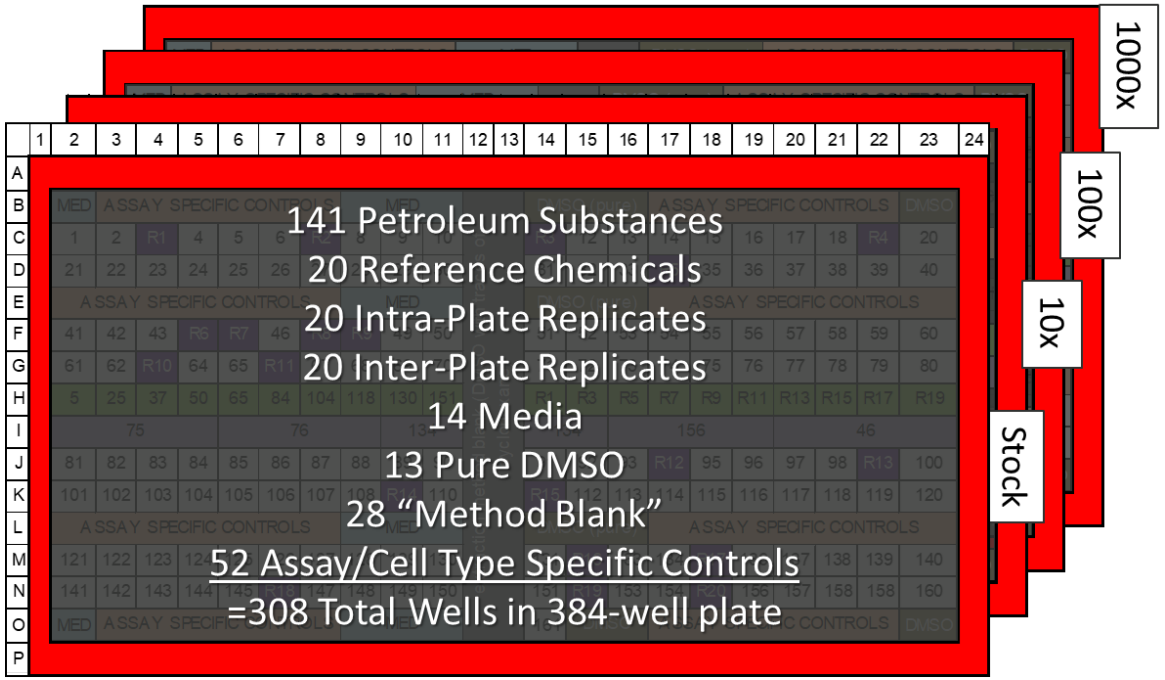
Work package 4a: Data Handling and Integration

Cat-App final event, Brussels, 6 September 2018

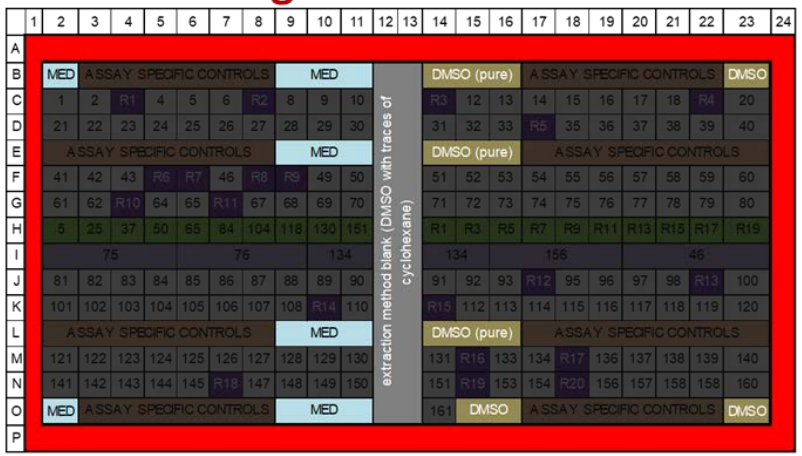
Fred Wright, John House, North Carolina State University

Bioactivity Profiling: Quality Control Considerations



- 4 plates (Stock Plate, 1:10, 1:100, 1:1,000 diluted) were prepared (5-point concentration-response)
- Extracts were diluted in "Method Blank" (cyclohexane-equilibrated DMSO)
- 4 sets of plates were prepared and sealed
- Plates were stored at -20°C before use
- Identical "master plates" were used at Texas A&M University and Public Health England laboratories

Negative Controls

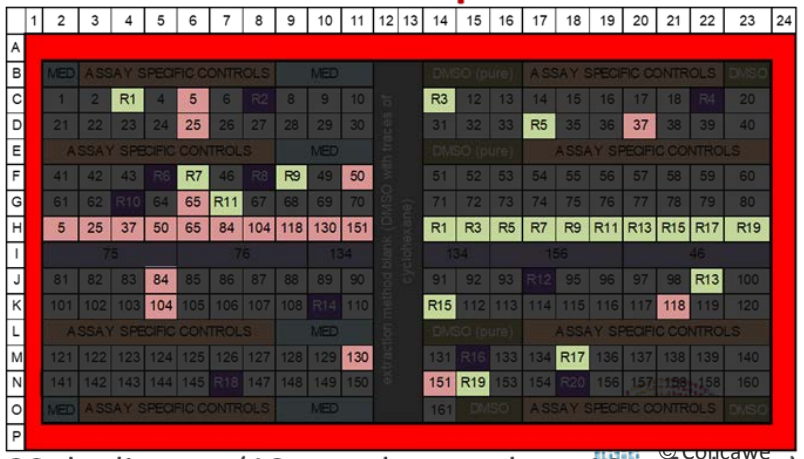


Media (14) Pure DMSO (13) "Method Blanks" (28)

Inter-Plate Replicates

Full concentration-response (5 petroleum subst.)

Intra-Plate Replicates

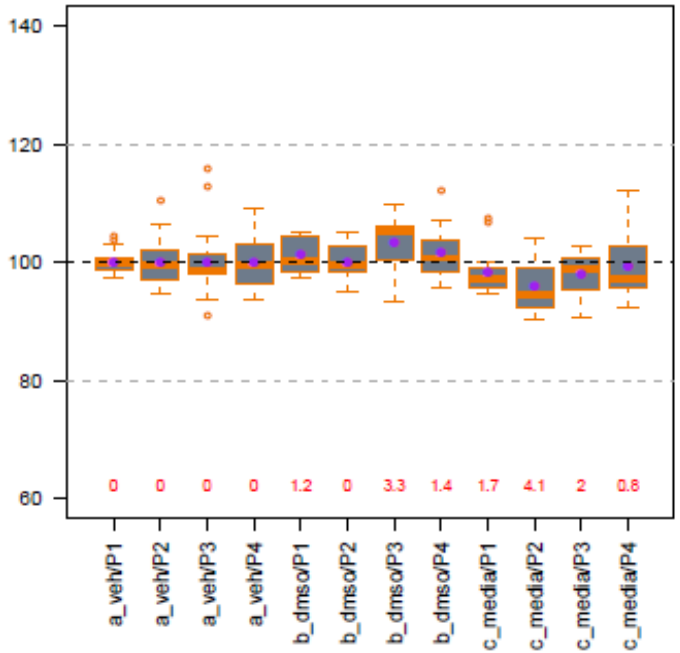


20 duplicates: (10 petroleum subst., 10 ref. chem.)

Bioactivity Profiling: QC – Negative Controls

	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																								
B	MED	ASSAY SPECIFIC CONTROLS							MED					DMSO (pure)	ASSAY SPECIFIC CONTROLS								DMSO	
C	1	2	R1	4	5	6	R2	8	9	10				R3	12	13	14	15	16	17	18	R4	20	
D	21	22	23	24	25	26	27	28	29	30				31	32	33	R5	35	36	37	38	39	40	
E	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)	ASSAY SPECIFIC CONTROLS									
F	41	42	43	R6	R7	46	R8	R9	49	50				51	52	53	54	55	56	57	58	59	60	
G	61	62	R10	64	65	R11	67	68	69	70				71	72	73	74	75	76	77	78	79	80	
H	5	25	37	50	65	84	104	118	130	151				R1	R3	R5	R7	R9	R11	R13	R15	R17	R19	
I	75				76				134					134		156				46				
J	81	82	83	84	85	86	87	88	89	90				91	92	93	R12	95	96	97	98	R13	100	
K	101	102	103	104	105	106	107	108	R14	110				R15	112	113	114	115	116	117	118	119	120	
L	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)	ASSAY SPECIFIC CONTROLS									
M	121	122	123	124	125	126	127	128	129	130				131	R16	133	134	R17	136	137	138	139	140	
N	141	142	143	144	145	R18	147	148	149	150				151	R19	153	154	R20	156	157	158	158	160	
O	MED	ASSAY SPECIFIC CONTROLS							MED					161	DMSO	ASSAY SPECIFIC CONTROLS							DMSO	
P																								

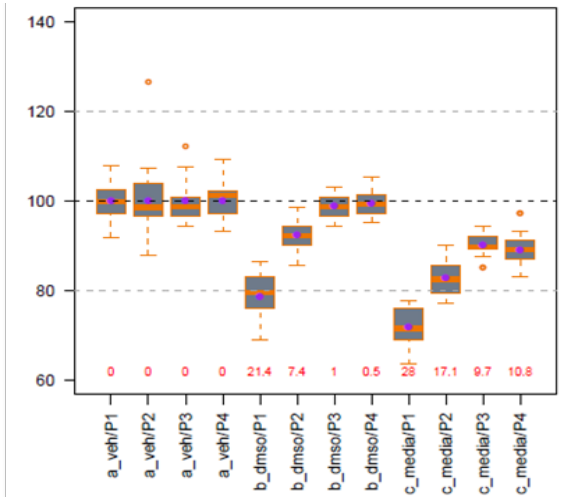
extraction method blank (DMSO with traces of cyclohexane)



