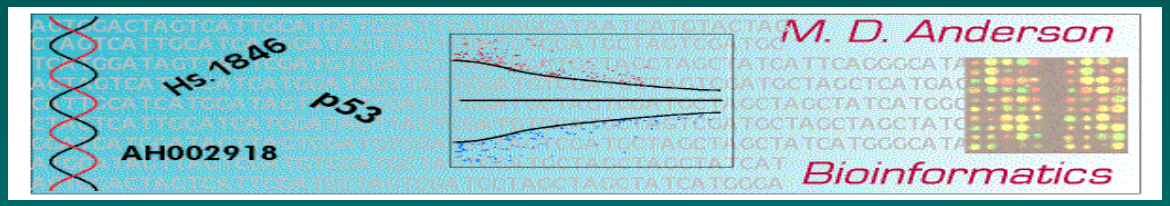
Learning from the Duke Case and the IOM Translational Omics Report: Context

Keith A. Baggerly
Bioinformatics and Computational Biology
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org

Committee on Responsible Science, July 9, 2012



Why is the Duke Case Relevant Here?

It's a recent, high profile case where several things went wrong, and sustained effort was required to correct matters.

It's high profile in part because
the stakes were high (patient therapy was involved),
it was prolonged, and
it required a "non-scientific" objection to end things.

More details:
http://bioinformatics.mdanderson.org/Supplements/ReproRsch-All/Modified/StarterSet

What Makes this Case Distinct?

This is not the first case of misconduct. How does this case differ from others (e.g., Darsee, Bezwoda, Schon)?

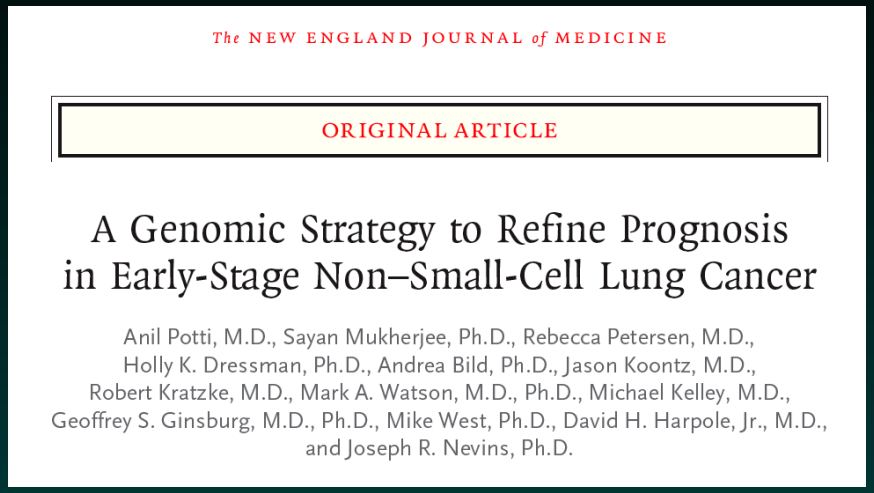
Semi-common themes –

* complex high-dimensional data
* incomplete descriptions of methods
* lack of code

Questionable research practices were established early on

Misconduct was not established for years
(to a degree, it's still being established)

2006: The Stage is Set



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Genomic Strategy to Refine Prognosis in Early-Stage Non–Small-Cell Lung Cancer

Anil Potti, M.D., Sayan Mukherjee, Ph.D., Rebecca Petersen, M.D., Holly K. Dressman, Ph.D., Andrea Bild, Ph.D., Jason Koontz, M.D., Robert Kratzke, M.D., Mark A. Watson, M.D., Ph.D., Michael Kelley, M.D., Geoffrey S. Ginsburg, M.D., Ph.D., Mike West, Ph.D., David H. Harpole, Jr., M.D., and Joseph R. Nevins, Ph.D.

