A Roadmap for Developing and Utilizing Therapeutically Relevant Genomic Classifiers

Richard Simon

National Cancer Institute

http://linus.nci.nih.gov/brb

http://linus.nci.nih.gov/brb

- http://linus.nci.nih.gov/brb
  - Powerpoint presentation
  - Reprints & Technical Reports
  - BRB-ArrayTools software
- *Design and Analysis of DNA Microarray Investigations*
  - R Simon, EL Korn, MD Radmacher, L McShane, G Wright, Y Zhao. Springer (2003)

Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the analysis of DNA microarray data: Class prediction methods. Journal of the National Cancer Institute 95:14-18, 2003.

Simon R. Using DNA microarrays for diagnostic and prognostic prediction. Expert Review of Molecular Diagnostics, 3(5) 587-595, 2003.

Simon R. Diagnostic and prognostic prediction using gene expression profiles in high dimensional microarray data. British Journal of Cancer 89:1599-1604, 2003.

Simon R, Supervised analysis when the number of candidate features (p) greatly exceeds the number of cases (n). Association for Computing Machinery SIGKDD Explorations 5:2:31-36, 2003.

Dobbin K and Simon R. Sample size determination in microarray experiments for class comparison and prognostic classification. Biostatistics 6:27-38, 2005.

Simon R and Maitnourim A. Evaluating the efficiency of targeted designs for randomized clinical trials. Clinical Cancer Research 10:6759-63, 2004..

Maitnourim A and Simon R. On the efficiency of targeted clinical trials. Statistics in Medicine 24:329-339, 2005.

Simon R. When is a genomic classifier ready for prime time? Nature Clinical Practice – Oncology 1:4-5, 2004.

Dobbin K, Beer DG, Meyerson M, et al. Inter-laboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clinical Cancer Research 11:565-572, 2005.

Simon R. An agenda for Clinical Trials: clinical trials in the genomic era. Clinical Trials 1:468-470, 2004.

Simon R. Development and Validation of Therapeutically Relevant Multi-gene Biomarker Classifiers. Journal of the National Cancer Institute 97:866-867, 2005.

Simon R. A roadmap for developing and validating therapeutically relevant genomic classifiers. Journal of Clinical Oncology (In Press).

“If new refrigerators hurt 7% of customers and failed to work for another one-third of them, customers would expect refunds.”

BJ Evans, DA Flockhart, EM Meslin Nature Med 10:1289, 2004

- Clinical trial for patients with breast cancer, without nodal or distant metastases, Estrogen receptor positive tumor
  - 5 year survival rate for control group (surgery + radiation + Tamoxifen) expected to be 90%
  - Size trial to detect 92% survival in group treated with control modalities plus chemotherapy

Using Genomics in Development of a New Therapeutic

- Develop a pharmacogenomic classifier
- Completely specify the classifier, translating platforms if necessary
- Establish reproducibility of measurement of the classifier
- Use the completely specified classifier to obtain definitive results about effectiveness of a new treatment in a population of patients defined by the classifier
  - Use of genomics in a hypothesis testing framework
  - Avoid endless exploratory analyses that never result in reliable results

