

Tox21 Federal Collaboration Cross Partner Project # 5: Project Update

Development of a Common Reference Chemical Dataset for Interpretation of High-Throughput Transcriptomics Screening Data



Disclaimer

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Project Background

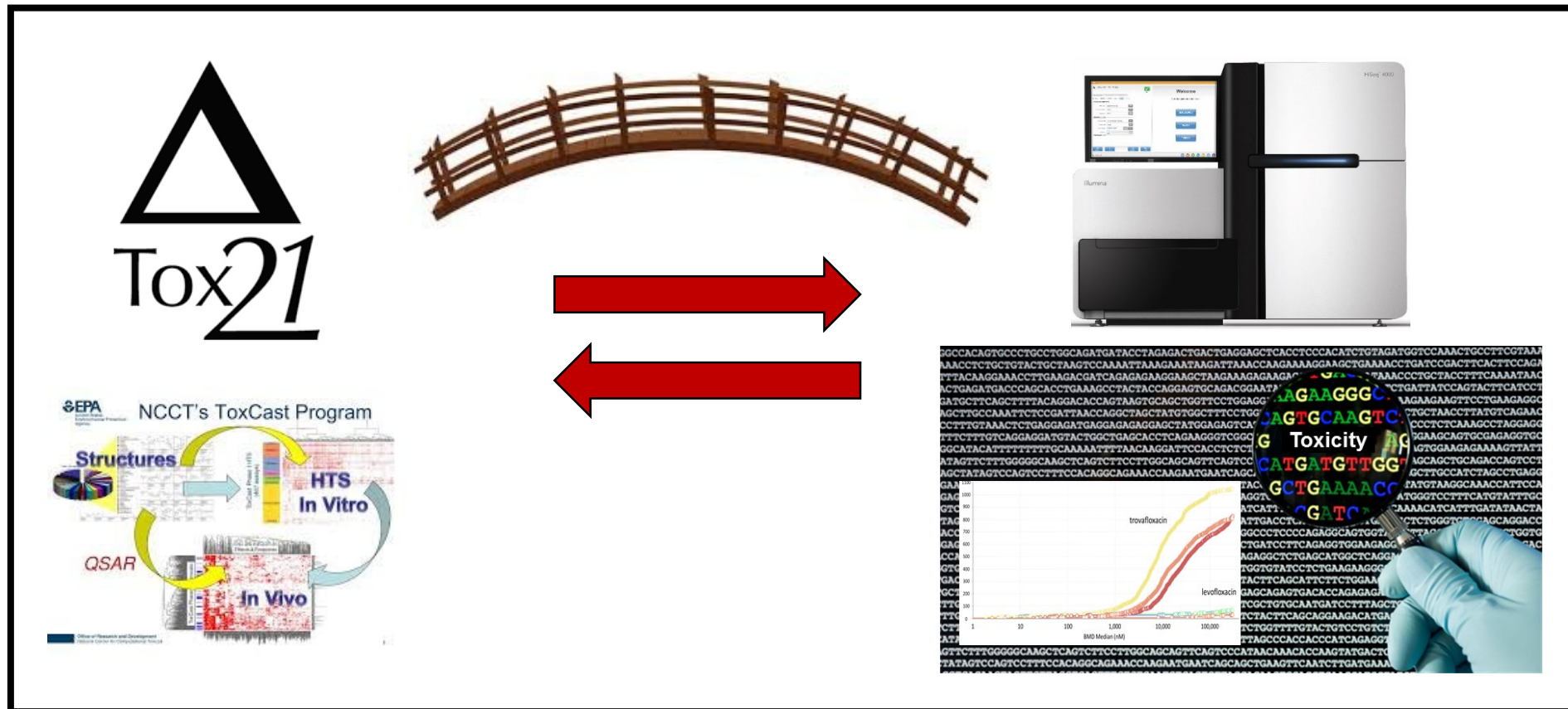
- CPP5 Partner organizations (USEPA, NTP) have both incorporated high-throughput transcriptomics (HTTr) screening into their research portfolios.
- To date, the cell models, gene panel and chemical sets evaluated by the respective organizations do not completely overlap.

Organization	Cell Models	Gene Panel	Chemical Sets
NTP	2-D HepaRG (Prolif & Diff), 3-D HepaRG	TempO-Seq S1500+	Liver Toxicants (mostly)
EPA	MCF-7, U-2 OS, 2-D HepaRG (Diff)	TempO-Seq whole transcriptome	ToxCast Chemicals

- It is anticipated that the diversity of cell models and chemical sets evaluated will increase over time for both organizations.
- There is a need for datasets and tools that can facilitate interpretation of HTTr data & identify putative molecular targets / mechanisms, especially for poorly characterized or unknown chemicals.

