United States Environmental Protection Agency

Accounting for Uncertainty in the Application of High Throughput Datasets

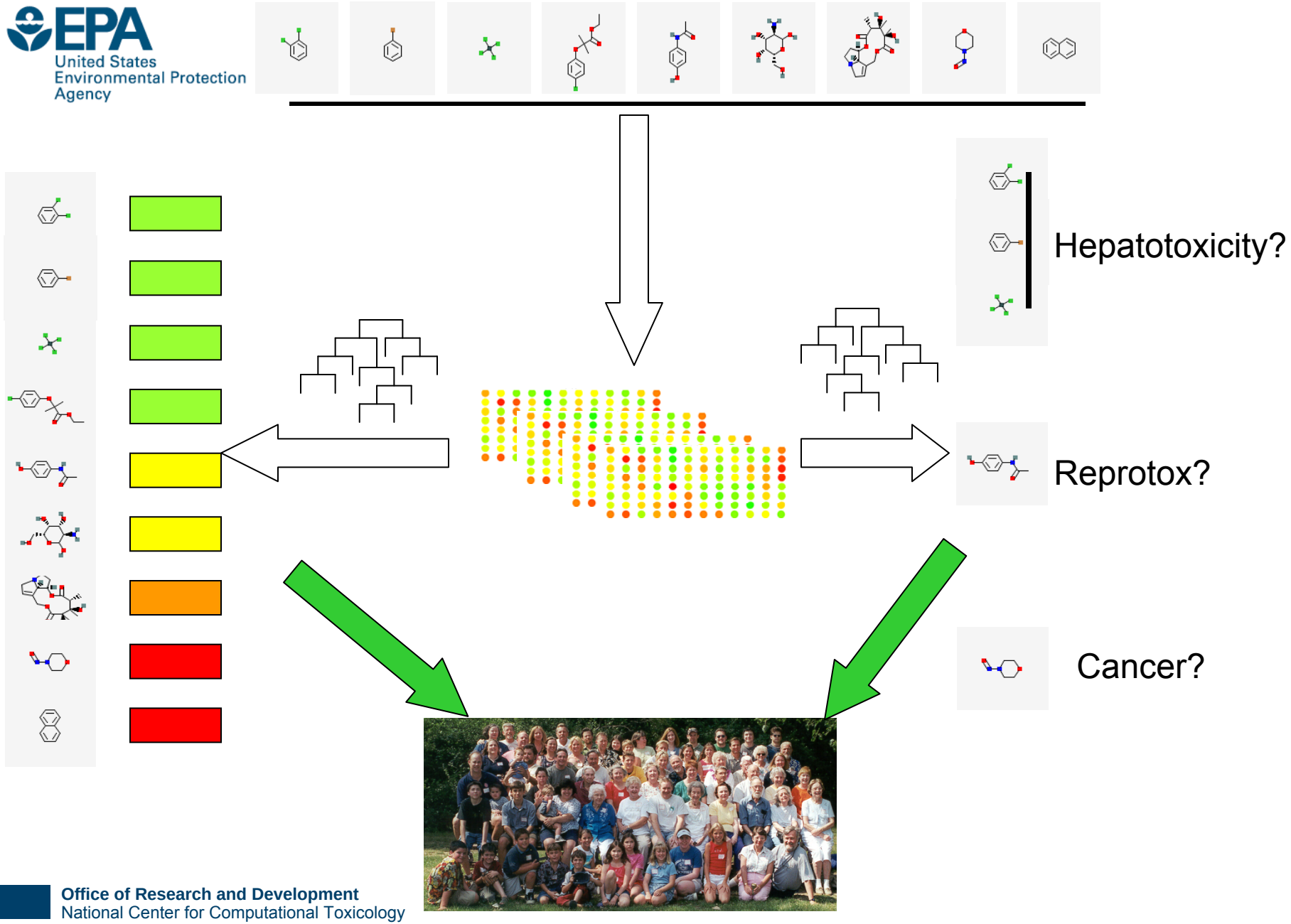
Society of Toxicology Annual Meeting, March 8, 2010

