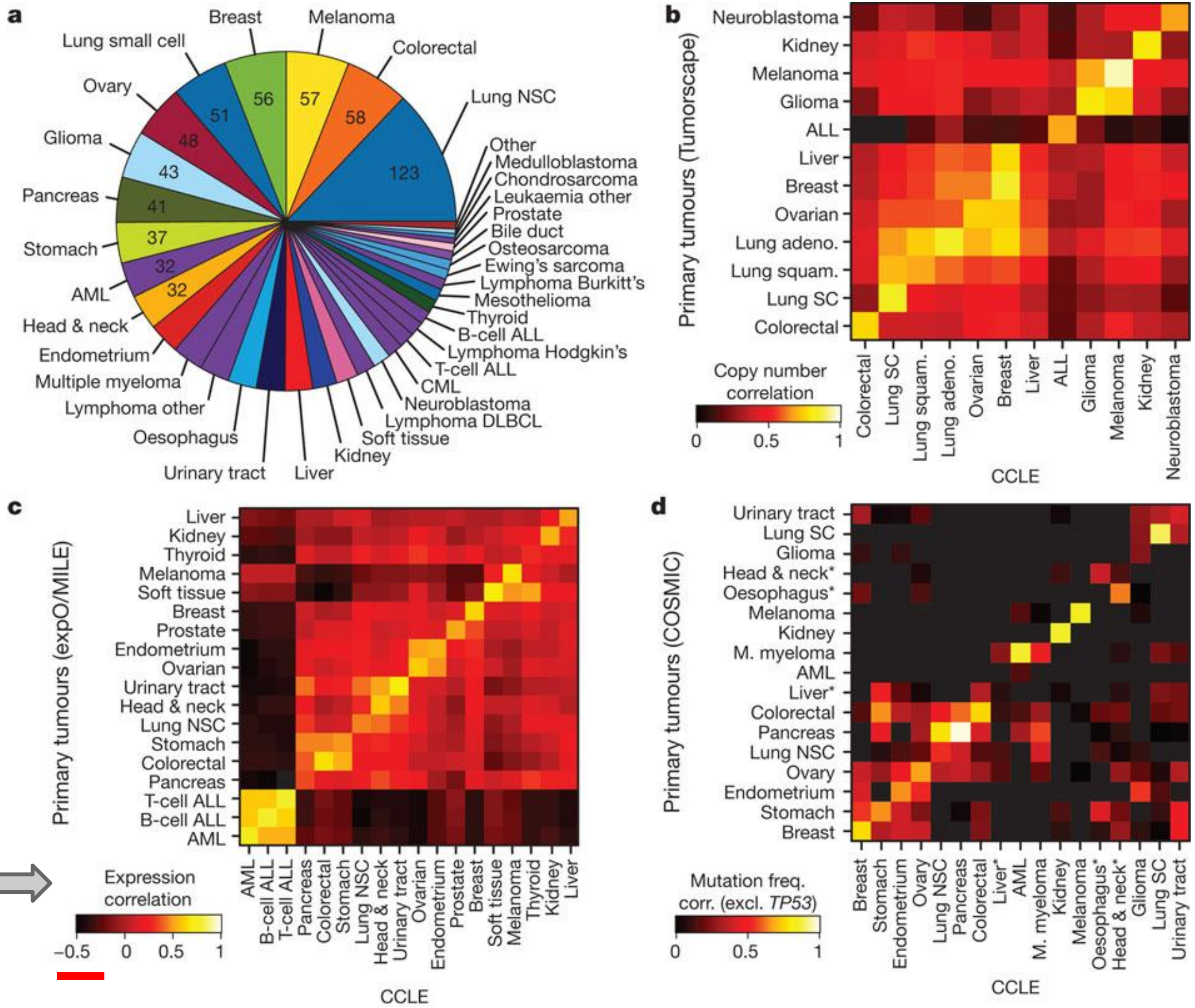
CCLE

Cancer Cell Line Encyclopedia

Data

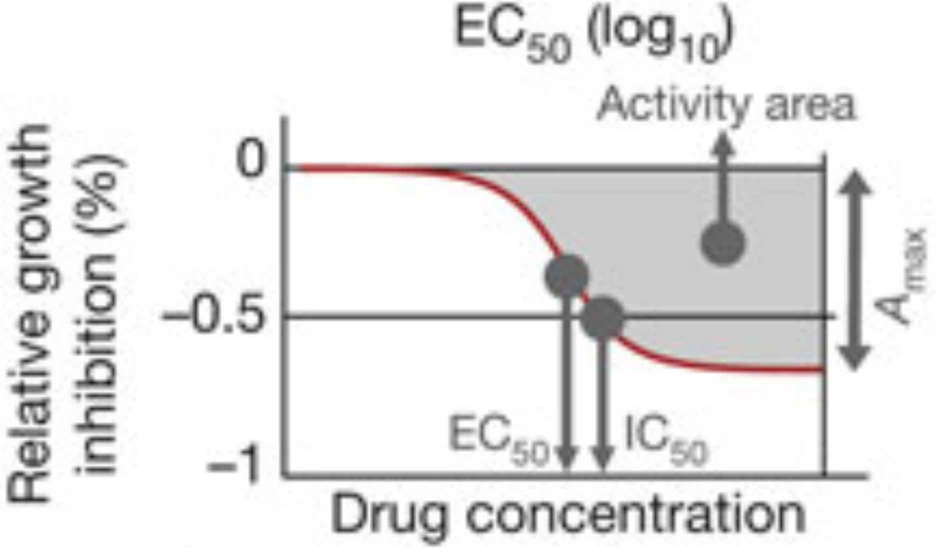
947 cancer cell lines with:

- ≈1600 genes from sequencing screen
- 392 common cancer genes mass spec’d
- mRNA expression
- drug-response profile of 24 compounds



Clustering from Tuesday

Dose-Response Curves

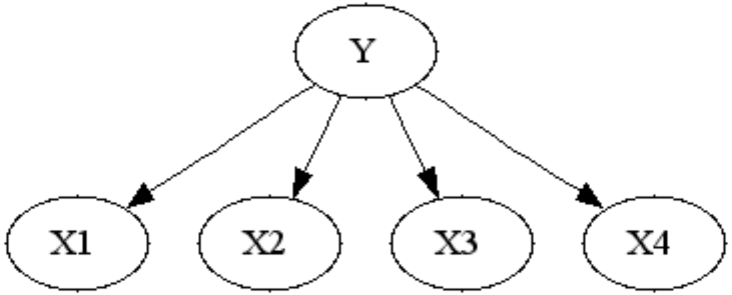


Drug Response Prediction

