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
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 $p \gg n$

\[
\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1
\]
## Leukemia classification example

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

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
Results (cont.)
Proof of Concept with AHR
Future Direction for TOP1 inhibitors

![Graphs and images related to TOP1 inhibitors and their expression and activity.](image-url)