ARTICLES

• Retracted •

nature medicine

# Genomic signatures to guide the use of chemotherapeutics

Anil Potti[1,2], Holly K Dressman[1,3], Andrea Bild[1,3], Richard F Riedel[1,2], Gina Chan[4], Robyn Sayer[4], Janiel Cragun[4], Hope Cottrill[4], Michael J Kelley[2], Rebecca Petersen[5], David Harpole[5], Jeffrey Marks[5], Andrew Berchuck[1,6], Geoffrey S Ginsburg[1,2], Phillip Febbo[1–3], Johnathan Lancaster[4] & Joseph R Nevins[1–3]

**Using *in vitro* drug sensitivity data coupled with Affymetrix microarray data, we developed gene expression signatures that predict sensitivity to individual chemotherapeutic drugs. Each signature was validated with response data from an independent set of cell line studies. We further show that many of these signatures can accurately predict clinical response in individuals treated with these drugs. Notably, signatures developed to predict response to individual agents, when combined, could also predict response to multidrug regimens. Finally, we integrated the chemotherapy response signatures with signatures of oncogenic pathway deregulation to identify new therapeutic strategies that make use of all available drugs. The development of gene expression profiles that can predict response to commonly used cytotoxic agents provides opportunities to better use these drugs, including using them in combination with existing targeted therapies.**

# Using Cell Lines to Predict Sensitivity

ture.com/naturemedicine

Genomic signatures to guide the use of chemotherapeutics

Anil Potti[1,2], Holly K Dressman[1,3], Andrea Bild[1,3], Richard F Riedel[1,2], Gina Chan[4], Robyn Sayer[4], Janiel Cragun[4], Hope Cottrill[4], Michael J Kelley[2], Rebecca Petersen[5], David Harpole[5], Jeffrey Marks[5], Andrew Berchuck[1,6], Geoffrey S Ginsburg[1,2], Phillip Febbo[1–3], Johnathan Lancaster[4] & Joseph R Nevins[1–3]

Potti et al (2006), Nature Medicine, 12:1294-1300.

The main conclusion is that we can use microarray data from cell lines (the NCI60) to define drug response "signatures", which can be used to predict whether patients will respond.

They provide examples using 7 commonly used agents.

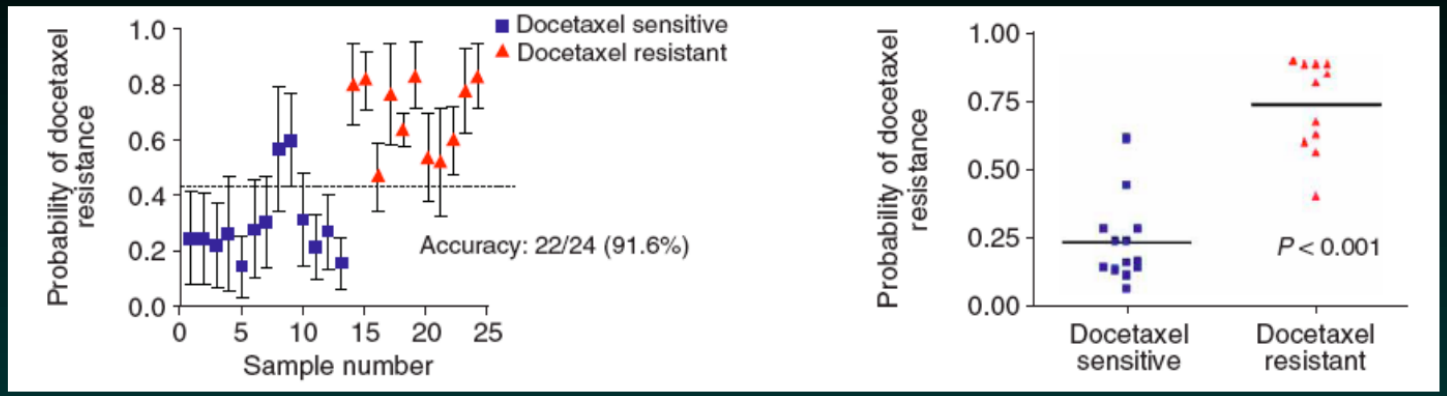
[ Keith Baggerly ]