Input: A feature vector for each gene, (mutation value, GISTIC results, “pathway activity score”)

Output: Predict a vector of dose-response statistics

Naive Bayes



Elastic Net

“It is like a stretchable fishing net that retains ‘all the big fish’”

Uses L1 and L2 regularization

Elastic Net (cont.)

Good when p>>n

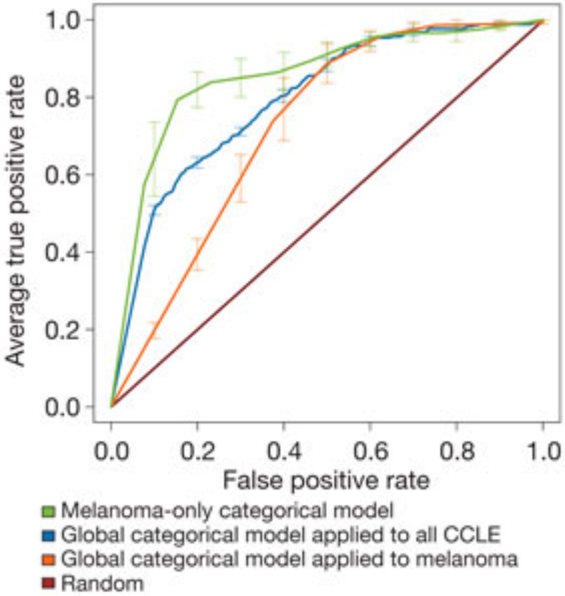
$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1$$

