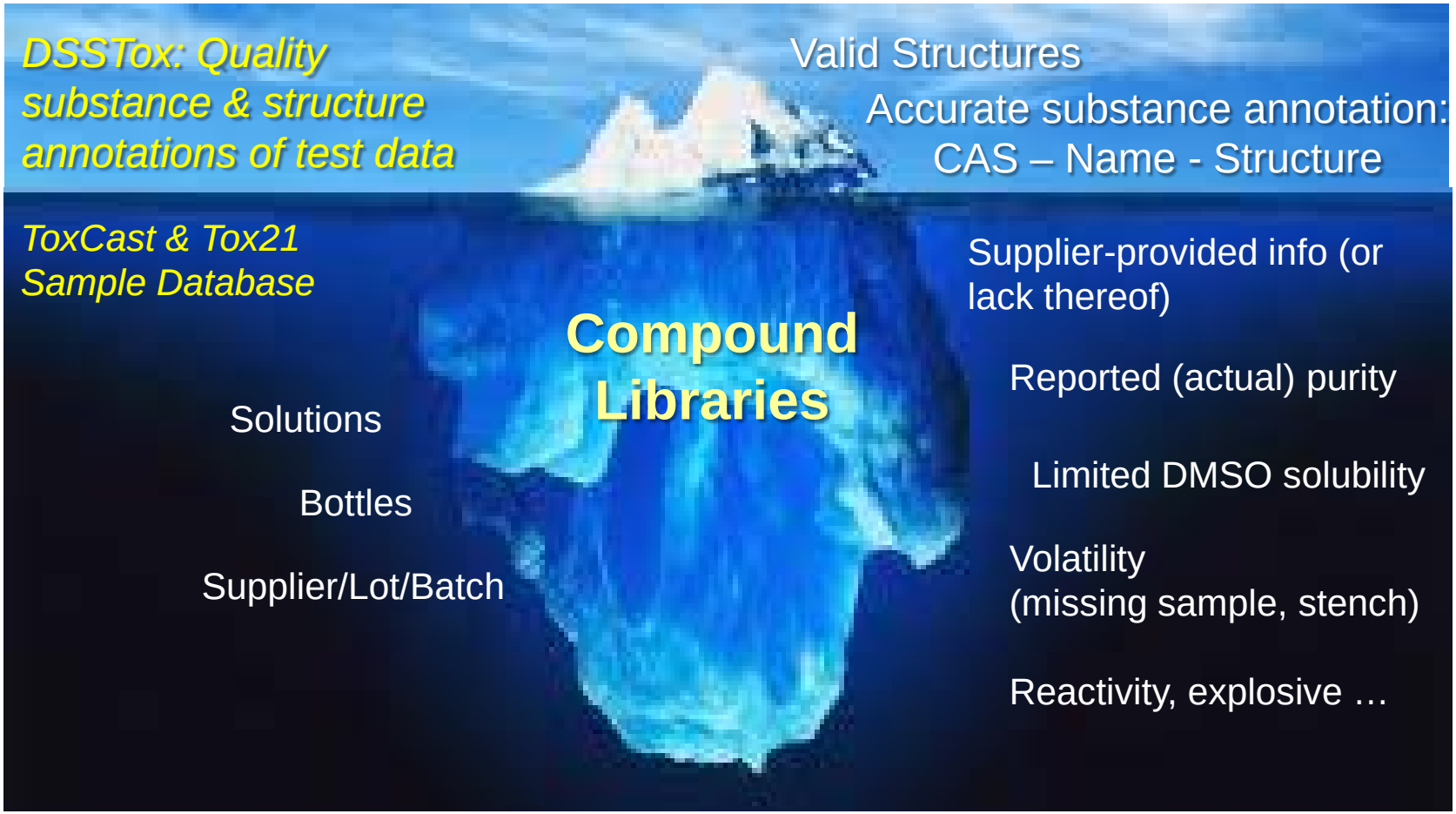



ToxCastPhI_293
ToxCastPhII_767
Tox21-EPA_2659
Phthalates
phthalate_metabolite
Green Plasticizers_Nisha
EPA_DfE
Exposure_Summit
IRIS+Nomin
Superfund_PPRTV
v-Embryo
v-Liver_22+14ref
OPPT_Categories
EPA_OPPT-RAD_All
EPA_Antimicrobials
EPA_DNT
EPA_OW_Priority2
EPA_PCCL
EPA_Fooduse_Inerts
FAR_EDSP_173
ECHA_organic_CMR_2010
LTKB_Priority1,2_PhaseII
NTP_Immunotox
OECD
PFCs
fragrance
FDA_GRAS
FDA_CFSAN_carc
FDA_CFSAN_chronic_subch...
FDA_CFSAN_repro_dev
Nongenotoxic_hepatocarci...
OPPIN_DER
NTP_DevTox
NTP_ImmunoTox
NTP_Phthalates_IUR
NTP_ReproTox
NTP_ToxCarcLong_Substan...
EPAFHM_276
CPDBAS_677
HPVCSI_1081
DRUGS...
0 50 100 150 200 250 3
0 100 200 300
0 100 200 300 400 500

Integrated Chemical-Assay Data Management



Chemical & Data Quality Issues

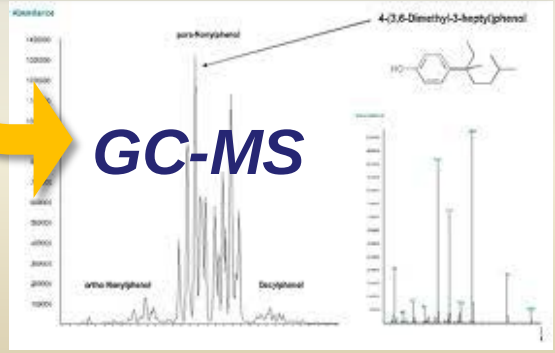


Tox21 Analytical QC

A copy of each parent Tox21 assay plate is subjected to analytical QC for assessing purity, identity, concentration, stability