After Developing the Classifier Comes

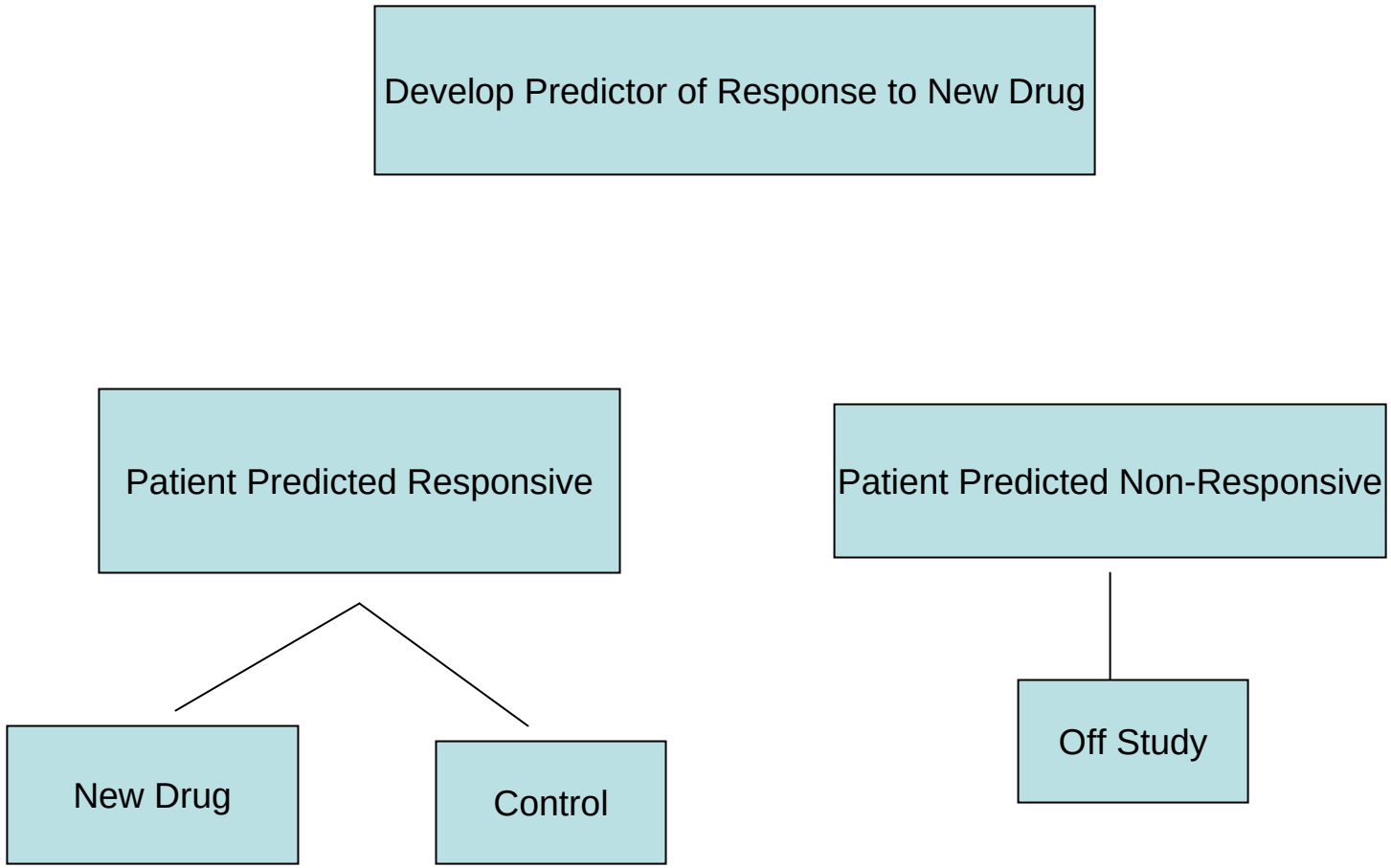
- Validation of the classifier?
- Use of the classifier to focus evaluation of the new treatment?

Biomarker validation vs pharmacogenomic classifier utilization

- Adoption of a pharmacogenomic classifier to restrict the use of a treatment in wide use should be based on adequate validation of the classifier
  - Validation means demonstrating that the classifier leads to better clinical outcome
- In new drug development, the role of a classifier is to select a target population for treatment
  - The focus should be on evaluating the new drug, not on validating the classifier
  - Whether one must establish that the new treatment is not effective for the classifier negative patients is unclear

Using Genomics in Development of a New Therapeutic (I)

- Develop a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Use the diagnostic as eligibility criteria in a prospectively planned evaluation of the new drug
- Demonstrate that the new drug is effective in a prospectively defined set of patients determined by the diagnostic
- Demonstrate that the diagnostic can be reproducibly measured
- Confirmatory phase III trial