R. Woodrow Setzer

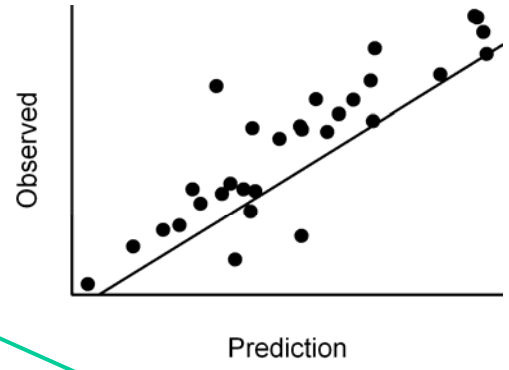
http://www.epa.gov/ncct/toxcast

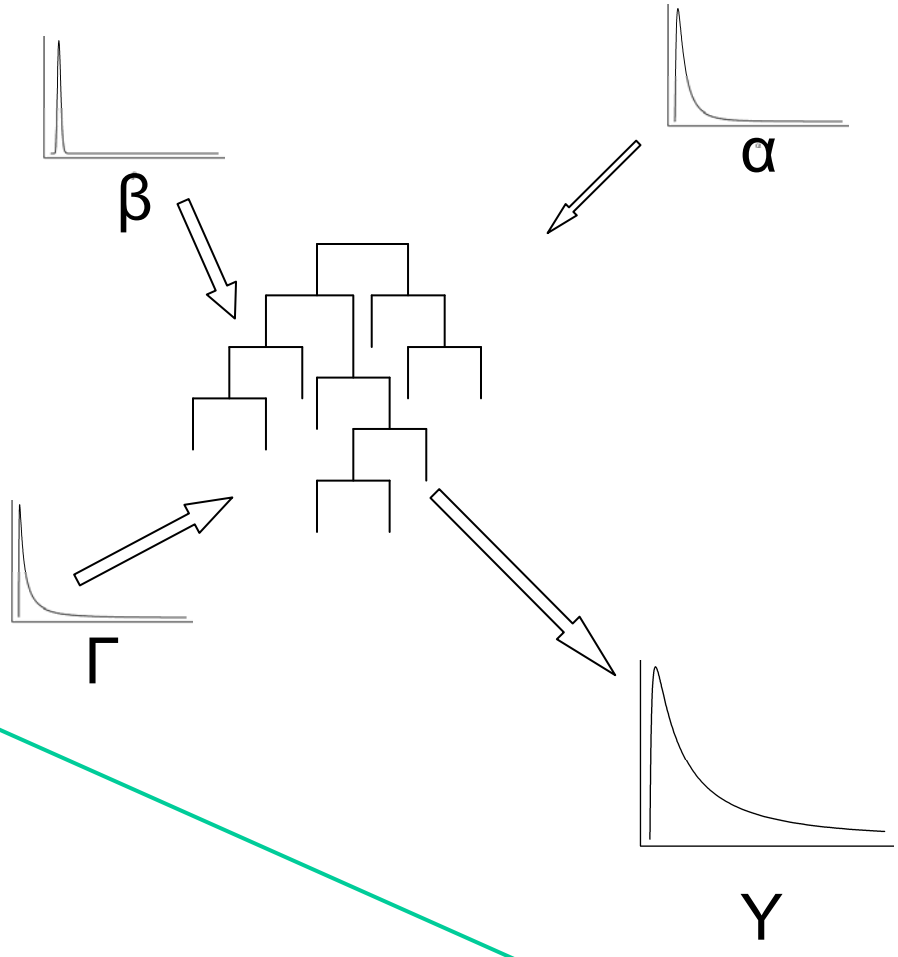
This work was reviewed by EPA and approved for presentation but does not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.

Office of Research and Development
National Center for Computational Toxicology

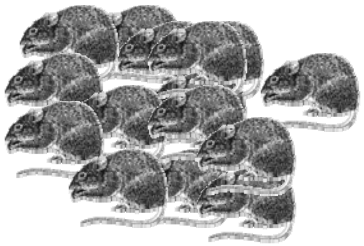

Hepatotoxicity?
Reprotox?
Cancer?

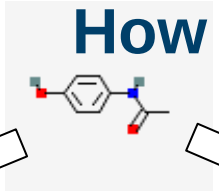
How Do You Quantify the Uncertainty of a Prediction?