PASS = Confirm parent ion peak and >90% purity

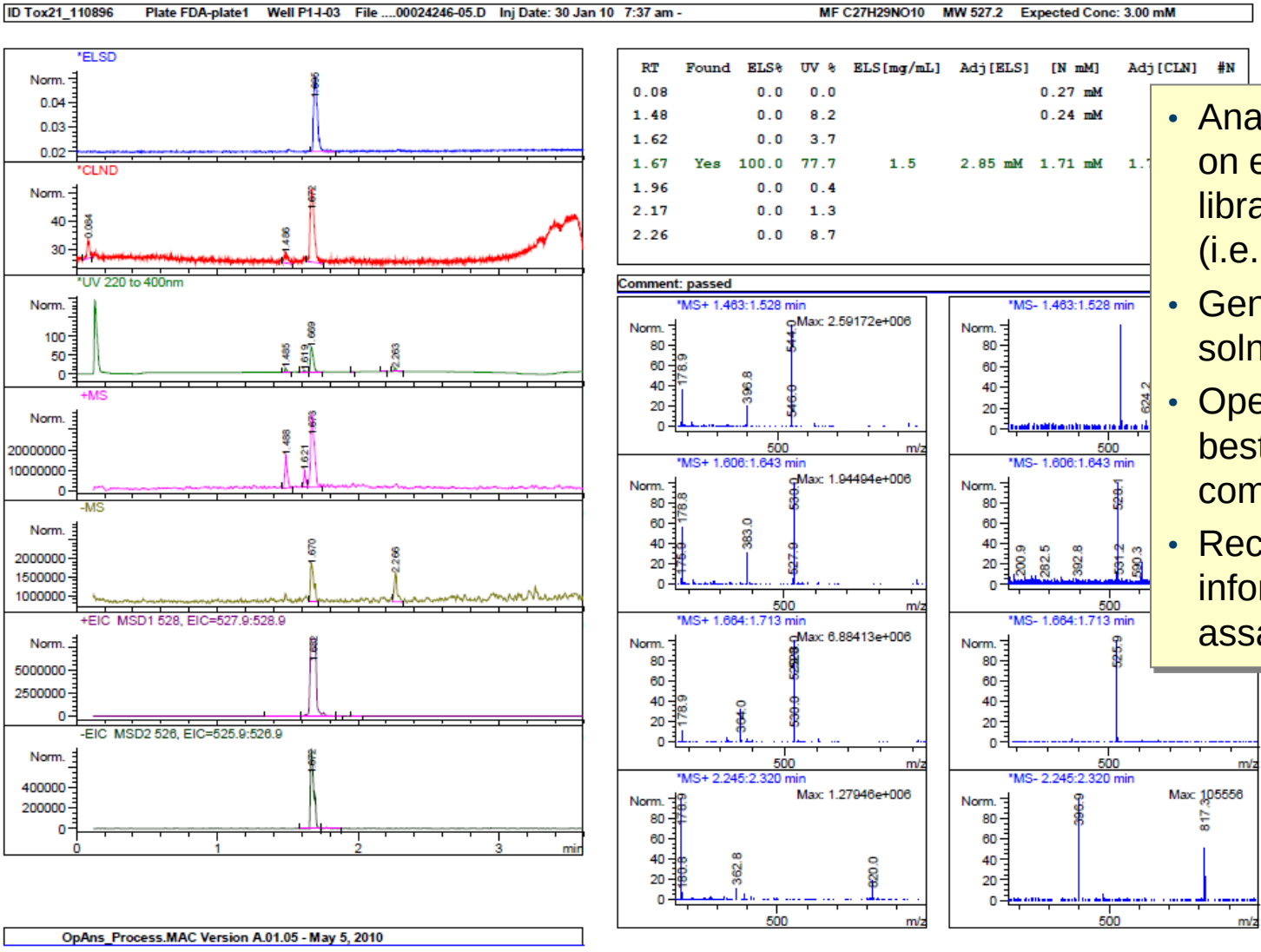
Fail, inconclusive or analytical method inappropriate



Retest at later time point under assay conditions for stability

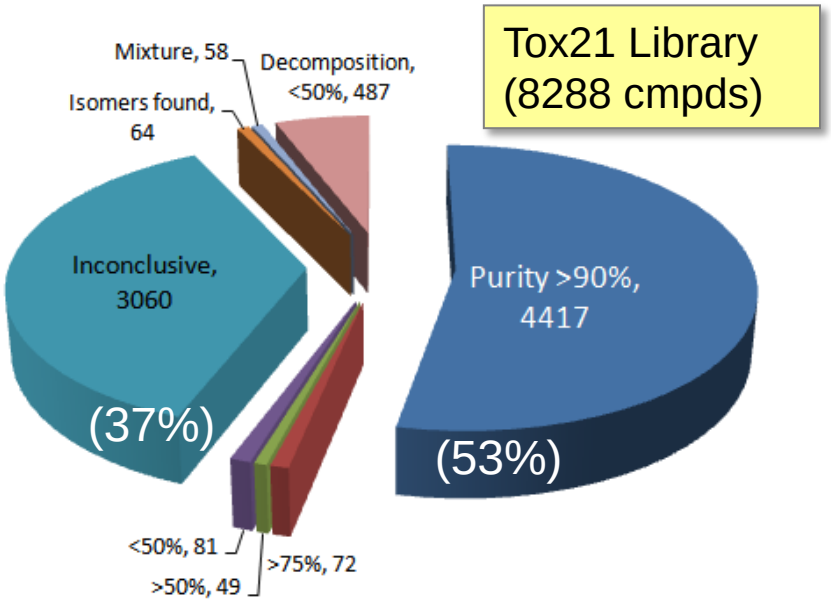
Publish QC summary results in association with assay data

