

Navigating through the domains of biology and chemistry

15 May, 2009

James F. Rathman

The Ohio State University

Chihae Yang

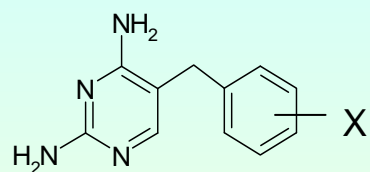
US FDA CFSAN OFAS

Landscape of ToxCAST data

- 309 unique chemicals
 - Mostly agrochemicals
- 524 *in vitro* bioassays
 - 9 *in vitro* assay providers
 - 285 cell-based, 239 cell-free
- 76 *in vivo* bioassays
 - ToxRefDB
 - Target organs (chronic), reproductive, developmental, carcinogenicity

Premise of Structure Activity Relationship

Only if we had that magical set of descriptors....



Molecular Structure



Activity

Properties

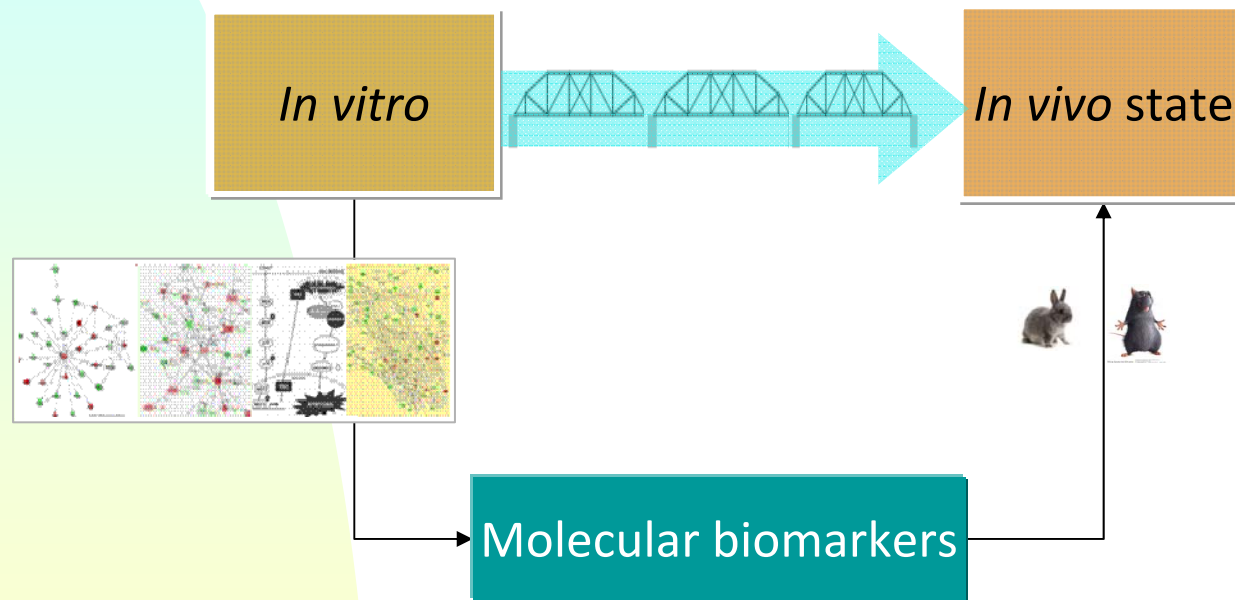


Molecular descriptors

“There just aren’t that many people interested in Chemistry” – Dave Weininger

Premise of in vitro predictions

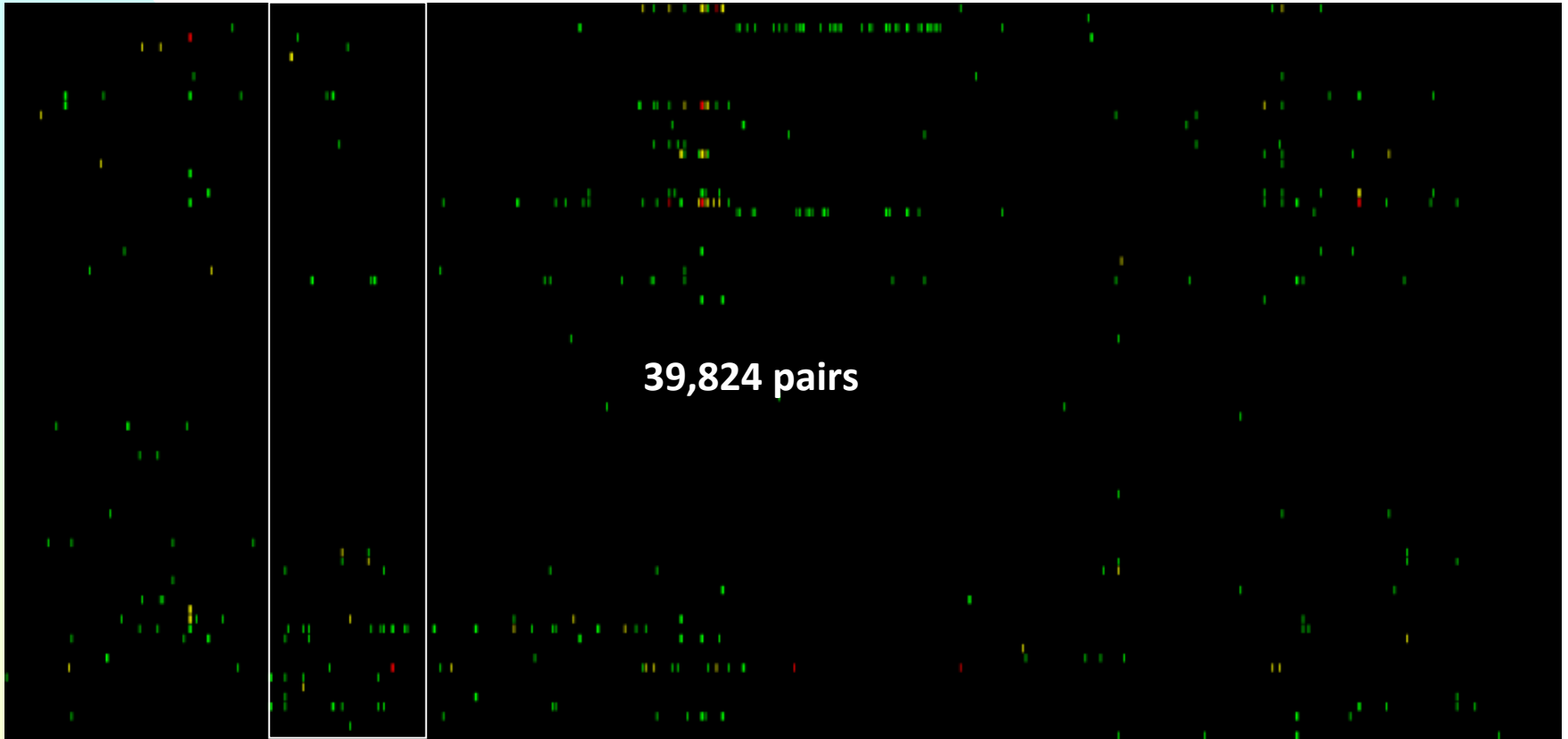
Only if we had that magical set of biomarkers, signatures....