Project Hypothesis

Development of a common set of transcriptional profiles from reference chemicals will allow more robust interpretation of high-throughput transcriptomic screens to link chemicals to biological-response pathways and molecular initiating events.



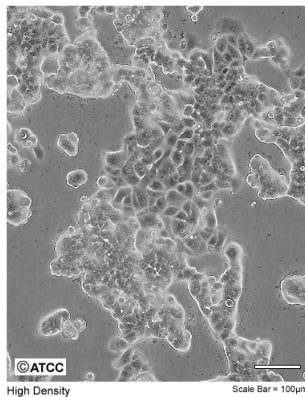
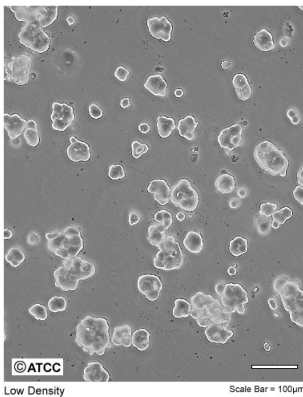
Project Goals

- Identify a diverse set of ~200-500 chemicals from the Tox21 chemical library that overlap with multiple annotated sources of reference chemicals (to biological pathways)
- Build a robust transcriptomics dataset of concentration-response biological perturbations to these reference chemicals using MCF-7 and HepaRG cultures employed by Tox21 partners in high-throughput screens
- Develop transcriptomic signatures that accurately identify specific molecular targets/biological pathways perturbed by the reference chemicals, and their respective 'firing orders' across the range of concentrations examined.

Step 1: Fit-for-purpose approach compatible with Tox21 partner capabilities

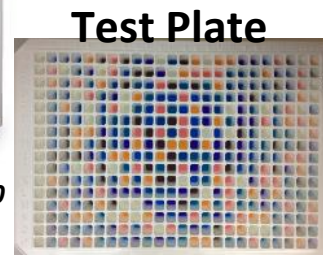
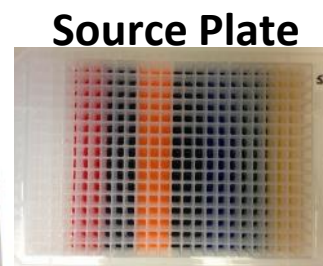
MCF-7

ATCC Number: **HTB-22**
Designation: **MCF-7**

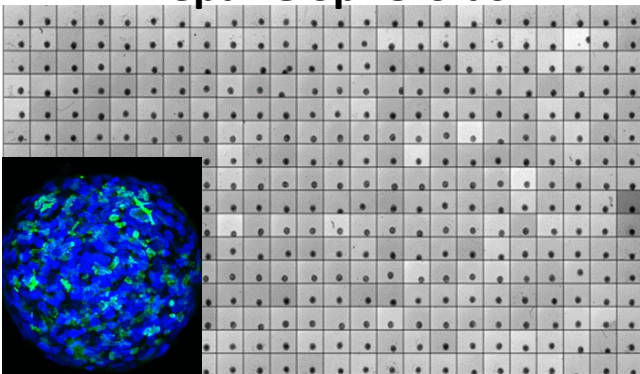


24h exposures

		QC Samples		Source Plate Map																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
NCCT QC Samples	1 A	UHRR	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	DMSO		
	2 B	UHRR	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	DMSO		
	3 C	HBRR	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	DMSO		
	4 D	HBRR	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	DMSO		
	5 E	Bulk Lysate_DMSO	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Ref Treatment 1		
	6 F	Bulk Lysate_DMSO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Ref Treatment 1	
	7 G	Bulk Lysate_TSA	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	Ref Treatment 1		
	8 H	Bulk Lysate_TSA	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	Ref Treatment 1		
	9 I	NTC	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Ref Treatment 2	
	10 J	NTC	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Ref Treatment 2		
BioSpyder QC Samples	11 K	UHRR	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Ref Treatment 2		
	12 L	UHRR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Ref Treatment 2		
	13 M	HBRR	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	Ref Treatment 3		
	14 N	HBRR	3	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	Ref Treatment 3		
	15 O	NTC	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	6	4	2	Ref Treatment 3		
	16 P	NTC	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	1	5	3	1	Ref Treatment 3		



