Is Expression Profiling Transforming Clinical Oncology?

Richard Simon, D.Sc.
Chief, Biometric Research Branch
National Cancer Institute
http://linus.nci.nih.gov/brb

Disclosures

- Financial relationships
  - None
- Off label or investigational usage of drugs
  - None

How Can I be Against Microarrays?

- *Design and Analysis of DNA Microarray Investigations*
  - R Simon, EL Korn, MD Radmacher, L McShane, G Wright, Y Zhao. Springer (2003)

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

- Encapsulates BRB experience in analysis of data and development of methods
- Educating biologists in microarray data analysis
- Easy user interface
  - Excel front-end
- Ease of data loading
  - integrated
- Drill-down linkage to genomic databases
- State-of-the-art analytic tools
  - Based on BRB critically evaluating literature
- Easily extensible
  - R add-ins
- Portable
  - Non-proprietary
  - Free for non-commercial use

Clinical Oncology Needs

- Better tools for utilization of existing treatments

- Better treatments

“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

- Better targeted therapies offer improved health quality and reduced waste of resources

Targeting of Treatment to Cancer Patient Based on Tumor Expression Profile in Broad Clinical Use

- Oncotype-Dx
  - Not yet

Pusztai et al. The Oncologist 8:252-8, 2003

- 939 articles on “prognostic markers” or “prognostic factors” in breast cancer in past 20 years
- ASCO guidelines only recommend routine testing for ER, PR and HER-2 in breast cancer
- “With the exception of ER or progesterone receptor expression and HER-2 gene amplification, there are no clinically useful molecular predictors of response to any form of anticancer therapy.”

Why Aren’t There More?

Limited by Appropriate Therapeutic Decision Contexts

- Potentially curative treatment for life threatening disease with no good alternative therapy
  - Not many curative treatments
  - Can rarely be sure that NPV is perfect

- Palliative treatment with low response rate and no good alternative treatments
  - Willingness of patients and physicians to use such treatments
- Expression profile based targeting have a stronger motivation in settings where several effective treatments are available

- Patients whose prognosis is so good without chemotherapy that it can be withheld
  - Unwillingness of physicians and patients to withhold treatment even if it’s chance for benefiting the patient is very low
  - Difficulty of finding a series of untreated patients with frozen tumor archived to develop expression based classifier

Assay Limitations of DNA Microarray Expression Profiling

- Need for fresh/frozen tumor
- Large number of cells needed
  - Heterogeneity of specimens
    - Tumor, stroma, endothelial cells, immune cells
- Variation due to heterogeneity of individual tumors
- Variation due to tissue handling
- Variation in assay
  - Batch effects, reagent effects, RNA labeling
- Cross-hybridization and errors in probe identification

- Some sources of variability are controlled within a study but limit the ability to accurately classify samples collected outside of study conditions
  - Internal validation
  - External validation

Design and Analysis Limitations of Existing Studies

- Retrospective analysis of a convenience sample of patients
  - Patients are too heterogeneous with regard to stage and treatment to constitute a therapeutically meaningful cohort
- Lack of prospective plan for analysis
  - Multiple analyses of patient subsets, endpoints, treatments, …
- Inadequate methods used to control for increased multiplicity problems of thousands of genes measured
- Inadequate number of patients
- Lack of prospective plan for internal validation

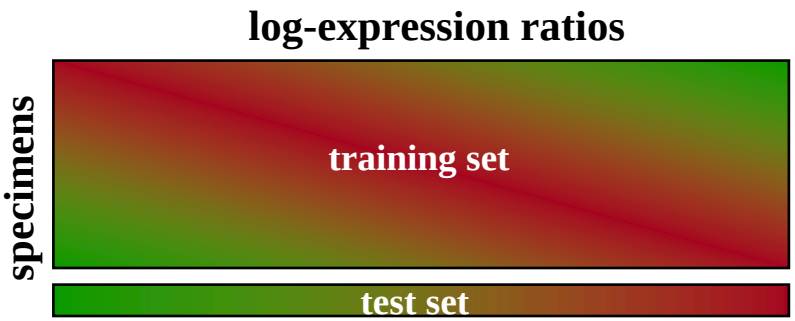
Internal Validation

- Fit of a model to the same data used to develop it is no evidence of prediction accuracy for independent data.
- When the number of candidate predictors (p) exceeds the number of cases (n), perfect prediction on the same data used to create the predictor is always possible

