IN THE SPOTLIGHT

Predicting Cancer Drug Response: Advancing the DREAM

Russ B. Altman

Summary: The DREAM challenge is a community effort to assess current capabilities in systems biology. Two recent challenges focus on cancer cell drug sensitivity and drug synergism, and highlight strengths and weaknesses of current approaches. *Cancer Discov; 5(3); 1–3. ©2015 AACR.*

Our ability to make "omics" measurements on groups of cells has exploded in the last 15 years. A particularly appropriate target for these measurements is cancer cells, where we can measure changes within the genome, transcriptome, and proteome. The knowledge gained from these measurements can be divided into (1) fundamental mechanistic insights into the biology of cancer, and (2) use of these data for practical applications, such as the selection and optimization of cancer treatments. The analyses of these data can also take two general tacks: (1) the data can be used to inform mechanistic models of cellular physiology that capture the structure and interactions of cellular processes, or (2) they can also be analyzed with machine learning methods to create associative or correlative predictions. In general, the mechanistic models achieve high prediction performance only when the molecular players and their relationships are comprehensively captured with high precision (1). On the other hand, the machine learning methods can sometimes achieve high prediction performance from the data with mathematical models that have unclear relationships to the underlying biological processes (Fig. 1). Recently, Costello and colleagues (2) reported a community-based effort to predict the sensitivity to 28 drugs of 18 cell lines from a multiomics data set; several participants achieved high predictive performance with diverse machine learning techniques. As a follow-on project, Bansal and colleagues (3) report another community effort to predict the activity of 91 pairs of compounds against a single cell line, with more modest results: Only three groups (of 31) had performance better than random. What are we to make of these findings, and is this the future of drug discovery and optimization?

The Costello article was focused on predicting the drug sensitivity of human breast cancer cell lines. The participants had access to genomic, epigenomic, and proteomic profiling data sets; they had a "training" set of 53 samples, including the response to 28 compounds for many of these samples, upon which to create their predictive methods. They then applied these methods to 18 additional cell lines for which they were provided the profiling data sets, but for which the sensitivities were unpublished and not known to them. The more ambitious Bansal article focused on predicting the degree of synergy (or antagonism) of 91 pairs of compounds in killing a human diffuse large B-cell lymphoma line. The participants were provided dose response curves for the individual 14 compounds, and gene expression and genetic profiles of the B-cell lines. They then created methods to predict the degree to which the two drugs work together: Synergism implies that the two drugs kill more than would be expected by adding their individual activity; antagonism implies that the two drugs kill less effectively than expected and in some way interfere with each other. Although some groups performed better than random guessing, the performance showed the need for significant improvement in these methods, and was not as reassuring as the results presented in the first article.

Both these papers are products of the Dialogue on Reverse Engineering Assessment and Methods (DREAM), a community of data scientists who set up, run, and evaluate challenge problems aimed at assessing current capabilities and shortcomings of systems biological analytic methods. Participants analyze a common data set and submit predictions blinded to the actual results, which are used by independent assessors to evaluate the predictions. The use of challenge problems to focus community attention on relatively unbiased computational method assessment has been used in three-dimensional (3D) structure prediction (4), protein function assessment (5), and other fields. Proponents argue that direct comparison and evaluation of methods on blinded data sets is the best way to assess value, and that it focuses a community on creating methods that have similar performance criteria. Critics argue that large amounts of intellectual capital are spent focusing on a small set of problems, leading to homogenization of the methods, decreased creativity, and a hegemony of the initially superior approaches.

To data scientists with affection for machine learning, these articles present a playground of technical lessons and issues. The designs of the individual algorithms are fascinating. Indeed, the design of a fair evaluation scheme is fascinating. The best-performing method in the Costello article (a group from Aalto University and the University of Helsinki) uses a Bayesian framework and several clever machine learning approaches to smooth and weight the data. One of the best-performing methods in the Bansal article (a group from the University of Texas) employs a model based on the similarity and subsequently complementary of the transcriptional changes induced by the individual drugs, to

Departments of Bioengineering and Genetics, Stanford University, Stanford, California