HepaRG Spheroids



		24h exposures																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC
B	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	SEQC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC	QC
C	RNA C1	RNA C1	1A	13A	25A	1B	13B	25B	1C	13C	25C	1D	13D	25D	DMSO	DMSO	1E	13E	25E	1F	13F	25F	RNA C1	RNA C1	RNA C1
D	RNA C2	RNA C2	2A	14A	26A	2B	14B	26B	2C	14C	26C	2D	14D	26D	Media	Media	2E	14E	26E	2F	14F	26F	RNA C2	RNA C2	RNA C2
E	4 Lysate C1	Lysate C1	3A	15A	27A	3B	15B	27B	3C	15C	27C	3D	15D	27D	DMSO	DMSO	3E	15E	27E	3F	15F	27F	Lysis buffer	Lysis buffer	Lysis buffer
F	4 Lysate C1	Lysate C1	4A	16A	28A	4B	16B	28B	4C	16C	28C	4D	16D	28D	Media	Media	4E	16E	28E	4F	16F	28F	Lysis buffer	Lysis buffer	Lysis buffer
G	4 Lysate C1	Lysate C2	5A	17A	29A	5B	17B	29B	5C	17C	29C	5D	17D	29D	DMSO	DMSO	5E	17E	29E	5F	17F	29F	Lysis buffer	Lysis buffer	Lysis buffer
H	4 Lysate C1	Lysate C2	6A	18A	30A	6B	18B	30B	6C	18C	30C	6D	18D	30D	Media	Media	6E	18E	30E	6F	18F	30F	Lysis buffer	Lysis buffer	Lysis buffer
I	4 Lysate C1	Lysate C3	7A	19A	31A	7B	19B	31B	7C	19C	31C	7D	19D	31D	DMSO	DMSO	7E	19E	31E	7F	19F	31F	Lysis buffer	Lysis buffer	Lysis buffer
J	4 Lysate C1	Lysate C3	8A	20A	32A	8B	20B	32B	8C	20C	32C	8D	20D	32D	Media	Media	8E	20E	32E	8F	20F	32F	Lysis buffer	Lysis buffer	Lysis buffer
K	4 Lysate C1	Lysate C4	9A	21A	33A	9B	21B	33B	9C	21C	33C	9D	21D	33D	DMSO	DMSO	9E	21E	33E	9F	21F	33F	Lysis buffer	Lysis buffer	Lysis buffer
L	4 Lysate C1	Lysate C4	10A	22A	34A	10B	22B	34B	10C	22C	34C	10D	22D	34D	Media	Media	10E	22E	34E	10F	22F	34F	Lysis buffer	Lysis buffer	Lysis buffer
M	4 Lysate C1	Lysate C5	11A	23A	35A	11B	23B	35B	11C	23C	35C	11D	23D	35D	DMSO	DMSO	11E	23E	35E	11F	23F	35F	Lysis buffer	Lysis buffer	Lysis buffer
N	4 Lysate C1	Lysate C5	12A	24A	36A	12B	24B	36B	12C	24C	36C	12D	24D	36D	Media	Media	12E	24E	36E	12F	24F	36F	Lysis buffer	Lysis buffer	Lysis buffer
O	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells
P	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells	outer wells
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24



- 4 independent experimental runs, 3 averaged for HTT
- 1 technical replicate per treatment group

- 6 chemical concentrations (half-log spacing)
- concentrations span reference target potencies from reporter assays (e.g., IC₅₀, EC₅₀)

Assays summary

- **Cell viability/Cytotoxicity**

- Cell viability will be evaluated based on each lab's respective established methods
 - NTP: Liver enzyme leakage assays (LDH-Glo) in spent culture media with photomicrographs for each well for 3D HepaRG spheroids
 - EPA: Employed a high content imaging (HCI)-based cell painting approach that assesses apoptosis and cell viability assay for use with MCF7 cells grown in monolayer culture.

- **High throughput transcriptomics**

- Cell cultures will be lysed with BioSpyder lysis buffer for 15m, frozen at -80°C, and shipped to BioSpyder for TempO-Seq analysis using the Human Whole Transcriptome Gene Set to be analyzed at a minimum average read depth of ~300-500 reads/transcript.
- Freezing leftover culture media for future contextual follow-up (e.g., chemical exposures, metabolomics) with 3D HepaRG spheroids

Identifying a diverse set of reference chemicals targeting specific biological targets (1)

•Goal:

- Develop a list of ~300 chemicals covering ~75-100 biological-response pathways (i.e., 3 or more for a given pathways) with well-annotated associations to specific molecular targets or biological-response pathways and 10-20 “negatives”

Identifying a diverse set of reference chemicals for specific biological targets (2)

RefChemDB

A database of ***chemical_target_mode_activity*** associations created by Judson et al. from information contained in the public domain.