Split-Sample Evaluation

- Divide cases into a training set and a testing set
- Do not permit any access to the testing set until a single fully specified classification model is developed on the training set
- A classification model is not just a set of genes
- The single fully specified classifier is simply applied to predict outcome for the cases in the testing set based on their expression profiles
  - No re-calibration or adjustment of the model is permitted
  - The number of prediction errors on the testing set is counted
- Used for Rosenwald et al. study of prognosis in DLBL lymphoma.
  - 200 cases training-set
  - 100 cases test-set

Cross-Validated Prediction (Leave-One-Out Method)



1. Full data set is divided into training and test sets (test set contains 1 specimen).
2. Prediction rule is built from scratch using the training set.
3. Rule is applied to the specimen in the test set for class prediction.
4. Process is repeated until each specimen has appeared once in the test set.

- Cross validation is only valid if the test set is not used in any way in the development of the model. Using the complete set of samples to select genes violates this assumption and invalidates cross-validation.
- With proper cross-validation, the model must be developed from scratch for each leave-one-out training set. This means that feature selection must be repeated for each leave-one-out training set.
- The cross-validated estimate of misclassification error is an estimate of the prediction error for model fit using specified algorithm to full dataset

- Split-sample validation is often applied with an inadequately small number of test samples

- Cross-validation is often applied invalidly, using the full dataset to select the genes for inclusion in the predictive model

Design and Analysis Limitations of Existing Studies

- Over-use of cluster analysis
- Identification of differentially expressed genes using significance thresholds that permit hundreds of false positives
  - 0.05*20,000=1000 false positives
- No or improperly applied internal validation
- Failure to demonstrate that the cross-validated misclassification rate is better than chance and better than that obtainable with standard clinical variables
- Mis-emphasis on statisical significance of genes rather than improvement in prediction accuracy

Studies Developing Gene Expression Profile Classifiers Should be Viewed as Analogous to Phase II Trials Requiring Phase III Validation

Limitations to Internal Validation

- Confounding by sample handling or assay effects
  - Cases collected and assayed at different times than controls
- Failure to incorporate important sources of variability
  - Assay variability
  - Tissue handling
  - Tumor heterogeneity
- Limited size of developmental study
- Problems of design and analysis in developmental study

Independent Data Validation

- From different clinical centers
- Specimens assayed at different time from training data
- Reproducibility of assay for individual tumors demonstrated to clinical reference laboratory standards
- Positive and negative samples collected in the same way
- Study sufficiently large to give precise estimates of sensitivity and specificity of the classifier
- The validation study is prospectively planned
  - patient selection pre-specified to address a therapeutically relevant question
  - endpoints and hypotheses pre-specified
  - predictor fully pre-specified
  - Study addresses assay reproducibility
  - Specimens may be either prospective or archived

Adequate External Validation Studies are Rarely Performed

- They are expensive and require multi-center cooperation
- They require demonstration of assay reproducibility
- The financial incentives for developing and validating PG classifiers of existing treatments are limited

Clinical Oncology Needs

- Better tools for utilization of existing treatments

- **Better treatments**

We know how to do good clinical trials and have a large clinical trial infrastructure waiting to get their hands on good compounds

“Expression profiling has given us hundreds of molecular targets, we just need better ways of validating them”

- Many genes are overexpressed or silenced in the tumor relative to normal tissue
- Most of these genes are poor therapeutic targets because they represent secondary and terciary events of proliferation and selection with a genomically unstable tumor.
- Many of the overexpressed and silenced genes may be deregulated only in a subset of the numerous sub-clones of the tumor
- The most promising molecular targets are those that represent initial mutations that are intrinsic to the tumor, drive the growth of the tumor, and exist throughout the tumor
- So far, expression profiling has not helped much to identify these key genes

- So many genes are differentially expressed in a tumor relative to a normal cell of the same tissue that it is difficult to identify the initiating events
- Microarrays are useful, in conjunction with gene silencing and genetic manipulation, to improve mapping of signaling pathways.
- In the future transcript expression profiling may become more useful for understanding key steps in oncogenesis

Clustering and pattern recognition techniques have only replaced microscopic phenomenology with molecular phenomenology and have not provided sufficient biological knowledge or a better understanding of the relationship between a drug and a target or disease

Richard Klausner

- Expression profiling is valuable for understanding the interactions of a candidate protein target and for understanding the multiple effects of inhibiting a protein