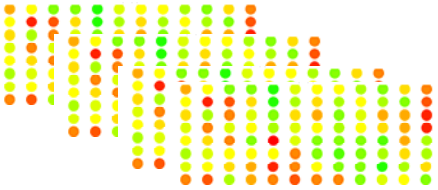


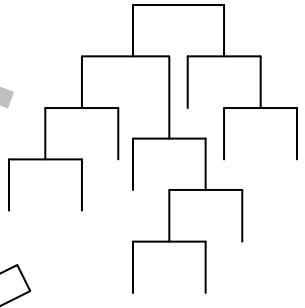


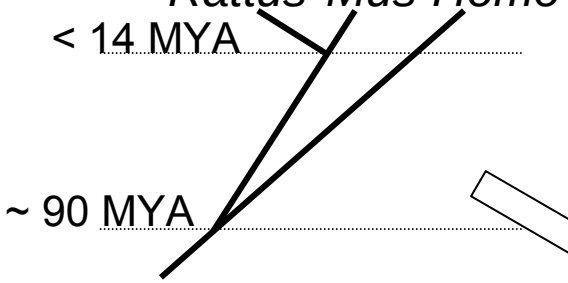
Toxicity Predictions Need Models (Regardless of How They are Done)



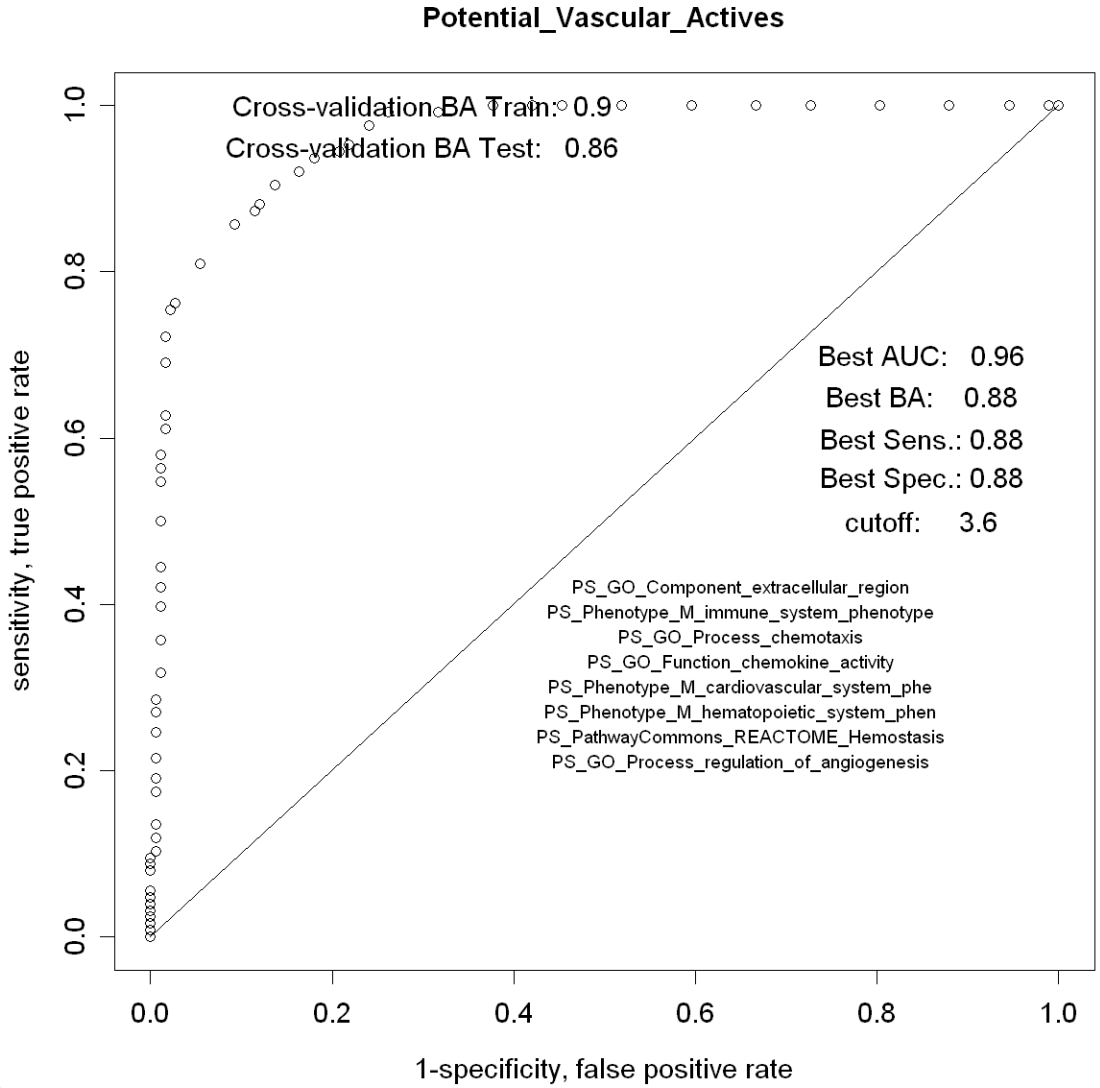




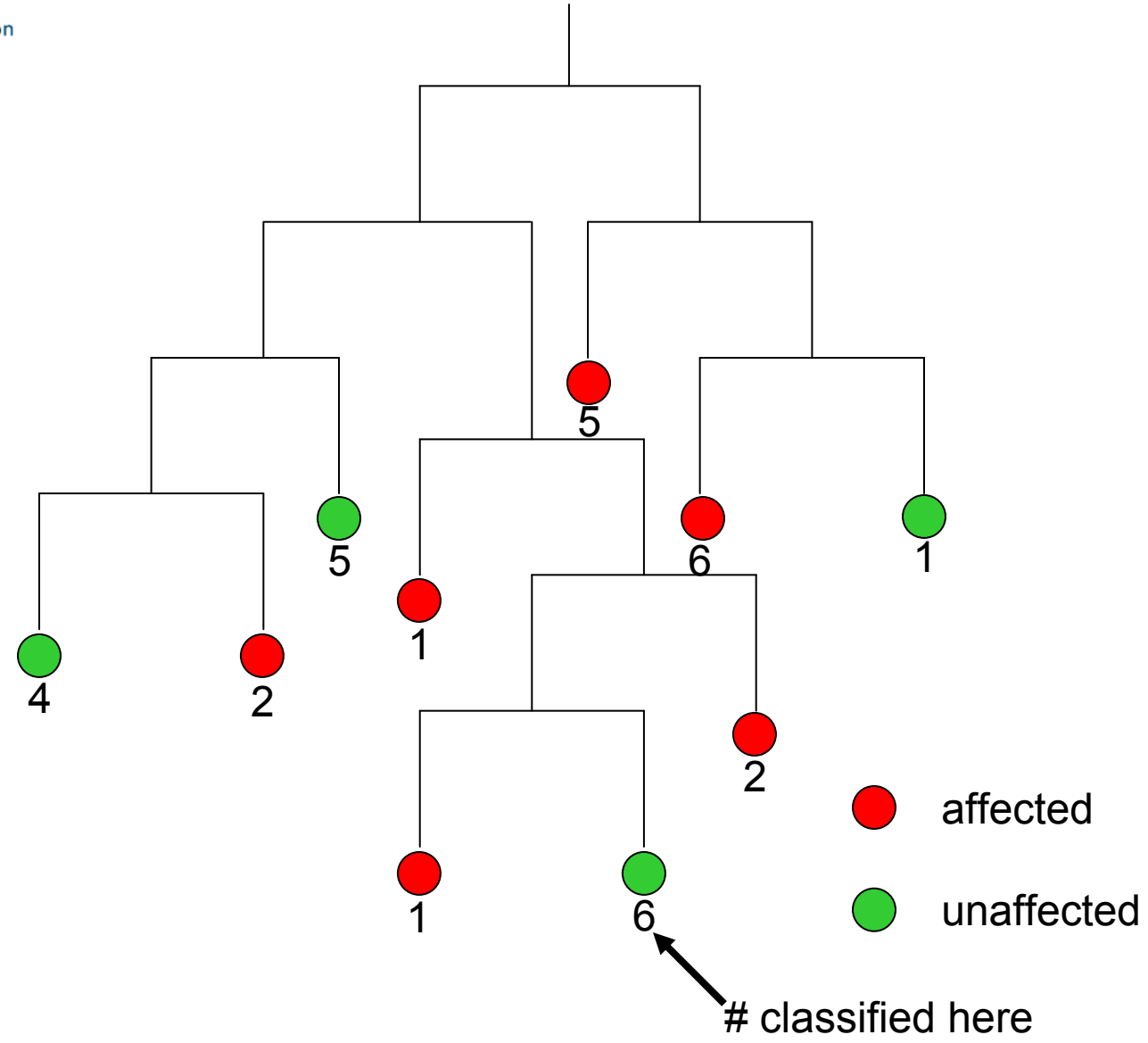


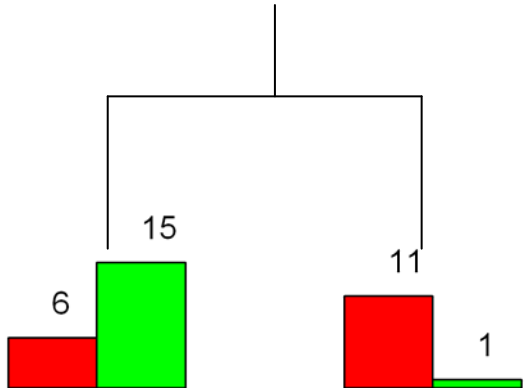


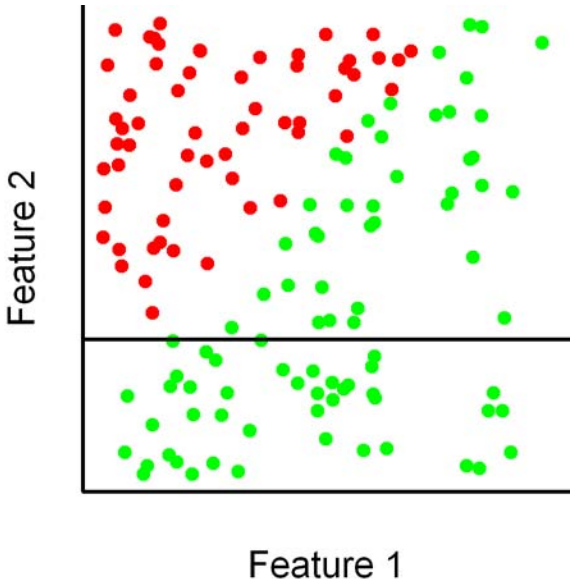
Pathway-level Signature: Toxicity Class Predictor