Corresponding Author: Russ B. Altman, Stanford University, Departments of Bioengineering and Genetics, 443 Via Ortega, Shriram Center, Room 309, MC: 4245, Stanford, CA 94305-4145. E-mail: russ.altman@stanford.edu
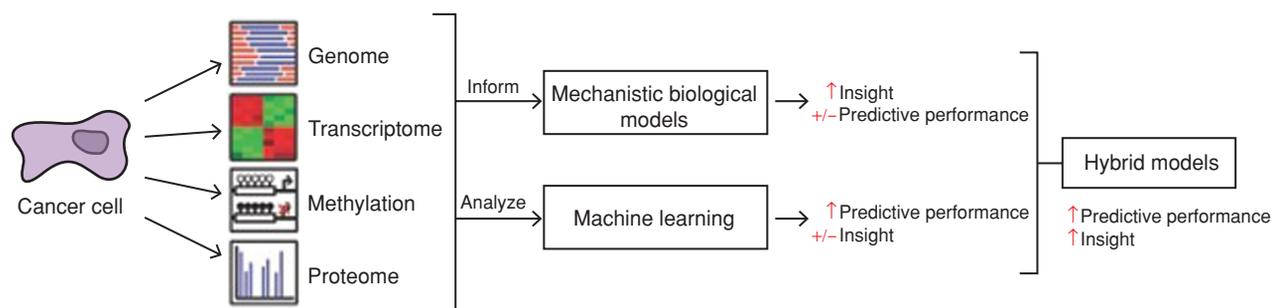
Figure 1. In two DREAM challenges, high throughput data characterizing cancer cells are used to build predictive models. Mechanistic models provide insight into the underlying biology, but do not take full advantage of the information within the data to achieve high performance. Machine learning methods are associative and extract maximum predictive value from the data, but do not always provide insight about mechanism. The future may bring hybrid models that combine the best of both approaches.

estimate their synergy—it is very good at predicting antagonistic (non-synergistic or "whole less than the sum of parts") relationships.

In many ways, however, the lessons from the Costello challenge for drug sensitivity explain the great difficulty in achieving strong performance in the Bansal study. Costello provides some clear lessons:

1. Integrating several data sources is advantageous: RNA expression data sets were complemented by other data sets (such as exome sequencing and methylation determinations). In addition, it generally helps to include external sources of knowledge, such as biological pathway information and other curated biological knowledge.
2. Providing a set of reference examples is critical: The sensitivity results for the 28 compounds to 35 cell lines were provided to allow for "training" of the systems before subjecting them to the prediction task on 18 other cell lines.
3. Machine learning methods that represented nonlinear relationships between variables performed best: Linear combinations of measured features are not sufficient, and (somewhat unsurprisingly) biological systems have substantial nonlinearities that need to be captured for good predictive performance.

The Bansal challenge was considerably more difficult: Participants received only RNA expression response data from a single cell line. They did not receive training data, and thus had to create at least a skeletal biological model of drug synergism. They received the inhibitory concentration of the drugs at two time points only, making complex nonlinear relationships difficult to extract.

Nonetheless, these challenges give us reason to be optimistic that predictive methods may provide powerful abilities to select drugs and drug combinations in the future. Of course, we must overcome several limitations. These challenges were performed on cell lines, and there will certainly be differences in the pharmacokinetics of these drugs within patients, who will absorb, distribute, metabolize, and excrete the compounds at different rates. The issue of dose will be paramount here, and we have only fragmentary information about dose in these studies. In addition, there will likely be pharmacodynamic differences given the heterogeneity of cancer cell populations and their adaptation over time to the selective pressure of toxic agents.

For the field of omics-based prediction, the fundamental challenge is to develop ways to marry the great predictive performance of modern machine learning methods with the specificity and insight gained from mechanistically based models. The Costello challenge provided a high ratio of training data to test data (complete data for 35 cell lines, predictions for 18), and thus it is not surprising that associative methods performed well. The Bansal challenge required the participants to construct a model for how synergy works biologically, before applying their machine learning methods—and that decision was critical because there was no "ground truth" training data provided. As we learn the complex processes that lead to biological phenotypes, a challenge going forward is to marry these models with our rich data sources to build predictors with high performance and high relevance to patient populations.

Disclosure of Potential Conflicts of Interest

R.B. Altman reports receiving a commercial research grant from Pfizer and has ownership interest in Personalis Inc.

Disclaimer

The author's laboratory is part of the broad DREAM community of investigators, and is listed as such on both papers. The laboratory submitted predictions to the challenge but was not involved in the evaluation.

REFERENCES

1. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, et al. A whole-cell computational model predicts phenotype from genotype. Cell 2012;150:389–401.

2. Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol 2014;32:1202–12.

3. Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, et al. A community computational challenge to predict the activity of pairs of compounds. Nat Biotechnol 2014;32:1213–22.

4. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)–round x. Proteins 2014;82(Suppl 2):1–6.

5. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. Nat Methods 2013;10:221–7.