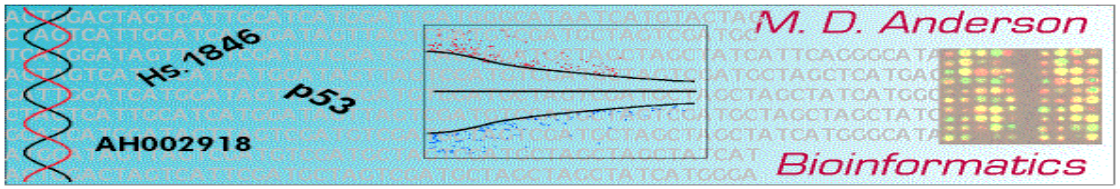
Reproducibility, Ethics, and Big Data: Lessons from a Train Wreck

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

SISBID, July 20, 2015



Why is Reproducible Research Important with Big Data?

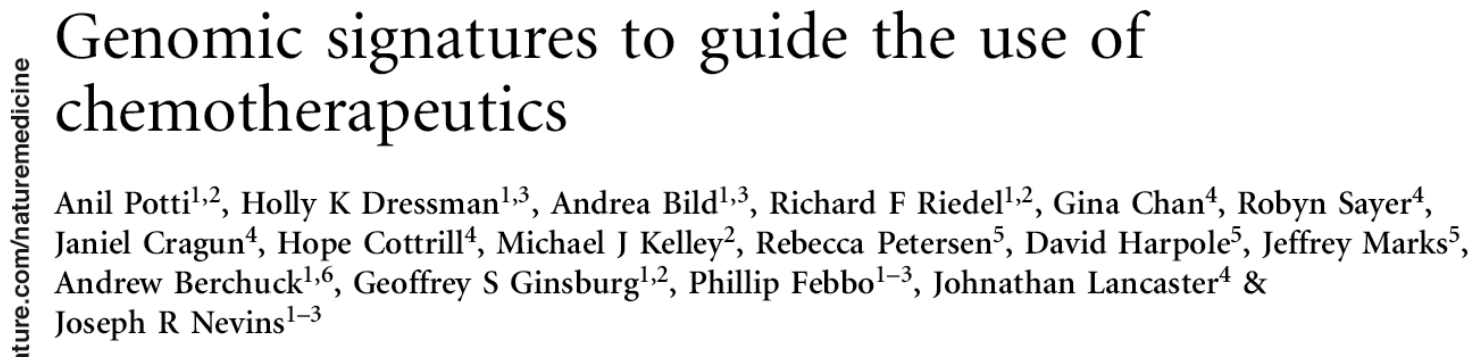
Our intuition about what "makes sense" is very poor in high dimensions.

To use "omics-based signatures" as biomarkers, we need to know they've been assembled correctly.

Without documentation, we may need to employ (lengthy!) *forensic bioinformatics* to infer what was done.

Let's look at examples in the context of a specific problem: *can we predict which patients will respond to which chemotherapeutics?*

Using Cell Lines to Predict Sensitivity



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

The main conclusion: we can use microarray data from cell lines (the NCI60) to define drug response “signatures”, which can predict whether patients will respond.

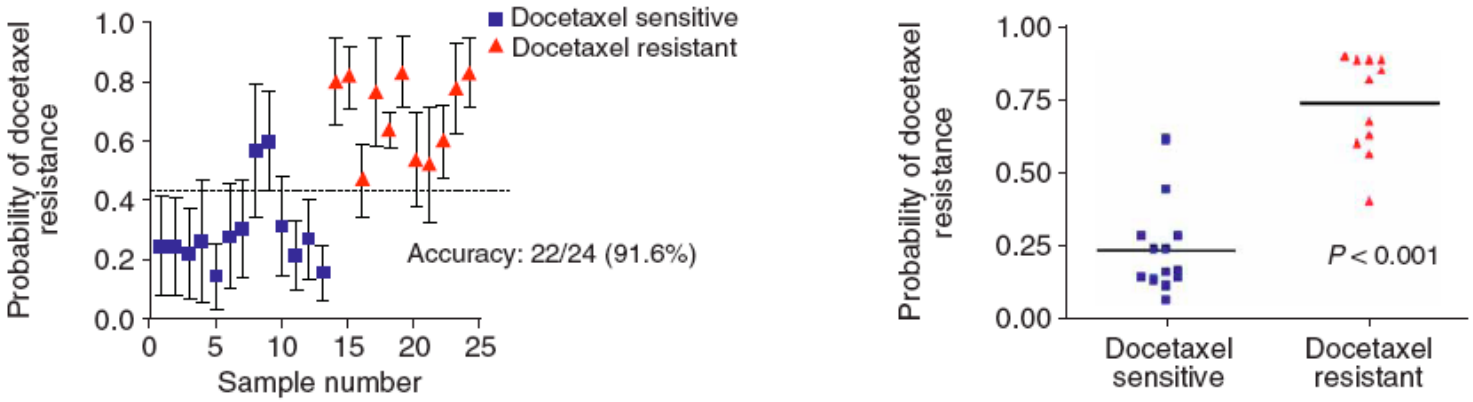
They provide examples using 7 commonly used agents.