# Their Gene List and Ours

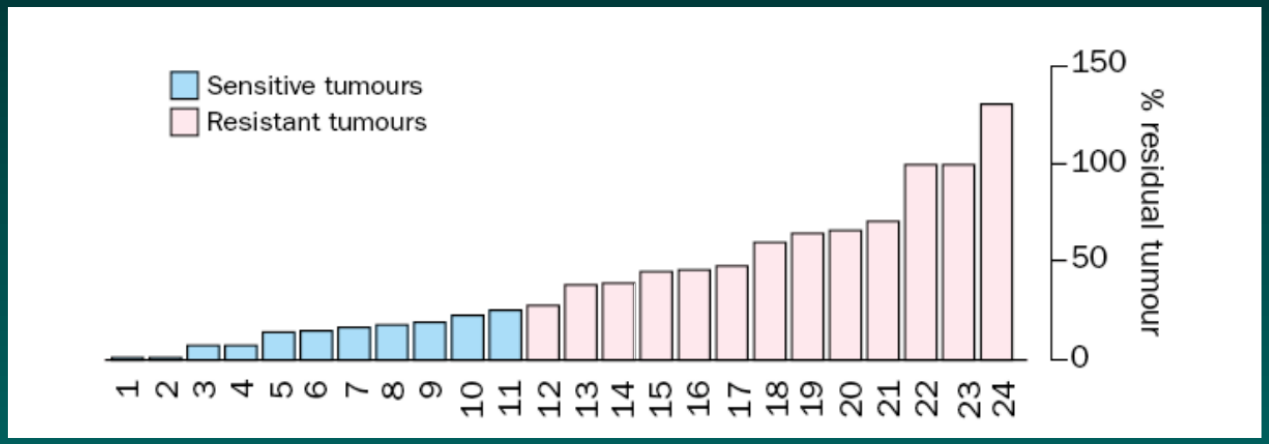
```
> temp <- cbind(
    sort(rownames(pottiUpdated)[fuRows]),
    sort(rownames(pottiUpdated)[
         fuTQNorm@p.values <= fuCut]);
> colnames(temp) <- c("Theirs", "Ours");
> temp
     Theirs         Ours
...
[3,] "1881_at"      "1882_g_at"
[4,] "31321_at"     "31322_at"
[5,] "31725_s_at"   "31726_at"
[6,] "32307_r_at"   "32308_r_at"
...
```

[ Keith Baggerly ]

# Predicting Docetaxel Response




Potti et al, Nat Med 2006, 12:1294-300, Fig 1d




Chang et al, Lancet 2003, 362:362-9, Fig 2 top

[ Keith Baggerly ]

# "Starter Set" Materials for the Saga

This web page derives from our web site supplement for the manuscript *Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology* by Keith A. Baggerly and Kevin R. Coombes. The main page is here.

Particularly since the story was covered by 60 Minutes, we've gotten requests for more details. A (noncomprehensive!) list of documents/links we've found ourselves suggesting frequently is given below. Hope some of these prove useful!

## 1. A Video of Us Telling the Story

There are a few videos out there of us giving talks on this story. The one that we'd recommend at present is one from Cambridge in late 2010. This is about 35 minutes long, but should convey the gist of the types of problems we were seeing and how we identified them. Fair warning -- one review of this on the web notes "*Be warned, Dr. Baggerly is a fast-talking nerdish PhD who thinks you understand what he's saying [which you likely won't totally get], but watch at least some of it to get the flavor of the genre*" which is probably fair ;).

## 2. The 60 Minutes Segment and Transcript

This is certainly how most people have encountered the story. The clip and transcript are available here. In addition to the segment that aired, there's a short (1:30) clip of Paul Goldberg (of the Cancer Letter) discussing the Rhodes scholar angle, which is well worth watching. We included both this clip and selected short bits from the main piece in talks we've given since the segment aired.

## 3. Slides from some Recent Talks

We try to update at least some of the slides in our talks, so more recent versions will at least mention later developments. The slides linked to here are from a talk I gave on Feb 15, 2012. We included clips from the 60 Minutes segment at the end of slide 27.

The slides linked to here are from a presentation I gave on Jul 9, 2012, where I used now-available documents to clarify more of who knew what *as things were going on*.

## 4. Our 2009 Annals of Applied Statistics Paper

This is where we detail the specific problems we encountered. This may be an atypical statistics paper in that we include all of 3 formulas, all of which are wrong. A copy of the paper is available here.