Potti et al (2006, Aug), NEJM, 355:570-80.

ture.com/naturemedicine

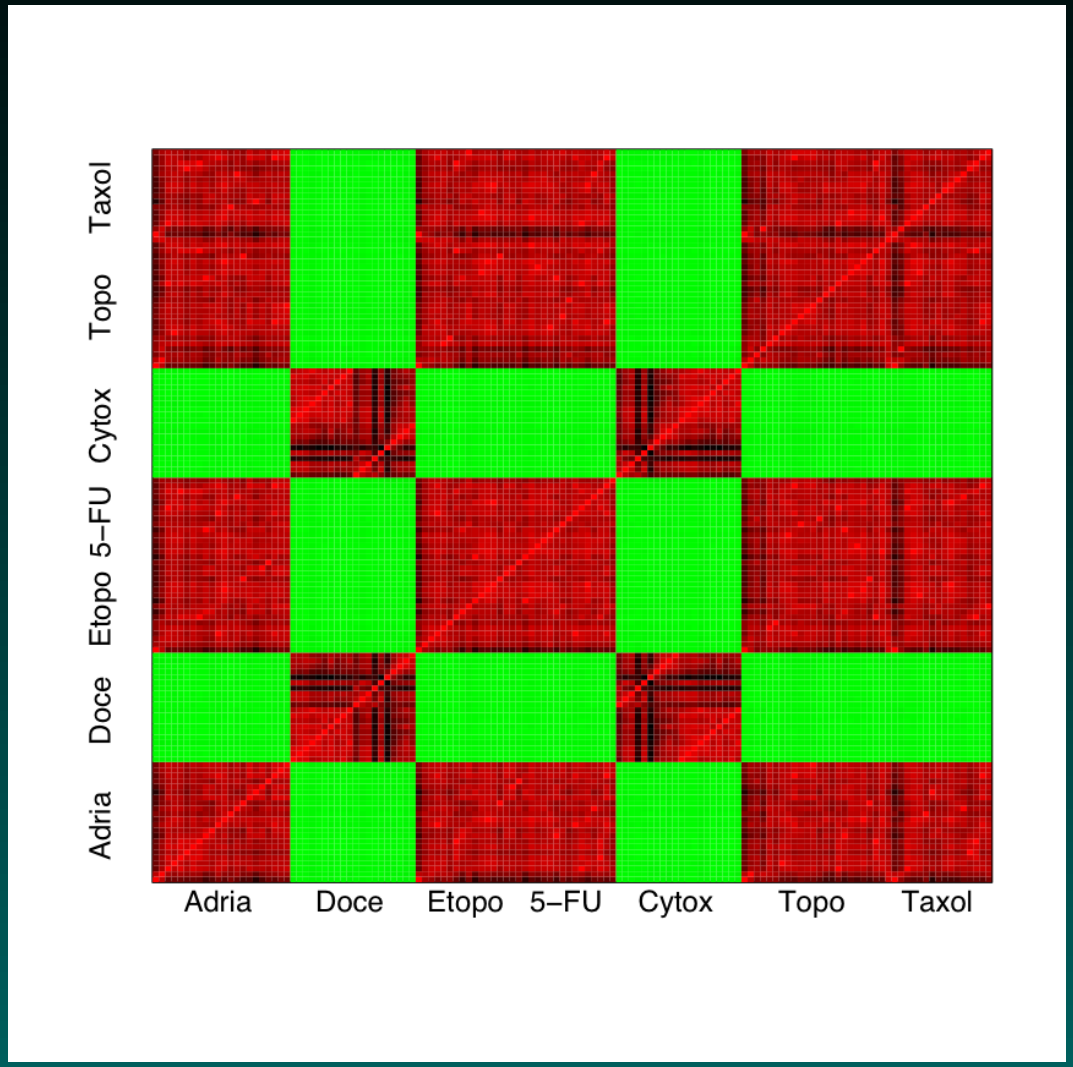
Genomic signatures to guide the use of chemotherapeutics

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, Nov), Nature Medicine, 12:1294-1300.

We begin communicating with Potti and Nevins.