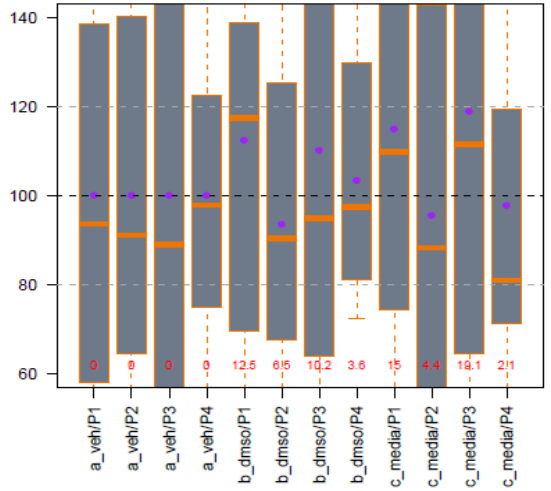
High concordance among negative controls

- Negative controls (media, DMSO, Method Blank) were arrayed on each of the plates
- Examine IQR of each type of negative controls
- Plate flagged if any negative control was IQR>20%
- Flagged if Method Blank IQR exceeded 25% indicating excessive variation between vehicle and media/DMSO

Discordance among controls



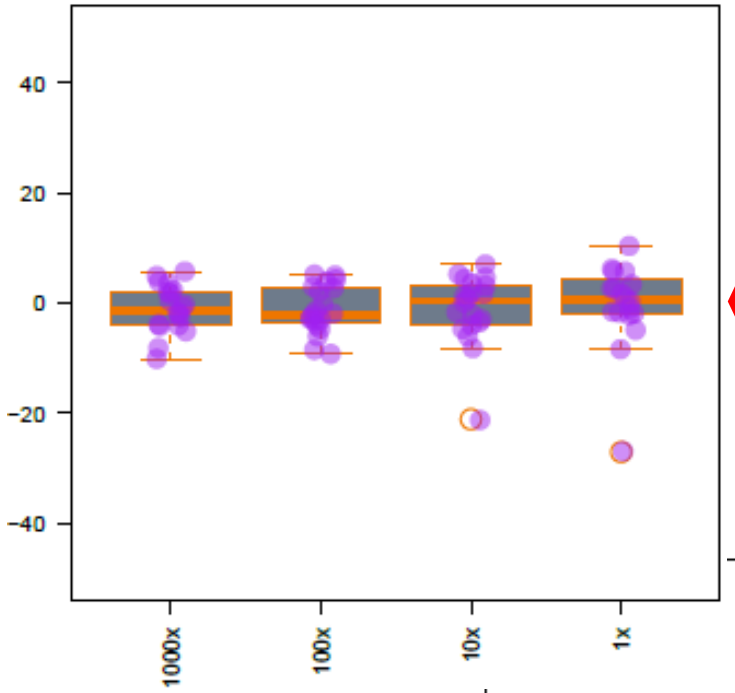
High variance in all controls



Bioactivity Profiling: QC – Intra-Plate Replicates

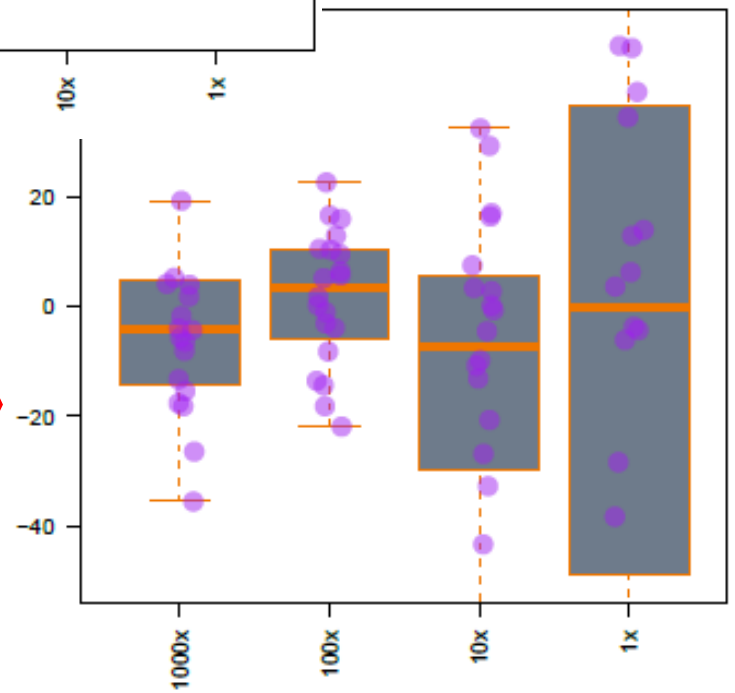
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A																								
B	MED	ASSAY SPECIFIC CONTROLS							MED			extraction method blank (DMSO with traces of cyclohexane)		DMSO (pure)			ASSAY SPECIFIC CONTROLS							DMSO
C	1	2	R1	4	5	6	R2	8	9	10				R3	12	13	14	15	16	17	18	R4	20	
D	21	22	23	24	25	26	27	28	29	30				31	32	33	R5	35	36	37	38	39	40	
E	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)			ASSAY SPECIFIC CONTROLS							
F	41	42	43	R6	R7	46	R8	R9	49	50				51	52	53	54	55	56	57	58	59	60	
G	61	62	R10	64	65	R11	67	68	69	70				71	72	73	74	75	76	77	78	79	80	
H	5	25	37	50	65	84	104	118	130	151				R1	R3	R5	R7	R9	R11	R13	R15	R17	R19	
I	75					76			134					134			156			46				
J	81	82	83	84	85	86	87	88	89	90				91	92	93	R12	95	96	97	98	R13	100	
K	101	102	103	104	105	106	107	108	R14	110				R15	112	113	114	115	116	117	118	119	120	
L	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)			ASSAY SPECIFIC CONTROLS							
M	121	122	123	124	125	126	127	128	129	130				131	R16	133	134	R17	136	137	138	139	140	
N	141	142	143	144	145	R18	147	148	149	150				151	R19	153	154	R20	156	157	158	158	160	
O	MED	ASSAY SPECIFIC CONTROLS							MED					161	DMSO		ASSAY SPECIFIC CONTROLS							DMSO
P																								

- 20 Substances duplicated on each plate
- Scaled metric [variance increases with concentration]
- Examine IQR of all 20 substances per plate in relation to IQR of their “Method Blank” controls
- Compare variation in these replicates to the variation in the vehicle controls



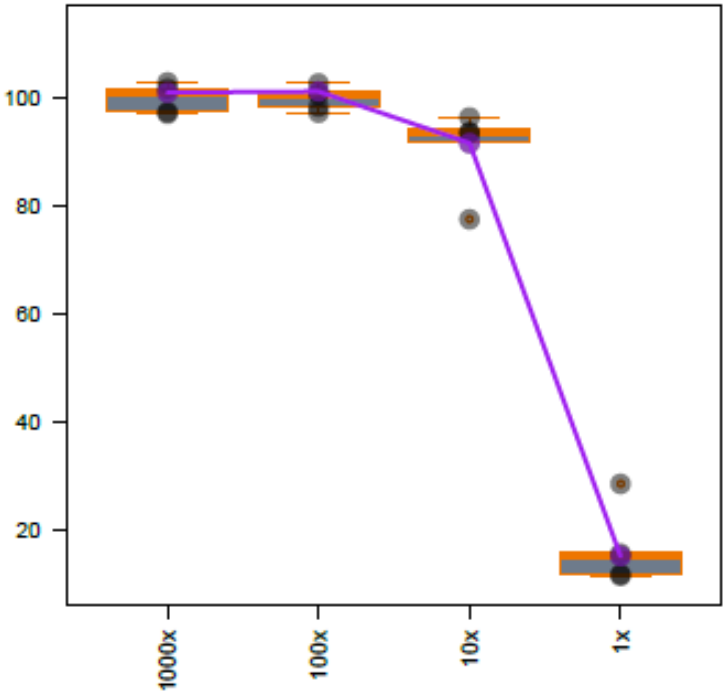
Excellent intra-plate replicability

Flagged intra-plate replicability



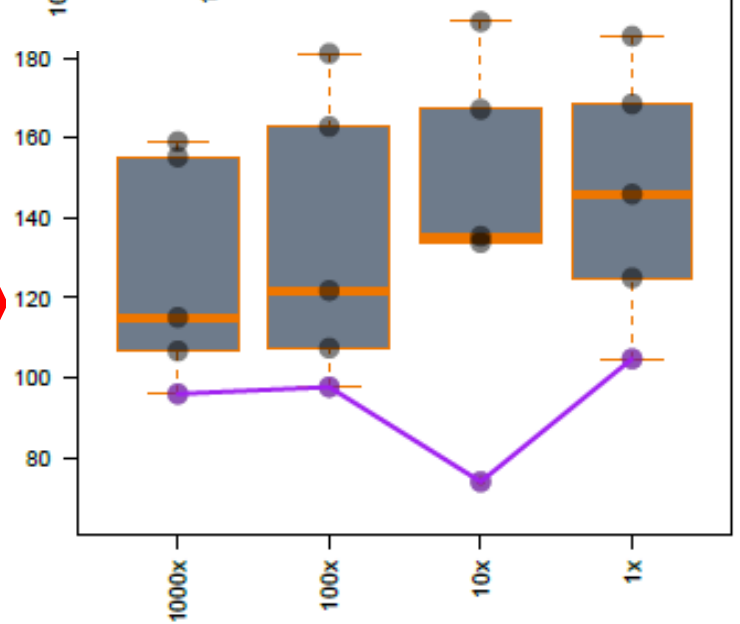
Bioactivity Profiling: QC – Inter-Plate Replicates