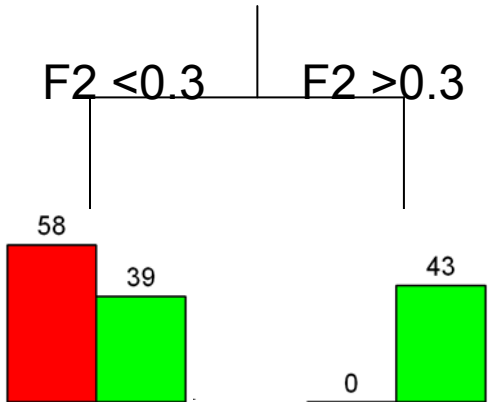


From Kleinstreuer, 2010

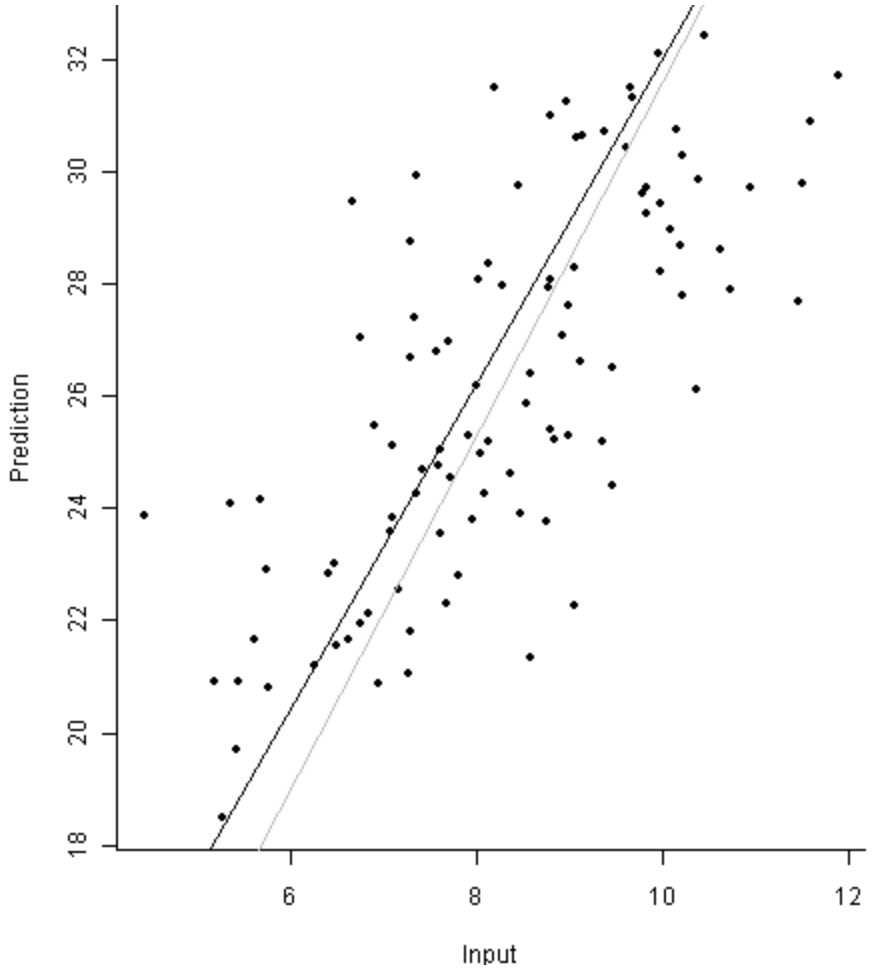

5
5
6
1
1
4
2
2
1
6
affected
unaffected
classified here


15
11
6
1

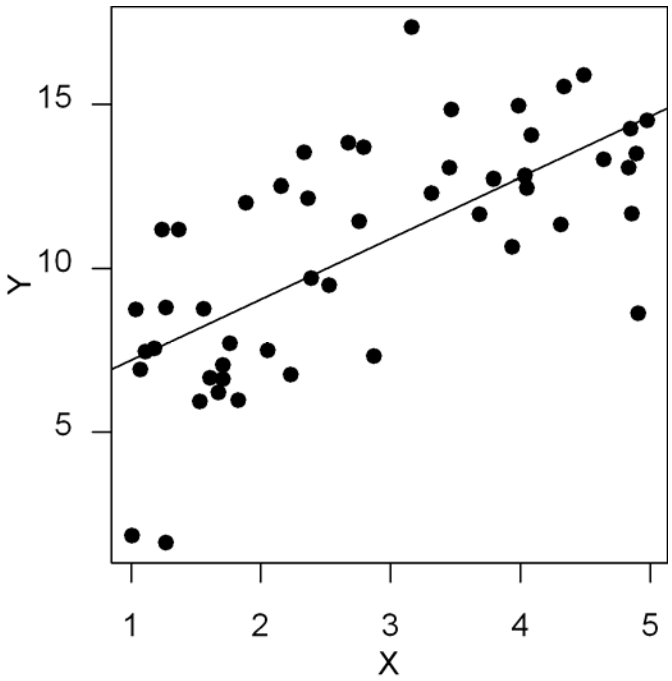

Feature 2
Feature 1

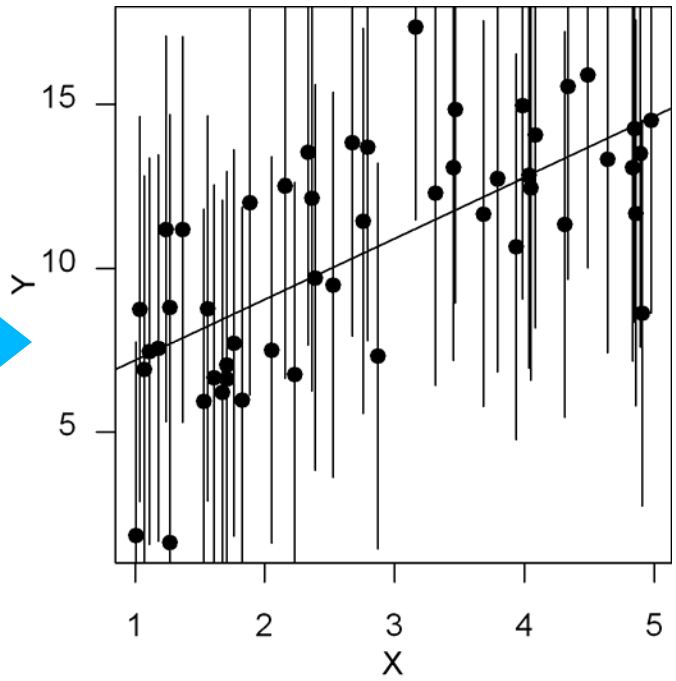

F2 <0.3
F2 >0.3
58
39
0
43

Adjusting for Input Uncertainty



Errors in the Target

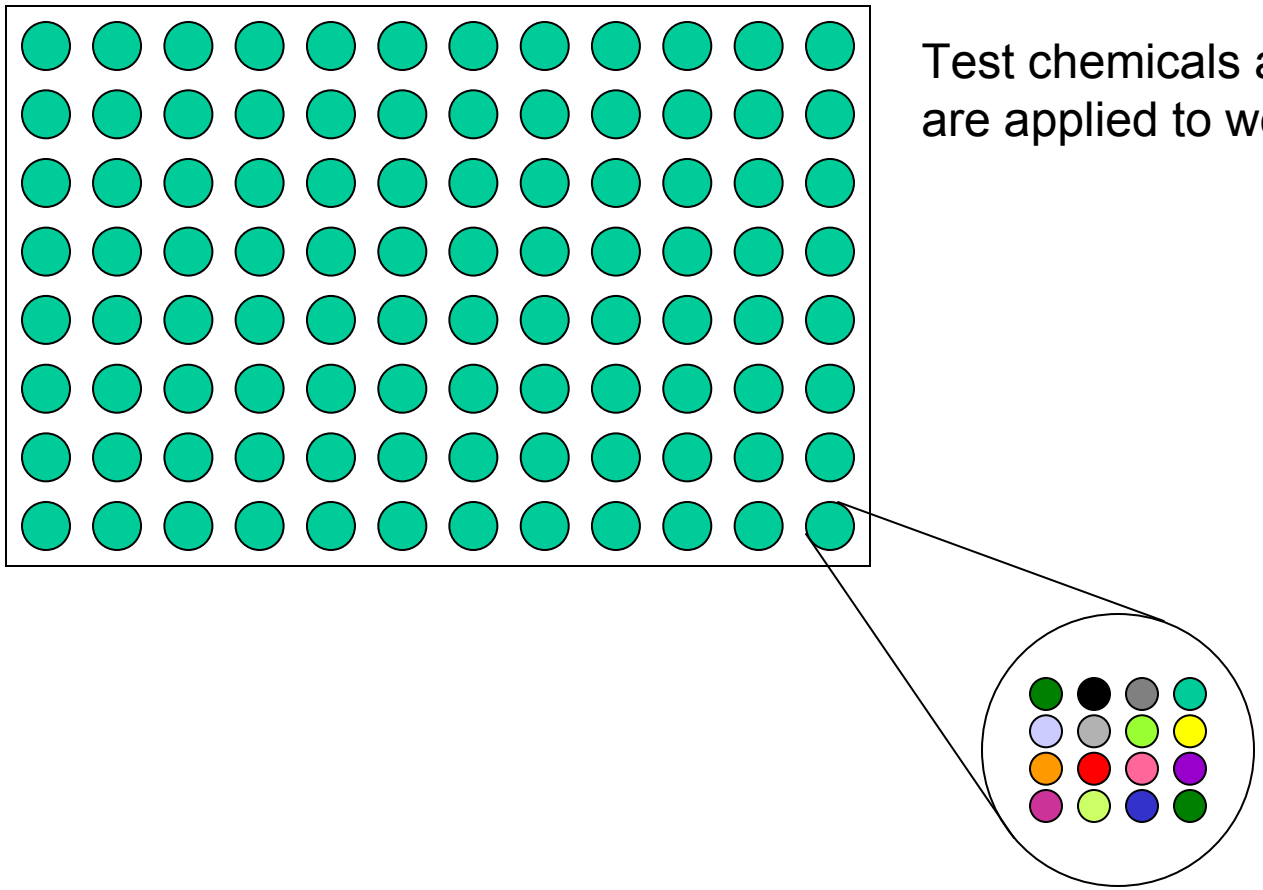




How is our predictor doing?

Maybe not so badly!

Some Statistical Issues in HTS: an Example



Test chemicals and controls are applied to wells

Multiplexed assay, 16 genes/well

United States Environmental Protection Agency

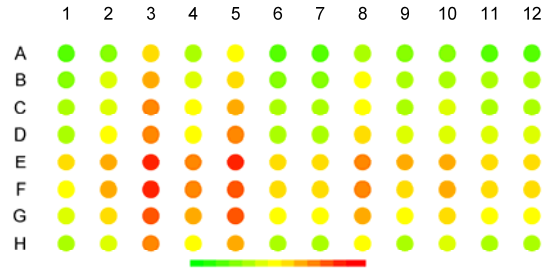
Statistical Models

- Some initial questions stemming from considering spatially arranged data:
  - Are there fixed position effects (persistent from plate to plate)?
  - How large is the variability among plates?
  - After everything else has been accounted for, are there patterns of autocorrelation among wells on the same plate?
  - How well does normalization by the intra-well ‘control’ values work? (Ideally, normalization should remove all the previous effects)
- Statistical methodology
  - Generalized least squares, as implemented in function gls() from package nlme in R
  - Work with log-transformed response – makes distributions of responses more symmetric, and variances more stable.