Data Inconsistencies



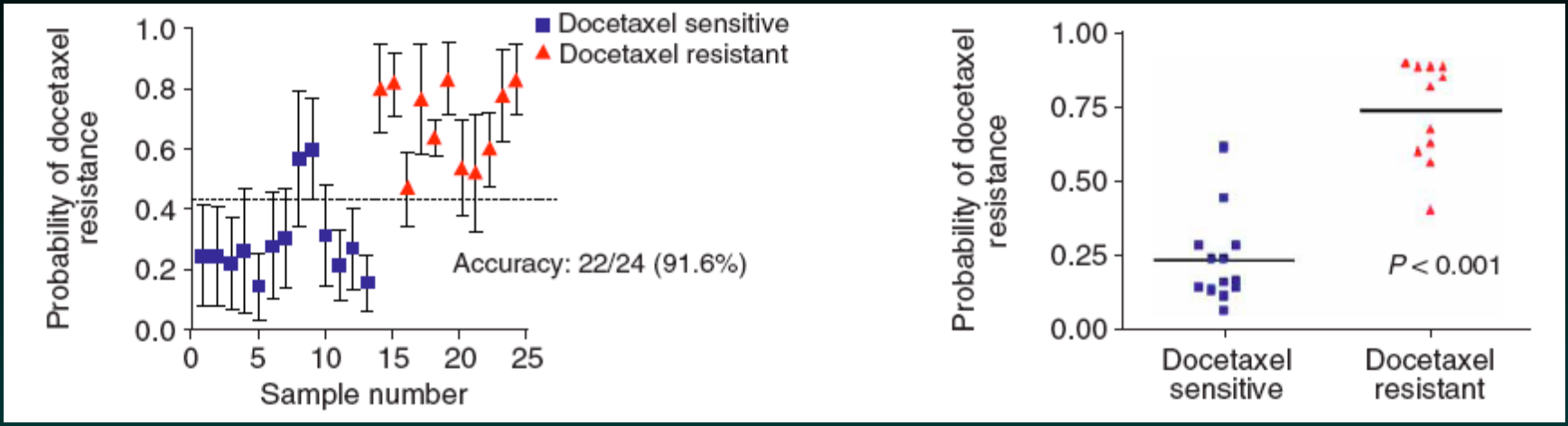
Nov 8, 16, 21: Some cell lines don't match.

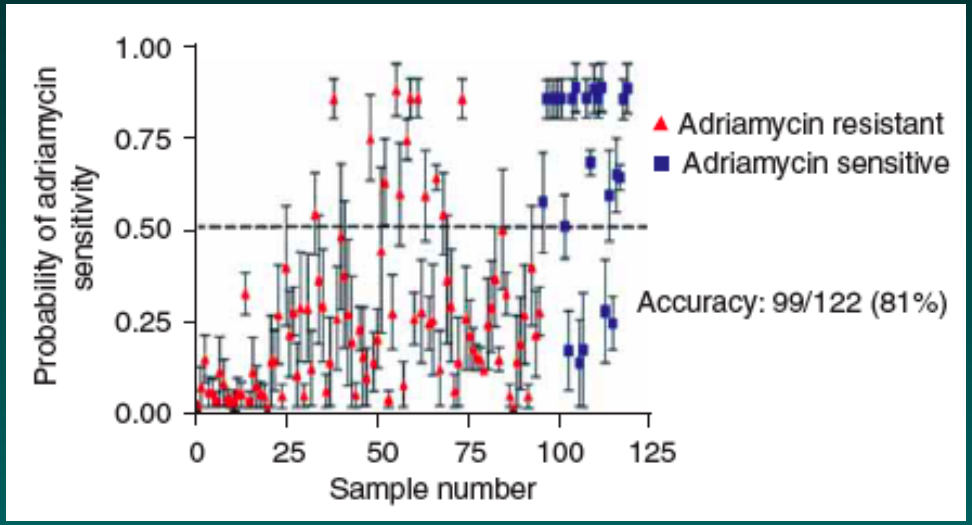
Off by One

```
> 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"
...
```

Dec 4

Sensitive/Resistant Labeling





Dec 13

Outliers

Even after accounting for off-by-one errors, there were some genes we couldn't match.

For docetaxel, we could explain 31/50.

The test data suppliers gave a list of 92 genes splitting responders from nonresponders.

14/19 genes we can't match are in this list of 92, as one contiguous block (entries 7-20).

The last 5/19 genes are those mentioned by name in the paper.

May 16, 2007

Open Conflict

Nov 2007

Nature Medicine correspondence and rebuttal.
We supply data and code.

2008-mid 2009

Iterative process, other papers

* some we could prove were wrong
* several attempts at journal correspondence were refused
* some we couldn't replicate
* have we shown misconduct, or just error?

Clinical Trials

Four clinical trials using the Potti et al Nat Med approach to choose patient therapy were started in 2007-8:

3 at Duke
1 at Moffitt

A fifth (larger) cooperative group trial (CALGB 30702) in lung cancer was proposed in 2009.

At the same time, a large cooperative group trial (CALGB 30506) testing the Lung Metagene Score (LMS) opened. The LMS was not guiding therapy.