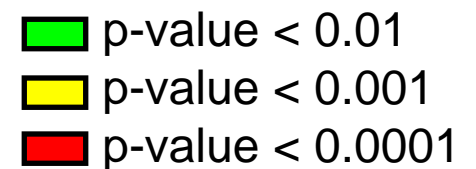
"Biology is incredibly complicated. Why?" – Richard Dawkins

Pairwise positive agreement for all in vivo vs. all in vitro assays ($\alpha = 0.01$)

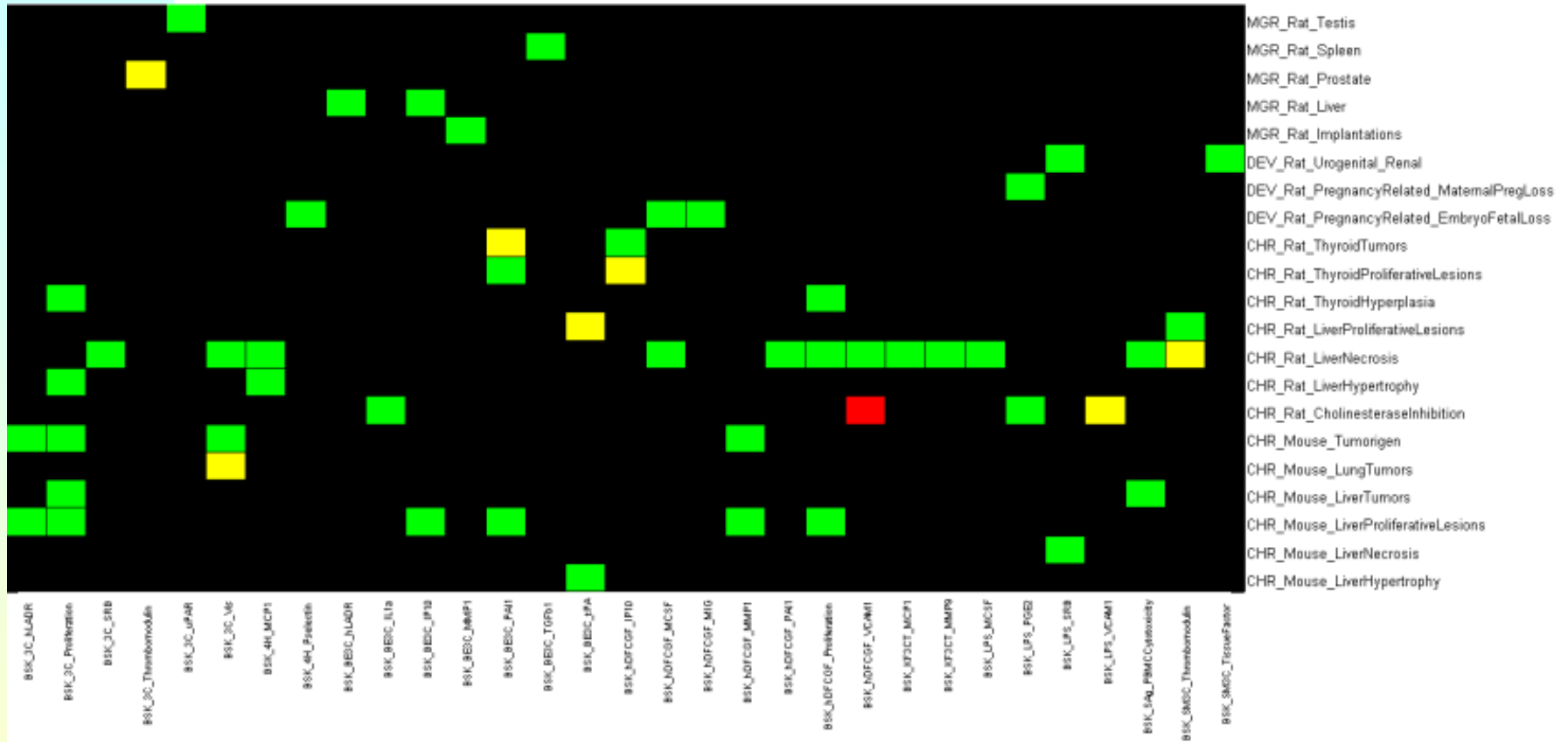
in vivo assays (76 total)



in vitro assays (524 total)