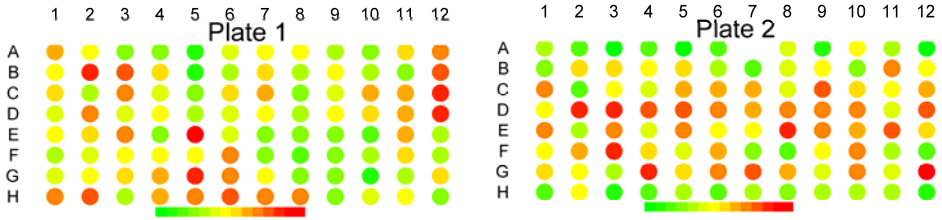
Position Effects, Autocorrelation, and Plate to Plate Variation in Controls

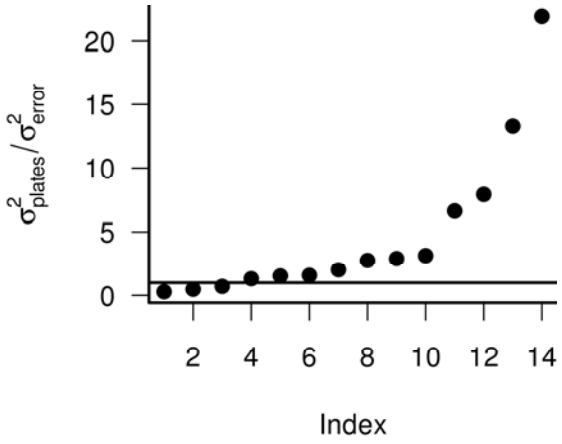
- After normalization: 6 of 14 endpoints have significant row effects and 2 of 14 significant column effects.
- All 14 show significant autocorrelation
- Most have substantial plate to plate variability: for all but 3, inter-plate variance exceeds the error variance.

Position Effects



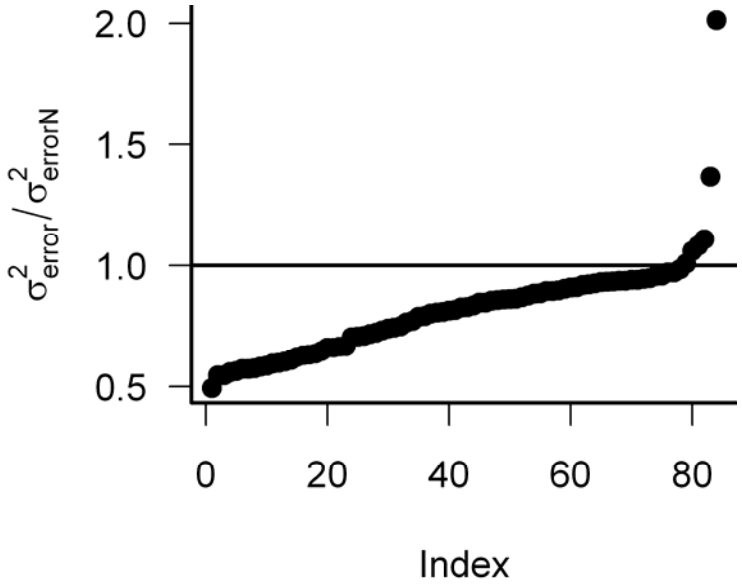
Residuals





Normalization and Sensitivity

- Error variance in normalized data greater than in non-normalized data
- No overall change in ‘sensitivity’, measured by lowest effective dose.



Some Tentative Conclusions

- Just as in agricultural studies, we potentially have to consider effects that stem from location
- The autocorrelation is probably not being modeled very well (and it is not yet clear how important that might be).
- Normalization does not make any of the complicating factors go away.
- Normalization **does** increase the noise (uncertainty) of treatment-specific estimates.

Summary

- Uncertainty in HTS assays considered in the context of their uses in predicting consequences of human exposure.
- Uncertainty in features and targets affect predictor construction differently.
- Predictor construction can be improved with proper quantification of input uncertainty.
- Statistical analysis of HTS data needs to consider the details of the assays, at the level of well and plate, to properly quantify resulting uncertainty.