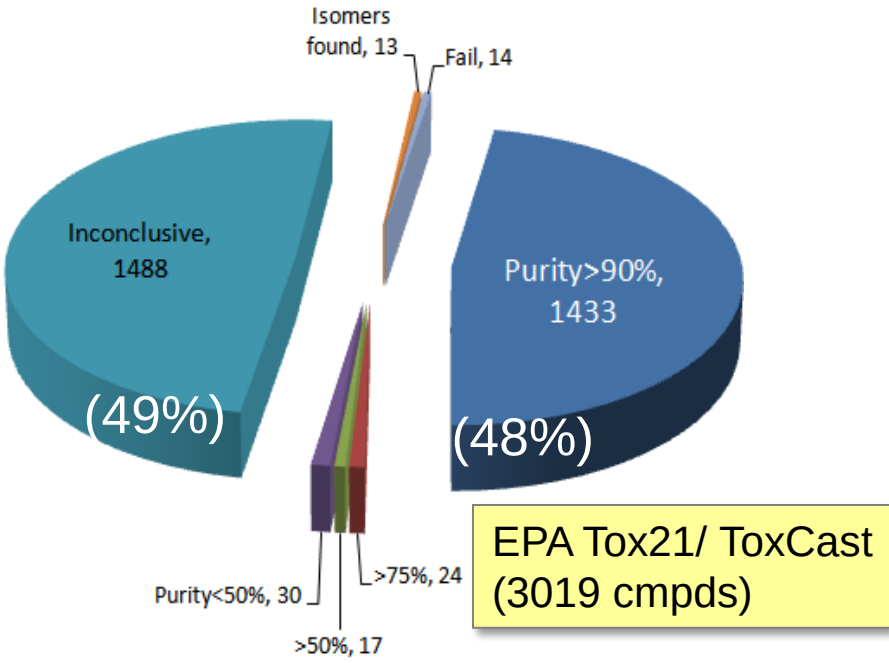
Tox21 Analytical QC



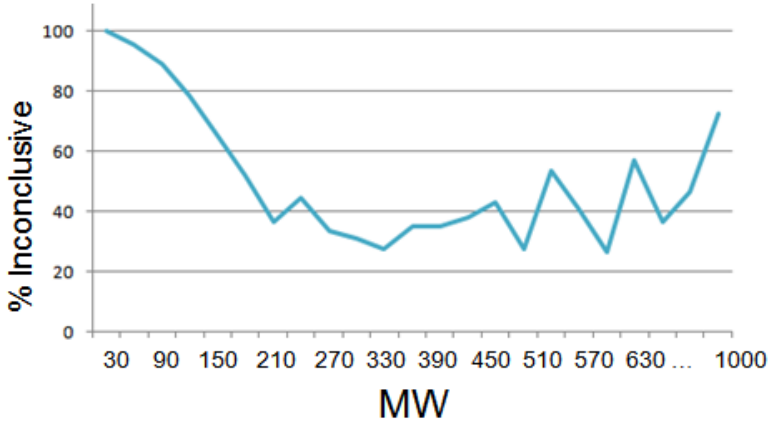
- Analytical QC performed on entire Tox21 chemical library at soln-well-level (i.e., test plate copy x time)
- Generate summary calls at soln and cmpd level
- Open discussion on how best to release and communicate QC results
- Recommend using to inform interpretation of assay results

Tox21/ToxCast Analytical QC

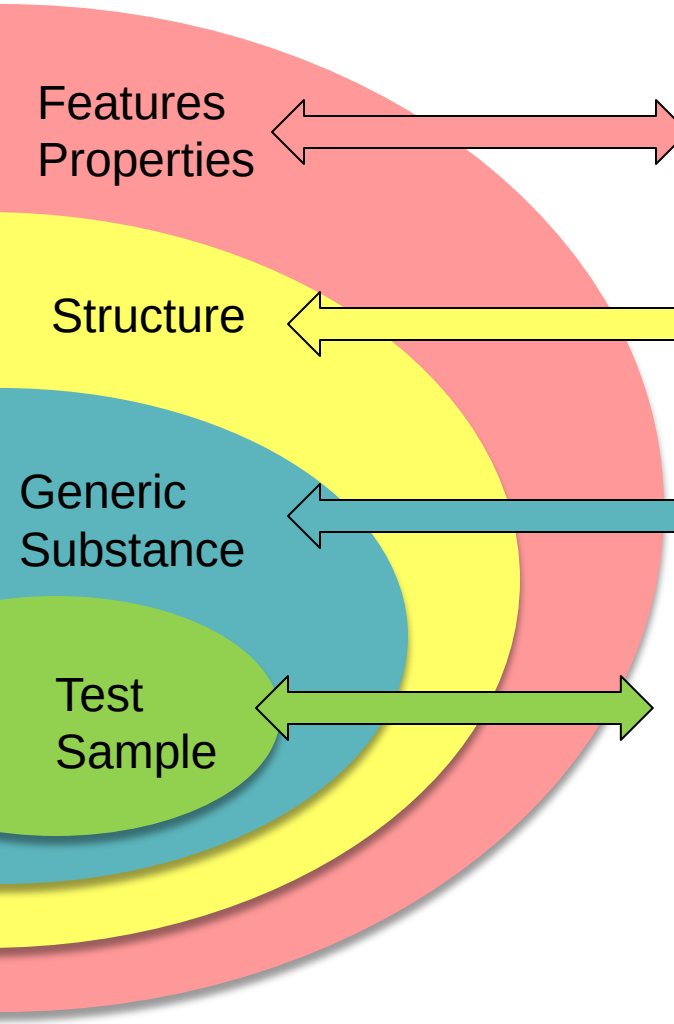




- "inconclusives" require additional follow-up & method development
- EPA's library contains higher proportion of inconclusives (e.g., low MW cmpds, dyes, metals)
- Overall very low "fail" rate