Mid-09, pt 1/3 (Us)

Jun: we learn of the trials

Sep: we publish objections in the *Annals of Applied Statistics*

Oct 2: story covered by *the Cancer Letter*

Mid-09, pt 2/3 (NCI)

* We spoke at NCI in Nov 2007 about the problems we'd seen in the Nature Medicine paper.

* In mid-Sept 2009, we sent copies of our *Annals* paper to people we knew at NCI.

* NCI didn't think it could intervene directly, since it wasn't funding any of the trials we identified. We'd hoped they could.

* But the NCI knew several things we didn't (which they couldn't tell us).

What NCI Knew

There were problems with the LMS trial (CALGB 30506)

* data for the NEJM paper were improperly obtained
* Duke's initial trial design used the LMS to guide therapy
* blinded validation of the LMS failed (hence no guidance)
* the NCI hadn't examined the data and code being used.

NCI was reviewing was a new lung cancer trial (CALGB 30702) using cisplatin and pemetrexed. There were similar problems in information submitted here.

NCI was funding a Moffitt trial to validate signatures before guiding therapy.
NCI didn't know use was underway.

What NCI Did

Sep 22

NCI contacts Duke to express concern.

NCI first contacted Kim Lyerly, assuming the trials were being run through Duke's Comprehensive Cancer Center.
Genomics was being treated differently.

After a phone conversation in which the problems were outlined, Duke agreed to suspend the trials and investigate.

Oct 8

NCI contacts the Moffitt PI to express concern.
The Moffitt trial was terminated the next day.

Mid-09, pt 3/3 (FDA)

How did trials start so quickly?

Per FDA, genomic signatures are medical devices.
FDA approval (IDE or IND) is required for investigational use.

Differential omics treatment meant this point was missed.

Aug: another IRB raises the IDE issue. Potti submits a pre-IDE to FDA asking if approval is required.

Oct: FDA says approval is required; requests an IND.

Dec: Potti sends some modifications (not an IND), and asks if approval is still needed. FDA has no record of this letter.

Oct 2009-Dec 2009, Duke's Investigation

* Experts identified with NCI help.

* Duke is not treating this as a misconduct investigation. (We mentioned possible patient harm, not misconduct.)

Oct 23: Potti/Nevins claim of blinded validation refuted.

Nov: new cisplatin/pemetrexed data posted. We send a report detailing new problems to Duke and NCI.

* Within Duke, Nevins argues our report is improper interference, and showing it to the experts would unfairly bias the review. *Duke withholds our report.*

Jan 2010

* Experts, *working with mislabeled data*, say the science looks plausible (exact reproduction of results was not attempted).

* Duke informs NCI it intends to restart trials.
Duke sends NCI a copy of the experts' report.

* Duke's IRB, knowing Potti sent something to FDA in Dec, assumes silence implies trials can proceed.
(This would hold with an IDE or IND, but not with the note actually written.)

Jan-Feb 2010 (Public)

* Duke announces successful completion of investigation and restarting of trials.

* We ask to see the data and report justifying the restarts, but are told these are confidential.

* We object to the restarts in *the Cancer Letter*, citing our Nov 2009 report, which we post. There is no response.

Feb: CancerGuide Dx (a company formed to commercialize the Potti/Nevins signature approach to response prediction) announces it has secured VC funding.

Jan-Mar 2010 (NCI, not Public)

* NCI audits LMS trial (CALGB 30506).

* NCI requests (and requires) the underlying data and code. Nobody else could do this.

* NCI discovers signature was never “locked down”, despite written statements to the contrary – running code twice on the same data gives different results (sometimes very different).

* NCI concludes this invalidates the LMS use in the trial (nothing is fixed), and (by implication) the NEJM paper.

* NCI informs PIs it intends to pull LMS from CALGB 30506.

Apr 2010 (NCI, not Public)

* NCI begins audit of R01 awarded to Anil Potti.

* R01 includes partial funding for a clinical trial, not mentioned on clinicaltrials.gov.

* NCI realizes it is providing partial support for one of the questioned trials.

* NCI requests and requires data and code showing how signatures were devised and validated.

May 2010 (Us, some Public)

* *The Cancer Letter* obtains a copy of the Duke investigation's experts' report from the NCI under FOIA. We (and others) note our Nov 2009 report is not mentioned.