	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																								
B	MED	ASSAY SPECIFIC CONTROLS							MED			extraction method blank (DMSO with traces of cyclohexane)		DMSO (pure)	ASSAY SPECIFIC CONTROLS									DMSO
C	1	2	R1	4	5	6	R2	8	9	10				R3	12	13	14	15	16	17	18	R4	20	
D	21	22	23	24	25	26	27	28	29	30				31	32	33	R5	35	36	37	38	39	40	
E	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)	ASSAY SPECIFIC CONTROLS									
F	41	42	43	R6	R7	46	R8	R9	49	50				51	52	53	54	55	56	57	58	59	60	
G	61	62	R10	64	65	R11	67	68	69	70				71	72	73	74	75	76	77	78	79	80	
H	5	25	37	50	65	84	104	118	130	151				R1	R3	R5	R7	R9	R11	R13	R15	R17	R19	
I	75				76				134					134			156			46				
J	81	82	83	84	85	86	87	88	89	90				91	92	93	R12	95	96	97	98	R13	100	
K	101	102	103	104	105	106	107	108	R14	110				R15	112	113	114	115	116	117	118	119	120	
L	ASSAY SPECIFIC CONTROLS								MED					DMSO (pure)	ASSAY SPECIFIC CONTROLS									
M	121	122	123	124	125	126	127	128	129	130				131	R16	133	134	R17	136	137	138	139	140	
N	141	142	143	144	145	R18	147	148	149	150				151	R19	153	154	R20	156	157	158	158	160	
O	MED	ASSAY SPECIFIC CONTROLS							MED					161	DMSO	ASSAY SPECIFIC CONTROLS							DMSO	
P																								



Excellent inter-plate replicability

Flagged inter-plate replicability

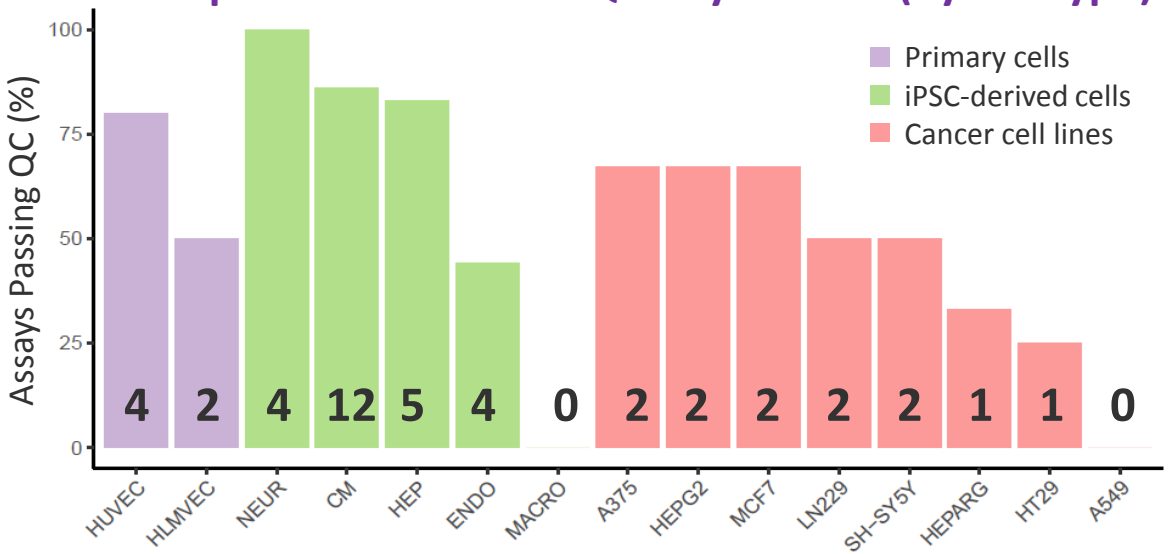


- One plate = one dilution [i.e., concentration]
- **Full dilution series** on every plate **for 5 substances**
- Examine variance of each plate “full dose response” vs a dose response from data on 4 separate plates
- Substance is flagged if multi-plate dose-response is >1.75 SD of any dose; assay flagged if 3+ chemicals flagged

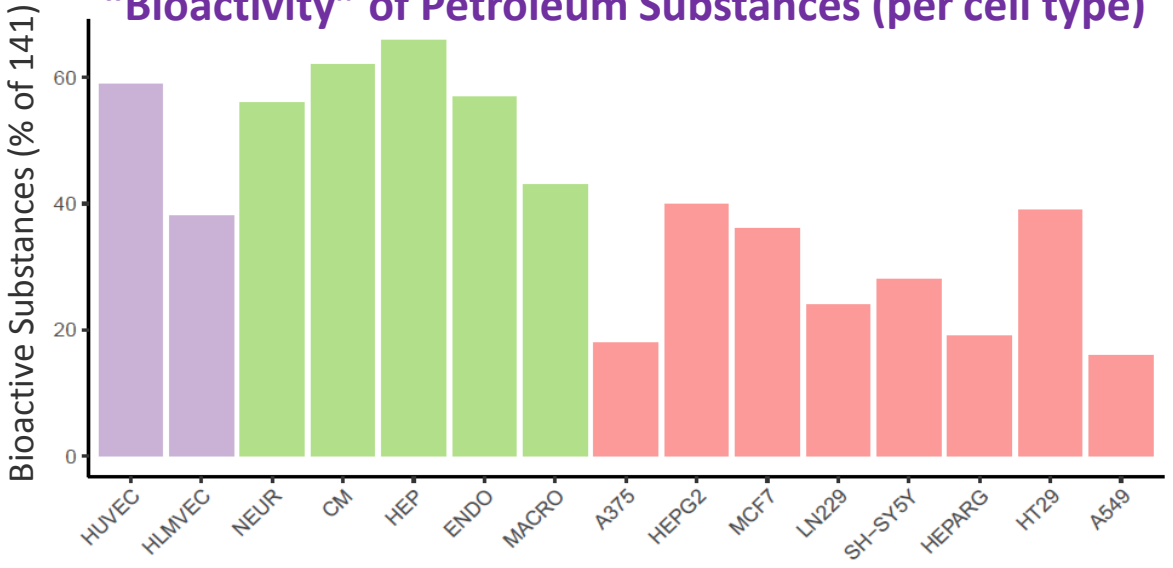
Bioactivity Profiling: QC – Summary

Cell Name	Cell Type	Total # of Endpoints	# of Endpoints with Flags
HUVEC Endothelial	Primary	5	1
HLMVEC Lung Endothelial	Primary	4	2
iCell Neurons	iPSC	4	0
iCell Cardiomyocytes	iPSC	14	2
iCell Hepatocytes	iPSC	6	1
iCell Endothelial	iPSC	9	5
iCell Macrophages	iPSC	1	1
A375 Melanoma	Cancer	3	1
HepG2 Hepatocytes	Cancer	3	1
MCF7 Breast Carcinoma	Cancer	3	1
LN229 Glioblastoma	Cancer	4	2
SH.SY5Y Neuroblastoma	Cancer	4	2
HepaRG Hepatocytes	Cancer	3	2
HT29 Colon Carcinoma	Cancer	4	3
A549 Lung Carcinoma	Cancer	3	3

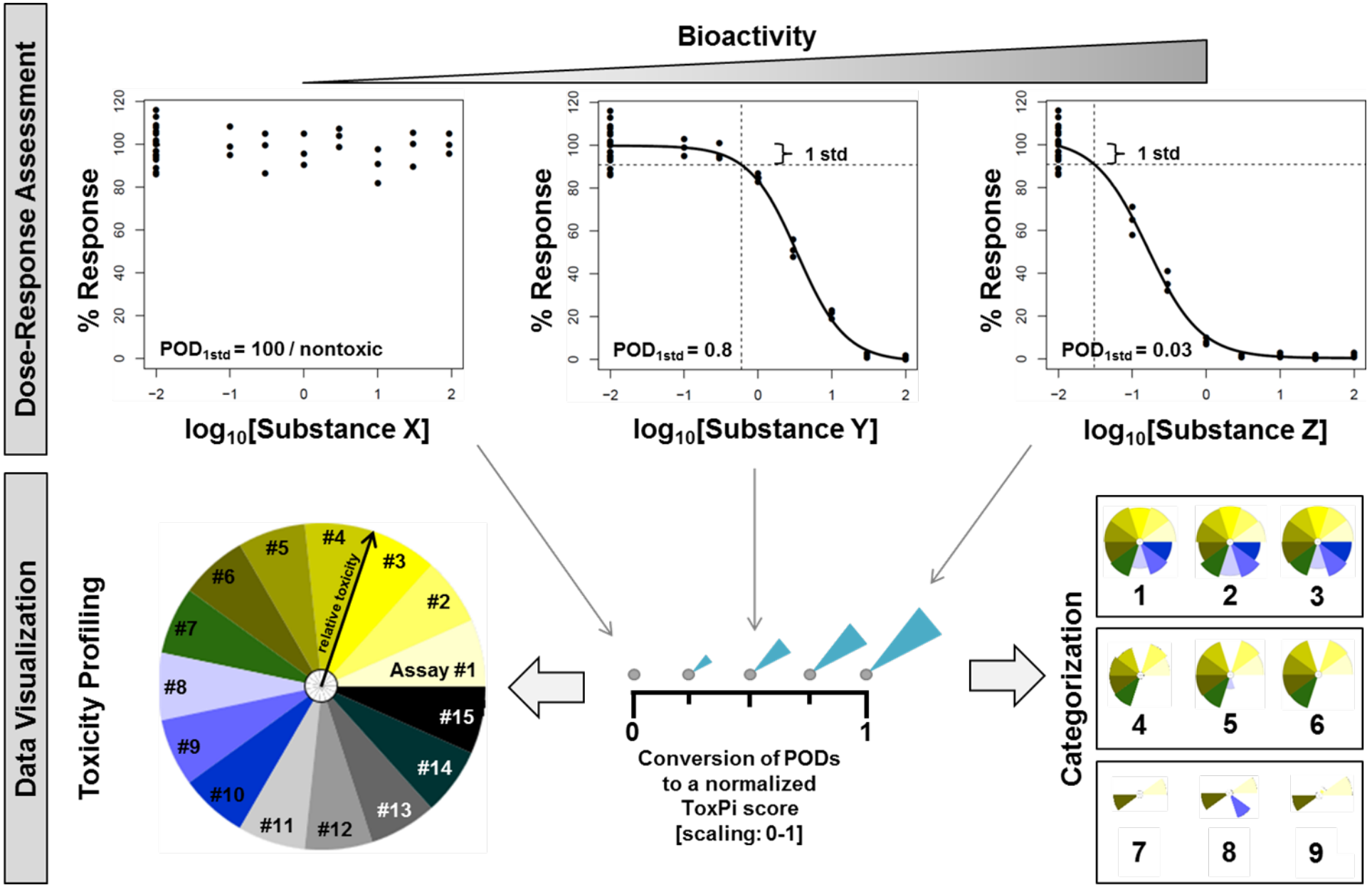
% of Endpoints that Passed Quality Control (by cell type)