In vivo and BioSeek assay pairs with significant positive agreement ($\alpha = 0.01$)



in vitro assays
(27 had at least one significant result)

- p-value < 0.01
- p-value < 0.001
- p-value < 0.0001

Pairwise agreement between *in vitro* and *in vivo* assay results

		in vivo	
		0	1
in vitro	0	n_{00}	n_{01}
	1	n_{10}	n_{11}

$H_0 : \theta = 1$ (independent)

$H_1 : \theta > 1$ (positive agreement)

where θ is the odds ratio: $\theta = \frac{P(0|0)P(1|1)}{P(1|0)P(0|1)}$

Fisher's exact test: $\hat{\theta} = \frac{n_{00} n_{11}}{n_{10} n_{01}}$ at a given significance level α

- Significant pairs:
- 55 at $\alpha = 0.001$
 - 336 at $\alpha = 0.01$

The "Good" News

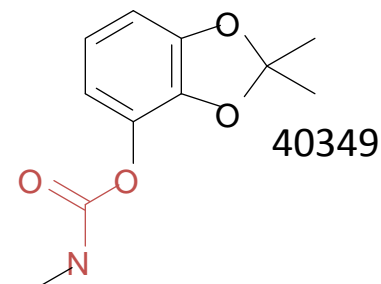
NVS_ENZ_rAChE

CHR_Rat_Cholinesterase Inhibition

	0	1
0	202	28
1	4	14

PNAS, March 18, 2008, Vol 105 (11), 4295.

p-value = 3.8×10^{-9}
concordance = 87%

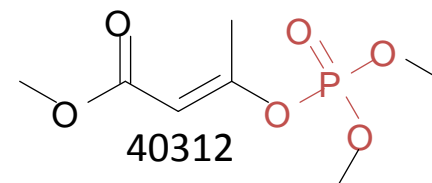


BSK_hDFCGF
_VCAM1

CHR_Rat_Cholinesterase Inhibition

	0	1
0	136	13
1	70	29

Chem. Res. Toxicol. 2009, 22, 633–638.



p-value = 2.8×10^{-5}
concordance = 67%

And the “Not so good news”

- The fraction of in vivo/in vitro assay pairs with statistically significant positive agreement is
 - 0.0084 at $\alpha = 0.01$
 - 0.0014 at $\alpha = 0.001$
- Type I error?
- Approximately half (50.8%) of the assay pairs have a p-value ≈ 1 due to very few (often zero) actives in one or both assays.

Implications - experimental factors

- Bioassays
 - detection limits and reproducibility...
- 309 Chemicals
 - some of them look quite reactive...
- ...



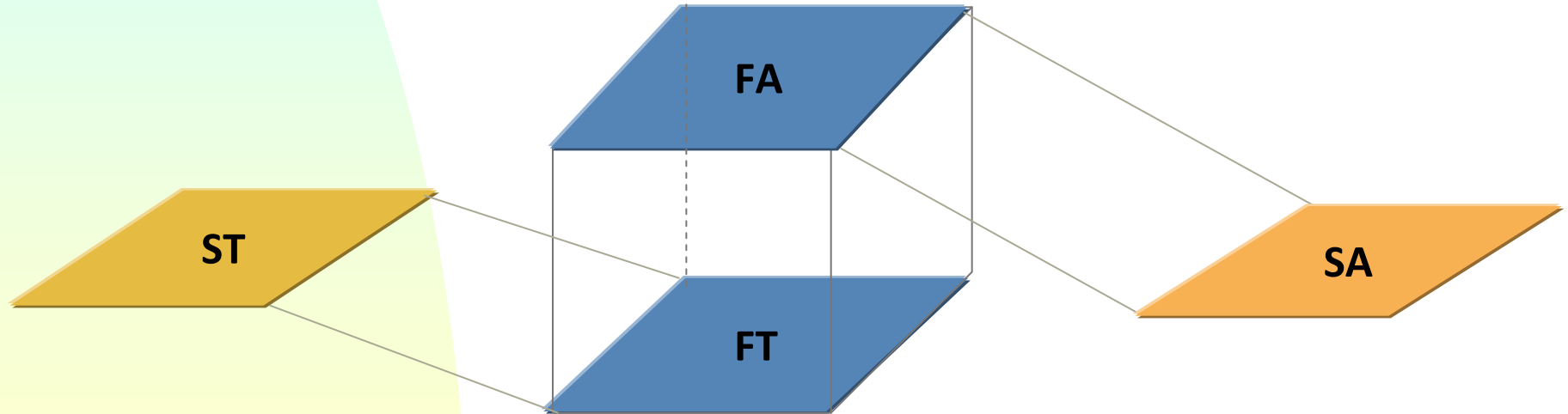
- Remove reactive chemicals from analysis
 - Acyl hydrazide, α -halo carbonyl, reactive alkyl halides, halo amine, α -halo ethers...
- Result: 288 chemicals
 - Large impact on the statistics of the agreement pairs
 - 30 pairs at $\alpha = 0.001$
 - 273 pairs at $\alpha = 0.01$