- > 2900 biological targets
- > 37,000 unique chemicals

Intended to be used in a semi-automated workflow for development ***candidate reference chemical lists*** for a molecular target

Candidate lists are then refined using expert knowledge.

For a given chemical, more than one literature source may support a given chemical_target_mode association (i.e. ***level of support***).

Research Article

ALTEX 36(2), 261-276. doi:10.14573/altex.1809281

Workflow for Defining Reference Chemicals for Assessing Performance of *In Vitro* Assays

Richard S. Judson¹, Russell S. Thomas¹, Nancy Baker², Anita Simha³, Xia Meng Howey³, Carmen Marable³, Nicole C. Kleinstreuer⁴ and Keith A. Houck¹

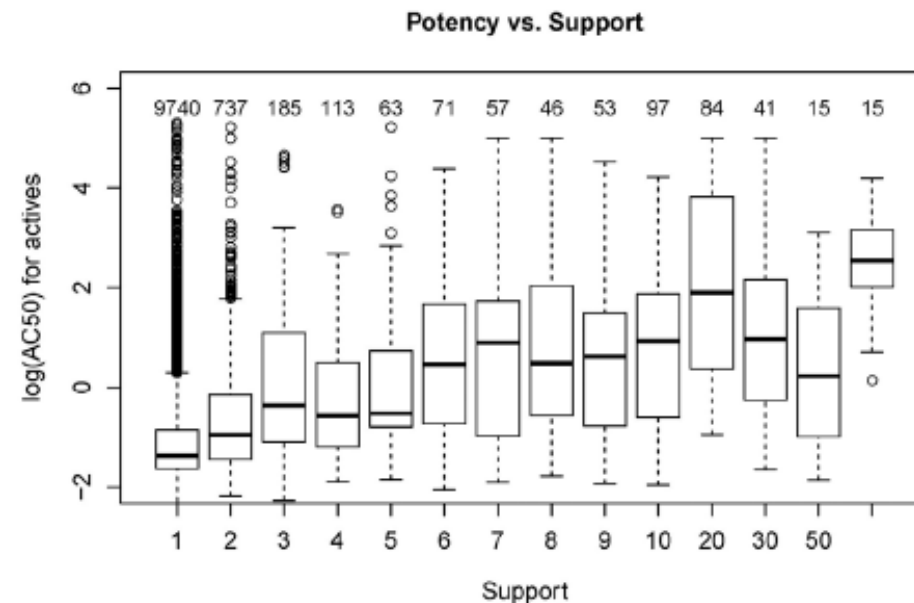
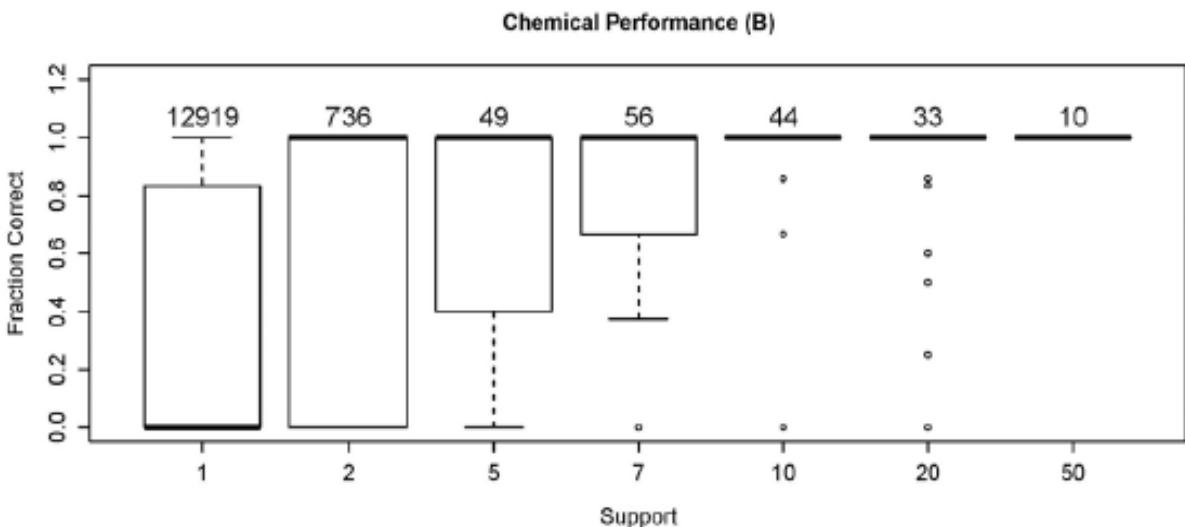
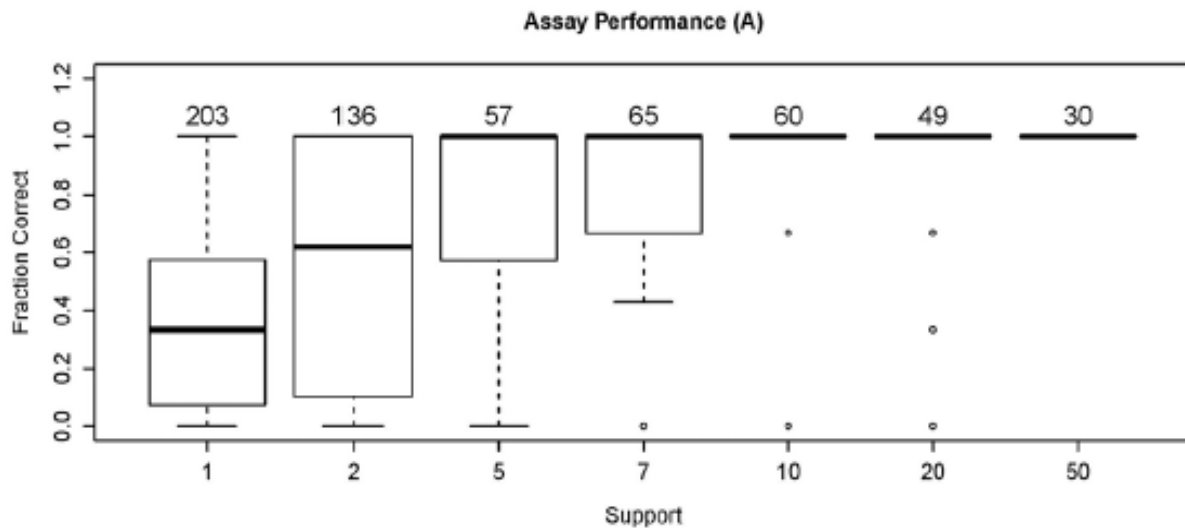
¹US EPA, National Center for Computational Toxicology, Research Triangle Park, NC, USA; ²Leidos, Inc., Research Triangle Park, NC, USA;