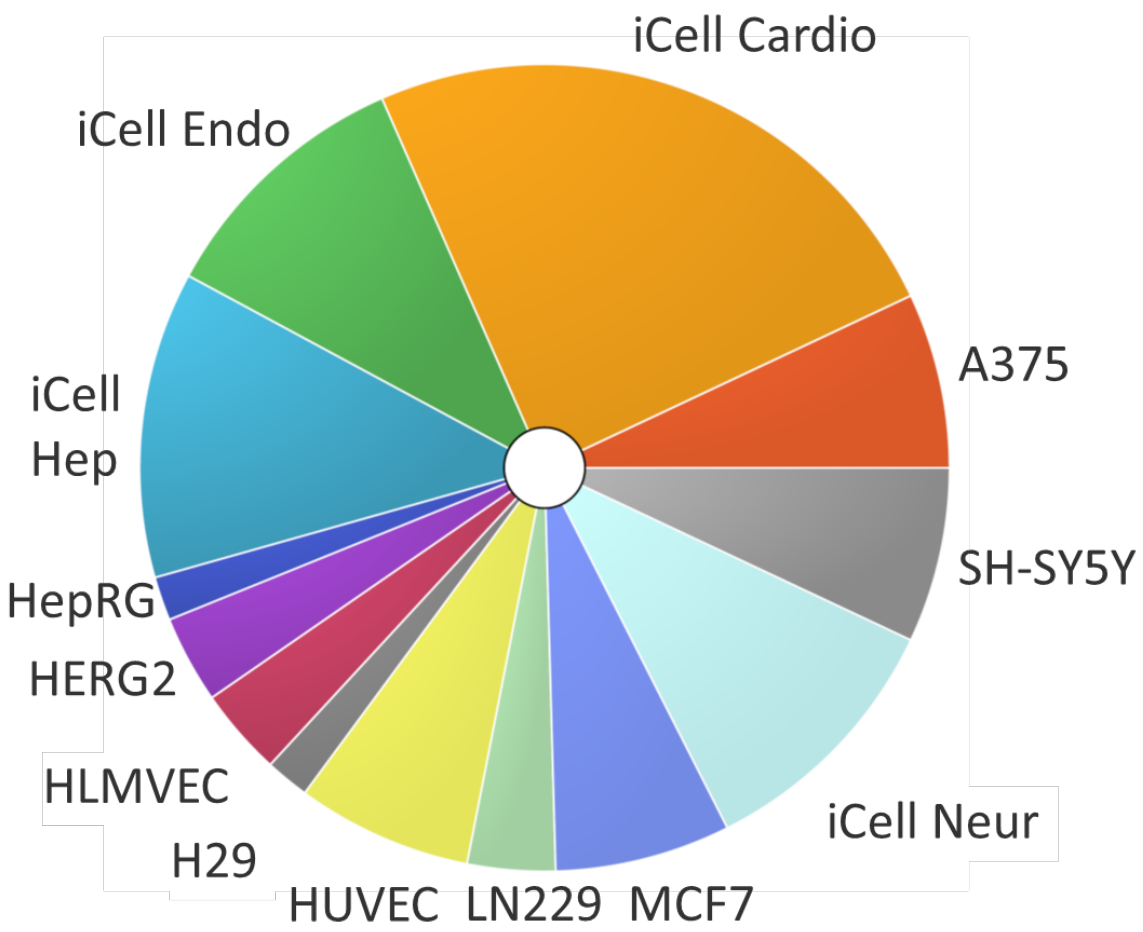
"Bioactivity" of Petroleum Substances (per cell type)



Using ToxPi for Grouping Petroleum Substances: Principle



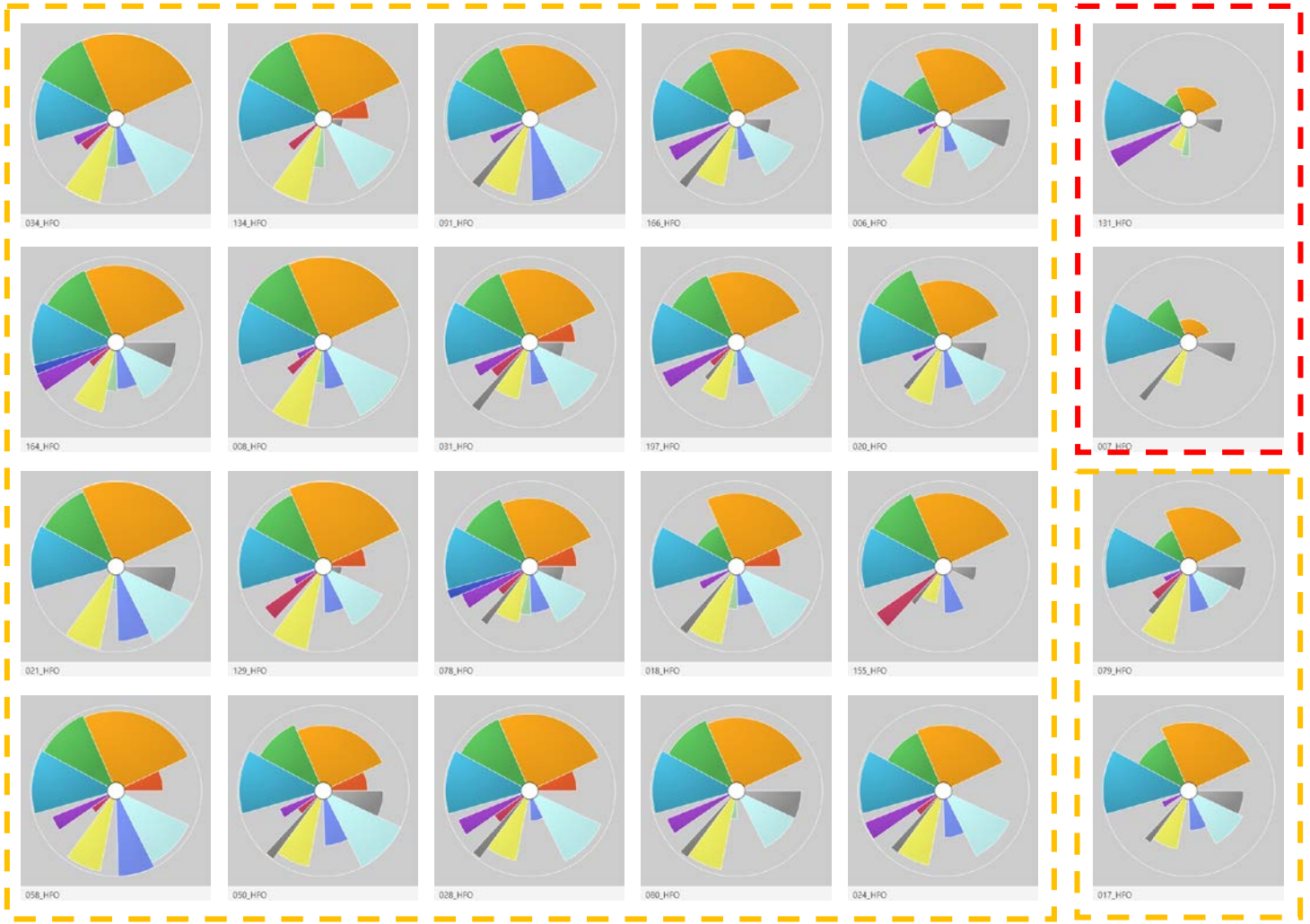
Using ToxPi for Grouping Petroleum Substances: Legend



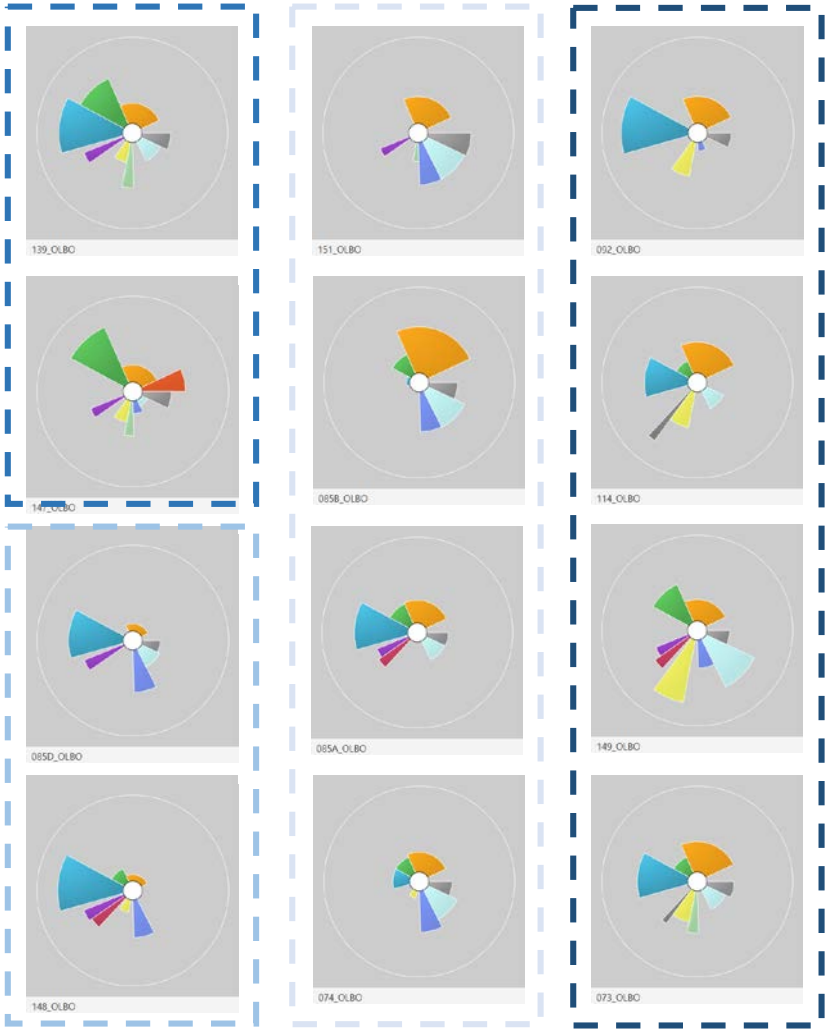
Name	Weight	# Metrics	Color
A375	4 (7.0%)	2	
CM	14 (24.6%)	12	
ENDO	6 (10.5%)	4	
HEP	7 (12.3%)	5	
HEPARG	1 (1.8%)	1	
HEPG2	2 (3.5%)	2	
HLMVEC	2 (3.5%)	2	
HT29	1 (1.8%)	1	
HUVEC	4 (7.0%)	4	
LN229	2 (3.5%)	2	
MCF7	4 (7.0%)	2	
NEUR	6 (10.5%)	4	
SH-SY5Y	4 (7.0%)	4	