Features dimension as a link

Structure-
In vivo assays

In vitro (FA) – in vivo (FT)

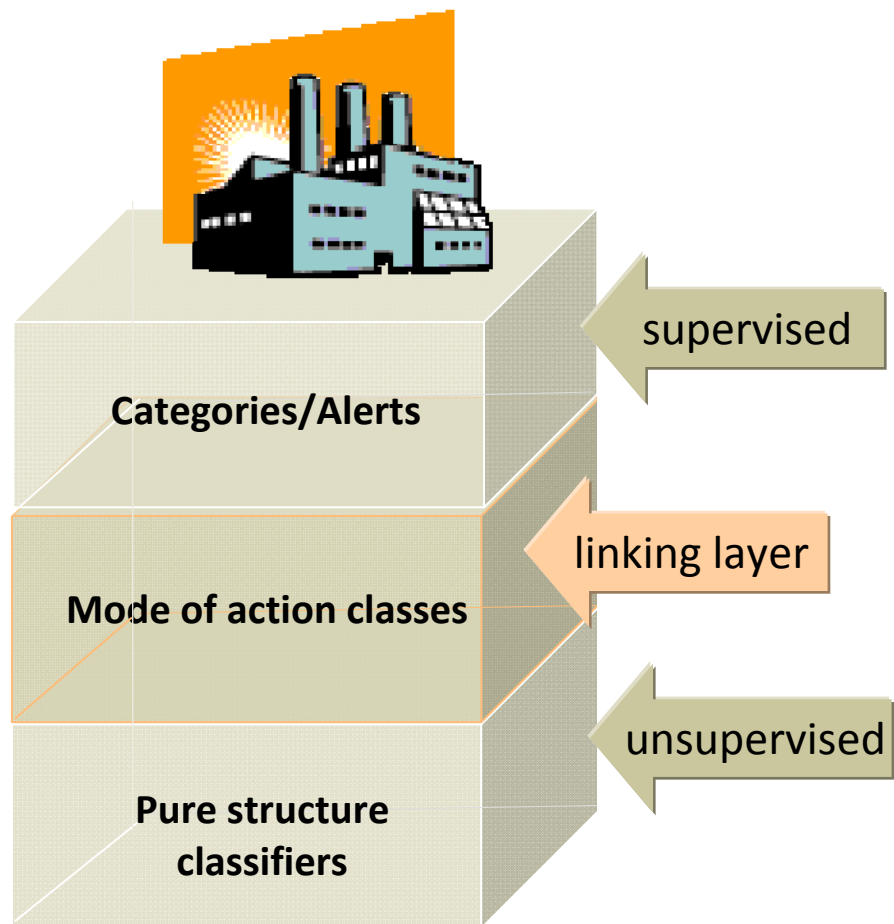
Structure-
In vitro assays



Presented in EPA Comp Tox Forum in May 2007

Conceptual world of chemical representation for toxicity predictions

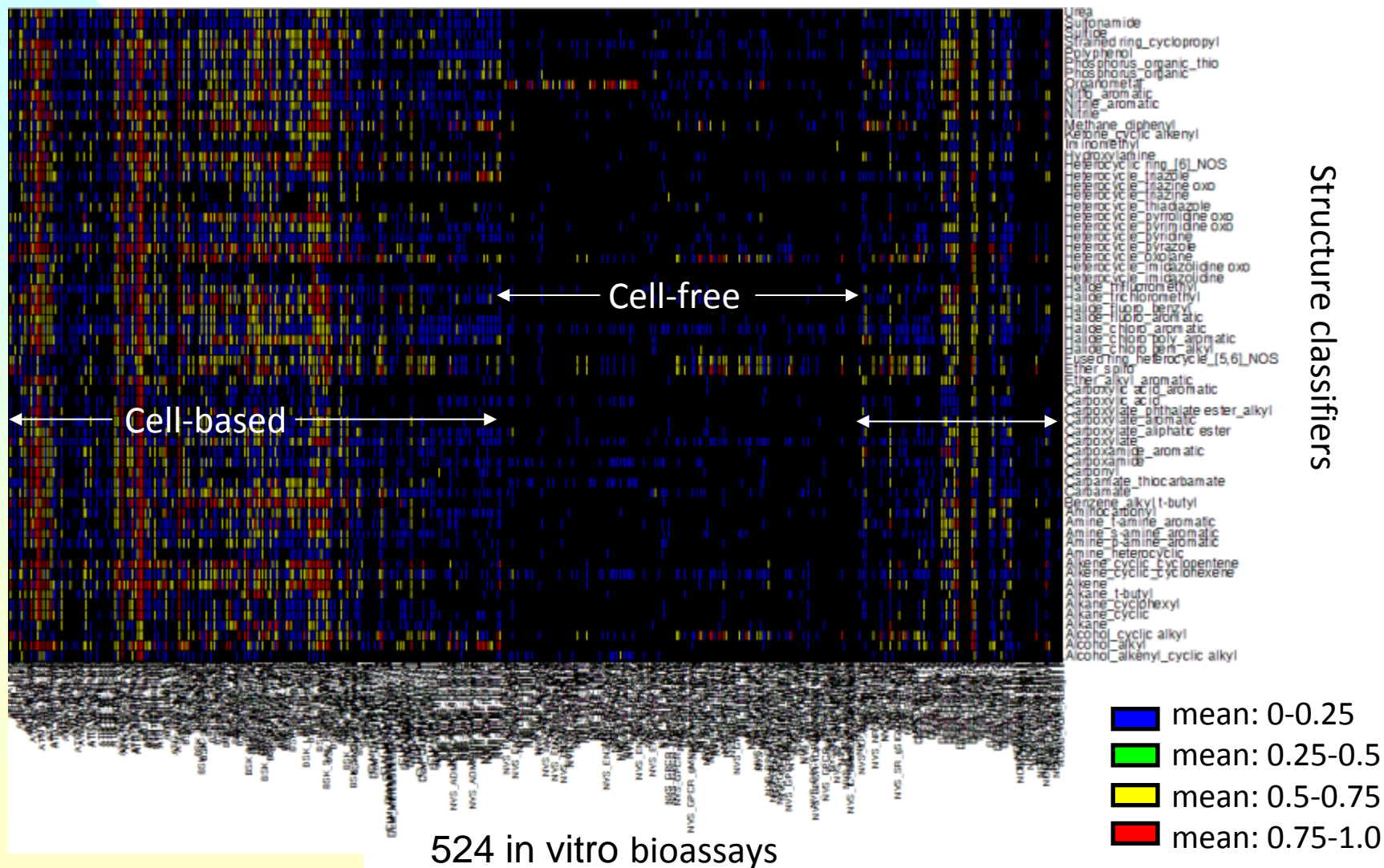
- Structural alerts
- OECD categories
- DSL groups
- EPA chemical classes
- FDA Redbook categories
- **Metabolic & chemical reactivity**
- PhysChem properties
- Calculated descriptors
- Structural keys, features