³ORAU, contractor to U.S. Environmental Protection Agency through the National Student Services Contract, Research Triangle Park, NC, USA;

⁴National Toxicology Program, Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC, USA

Source	Chemicals	Targets	Chemical-target-mode-activity combinations	Mean multiplicity	PMIDs
ChEMBL	28,832	2,238	310,984	1.16	11,520
ChEMBL Drug	1,187	738	4,099	1	0
CTD	2,317	7,904	25,606	1.22	5,280
DrugBank	1,630	1,169	3,623	3.41	6,274
Eurofins Biochemical	206	570	925	1	0
Eurofins Functional	211	239	706	1	0
Iuphar BPS	1,860	941	5,081	1	0
KEGG Drug	661	263	1,201	1	0
KIDB	535	450	6,532	1	0
KInaseDB	133	168	676	1	1
LitDB	2,654	88	8,348	4.94	27,909
Open Targets	1,031	820	3,973	1	0
Prodrug	41	33	41	1	1
Repurposing Hub	2,279	2,172	10,209	1	0
ToxCast	9,136	343	852,470	1.03	0
TTD	3,916	1,575	11,557	1.00	0
Web Curation	3,940	1,059	5,617	1.01	0
Total	37,301	11,055	123,4580	1.02	49,883

Identifying a diverse set of reference chemicals for specific biological targets (3)



Higher level-of-support is associated with greater reliability of chemical_target_model predictions using the ToxCast/Tox21 assay results as a benchmark.

Chemical_target_mode predictions with level of support ≤ 5 are suspect.

Higher level-of-support is also associated with greater potency at the predicted molecular target.