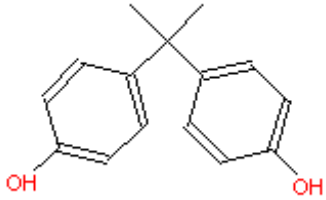


Chemical Elements to Data Integration: *Chemical representations → Uses*



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

Chemical analogs, Read-across, SAR modeling



SMILES
InChI

Structure searching & modeling

Chemical Name
CASRN

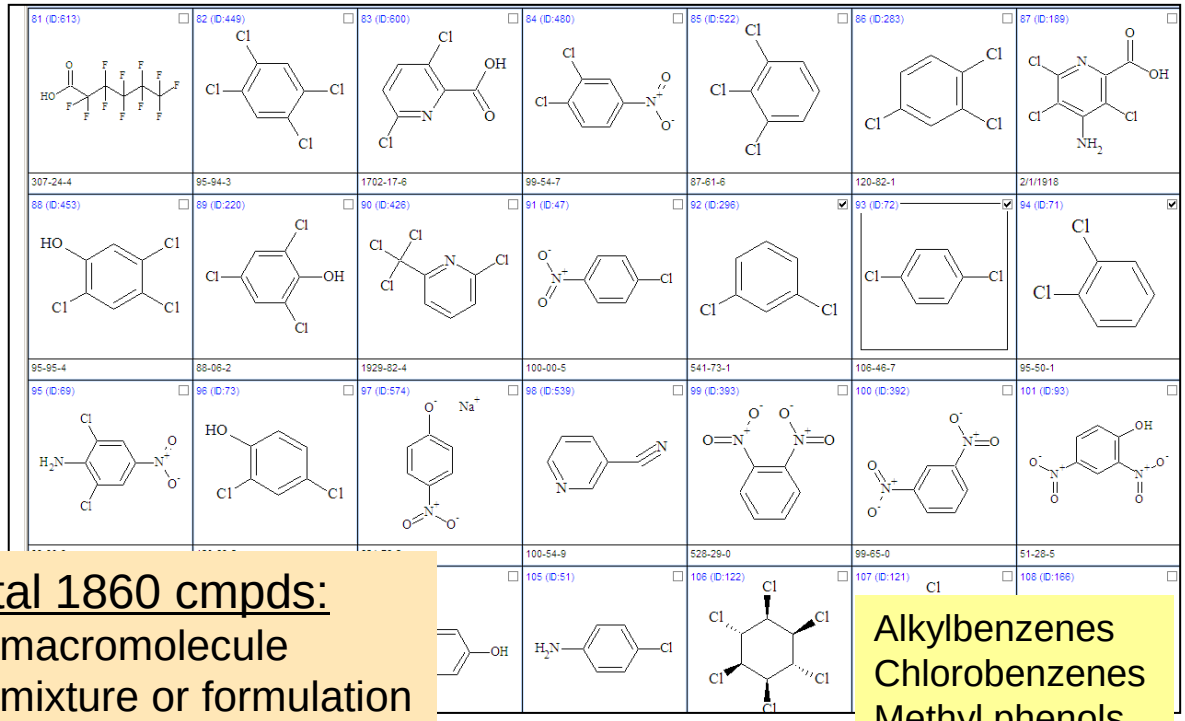
Public toxicity datasets

Supplier, Lot/Batch, physical description

Experimental Endpoint Data

DSSTox TOXCST (PhI,II,E1K) Structure File

DSSTox_RID
DSSTox_GSID
DSSTox_CID
DSSTox_FileID
TestSubstance_ChemicalName
TestSubstance_CASRN
TestSubstance_Description
ChemicalNote
STRUCTURE_Shown
STRUCTURE_Formula
STRUCTURE_MolecularWeight
STRUCTURE_ChemicalType
STRUCTURE_TestedForm_DefinedOrganic
STRUCTURE_ChemicalName_IUPAC
STRUCTURE_SMILES
STRUCTURE_Parent_SMILES
STRUCTURE_InChIS
STRUCTURE_InChIKey
Substance_modify_yyyymmdd
ToxCast_TestingStatus
ToxCast_PhI_v1
ToxCast_PhI_v2
ToxCast_PhII
ToxCast_PhII_DonatedPharma
ToxCast_e1k
ToxCast_Tox21
AnalyticalQC_Problems
Note_TOXCST

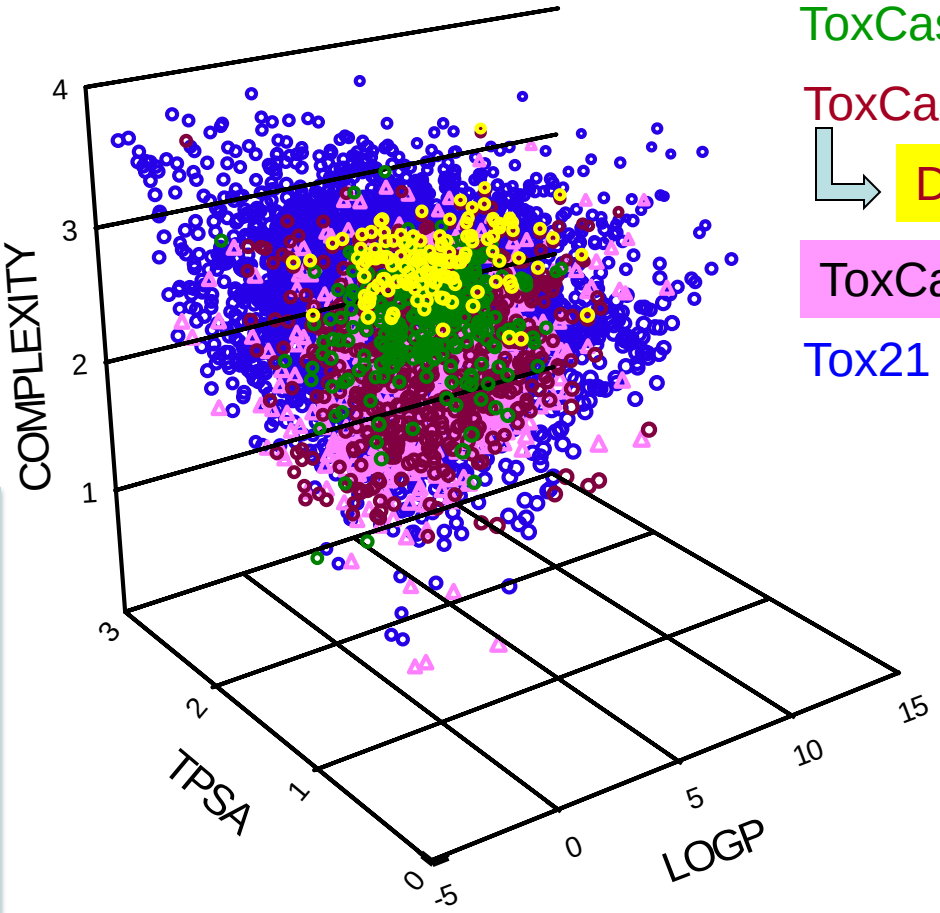


Total 1860 cmpds:
10 macromolecule
96 mixture or formulation
17 inorganics
43 organometallics
84 complexes
136 salts