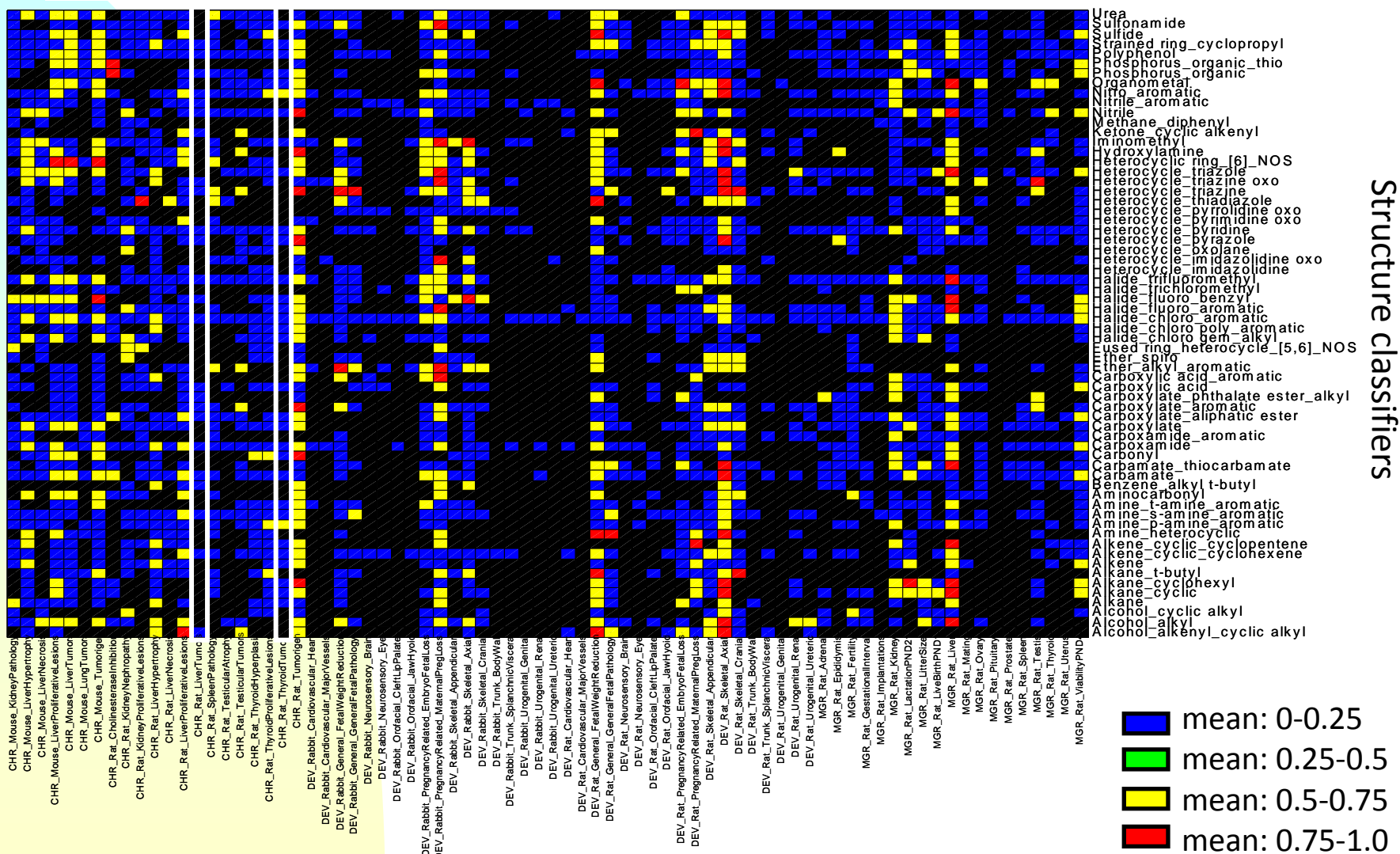
ToxCAST dataset - “Structural Classifiers”