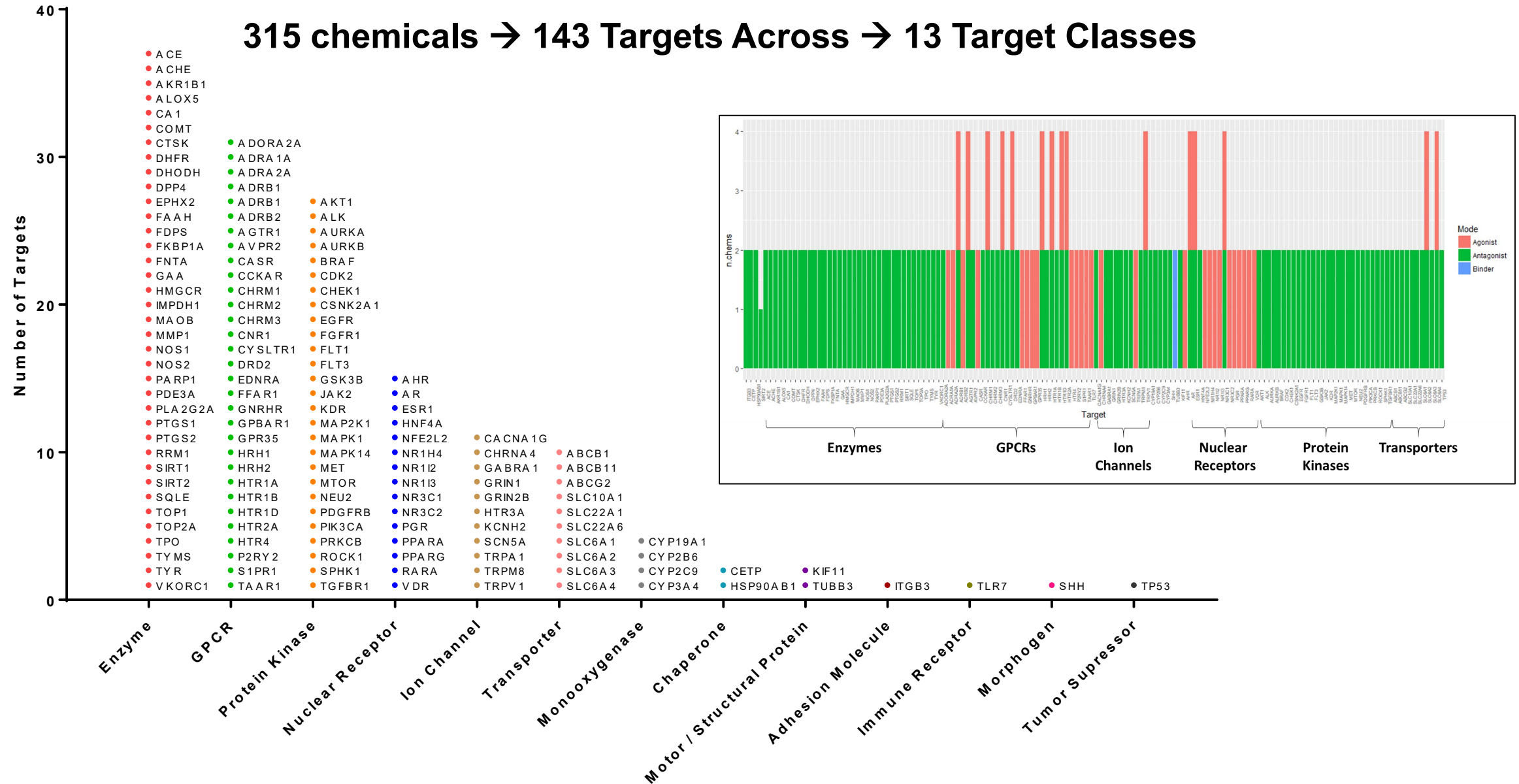
Identifying a diverse set of reference chemicals targeting specific biological targets (4)

• Approach:

- RefChemDB filtered for chemicals with > 5 chemical_target_mode associations for a given pathway (n = 3263 chemicals)
- An additional filter was applied to remove “promiscuous” chemicals that map to > 3 targets.
- Two agonists / antagonists were then selected (if available) for each remaining target with preference toward more selective chemicals (n = 330 chemical candidate list).
- The list was then manually curated to confirm target / mode associations, remove spurious target/mode associations and identify targets / pathways of interest that were not selected using the automated methods.
- The candidate list includes chemicals that target classical signal transduction pathways (e.g., nuclear receptors) and upstream / downstream proteins (e.g., GPCRs, enzymes, transporters) as “pathways”.