Alkylbenzenes
Chlorobenzenes
Methyl phenols
Nitrobenzenes
Aromatic amines
Phthalates
Perfluorinates…

MW ranges from 30 (Formalin) to 1700 (Tannic acid)

ToxCast & Tox21 Property Space



ToxCast PhaseI (293)

ToxCast PhaseII (767)

Donated Pharma (135)

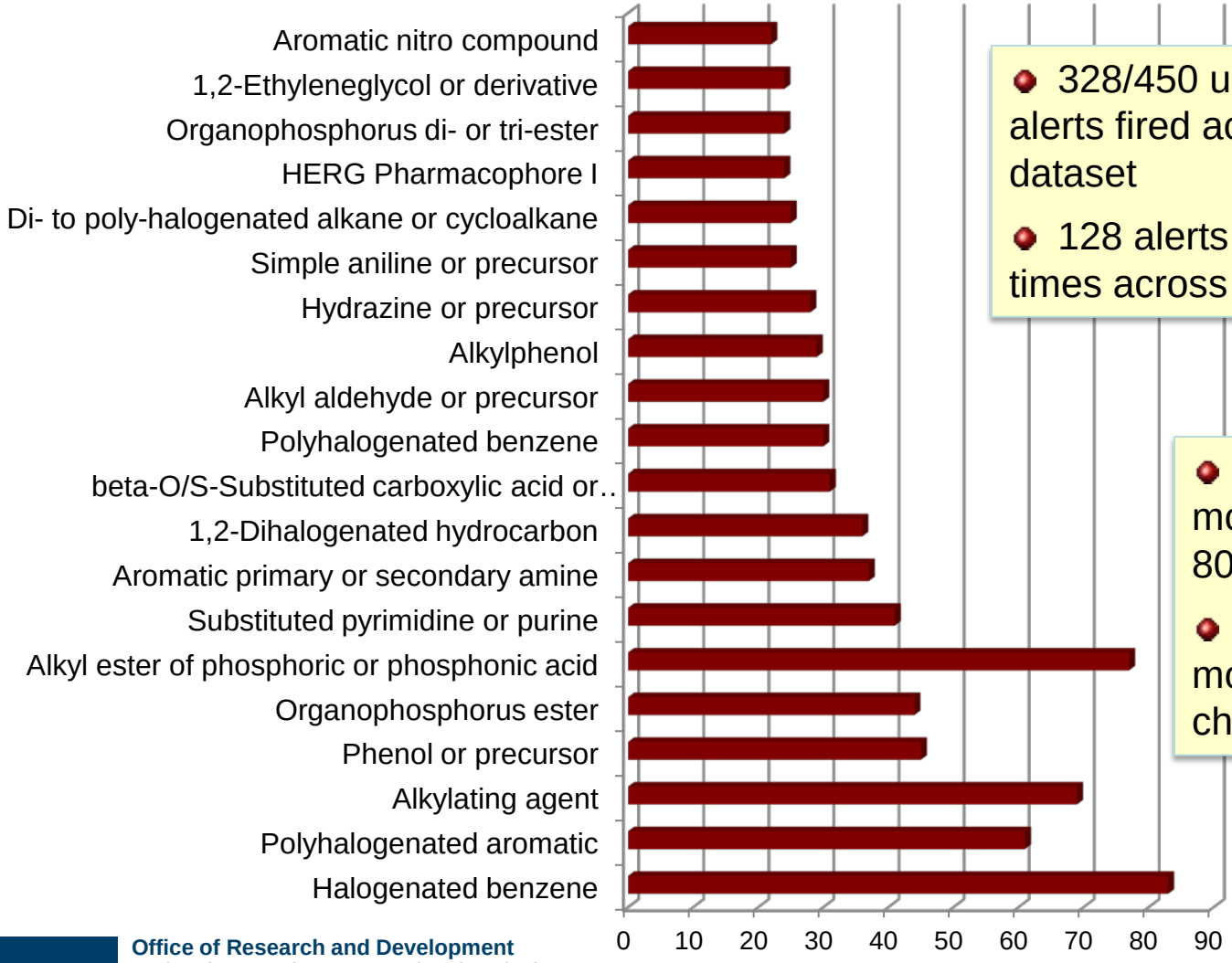
ToxCast e1k (+800)

Tox21 (7324 unique)

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

Chemical properties computed using "Adrianna" software by Molecular Networks (P. Volarath)

Estimating Toxicity Mechanism Coverage:
DEREK (LHASA) Predictions for ToxCast PhII (1060)



- 328/450 unique DEREK alerts fired across entire dataset
- 128 alerts fired 5 or more times across dataset

- DEREK predicts 1 or more toxicity endpoints for 80% of chemicals
- DEREK predicts 3 or more endpoints for 40% chemicals