- To expand the structural categories defined in US FDA Redbook
 - To participate in the chemical ontology movement
 - To describe chemicals in FDA and ToxRef databases
- Chains
 - aliphatic, long-alkyl, alkenyl, alkynyl, alkyl_c9:c10_alkenyl...
 - Rings
 - aromatic, carbocyclic, heterocyclic, fused (shapes), strained...
 - Functional groups
 - alcohol, amine, carboxylic acid, halide_alkyl, halide_aromatic...
 - Coordination chemistry
 - chelating ligands, metal environments

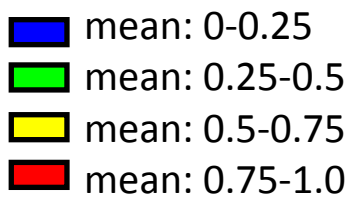
All in vitro assays against selected “Structural Classifiers”



All 76 in vivo assays against the selected “Structural Classifiers”



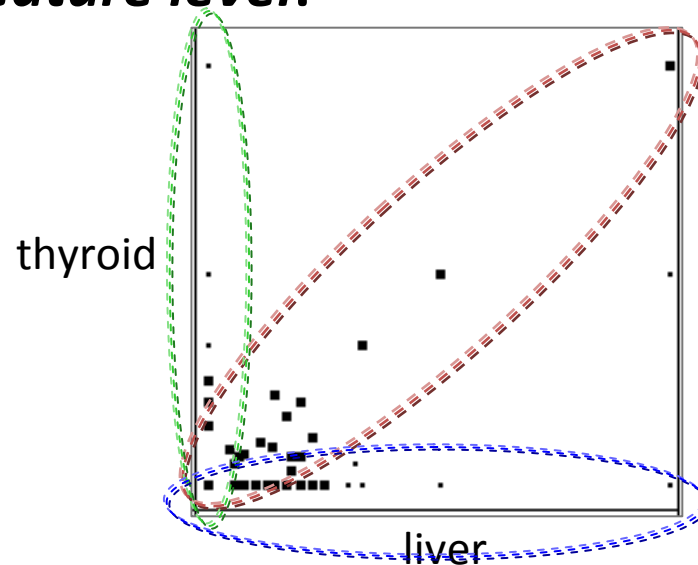
Structure classifiers



Tumorigenic and proliferative lesions – liver and thyroid

- Due to TSH mechanism, thyroid tumors are usually correlated with liver lesions.
- Multivariate plots using structural classes again show the relationship.

Feature level:

