Building patient-specific predictors of drug responses from cell line genomics

Christopher W. Szeto Ph.D., Stephen C. Benz Ph.D., Charles J. Vaske Ph.D., Shahrooz Rabizadeh Ph.D., Patrick Soon-Shiong M.D.

Abstract

Here we demonstrate a method for using cell line genomics and drug response data to build robust therapy outcome predictors. We present here a case study of using this system to predict Dasatinib response in glioblastoma multiforme (GBM) patients. First, we infer pathway-level knowledge of cancer cell lines by integrating multiple genome wide assays into curated pathways (PARADIGM). Next, we use a high-throughput machine-learning library (topmodel) to build, analyze, and rank, thousands of candidate predictive models of drug response. Then, we ensure our most accurate predictive models can be extrapolated to patient samples by using statistical methods to bound our predictions with confidence intervals. Using these methods we identify a predictive, that Dasatinib sensitivity prediction, that is especially predictable using genome-wide assays (i.e., 77% accuracy in cross-validation). Among the datasets used to predict Dasatinib sensitivity, PARADIGM inferred pathway activities are more predictive than other data types. A statically significant proportion of the cell line data that scored most highly sensitive were neural cell lines, suggesting some subset of gliomas may be uniquely responsive to Dasatinib.

We show that there is a proportional number of GBM patients that do conform to the Dasatinib-sensitive profile derived in cell lines.

Datasets & Tools

<table>
<thead>
<tr>
<th>Datasets &amp; Tools</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic datasets</td>
<td>(830 samples)</td>
</tr>
<tr>
<td>Drugs</td>
<td>55</td>
</tr>
<tr>
<td>Classifiers</td>
<td>13</td>
</tr>
<tr>
<td>Prediction models</td>
<td>210</td>
</tr>
<tr>
<td>Patient selection</td>
<td>4</td>
</tr>
</tbody>
</table>

Machine Learning Pipeline

Learning pathway activities

Building response predictors

Results

Drug predictability: Max. avg., and this accuracy gain over the random classifier for each drug, central straight line is avg. accuracy. Drugs to the left are more consistently accurately predicted. The most consistently correctly predicted drug is Dasatinib. X-axis represents the drug expression use consistently the most consistently correctly predicted drug is Dasatinb.

Discussion

Presented here is a rational, data-driven method for stratifying individual patients into responders and non-responders to oncotherapeutics.

This approach:
- uses best-in-class genome-wide assays, pathway analysis, and machine learning techniques in combination
- is not rate-limited by laboriously identifying drug-target interactions biochemically
- is agnostic to tissue of origin; Potential for rational drug reuse in novel contexts
- can recognizing when a prediction challenge is not surmounted by the given data.

We present this approach with a potentially clinically-actionable demonstration: identifying a subset of GBM patients who may respond to Dasatinib. Dasatinib sensitivity in GBM xenografts (in combination with bevacizumab) has been demonstrated by others[2], leading to clinical trials in GBM patients being conducted. At the time of writing 11 clinical trials are being conducted testing Dasatinib response in glioma patients. To the author’s knowledge none of these trials use any biomarkers to recruit or stratify participants.

One such trial recently showed a very small proportion of GBM patients (3/50) has an increased fm PBS3[3].

References