New Publicly Available Resources: Chemotyper & ToxPrint Chemotypes

Chemotyper: http://www.chemotyper.org

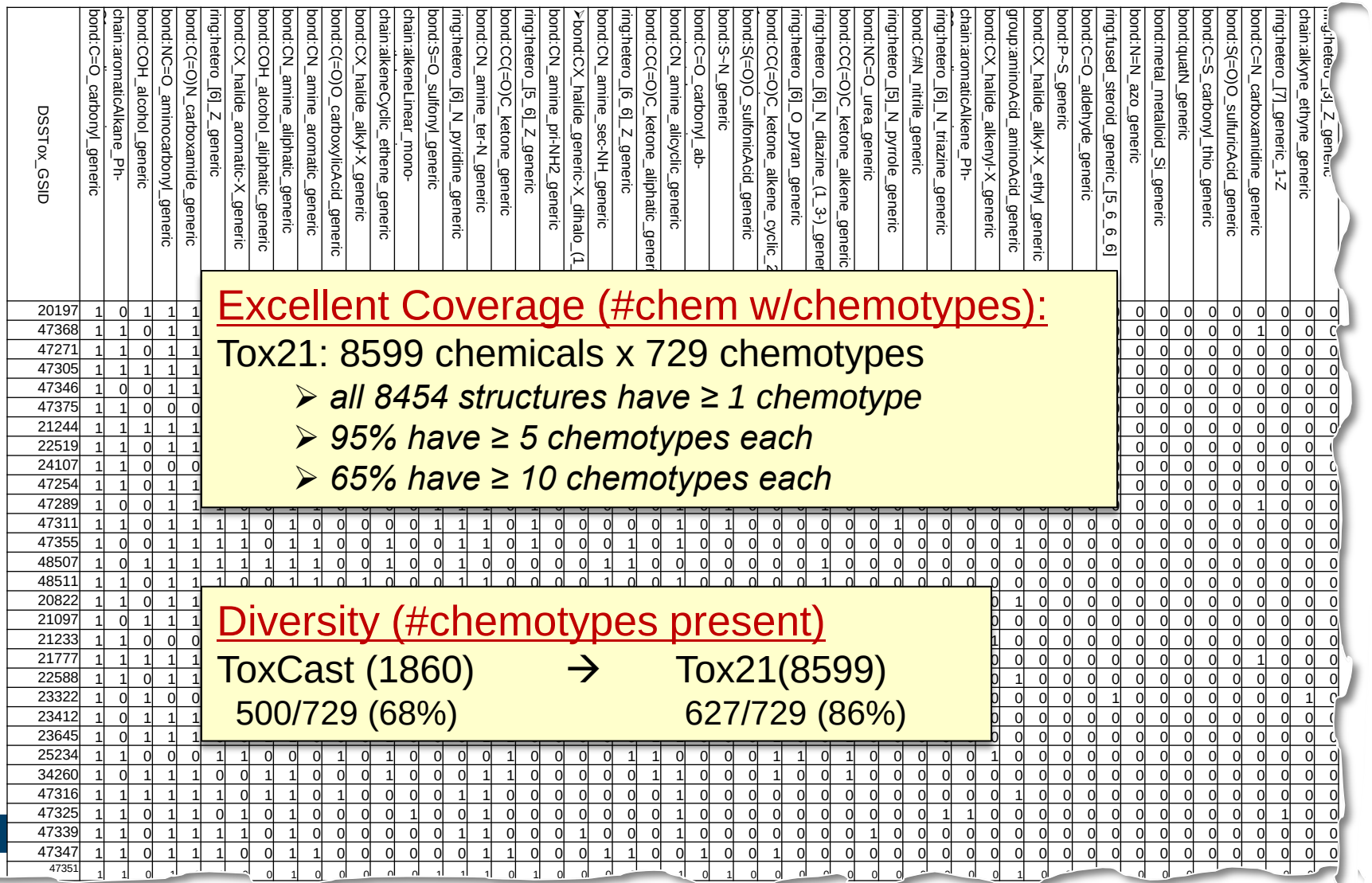
- MS Windows application allows for searching and highlighting of chemical chemotypes (chemical substructures or subgraphs) in structure files.
- Developed by *Molecular Networks GmbH (MN)* under contract from FDA CFSAN to house the "ToxPrint" chemotypes

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 rules (Ashby, J; Tennant, RW, 1988)*
  - *cancer TTC categories (Kroes, R. et. al. 2004)*

ToxCast ToxPrint Chemotype "Fingerprints"

"toxprint_v2_vs_TOX21S_v4a_8599_03Dec2013.csv"



Excellent Coverage (#chem w/chemotypes):

Tox21: 8599 chemicals x 729 chemotypes

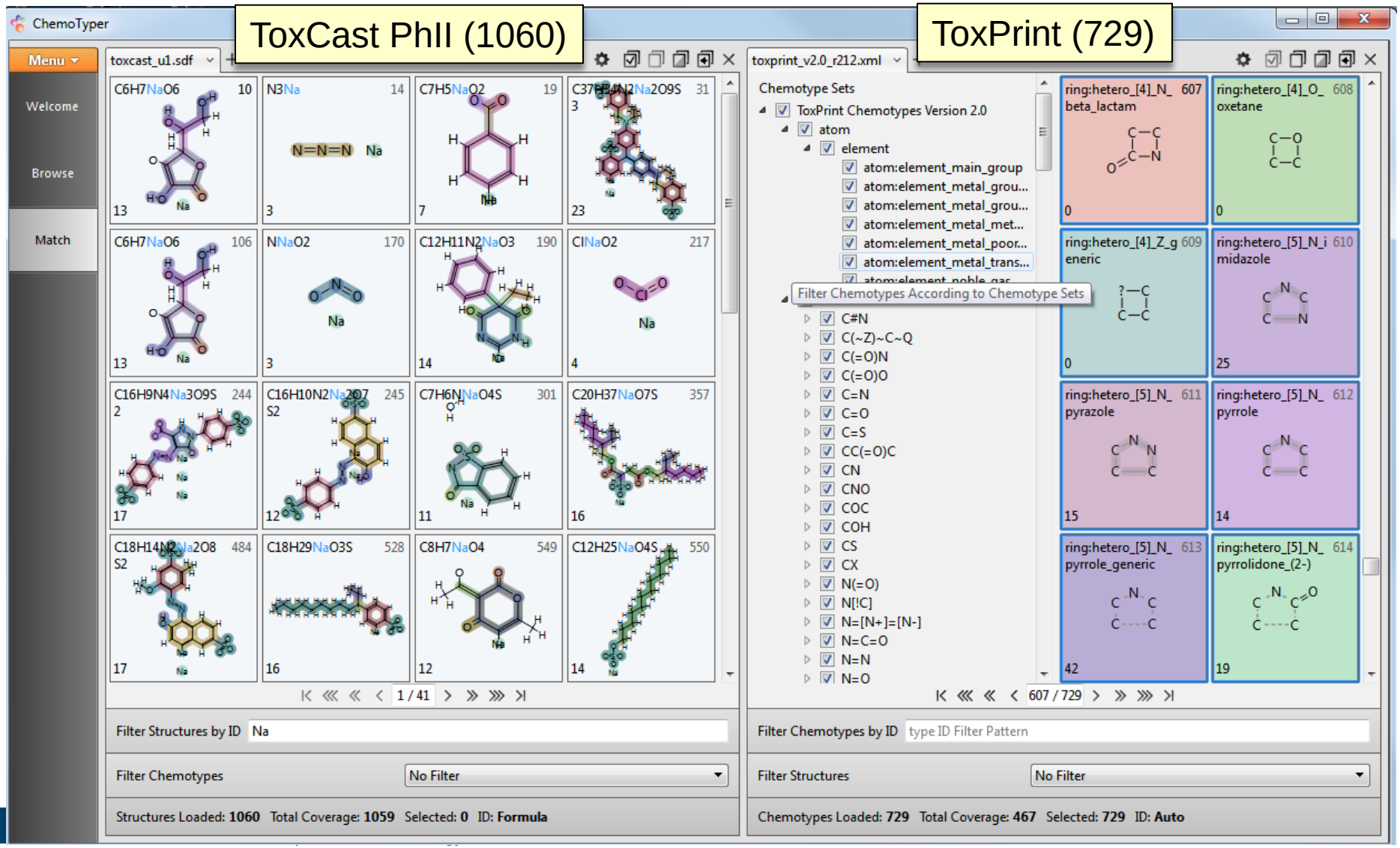
- *all 8454 structures have ≥ 1 chemotype*
- *95% have ≥ 5 chemotypes each*
- *65% have ≥ 10 chemotypes each*