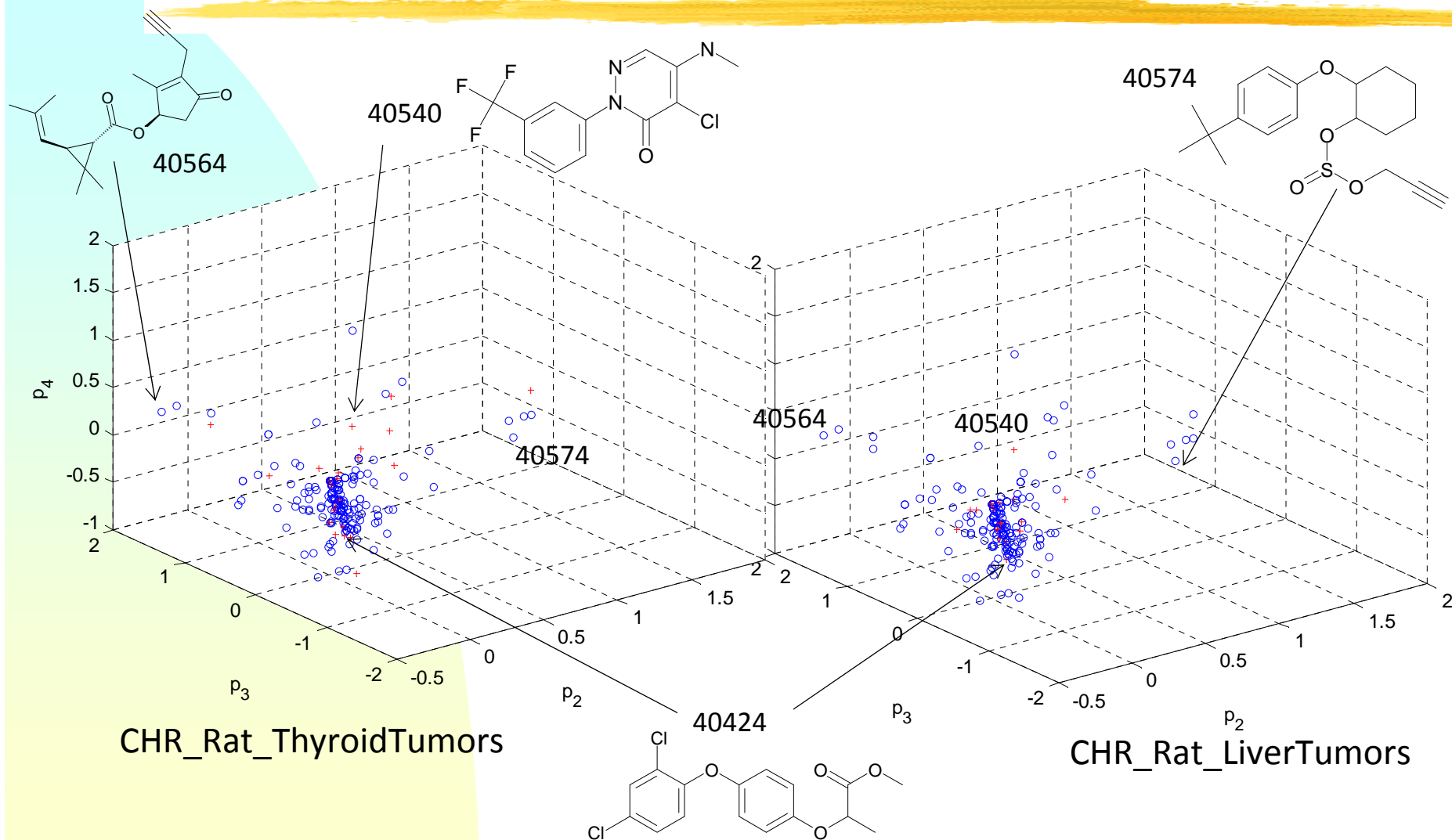


Used all 243
structural classifiers

Compound level:

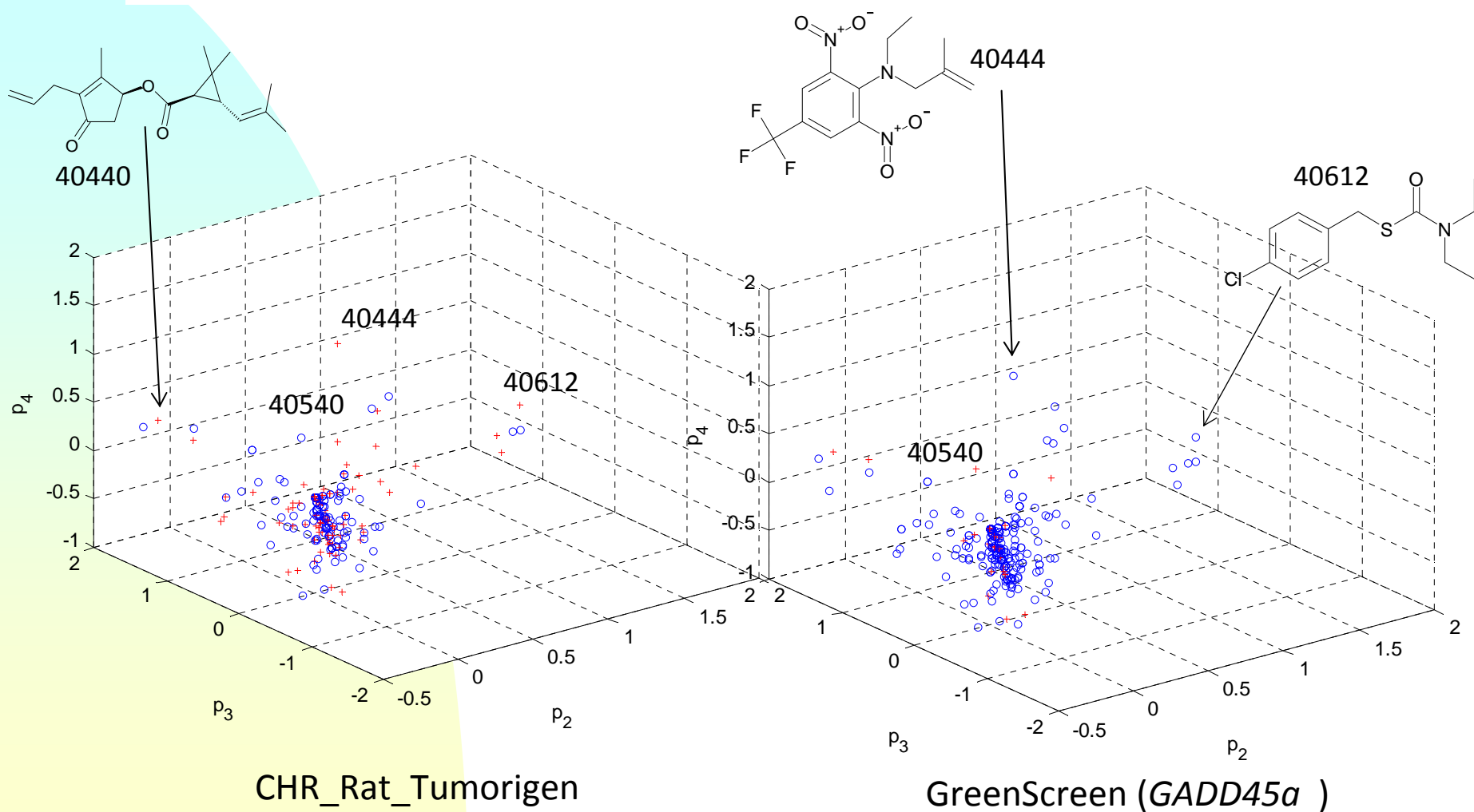
Sensitivity = 39 %
Specificity = 90 %

PC projections of chemicals using the “StructureClassifiers”