Using ToxPi for Grouping Petroleum Substances: Supervised

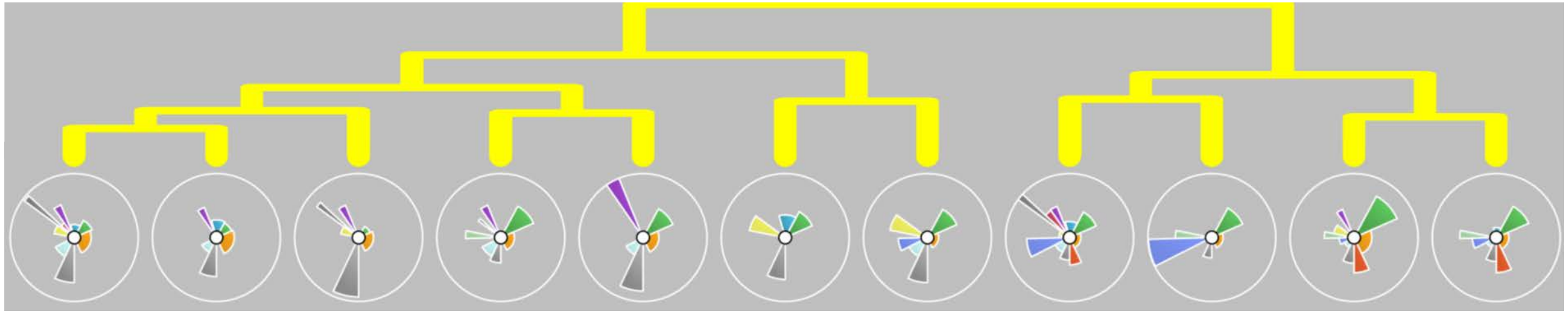
Heavy Fuel Oils



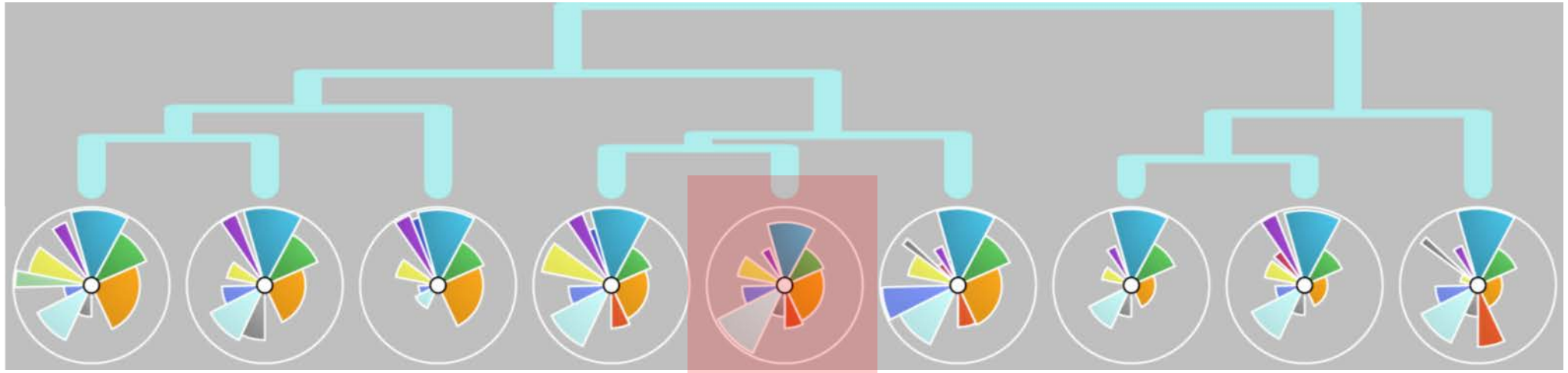
Other Lubricating Base Oils



Using ToxPi for Grouping Petroleum Substances: Un-Supervised

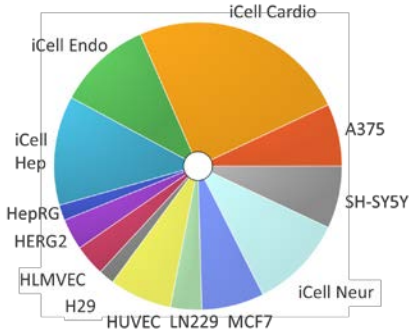


OLBO OLBO Wax FO Wax Wax OLBO OLBO Petrolatum OLBO OLBO

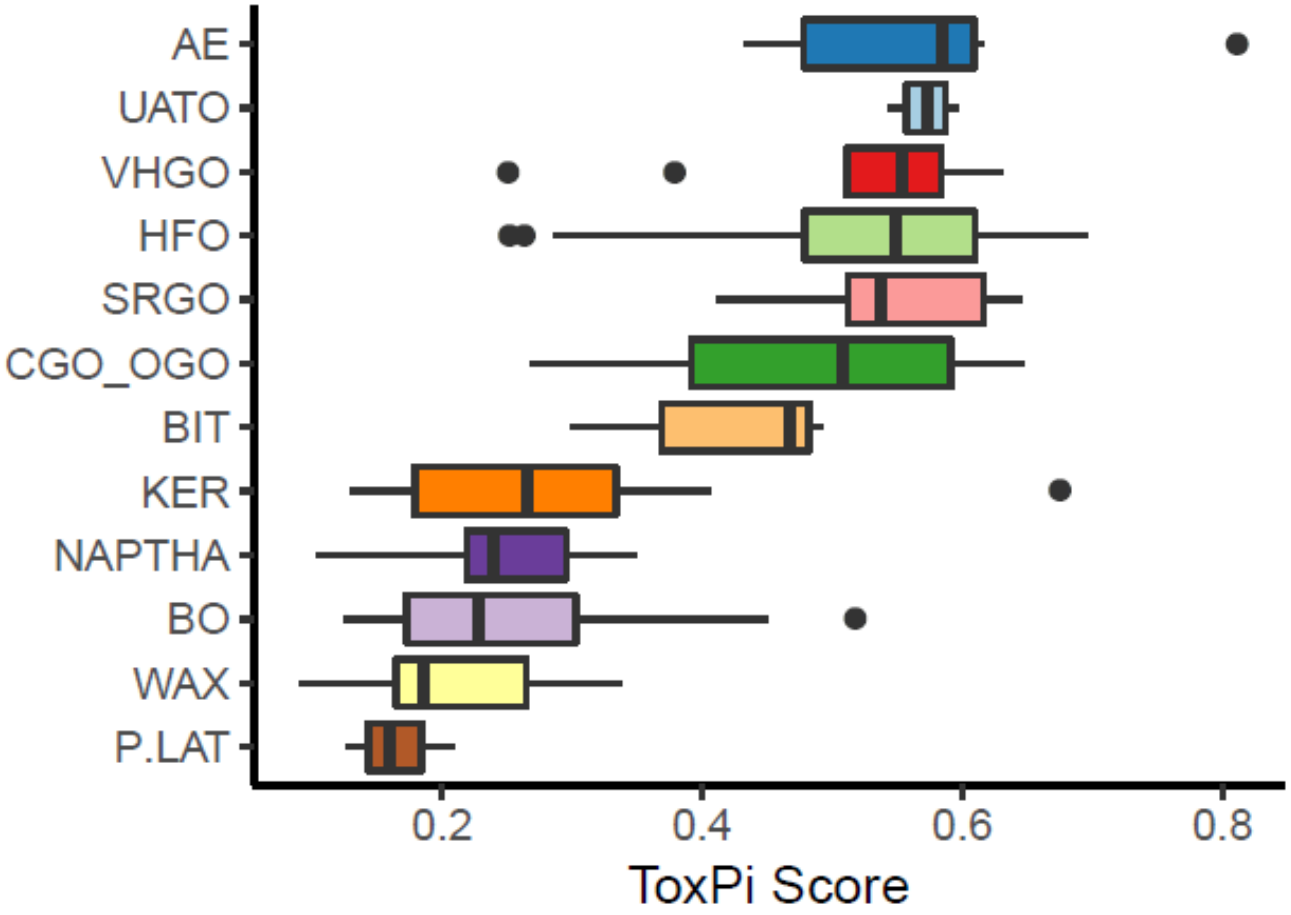


OxAsph BIT AE AE OLBO BIT BIT BIT AE

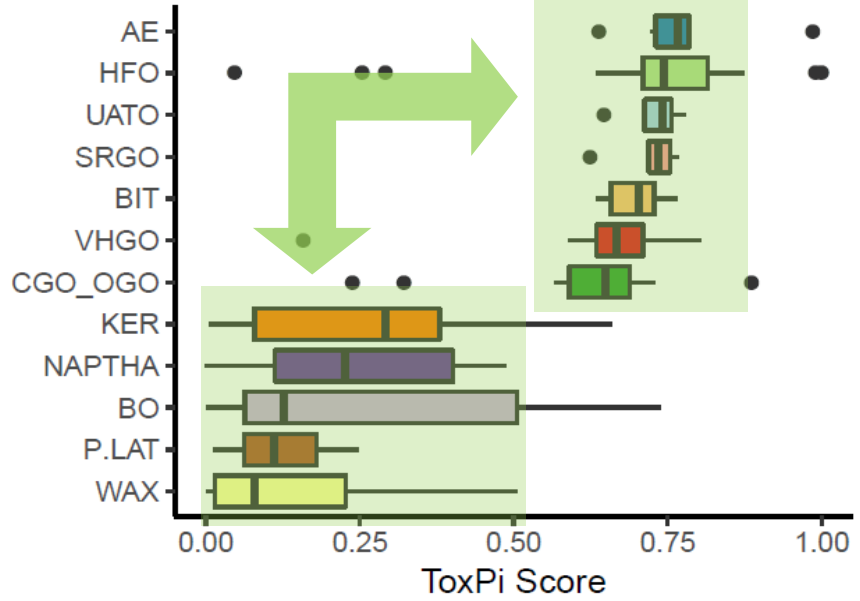
Using ToxPi for Grouping Petroleum Substances: Cell Types