Develop Predictor of Response to New Drug
Patient Predicted Responsive
Patient Predicted Non-Responsive
New Drug
Control
Off Study

Development of PG Classifier of Tumor Sensitivity to Drug

- Can immensely improve the efficiency of phase III trials
  - Select patients predicted to be most sensitive
- Enables patients to be treated with drugs that actually work for them
- Avoids false negative trials for heterogeneous populations
- Avoids erroneous generalizations of conclusions from positive trials

Tumors of a given primary site are often heterogeneous with regard to oncogenesis. These tumors may represent different diseases

- “Hypertension is not one single entity, neither is schizophrenia. It is likely that we will find 10 if we are lucky, or 50, if we are not very lucky, different disorders masquerading under the umbrella of hypertension. I don’t see how once we have that knowledge, we are not going to use it to genotype individuals and try to tailor therapies, because if they are that different, then they’re likely fundamentally … different problems…”
  - George Poste

Randomized Clinical Trials Targeted to Patients Predicted to be Responsive to the New Treatment Can Be Much More Efficient than Traditional Untargeted Designs

- Simon R and Maitnourim A. Evaluating the efficiency of targeted designs for randomized clinical trials. Clinical Cancer Research 10:6759-63, 2004.
- Maitnourim A and Simon R. On the efficiency of targeted clinical trials. Statistics in Medicine 24:329-339, 2005.
- reprints at http://linus.nci.nih.gov/brb

Two Clinical Trial Designs

- Un-targeted design
  - Randomized comparison of E to C without screening for probability of benefit from E
- Targeted design
  - Classify patients based on probability of benefit from E
  - Randomize only patients likely to benefit

- Compare the two designs with regard to the number of patients required to achieve a fixed statistical power for detecting treatment effectiveness and the number of patients needed for screening

Pharmacogenomic Model for Two Treatments With Binary Response

- Molecularly targeted treatment E
- Control treatment C
- λ Proportion of patients predicted responsive (Assay+)
- p_c control response probability
- response probability for Assay+ patients receiving E is $(p_c + \delta_1)$
- Response probability for Assay- patients receiving E is $(p_c + \delta_0)$

Randomized Ratio
(normal approximation)

- RandRat = $n_{untargeted}/n_{targeted}$

$$RandRat \approx \left(\frac{\delta_1}{\lambda\delta_1 + (1-\lambda)\delta_0} \right)^2$$

- If δ_0=0, RandRat = $1/\lambda^2$
- If $\delta_0 = \delta_1/2$, RandRat = $4/(\lambda+1)^2$

Randomized Ratio $n_{untargeted}/n_{targeted}$

λ Assay+	δ_0=0	$\delta_0 = \delta_1/2$
0.75	1.78	1.31
0.5	4	1.78
0.25	16	2.56

Screened Ratio

- $N_{untargeted} = n_{untargeted}$
- $N_{targeted} = n_{targeted}/\lambda$
- $ScreenRat = N_{untargeted}/N_{targeted} = \lambda RandRat$