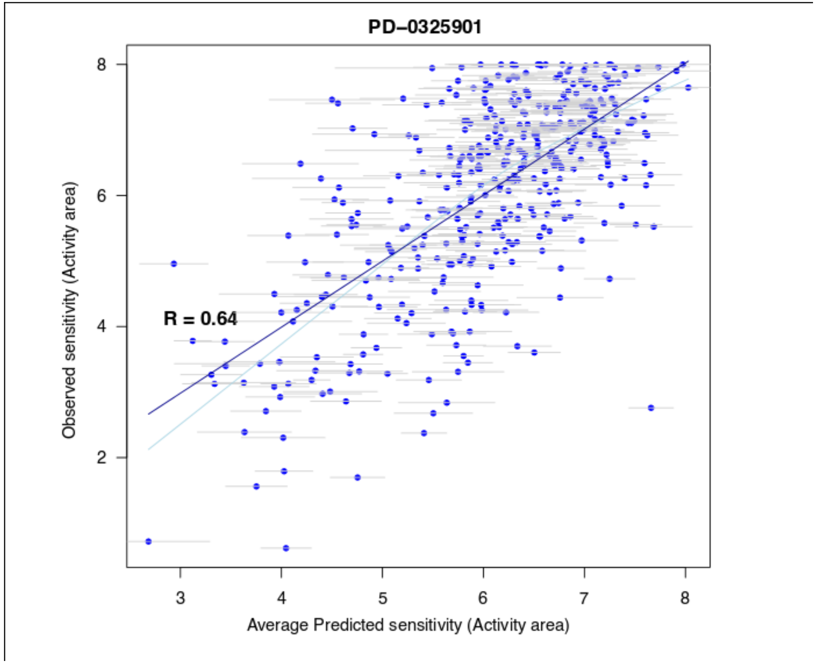
Leukemia classification example

Method	10-fold CV error	Test error	No. of genes
Golub UR	3/38	4/34	50
SVM RFE	2/38	1/34	31
PLR RFE	2/38	1/34	26
NSC	2/38	2/34	21
Elastic Net	2/38	0/34	45

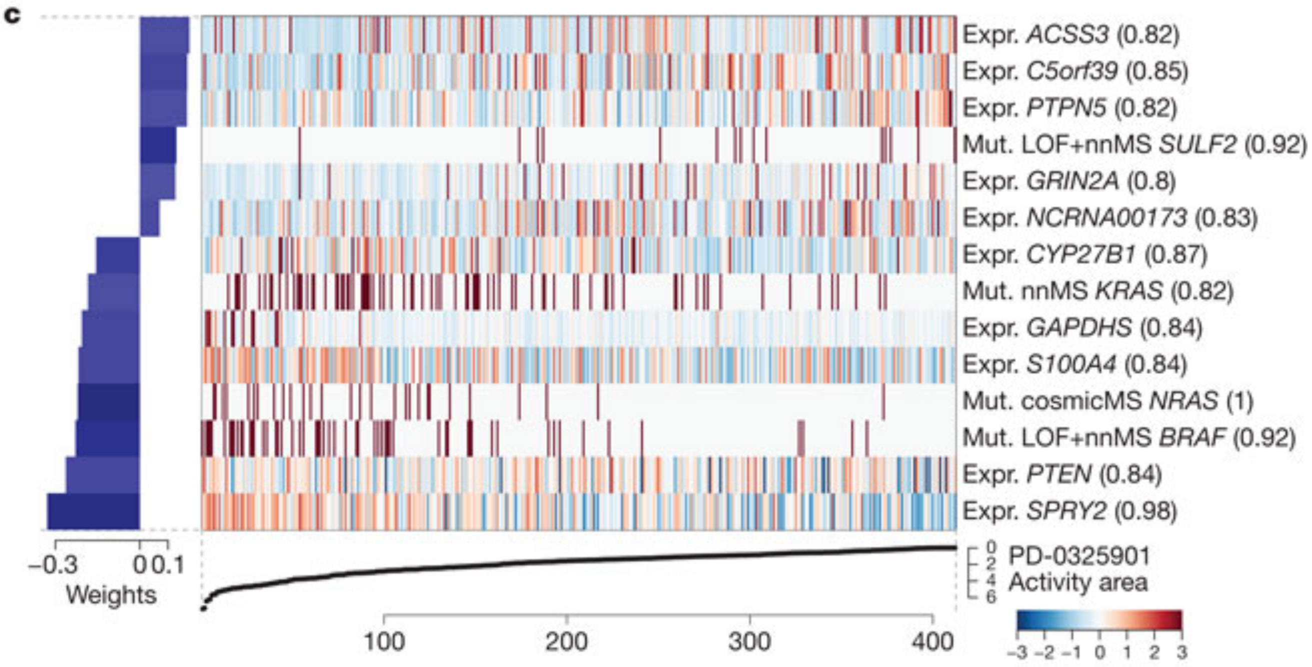
UR: univariate ranking (Golub et al. 1999)
RFE: recursive feature elimination (Guyon et al. 2002)
SVM: support vector machine (Guyon et al. 2002)
PLR: penalized logistic regression (Zhu and Hastie 2004)
NSC: nearest shrunken centroids (Tibshirani et al. 2002)