* NCI announces yanking LMS from CALGB 30506.
No detailed explanation is given.

* We talk with ORI.
Can we prove fraud? patient harm?
Is what we can prove misconduct vs questionable practice?

Jun 2010 (NCI/Duke, not Public)

Jun 29

* NCI meets with Duke officials and PIs in Bethesda.

* NCI: data and code supplied do not justify the trial being run.

* NCI mandates PIs search for original data claimed to justify the trial.

* Per testimony, Duke officials first appreciate the nature and extent of NCI's concerns at this meeting.

Jul 2010 – the Dam Breaks

Fri, Jul 16: *the Cancer Letter* breaks Rhodes Scholar story; Potti placed on administrative leave.

This *Cancer Letter* issue is widely circulated to members of the biostat/bioinformatics community that evening.

Mon, Jul 19: Letter from 31 prominent Biostatisticians to NCI Director Harold Varmus requesting trials be suspended until science is publicly clarified.

* Duke trials are resuspended
* NCI, Duke request IOM involvement
* Credentialing investigation initiated

Later 2010

Oct: first call for paper retractions following NCI mandated data recheck. Errors match those we reported in Nov 2009.

Nov: Duke trials officially terminated.
CancerGuide Dx folds.

Dec: IOM review begins (Gil Omenn chairing)
Lisa McShane presents NCI view, NCI releases documents

* Duke TMQF Initiated (Rob Califf chairing)

2011

Jan: FDA inspection of Duke trials.

Mar: “nonrandom data corruption” acknowledged.
Duke TMQF draft out for comment.

Jun: NCI workshop on using Omics-Based predictors in clinical trials.

2012, Summary, and Going Forward

Feb: 60 Minutes – Nevins claims fraud is “abundantly clear”.

Mar: IOM report released

Apr: AACR

Jun: ASCO

* 117 patients enrolled in the three Duke trials

* 10 papers retracted (incl NEJM, Nat Med, JAMA, Lancet Oncology, PLoS One, JCO)

* Duke TMQF implementation in Development

Q1: What are the key lessons?

* We Need Data, Code, and Metadata.

* Reproducible Research (reproducing analyses from the same data) needs to be become more of an assumed norm.

* Given data complexity, many problems will not be caught in initial review.

* Journals can potentially sample, say, 1 paper a quarter for an analysis recheck, to be published in the journal.

* Upping the Stakes, Lowering the Bar: Tools are better.

2009: Sweave

2012: knitr, markdown, RStudio, git

Q2: Reaction to the Duke case/IOM report? Progress in implementation?

* Clarified FDA guidance on IDEs/INDs
(esp as they apply to companion diagnostics)

* NCI adopting recs for cooperative group trials

* *Science* Dec 2 2011: "Data Replication and Reproducibility"
* *Nature* advertising for "Chief Editor of Data" (Apr 2012)

* Institutional Compliance Officers and Journal Editors are listening (I've given talks to same).

* Professional Societies examining feasibility.

Q3: To what extent is omics a special case?

In terms of this committee, some of the issues associated with how clinical trials are conducted are a special case.

The issues of increasing data volume and complexity, and the associated need for better tracking, are not unique to omics.

* Ioannidis et al. (2009), *Nat. Gen.*, **41**:149-55.
* Begley and Ellis (2012), *Nature*, **483**:531-3.
* Retraction Watch

Q4: What questions does this case raise about federal agency involvement in dealing with potential misconduct?

What information can be shared, and with whom?

* NCI/Institution
* NCI/DoD
* NCI/ORI
* ORI: when should it be involved?

Other Points/Questions

Misconduct and Questionable Research Practices

Irreproducibility can shift from being questionable to being misconduct. This shift may be linked to the resources involved.

Investigating at Misconduct-level Stringency

Errors are far easier to prove and establish than misconduct. If errors are sufficiently numerous, then checking should involve documented reproduction as well.

What Would We Do Differently Today?

Primarily, we'd be more active about seeking direct contact – e.g., with editors about why we're writing correspondences, with deans about what our objections are.

Acknowledgements

Kevin Coombes

Shannon Neeley, Jing Wang
David Ransohoff, Gordon Mills
Jane Fridlyand, Lajos Pusztai, Zoltan Szallasi

M.D. Anderson Ovarian, Lung and Breast SPOREs

For updates:
http://bioinformatics.mdanderson.org/Supplements/ReproRsch-All/Modified