Screened Ratio

λ Assay+	δ_0=0	$\delta_0 = \delta_1/2$
0.75	1.33	0.98
0.5	2	0.89
0.25	4	0.64

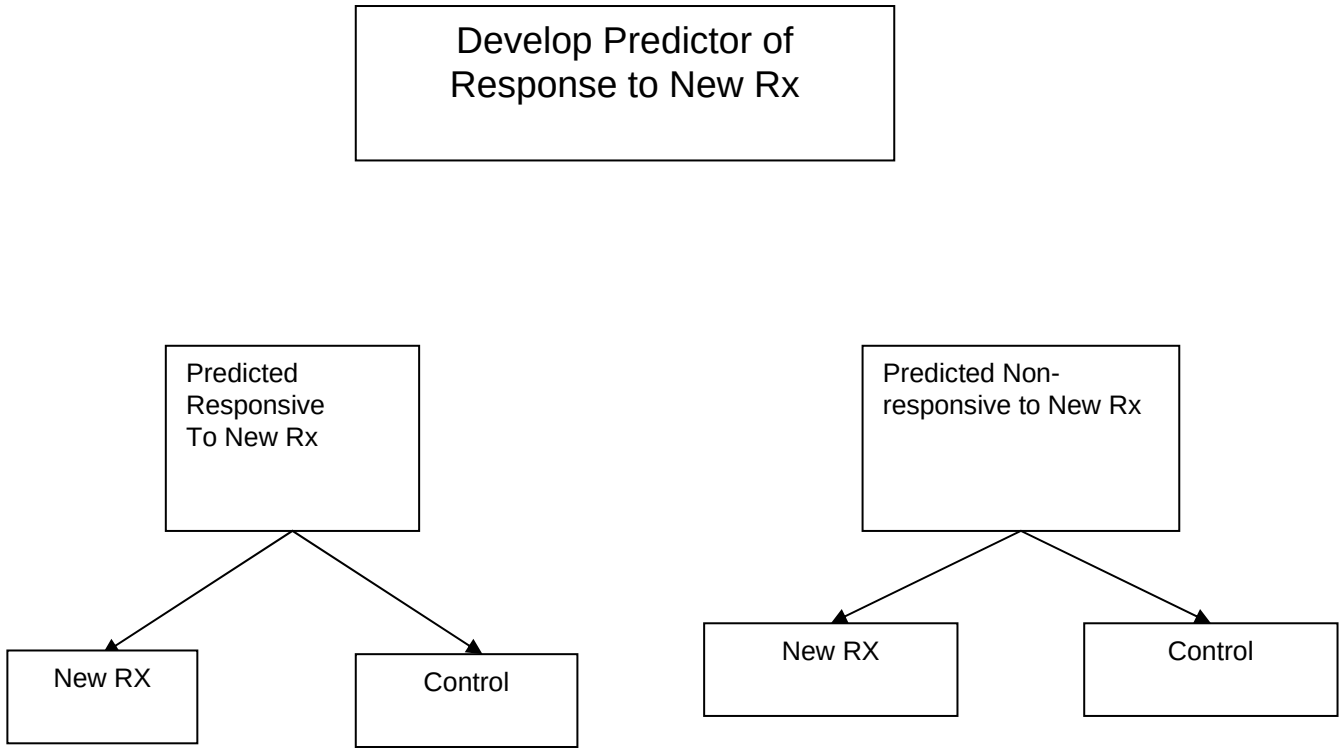
- For a drug like Iressa in lung cancer
  - 10% response rate
  - If only responders benefit, untargeted designs are very inefficient, even with 1000 patients randomized
  - More effort should be placed in finding predictors of response based on phase II data
    - Sequencing key genes
    - Expression profiling

- For Herceptin, even a relatively poor assay enabled conduct of a targeted phase III trial which was crucial for establishing effectiveness
- In many cases, the assay based on the presumed mechanism of action will not correlate with response and it may be more effective to let the data develop the assay via expression profiling

Using Genomics in Development of a New Therapeutic (II)

- Develop a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Perform separate adequate randomized clinical trials for classifier + and classifier – patients.
- Demonstrate that the diagnostic can be reproducibly measured
- Confirmatory phase III trial

Using PG Classifiers to Select Patients for Phase III Trials



Using Genomics in Development of a New Therapeutic (III)

- Develop a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Do not use the diagnostic to restrict eligibility, but rather to structure a prospectively planned analysis strategy of a randomized trial of the new drug.
- Compare the new drug to the control overall for all patients ignoring the classifier.
  - If the treatment effect on the primary pre-specified endpoint is significant at the 0.04 level, then claim effectiveness for the eligible population as a whole.
- If the overall test is not significant at the 0.04 level, then perform a single subset analysis evaluating the new drug in the classifier + patients.
  - If the treatment effect is significant at the 0.01 level, then claim effectiveness for the classifier + patients.
- Demonstrate that the diagnostic can be reproducibly measured
- Confirmatory phase III trial