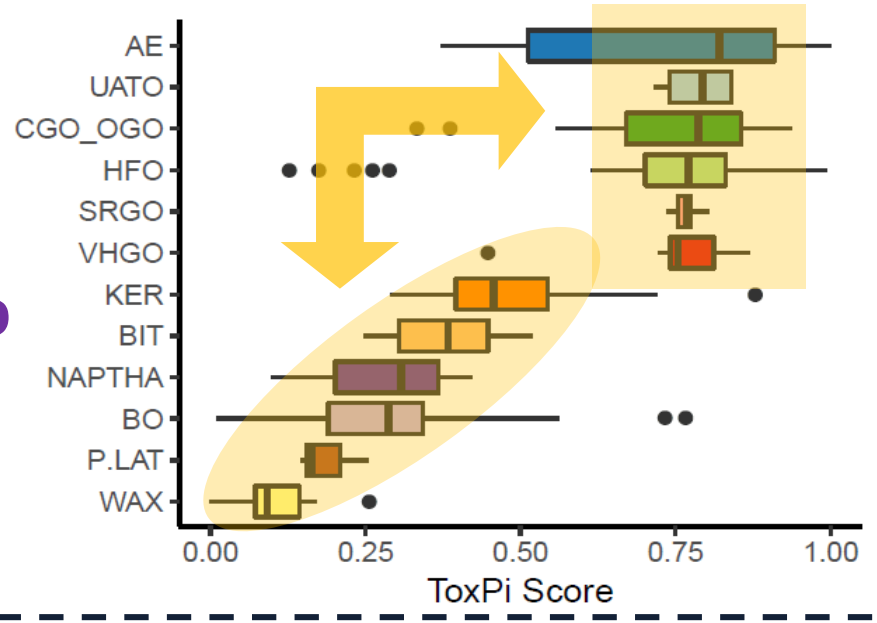
Combined data for 15 cell types



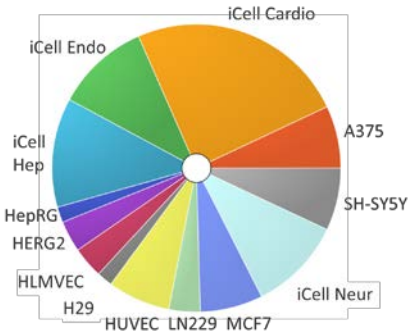
iCell Heps



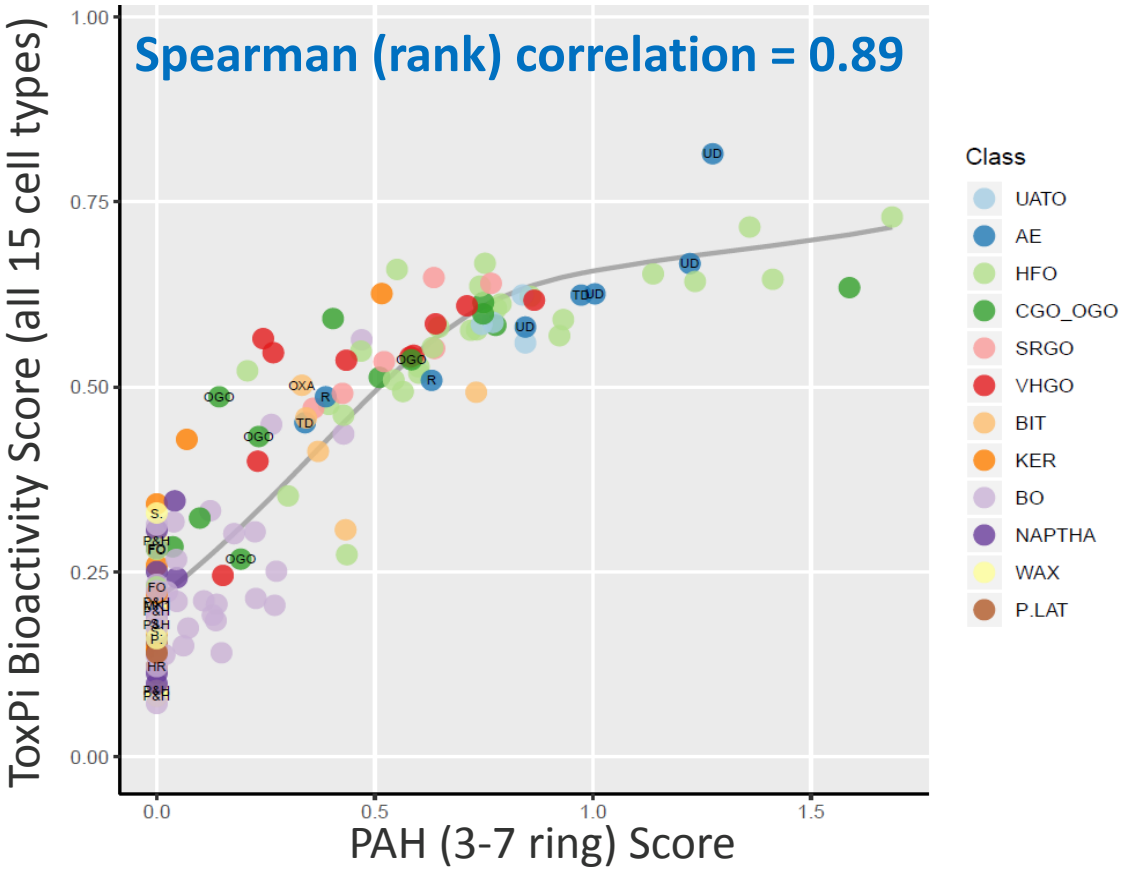
iCell Cardio



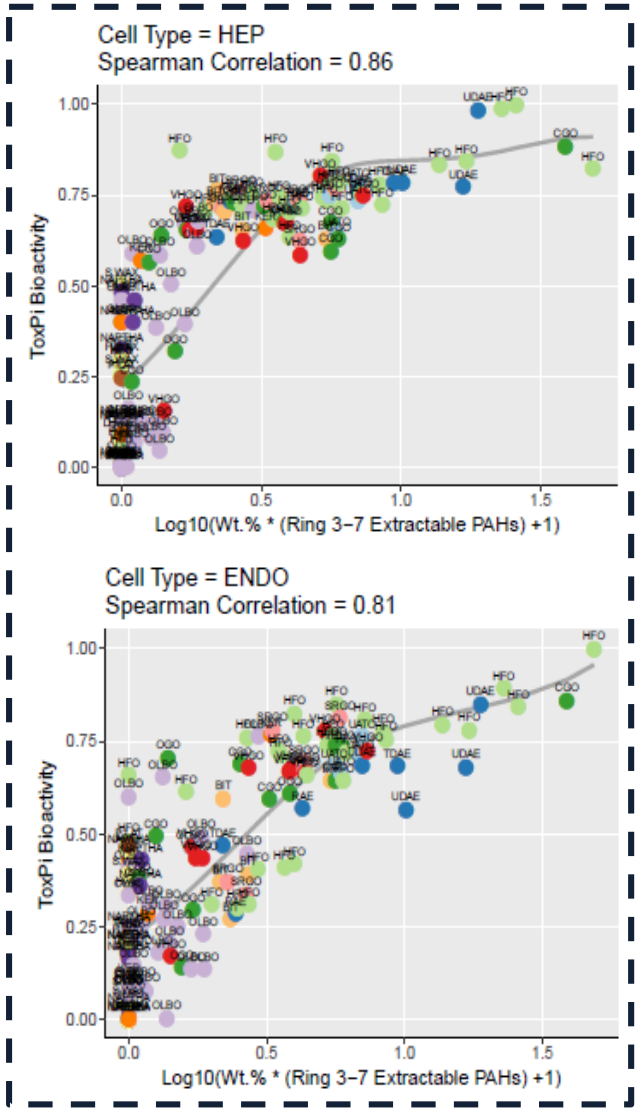
Using ToxPi for Grouping Petroleum Substances: PAH Content