Microarrays are Valuable in the Clinical Development of New Targeted Drugs

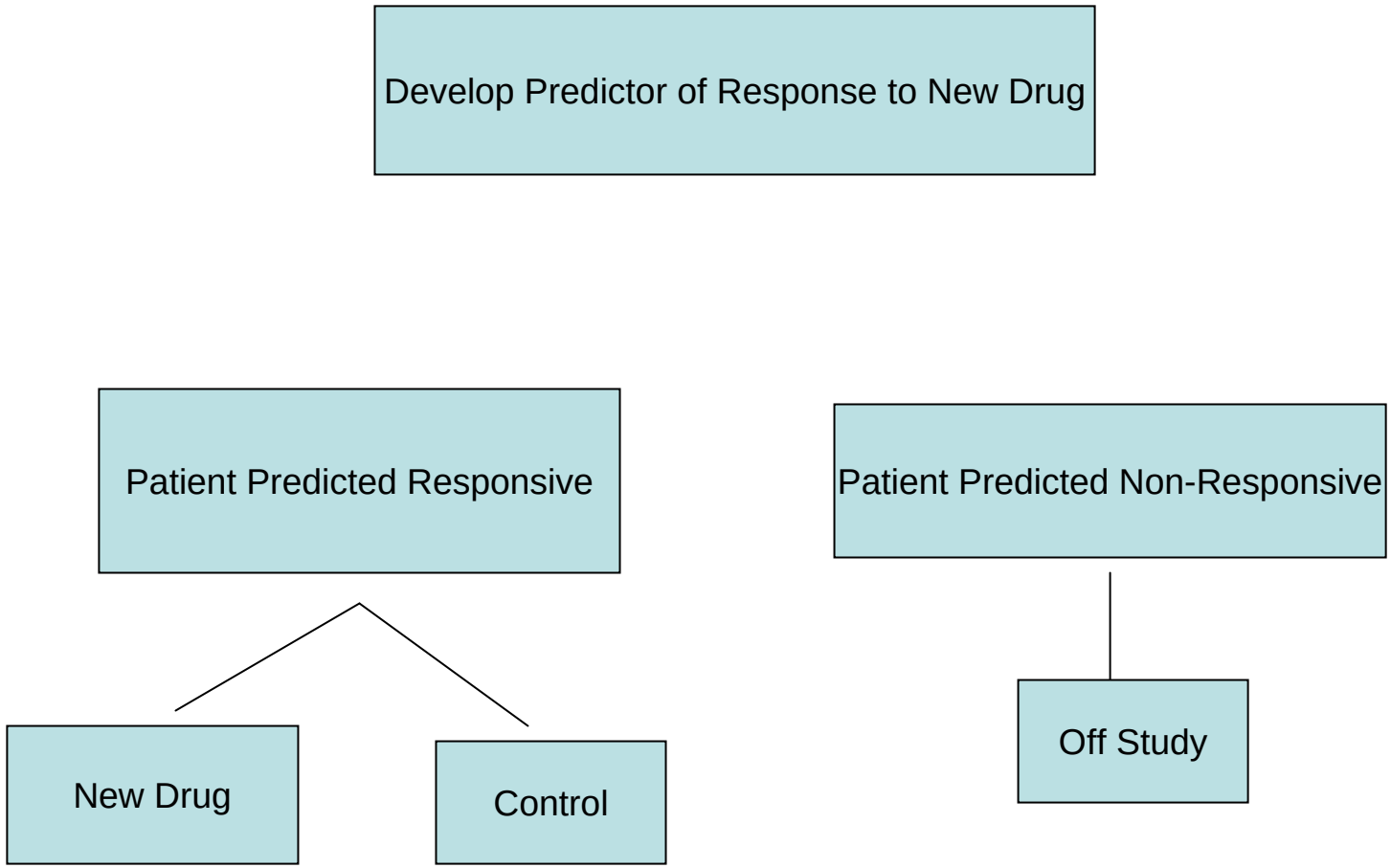
- Does the drug inhibit the target
- Does the target drive the growth of the tumor