## 5. A 2011 Editorial from Clinical Chemistry (Subscription Required)

This is a short (2.5pg) piece we wrote after listening to representatives from both NCI (Lisa McShane) and FDA (Robert Becker) give testimony to the Institute of Medicine (IOM). We use extracts from their talks to emphasize just what information should be required to support clinical "omics" publications. The piece (subscription required) is here.

## 6. Various Notes from the Institute of Medicine Open Sessions

Sparked by this case, the IOM reviewed the level of evidence that should be required before "omics"-based assays are used to guide patient therapy in clinical trials. This committee began meeting in December of 2010, and issued its report on March 23, 2012. Many of the meetings were open and recorded (audio only, but accompanying slides are typically available). Most of these are linked to from here.

We'd probably start with the testimony we gave on March 31, 2011. Be warned, this segment wound up being nearly 3 hours long. The other one we'd recommend listening to early is Lisa McShane's (biostatistician from the NCI) from December 20, 2010, where she laid out much of what the NCI knew and was doing behind the scenes while all of this was going on. Our annotation of the 550 pages (!) of documents the NCI released at this session is available from the link above. Our summary is about 15 pages. The audio of Lisa McShane's presentation (about an hour) is available from The Cancer Letter linked to their Jan 28, 2011 issue.

## 7. The IOM Omics Report, and Some Subsequent Presentations by IOM Committee Members

We debated a bit about this, because the full Omics report (at 274p!) isn't really *starter* material. It is, however, a very thorough exploration of how studies should be performed if the goal is to translate the omics-based tests into clinical use. That said, a "report brief" (5p) is here, and if you really understand the figure on the last page, you're essentially there. I suspect, however, that you might not fully understand it. I thought I did, but listening to some of the recorded presentations by committee members -- Gil Omenn, Joe Gray, Dan Hayes and Daniela Witten at AACR (Apr 3), and Daniela Witten and Larry Kessler at a U Washington session on research ethics (Jul 19) -- added further detail for me. If you're familiar with the background now, I think I'd point you to the video of Kessler's summary of the recommendations first and suggest expanding from there.

| | |
|---|---|
| NORTH CAROLINA<br><br>DURHAM COUNTY | IN THE GENERAL COURT OF JUSTICE<br>SUPERIOR COURT DIVISION<br>1 CVS 4721 |



Richard Aiken, Jean K. Carroll, Executrix of the Estate of Harold G. Carroll, Jean K. Carroll, Individually, Peggy Cox, as Administratrix of the Estate of Paul F. Cox, Peggy Cox, Individually, Helene L. Fligel, Jason Gannon, as Personal Representative of the Estate of Jennifer L. Gannon, John Haddock, as Executor of the Estate of Karen Heath, Walter Jacobs, as Executor of the Estate of Juliet J. Jacobs, Walter Jacobs, Individually, Polly Johnson, as Executor of the Estate of Malcom W. Johnson, and Polly Johnson, Individually,

Plaintiffs

vs.

Duke University, Duke University Health System, Inc., Private Diagnostic Clinic, PLLC, Joseph Nevins, Ph.D., Anil Potti, M.D., Michael Cuffe, M.D., Sally Kornbluth, M.D., John M. Harrelson, M.D., and CancerGuide Diagnostics, Inc. f/k/a Oncogenomics, Inc,