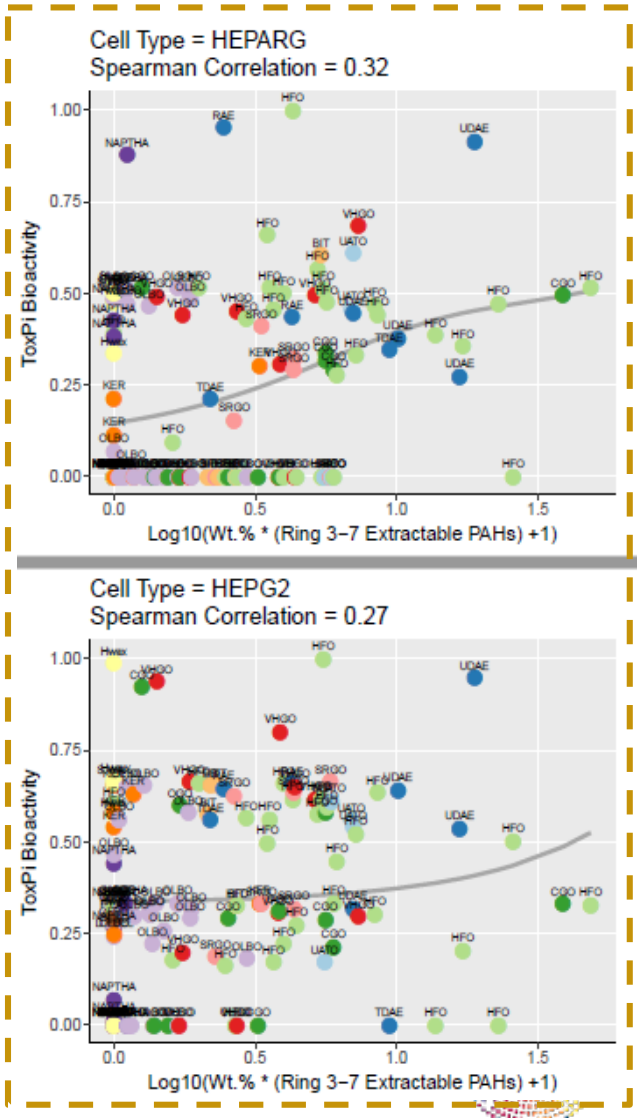
Combined data for 15 cell types



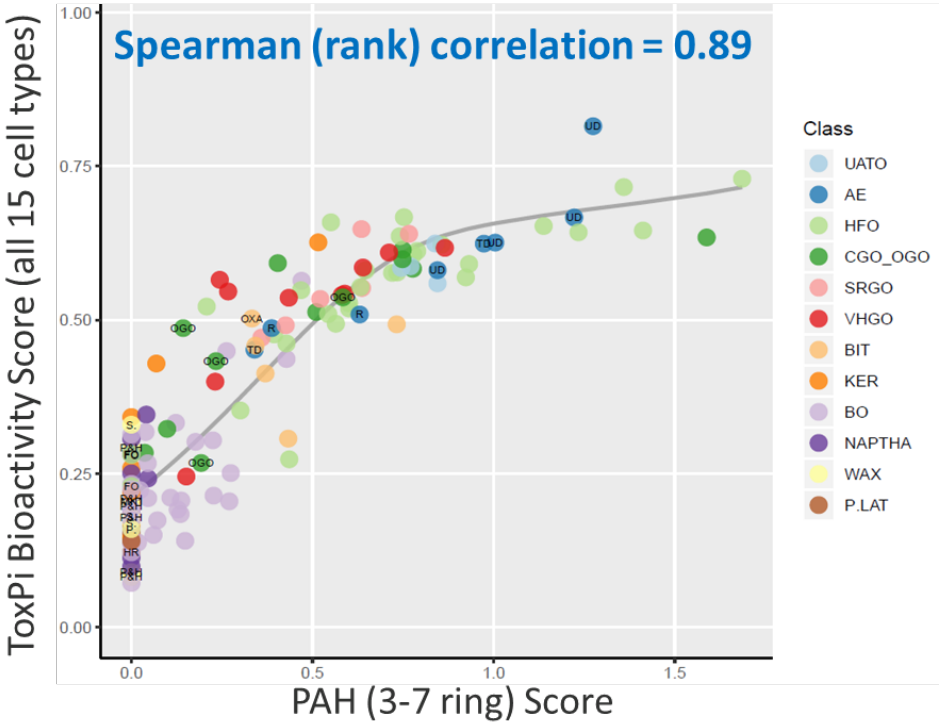
"Well correlated" cell types



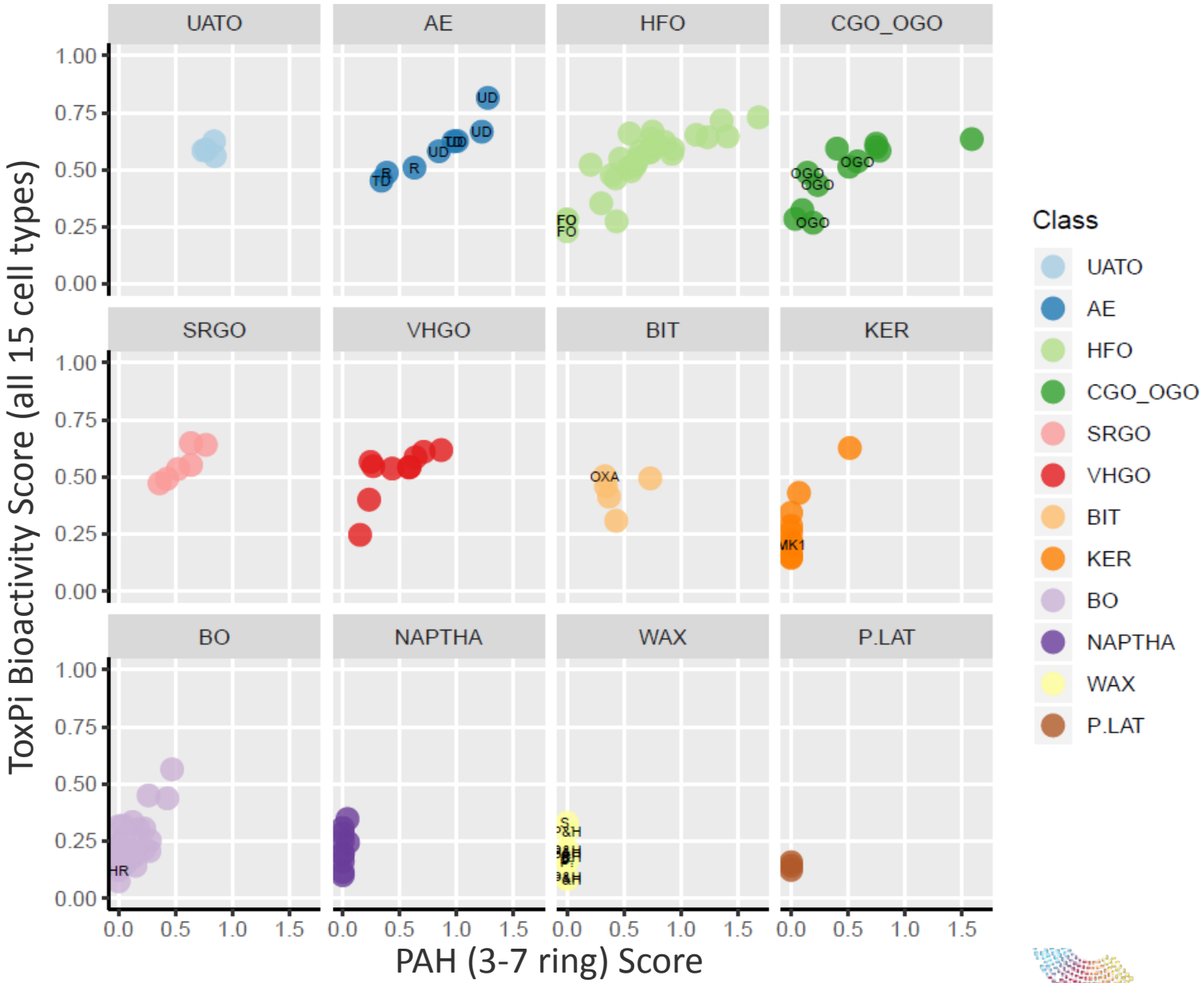
"Poorly correlated" cell types



Using ToxPi for Grouping Petroleum Substances: PAH Content



PAH Content	Correlation with ToxPi Bioactivity
Rings 3-7	**0.89**
Rings 4-7	0.70
Rings 5-7	0.51
Rings 1-2	0.36
3 Ring	**0.84**
4 Ring	0.73
5 Ring	0.55
6 Ring	0.43
7 Ring	0.29