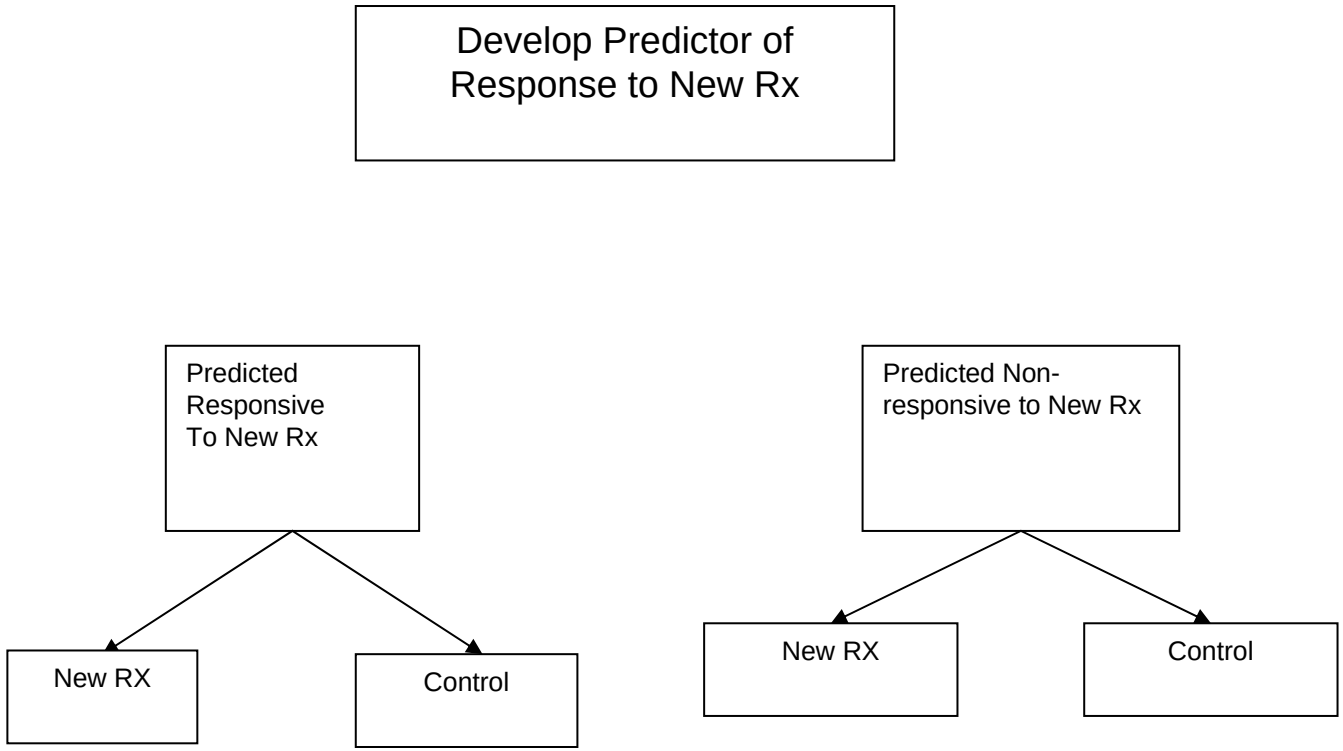
For which tumors is the target of key importance

- Usually difficult to answer
  - ABL in CML
  - BRAF in melanoma
- Over-expression of the drug target does not necessarily mean the target is important

Microarrays in Clinical Development of New Targeted Drugs

- During enlarged phase II, develop expression signature that can be used to identify responders to the drug
- Translate the expression profile to a platform that can be applied during phase III
- Use classifier in the design and analysis of phase III trial


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


Develop Predictor of
Response to New Rx
Predicted
Responsive
To New Rx
New RX
Control
Predicted Non-
responsive to New Rx
New RX
Control

Perspective

Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials

Richard Simon and Aboubakar Maitournam
Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

ABSTRACT

Purpose: **Genomic technologies make it increasingly possible to identify patients most likely to benefit from a molecularly targeted drug. This creates the opportunity to conduct targeted clinical trials with eligibility restricted to patients predicted to be responsive to the drug.**

Experimental Design: **We evaluated the relative efficiency of a targeted clinical trial design to an untargeted design for a randomized clinical trial comparing a new treatment to a control. Efficiency was evaluated with regard to number of patients required for randomization and number required for screening.**

Results: **The effectiveness of this design, relative to the more traditional design with broader eligibility, depends on multiple factors, including the proportion of responsive patients, the accuracy of the assay for predicting responsiveness, and the degree to which the mechanism of action of the drug is understood. Explicit formulas were derived for computing the relative efficiency of targeted *versus* untargeted designs.**

Conclusions: **Targeted clinical trials can dramatically reduce the number of patients required for study in cases where the mechanism of action of the drug is understood and an accurate assay for responsiveness is available.**