Diversity (#chemotypes present)

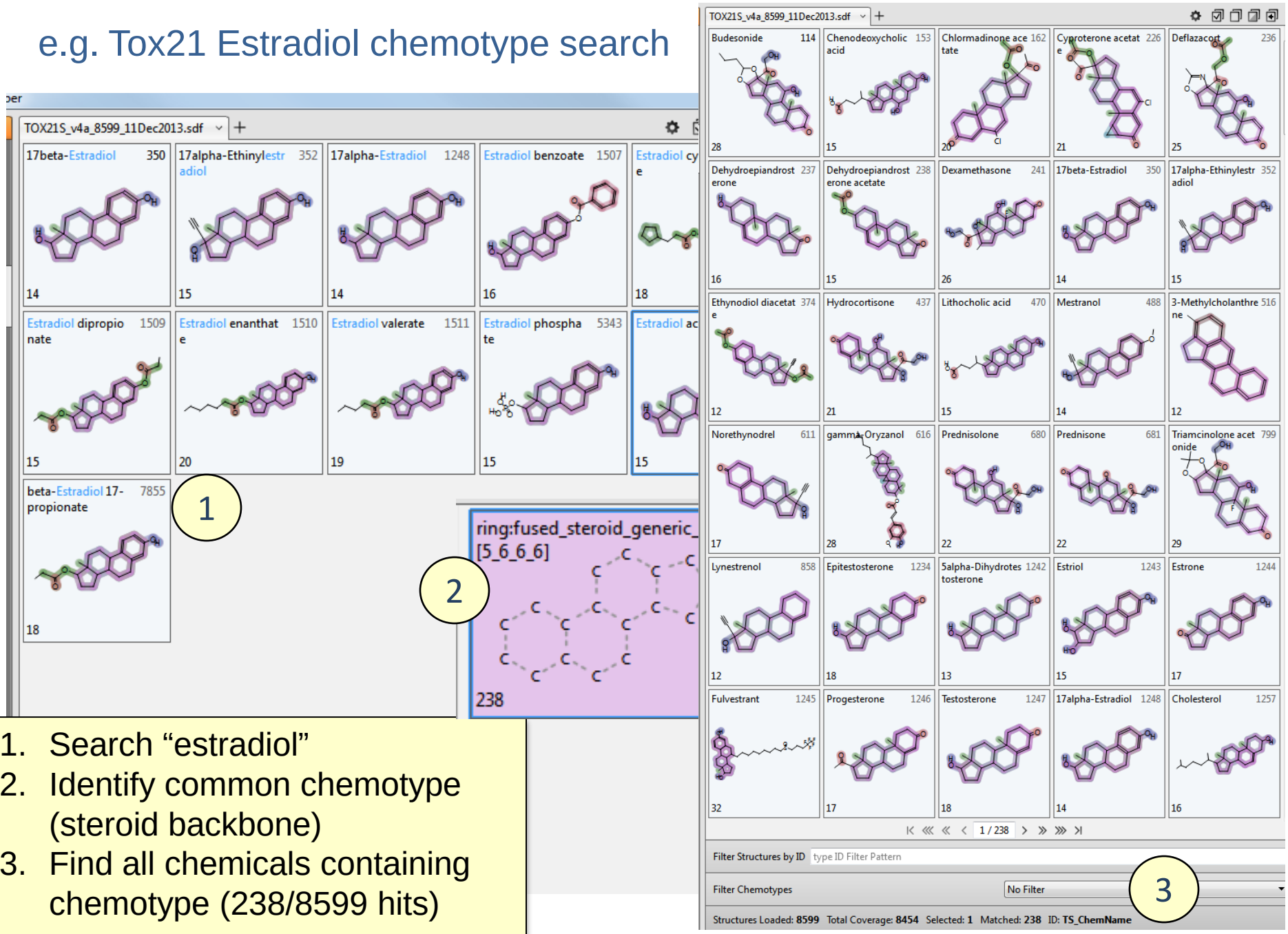
ToxCast (1860) → Tox21(8599)

500/729 (68%) 627/729 (86%)