These Strategies Require

- The data used to develop the classifier must be distinct from the data used to test hypotheses about treatment effect in subsets determined by the classifier
  - Developmental studies are exploratory
  - Studies on which treatment effectiveness claims are to be based should be hypothesis testing studies based on completely pre-specified classifiers

Pharmacogenomic Classifier
Composite Biomarker
Genomic Signature

- A set of genes is not a classifier

Strategies for Development of a Genomic Classifier

- During phase I/II development
  - Extended phase II
- After failed phase III trial using archived specimens
- “Prospectively” during phase III

Phase III Treatment Evaluations in Classifier Determined Subsets

- Phase III trials of new patients in which PG classifier is measured prior to randomization

- Previously conducted randomized phase III trials in which specimens were archived

Phase III Treatment Evaluations in Previously Conducted Randomized Phase III Trials

- Data not used in development of classifier
- Prospective analysis plan based on completely specified classifier
- Completeness of specimen archive
  - What percentage of patients would not agree to specimen collection in new trial?

Developing Composite Genomic Classifiers

- Classifiers should classify accurately
- Composite classifiers incorporate the contributions of multiple single-gene features
- The single gene feature are usually selected based on their “informativeness” for distinguishing patients likely to respond to the new rx from patients not likely to respond
- The single gene features can be selected based on informativeness in identifying patients more likely to respond to a new treatment than to a control treatment

Developing Composite Genomic Classifiers

- Classifiers should classify accurately
- To classify accurately, it is much more important that informative features not be excluded
- To classify accurately, it is less important that noise features be excluded
- If we wished to “validate” a classifier, we should validate it’s predictions, not that the same features (genes) are included in a classifier developed on independent data

Genomic Classifiers Used for Targeting Patients in Drug Development

- The classifier can be considered a composite biomarker, but the components should not have to be “valid disease biomarkers” in the FDA sense

Biomarker

- “Any biological measurement that provides actionable information regarding disease progression, pharmacology, or safety that can be used as a basis for decision making in drug development.”
  - J. Boguslavsky

- “I don’t know what ‘clinical validation’ [of a biomarker] means. The first thing you have to do is define a purpose for the biomarker. Validation is all about demonstrating fitness for purpose.”
  - Dr. Stephen Williams, Pfizer

Conclusions

- New technology and biological knowledge makes is increasingly feasible to identify which patients are most likely to benefit from a new treatment
- Targeting treatment can make it much easier to convincingly demonstrate treatment effectiveness
- Targeting treatment can greatly improve the therapeutic ratio of benefit to adverse effects, the proportion of treated patients who benefit

Conclusions

- Effectively defining and utilizing PG classifiers in drug development offers multiple challenges
- Much of the conventional wisdom about how to develop and utilize biomarkers is flawed and does not lead to definitive evidence of treatment benefit for a well defined population

Conclusions

- With careful prospective planning, genomic classifiers can be used in a manner that provides definitive evidence of treatment effect
  - Trial designs are available that will support broad labeling indications in cases where drug activity is sufficient, and the opportunity to obtain strong evidence of effectiveness in a well defined subset where overall effectiveness is not established

Conclusions

- Prospectively specified analysis plans for phase III data are essential to achieve reliable results
  - Biomarker analysis does not mean exploratory analysis except in developmental studies
  - Biomarker classifiers used in phase III evaluations should be completely specified based on external data
- In some cases, definitive evidence can be achieved from prospective analysis of patients in previously conducted clinical trials with extensive archival of pre-treatment specimens

Acknowledgements

- Boris Freidlin
- Aboubakar Maitournam
- Sudhir Varma
- Sue-Jane Wang