Results

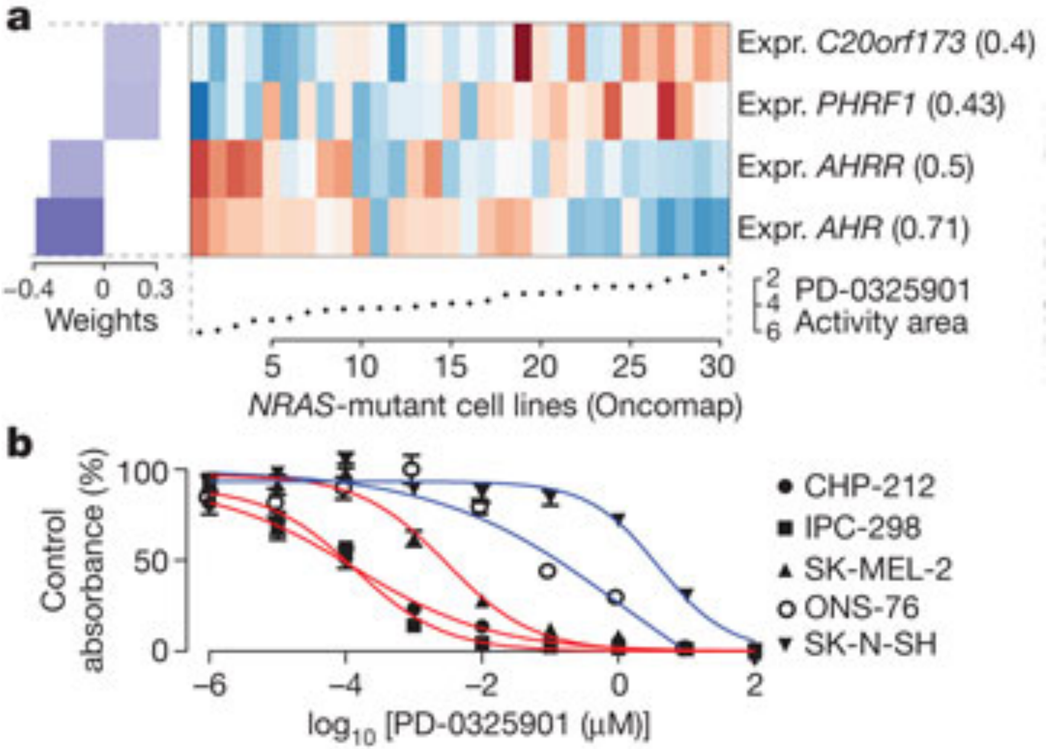

1.0
0.8
0.6
0.4
0.2
0.0
Average true positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
🟩 Melanoma-only categorical model
🟦 Global categorical model applied to all CCLE
🟧 Global categorical model applied to melanoma
🟥 Random


PD–0325901
8
6
4
2
Observed sensitivity (Activity area)
R = 0.64
3
4
5
6
7
8
Average Predicted sensitivity (Activity area)

Results (cont.)



Proof of Concept with AHR



Future Direction for TOP1 inhibitors