Defendants

**COMPLAINT**
**(JURY TRIAL DEMANDED)**

# The Real Reason Reproducible Research is Important

Posted on June 6, 2014 by Roger Peng

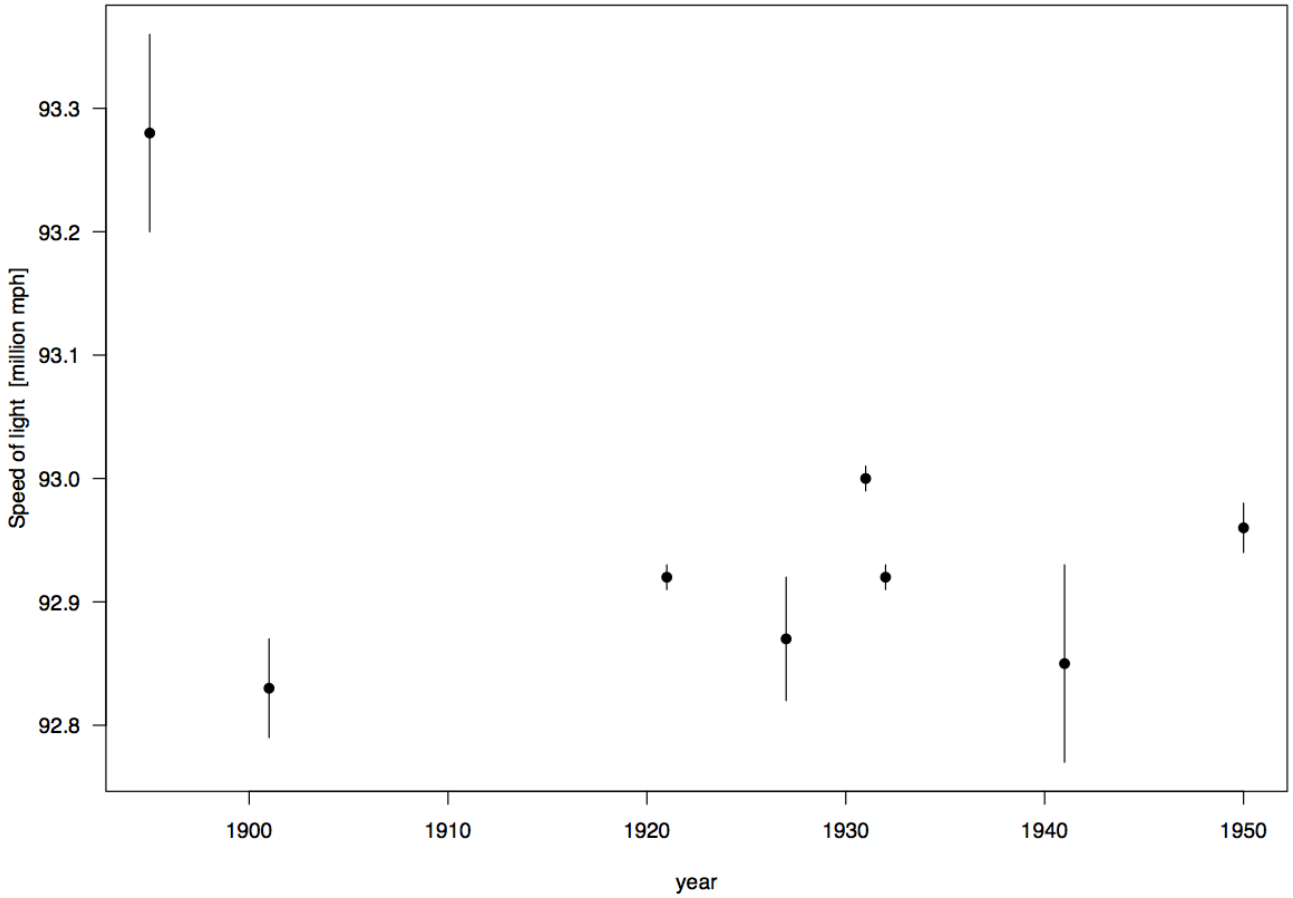
Reproducible research has been on my mind a bit these days, partly because it has been in the news with the Piketty stuff, and also perhaps because I just published a book on it and I'm teaching a class on it as we speak (as well as next month and the month after...).

However, as I watch and read many discussions over the role of reproducibility in science, I often feel that many people miss the point. Now, just to be clear, when I use the word "reproducibility" or say that a study is reproducible, I do not mean "independent verification" as in a separate investigator conducted an independent study and came to the same conclusion as the original study (that is what I refer to as "replication"). By using the word reproducible, I mean that the original data (and original computer code) can be analyzed (by an independent investigator) to obtain the same results of the original study. In essence, it is the notion that the *data analysis* can be successfully repeated. Reproducibility is particularly important in large computational studies where the data analysis can often play an outsized role in supporting the ultimate conclusions.

Many people seem to conflate the ideas of reproducible and correctness, but they are not the same thing. One must always remember that **a study can be reproducible and still be wrong**. By "wrong", I mean that the conclusion or claim can be wrong. If I claim that X causes Y (think "sugar causes cancer"), my data analysis might be reproducible, but my claim might ultimately be incorrect for a variety of reasons. If my claim has any value, then others will attempt to replicate it and the correctness of the claim will be determined by whether others come to similar conclusions.

Then why is reproducibility so important? Reproducibility is important because **it is the only thing that an investigator can guarantee about a study**.

Estimates of the speed of light, with “confidence intervals”.




Youden W (1972). *Technometrics* 14: 1-11.