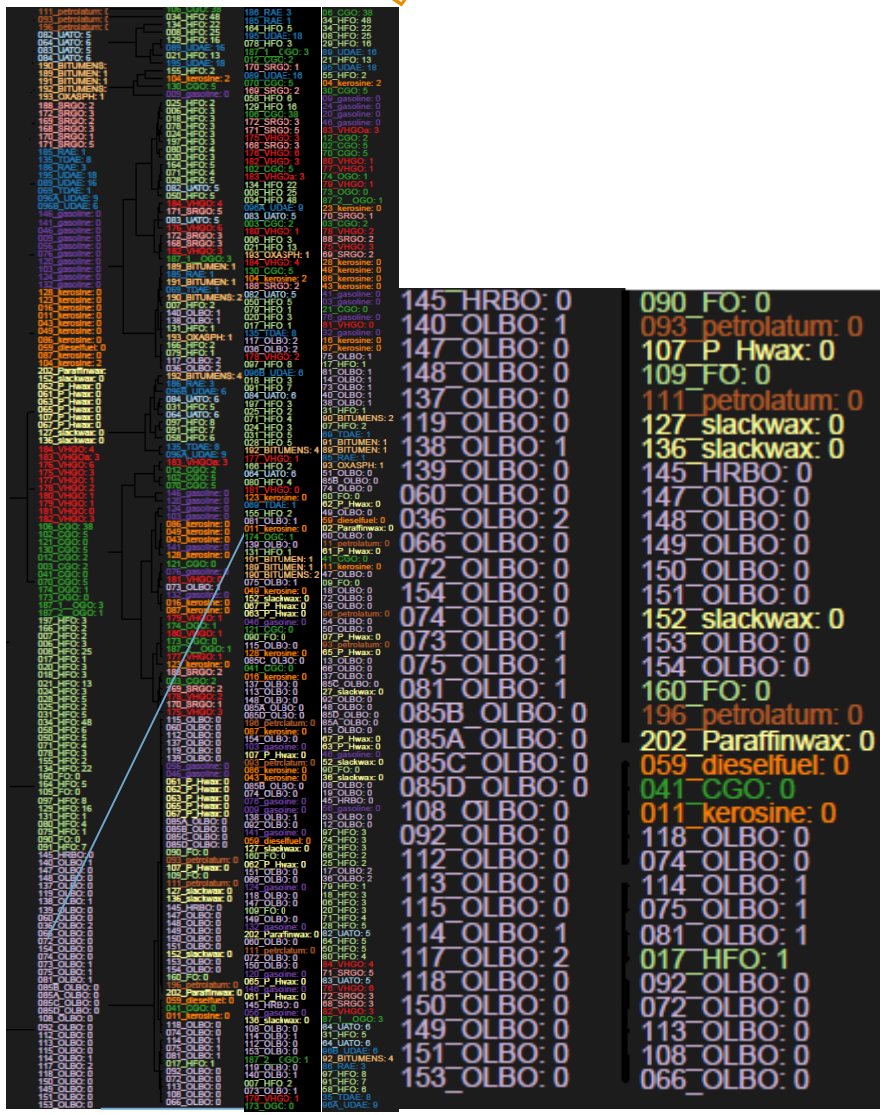


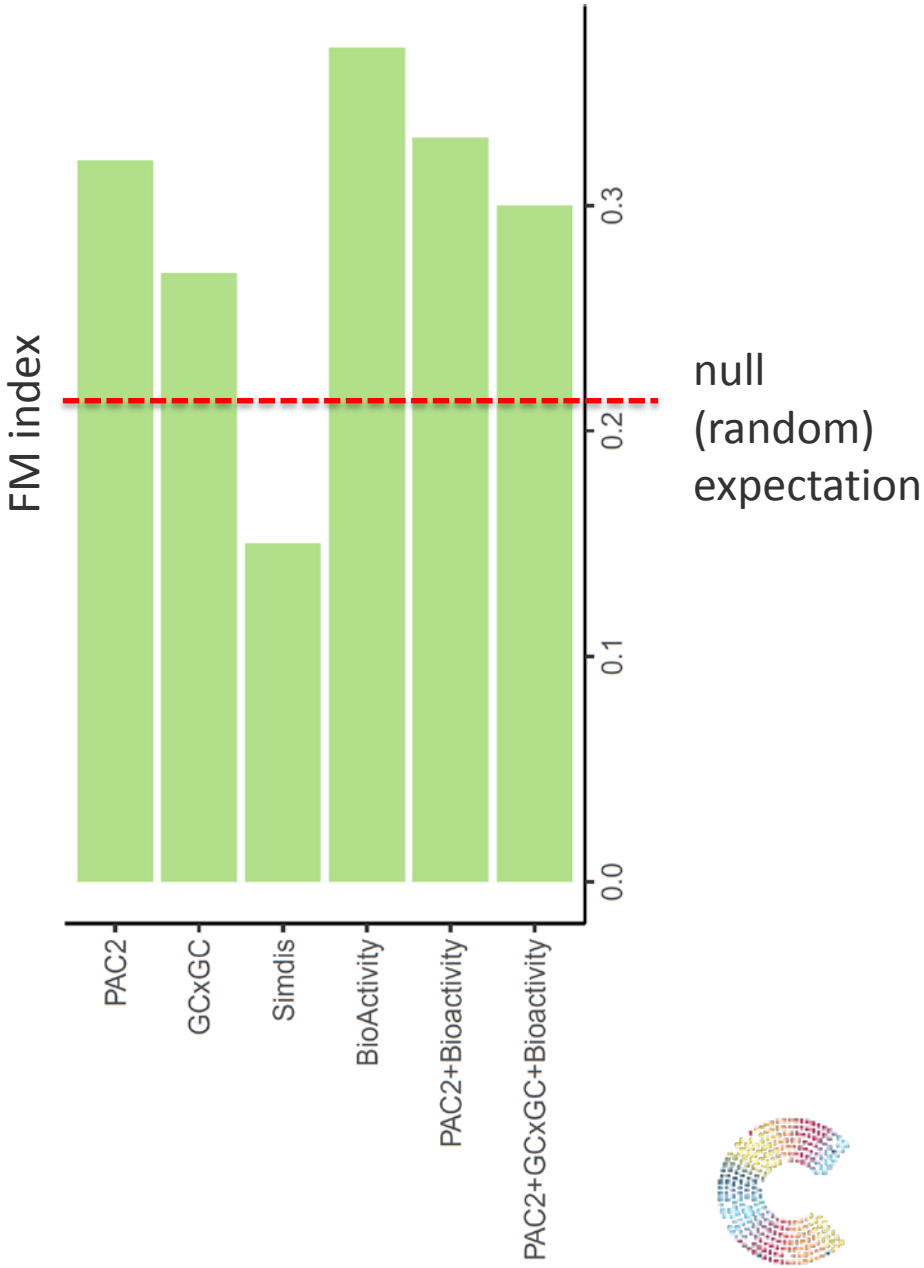
Grouping Petroleum Substances: Unsupervised clustering

- Clustering methods find groups of substances that are derived from the data (***analytical***, ***bioactivity***, or ***combination***), with no "bias" of predefined (*e.g.*, manufacturing stream) grouping
- Can be useful in identifying both "groups" and "fingerprints"
- Can also be affected by extraneous data features
- Comparisons of groupings by Fowlkes-Mallows (FM) Index
- FM ranges from 0.0 (no correspondence) to 1.0 (perfect correspondence)

Grouping Petroleum Substances: Unsupervised “grouping”

AE
UATO
HFO
VHGO
SRGO
CGO_OGO
BIT
KER
NAPTHA
BO
WAX
P.LAT



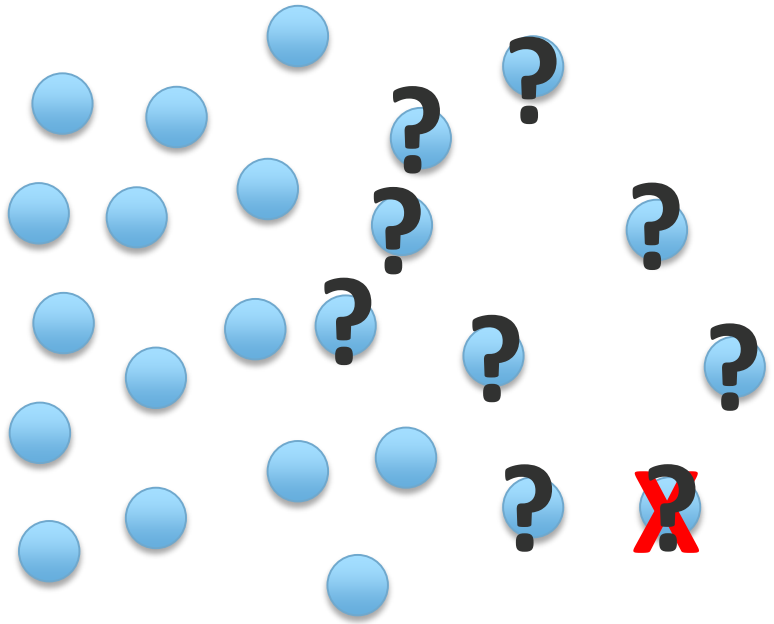


Grouping Petroleum Substances: “Fingerprint” selection

- Automatic supervised analyses can be used to identify features that are characteristic of existing categories
- Misclassifications can identify substances that i) do not match most other members of the category; ii) match another category closely
- **A**nalytical: (PAC2 analysis/Phys-chem properties-derived features: 7 individual classes, PAH 3-7 grouped, PAH 4-7 grouped, Final Boiling Point)
- **B**iological activity: QCed bioactivity endpoints (43 features)
- **A**+**B**=53 features

Grouping Petroleum Substances: “Fingerprint” selection

- Start with 141 UVCBs
- Leave one out, build a classifier model using remaining substances (for either 12 or 16 predefined petroleum substance categories)
- Classify the left out substance using the model
- Repeat procedure 140 more times



141 UVCBs

Analytical data only

Predicted

“True”

	P.LAT	WAX	BO	NAPTHA	KER	BIT	CGO_OGO	SRGO	VHGO	RFO	UATO	AE
P.LAT			3									
WAX			10									
BO			**19**				4	3		7		
NAPTHA			2	**7**			1					
KER				2	**6**		2					
BIT						**3**				2		
CGO_OGO			3				**9**					
SRGO			3				3					
VHGO			4				6					
RFO			9	1	2					**18**		
UATO			1							3		
AE						3				5		

Near matches

Distant matches

Grouping Petroleum Substances: “Fingerprint” selection

A=Analytical Data

A (44%)	P.LAT	WAX	BO	NAPTHA	KER	BIT	CGO_OGO	SRGO	VHGO	HFO	UATO	AE
P.LAT			3									
WAX			10									
BO			**19**			4	3			7		
NAPTHA			2	**7**			1					
KER				2	**6**		2					
BIT						**3**				2		
CGO_OGO			3				**9**					
SRGO			3				3					
VHGO			4				6					
HFO			9		1	2				**18**		
UATO			1							3		
AE						3				5		

B=Biological “endpoint” Data

B (38%)	P.LAT	WAX	BO	NAPTHA	KER	BIT	CGO_OGO	SRGO	VHGO	HFO	UATO	AE
P.LAT			3									
WAX			10									
BO		1	**29**							3		
NAPTHA			10									
KER			8							2		
BIT			1							4		
CGO_OGO			3							9		
SRGO										6		
VHGO			1							9		
HFO			5							**25**		
UATO										4		
AE										8		

A+B=Analytical+Biological Data

A+B (55%)	P.LAT	WAX	BO	NAPTHA	KER	BIT	CGO_OGO	SRGO	VHGO	HFO	UATO	AE
P.LAT		2	1									
WAX		**8**	2									
BO		4	**16**			4	3	1		3	1	1
NAPTHA		2		**7**			1					
KER				2	**6**		2					
BIT						**4**						1
CGO_OGO							**10**	1		1		
SRGO							3	**3**				
VHGO							7	2		1		
HFO		3			1	2				**23**		1
UATO								1		2		1
AE						4				4		

Conclusions

- Not all cell lines/assays are informative for bioactivity-based grouping
  - Careful quality control of bioactivity data is needed to increase confidence
  - These studies will guide future use of certain cell types and assays

- Bioactivity data are useful for grouping and “fingerprint” selection
  - Clear differences among some of the groups of substances was evident
  - Grouping based on bioactivity alone is not matching every substance to a pre-determined category

- Petroleum substance bioactivity shows variability within a group

- Further analyses are needed to define what bioactivity and analytical “features” are most informative for grouping of petroleum substances