Targets by Target_Class

315 chemicals → 143 Targets Across → 13 Target Classes



Chemical Sourcing Summary / Progress (1)

Procurement Plan:

- 58 chemicals available in existing ToxCast library
- 168 chemicals obtained by EPA for CPP5 and other projects.
- 89 compounds designated for procurement by NTP.
 - 8 of 89 chemicals were present in the MRI chemical inventory
 - Initial sourcing of the remaining 81 chemicals targeted 5 g procurements of each to provide enough for preparation of 20 mM solutions in DMSO plus 1 g of neat material.
 - Initial sourcing cost was prohibitively high (~\$1.4M) 😞

Chemical Sourcing Summary / Progress (2)

Procurement Progress:

- NTP Chemistry worked with Steve Ferguson and Josh Harrill to identify substitutes for the most expensive chemicals on the sourcing list.
- Replacement chemicals:
 - Selected to target the same biomolecular pathways(s) and mode as the original chemicals.
 - Sourced and CAS number/chemical name/chemical form issues were resolved.
 - New sourcing cost: **\$181,400** 😊
- Chemicals were ordered through MRI Global.
- **2019-12-23:** All non-DEA schedule CPP5 compounds are in house at MRI and EPA-required paperwork has been completed. Project all non-DEA solubilized chemicals will be received by Evotec in Feb-March 2020.

Examples of manually curated chemical substitutions

Preferred_Name	Original_Substance_Requested	Target_Mode	Reason for substitution
Citronellol	Substitute for JYL-1511 (623166-14-3)	TRPV1 _ Agonist,TRPV1 _ Antagonist	Already in NTP inventory
Proguanil hydrochloride	Substitute for Piritrexim (72732-56-0)	DHFR _ Antagonist	Already in NTP inventory
Anisotropine methylbromide	Substitute for Biperiden (514-65-8)	CHRM1 _ Antagonist	Already in NTP inventory
Fludrocortisone acetate	Substitute for Fludrocortisone (127-31-1)	NR3C2 _ Agonist	Already in NTP inventory
Solifenacin	Substitute for 4-Diphenylacetoxy-1,1-dimethylpiperidinium (81405-11-0)	CHRM3 _ Antagonist	Already in NTP inventory
SB202190	Substitute for SB202190 (152121-30-7)	MAPK1 _ Antagonist,MAPK14 _ Antagonist	Already in NTP inventory
Montelukast sodium	Substitute for Montelukast (158966-92-8)	CYSLTR1 _ Antagonist,CYSLTR1 _ Binder	Already in NTP inventory
Rabusertib	Substitute for granulatimide (219828-99-6)	CHEK1 _ Antagonist	Already in NTP inventory
SU9516	Substitute for 3-[(1H-Imidazol-5-yl)methylidene]-5-methoxy-1,3-dihydro-2H-indol-2-one (377090-84-1)	CDK2 _ Antagonist	Already in NTP inventory
Apatinib myselate	Substitute for CP 547632 (252003-65-9)	KDR _ Antagonist	Already in NTP inventory
Ifenprodil (+)-tartrate salt	Substitute for Ifenprodil (23210-56-2)	GRIN2B _ Antagonist	Already in NTP inventory
Scopolamine hydrochloride	Substitute for Methoctramine (104807-46-7)	CHRM2 _ Antagonist	Already in ToxCast
Caffeine	Substitute for ST-1535 (496955-42-1)	ADORA2A _ Antagonist	Already in ToxCast
ZM-447439	Substitute for Barasertib (722543-31-9)	AURKB _ Antagonist	Cost savings
Leflunomide	Substitute for Brequinar (96187-53-0)	DHODH _ Antagonist	Cost savings
Doxycycline hydrochloride	Substitute for Prinomastat (192329-42-3)	MMP1 _ Antagonist,MMP2 _ Antagonist	Cost savings
Picrotoxin	Substitute for Mephobarbital (115-38-8)	GABRA1 _ Antagonist	Cost savings
Flt-3 inhibitor (TCS 359)	Substitute for Lestaurtinib (111358-88-4).	FLT3 _ Antagonist	Cost savings
Mebendazole	Substitute for MMAD (Monomethylauristatin D) (203849-91-6).	tubulin	Cost savings
AZD8055	Substitute for Temsirolimus (162635-04-3).	MTOR _ Antagonist	Cost savings
SCH-900776	Substitute for isogranulatimide (219829-00-2)	CHEK1 _ Antagonist	Cost savings
Ellagic acid	Substitute for 4,5,6,7-tetrabromobenzimidazole (577779-57-8)	CSNK2A1 _ Antagonist,CSNK2A2 _ Antagonist	Cost savings
Cetorelix acetate	Substitute for Cetorelix (120287-85-6)	GNRHR _ Agonist,GNRHR _ Antagonist	Cost savings
Tyramine	Substitute for 3-iodothyronamine (712349-95-6).	TAAR1 _ Agonist	Cost savings
Uridine 5'-triphosphate	Substitute for Diquafosol (59985-21-6)	P2RY2 _ Agonist	Cost savings
Zopolrestat	Substitute for Sorbinil (68367-52-2)	AKR1B1 _ Antagonist	Cost savings
Cetirizine	Substitute for Levocabastine hydrochloride (79547-78-7)	HRH1 _ Agonist	Cost savings
Zanamivir	Substitute for 2-Deoxy-2,3-dehydro-N-acetylneuraminic acid (24967-27-9)	NEU2 _ Antagonist	Cost savings
HU 210	Substitute for Tetrahydrocannabinol (1972-08-3). Class I DEA Compound	CNR1 _ Agonist,CNR1 _ Binder,CNR2 _ Binder	DEA Schedule I substance
Ergotamine tartrate	Substitute for Ergotamine (113-15-5). Class I DEA compound	HTR1D _ Agonist	DEA Schedule I substance
Milnacipran	Substitute for MDMA (42542-10-9). Class I DEA Compound	SLC6A4 _ Agonist	DEA Schedule I substance