Chemotyper Application: ToxCast x ToxPrint

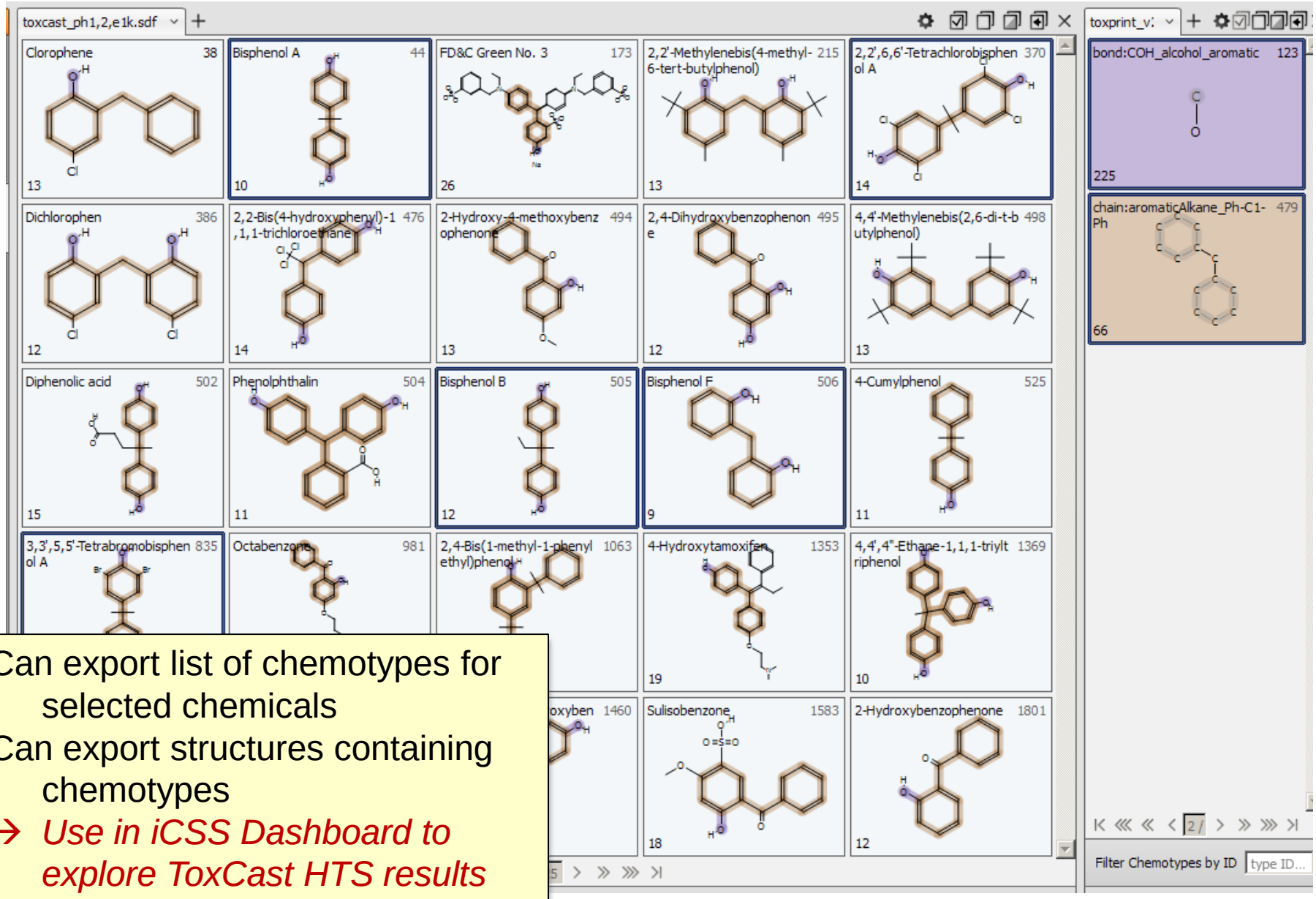


e.g. Tox21 Estradiol chemotype search



1. Search “estradiol”
2. Identify common chemotype (steroid backbone)
3. Find all chemicals containing chemotype (238/8599 hits)

e.g. ToxCast (1860) “Bisphenol A” chemotype search (25 hits)



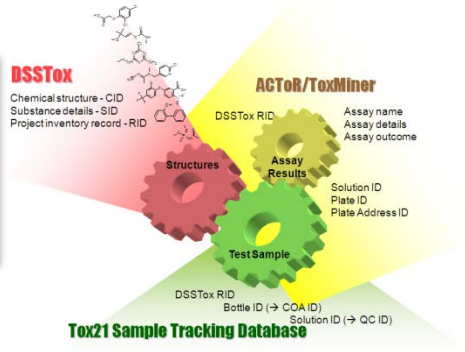
Can export list of chemotypes for selected chemicals

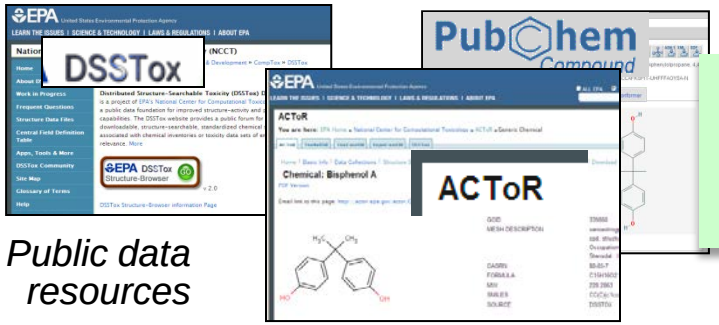
Can export structures containing chemotypes

→ *Use in iCSS Dashboard to explore ToxCast HTS results*

Chemistry: What's needed

- Accurate chemical annotations of testing libraries (e.g., ToxCast & Tox21), transparency, & reporting of error sources
- Cheminformatics foundation to enable structure modeling

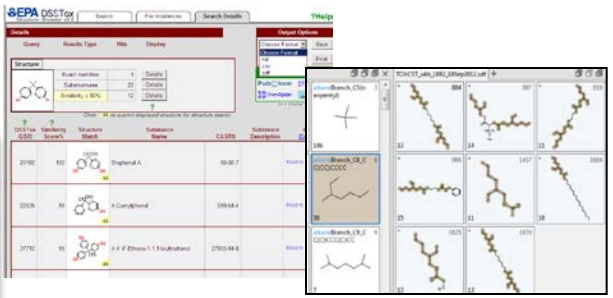




Public data resources

Public data release: ability of chemists & non-chemists (biologists, statisticians) to access & utilize chemical information

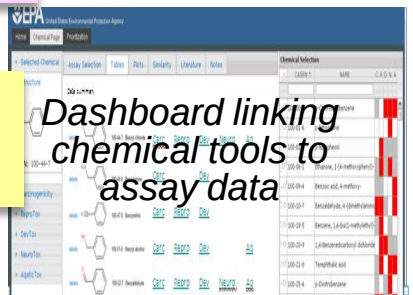
Structure-similarity searching



Informed feature-grouping & highlighting

Use all available data (HTS+chemistry) to form hypotheses, guide & inform analog selection, and improve prediction models

Incorporate chemical information into usable tools for chemical prioritization & safety assessments



Dashboard linking chemical tools to assay data

United States Environmental Protection Agency

Questions?