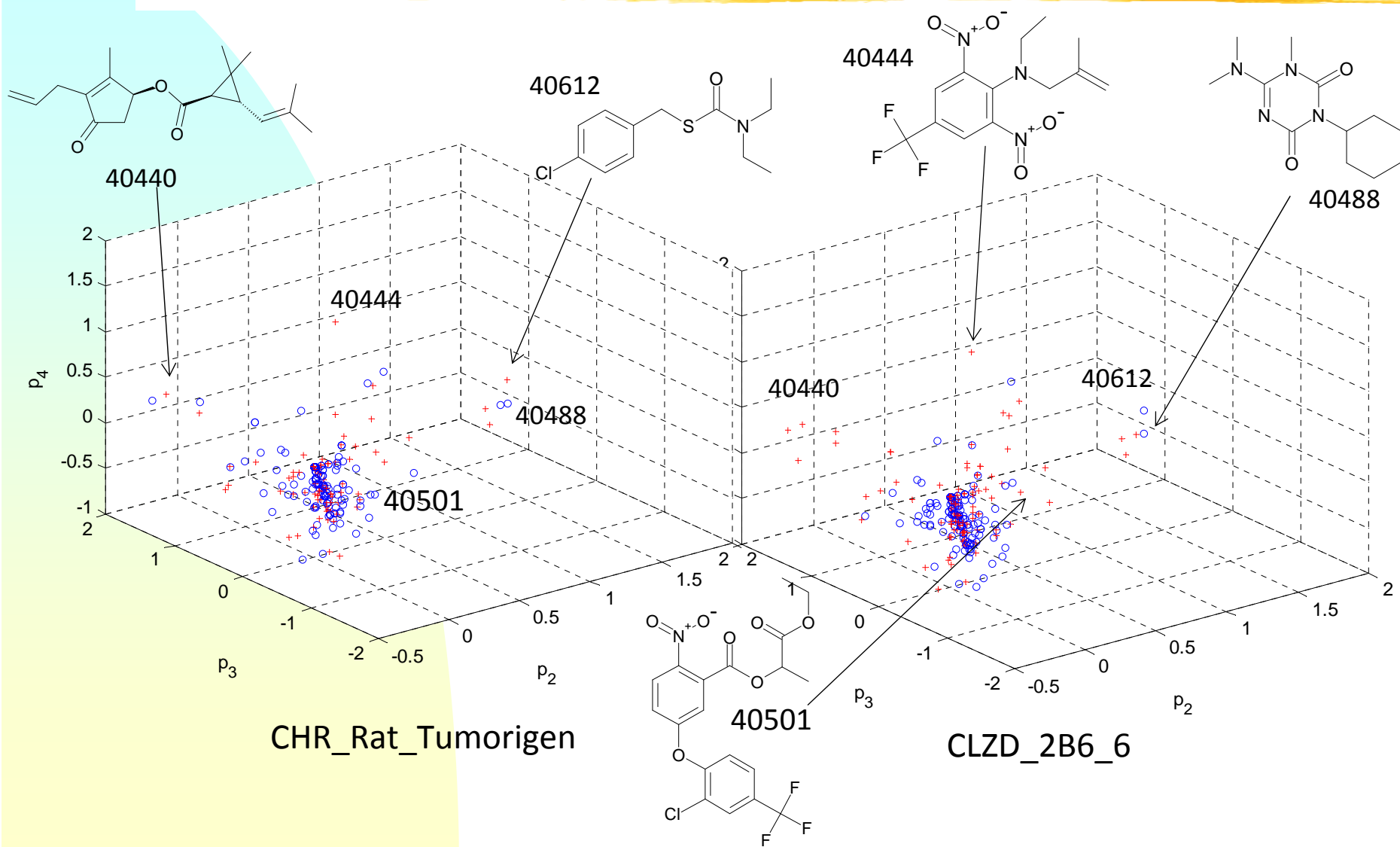
230 chemicals with both liver and thyroid tumors are plotted.

A rodent bioassay vs. an in vitro genetox assay



Andrew Knight, Steve Little et. al, manuscript submitted, 2009

A rodent bioassay vs. an in vitro assay



Summary

- Finding signatures systematically from a variety of in vitro assays relating to in vivo phenotypic effects may be possible.
 - Data mine the 30 significant agreement pairs of in vitro and in vivo assays
 - Data mine the correlation of [AT]

$$[FA]^T [FT] = [AT]$$

[FA]: features vs. in vitro activity

[FT]: feature vs. in vivo activity

From the vantage point of a pragmatist

- Are these predictions better than QSAR models and experts' rules?
 - e.g., a weight of evidence model for rat tumorigens (230)
 - 4 partial logistic regression models based on “structure classifiers” and whole molecule properties are optimized to give a final result.
 - sensitivity:75%; specificity:90%; ROC (true positive/false positive):3.1
- Will the bioassays help build mode-of-action models?
 - e.g., aromatic halides for liver necrosis or pyridines for kidney nephropathy
- How practical is it to use bioassays as predictors?

Integrated Testing Strategies

- How do we apply what we learned from ToxCAST analysis?
 - Our perpetual wish list: better assays and better descriptors
- How do we integrate actual testing to improve predictions during the prioritization cycle?
 - select assays and descriptors
 - link experiment to predictions/prioritizations

Acknowledgment

Collaborators for this paper

- Kirk Arvidson
 - US FDA CFSAN
- Mitchell A. Cheeseman
 - US FDA CFSAN
- Johann Gasteiger
 - Molecular Networks
- Ann Richard
 - US EPA NCCT

Software providers

- Molecular Network
- Leadscope, inc.
- Lhasa, Limited