CompTox Chemistry Dashboard Registration

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Tox21 CPP5 – Reference chemicals for high-throughput transcriptomics

Search TOX21TRANS Chemicals

Identifier substring search

List Details

Description: Tox21 CPP5 – Reference chemicals for high-throughput transcriptomics
Number of Chemicals: 312

312 chemicals

Select all Download Send to Batch Search Default

DTXSID CASRN TOXCAST

Hide chemicals that are: Filter by Name or CASRN

C=CC=C=S

Allyl isothiocyanate
DTXSID:DTXSID3020047
CASRN:57-06-7
TOXCAST:6/426

O=C1C=CC(=C2C=C1C(=C3C=CC(=C2)N3)C4=CC=CC=C4)C5=CC=CC=C5

Astemizole
DTXSID:DTXSID9020110
CASRN:68844-77-9
TOXCAST:229/612

CN1C=NC2=C1C(=O)N(C)C(=O)N2C

Caffeine
DTXSID:DTXSID0020232
CASRN:58-08-2
TOXCAST:53/676

O=C1C=CC(=C2C=C1C(=C3C=CC(=C2)N3)C4=CC=CC=C4)C5=CC=CC=C5

Capsaicin
DTXSID:DTXSID9020241
CASRN:404-86-4
TOXCAST:31/235

CC12CCC3C(C1CC4=C2C(=C(C=C4)O)C(=O)O)C(=O)O

Chenodeoxycholic acid
DTXSID:DTXSID2020260
CASRN:474-25-9
TOXCAST:50/235

CC1(C)C(=O)OC1C2=CC=C(C=C2)C3(Cl)CC3

Ciprofibrate
DTXSID:DTXSID8020331
CASRN:52214-84-3
TOXCAST:9/402

CC(=O)OC1C=CC2=C1C(=C(C=C2)C3=C(C=C1)C(=O)OC3C)C4=CC=CC=C4

Cyproterone acetate
DTXSID:DTXSID5020366
CASRN:427-51-0
TOXCAST:131/720

CC12CCC3C(C1CC4=C2C(=C(C=C4)O)C(=O)O)C(=O)O

Dexamethasone
DTXSID:DTXSID3020384
CASRN:50-02-2
TOXCAST:16/239

2:58 PM
1/11/2020

HTTr Reference chemical list registered in EPA Chemicals CompTox Dashboard. Not public (...yet).

Hypothesis & Specific Aims

- Project Goals (Initiated January 2018)



1. Design a fit-for-purpose approach to link reference chemicals for specific molecular targets (e.g., nuclear receptors, kinases) and transcriptomic signatures that is compatible with Tox21 partner capabilities



2. Identify a diverse set of reference chemicals targeting relevant biological-response pathways



3. Procure reference chemicals and prepare stock solutions

June
2020

4. Build a robust reference transcriptomics dataset with selected reference chemicals in 2 cell culture models (MCF-7 and 3D HepaRG) currently employed by Tox21 partners (384-well)

Oct
2020

5. Develop transcriptomic signatures that characterize cellular responses to reference chemicals & the utility of a 3-concentration approach to screen larger numbers of Tox21 compounds & untested chemicals

Dec
2020

6. Relate transcriptomic responses to reference chemicals (e.g., DEGs, BMCs) with established responses for reference biological targets (e.g., IC_{50} , K_i , EC_{50})

2021

7. Extend approach to thousands of compounds in a data-driven manner