INTRODUCTION

Many cancer therapeutics benefit only a subset of treated patients. Genomic technologies such as DNA microarray expression profiling are providing biomarkers that facilitate the prediction of which patients are most likely to respond to a given regimen (1, 2). Molecularly targeted drugs are of increasing importance in cancer therapeutics, and such drugs are only expected to be effective for patients whose tumors express the target (3, 4). Thus, clinical trials may be increasingly tailored for patients who are predicted to respond to therapy (5). We call these targeted designs. As discussed in this article, we studied the efficiency of targeted designs in comparison with traditional randomized designs with broader eligibility criteria. We evaluated efficiency in the context of a binary outcome end point. Although many clinical trials use survival or time-to-progression end points, the binary end point setting is more tractable, and we obtained results that are intuitive and should be useful in understanding the factors that effect efficiency generally. For the untargeted and targeted design, we considered the comparison of a control *versus* experimental treatment with the same number of randomized patients in the two groups.

We compared the two designs with regard to the number of randomized patients required. We also compared the number of randomized patients for the untargeted design to the number of screened patients required for the targeted design. We assume that in the targeted design patients are screened using an assay that indicates whether the patient is likely to benefit from the new treatment. If the control arm is an active treatment, then the screening classifier should provide an indication of whether the patient is more likely to respond to the new regimen than to the control arm. Our efficiency comparisons are based on using the formula of Ury and Fleiss (6) for planning sample size for comparing proportions because of its known accuracy for approximating the tables of Casagrande, Pike, and Smith for the power of Fisher's exact test (7).

MATERIALS AND METHODS

We considered a population of patients consisting of an R+ portion who were predicted to be responsive to the new treatment and a remainder portion R−. The R− strata constituted a proportion γ of the population. Patients were randomized between the control and the experimental groups. p_c denotes the response probability in control group and was assumed to be the same for R− and R+ patients. The response probability in the treatment group was $p_c + \delta_0$ and $p_c + \delta_1$ for the R− and R+ patients, respectively. The response probability p_e for the experimental treatment group in the untargeted design was a weighted average of $p_c + \delta_0$ and $p_c + \delta_1$ with weights γ and 1-γ, respectively.

For the targeted design we added the symbol T. The response probability in the experimental group was $p_e^T = p_c + \delta_1$. We consider the one-sided test of the null hypothesis $p_c = p_e$ against the alternative hypothesis $p_e > p_c$.

Let n and n^T denote the number of patients needed to randomize in the untargeted and targeted design respectively to achieve the same statistical power for testing the null hypothesis. The expressions for n and n^T are indicated in the Appendix. The relative efficiency of the untargeted and the targeted designs can be expressed in the form:

$$n/n^T = \left[\frac{\delta_1}{\gamma\delta_0 + (1-\gamma)\delta_1}\right]^2 f \qquad \text{(A)}$$

Received 3/11/04; revised 5/14/04; accepted 5/19/04.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article may be found at http://linus.nci.nih.gov/~brb/TechReport.htm.
Requests for reprints: Richard Simon, National Cancer Institute, 9000 Rockville Pike, Bethesda MD 20892-7434. Phone: 301-496-0975; Fax: 301-402-0560; E-mail: rsimon@nih.gov.

Summary

- Transcript expression profiling by microarrays is a powerful technology for helping to elucidate biological mechanisms when used in conjunction with clever experimentation
- So far, transcript profiling has not been successful in identifying the initiating genomic changes responsible for oncogenesis or for identifying key molecular targets
  - This will hopefully change with more sophisticated use of microarrays with silencing reagents and better mapping of signaling pathways

Summary

- There is a great degree of over-treatment in oncology, resulting in economic waste and patient toxicity without benefit.
- There is a need to better target existing treatments to those patients that have a reasonable likelihood of benefiting from them.

Summary

- Transcript expression profiling can provide effective tools for such targeting but there are serious barriers to the development of widely used gene expression based targeting tools.
  - Therapeutic context and willingness of oncologists and patients to withhold treatment
  - Assay variability
  - Limitations in broad use of an RNA based assay requiring fresh/frozen tissue
  - Serious problems with the design and analysis of some expression profiling studies and inadequacy of journal review to spot problems, making some results unreliable
  - Difficulty and expense of conducting adequate validation studies and developing assays to clinical reference laboratory standards of reproducibility

Summary

- Microarray expression profiling has an important role to play in the pre-clinical and clinical development of new therapeutics
  - Identification of a signature for targeting the drug to those patients most likely to benefit
  - Establishing a reliable clinically applicable assay for the signature