This got people at MDA very excited.

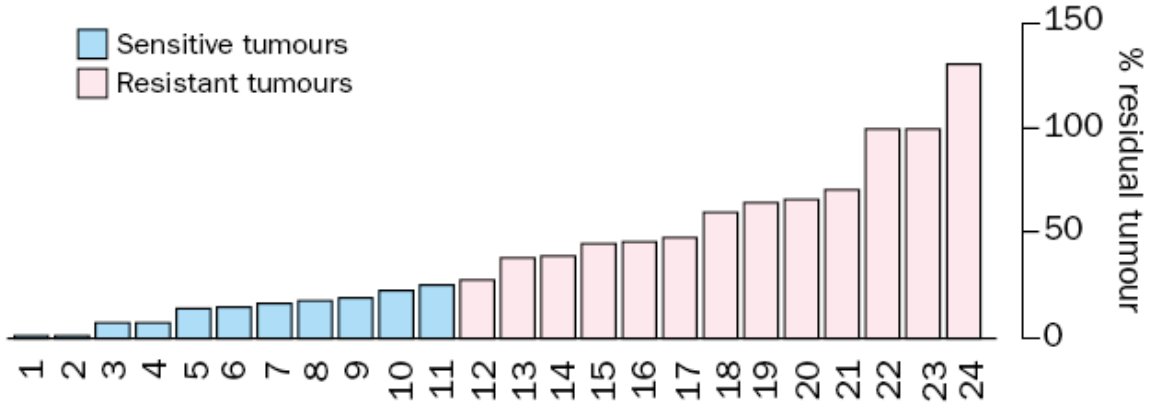
Their Gene List and Ours

```
> 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"
...
```

Predicting Response: Docetaxel

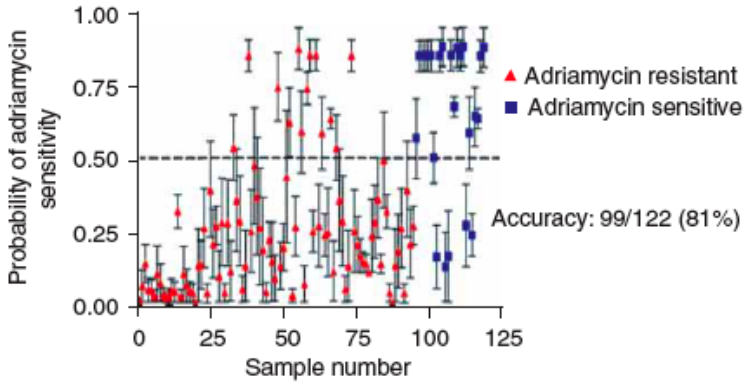


Potti et al, Nat Med 2006, 12:1294-300, Fig 1d

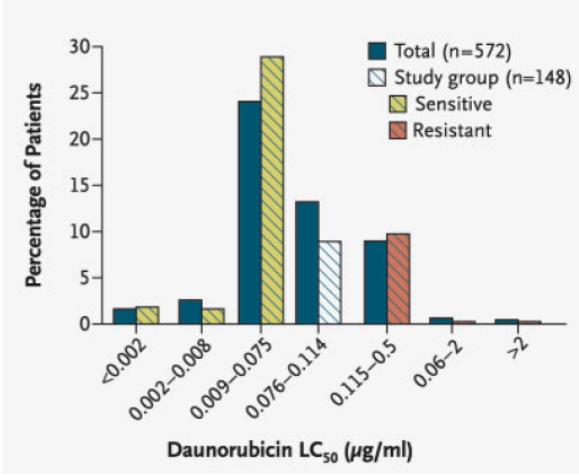


Chang et al, Lancet 2003, 362:362-9, Fig 2 top

Predicting Response: Adriamycin

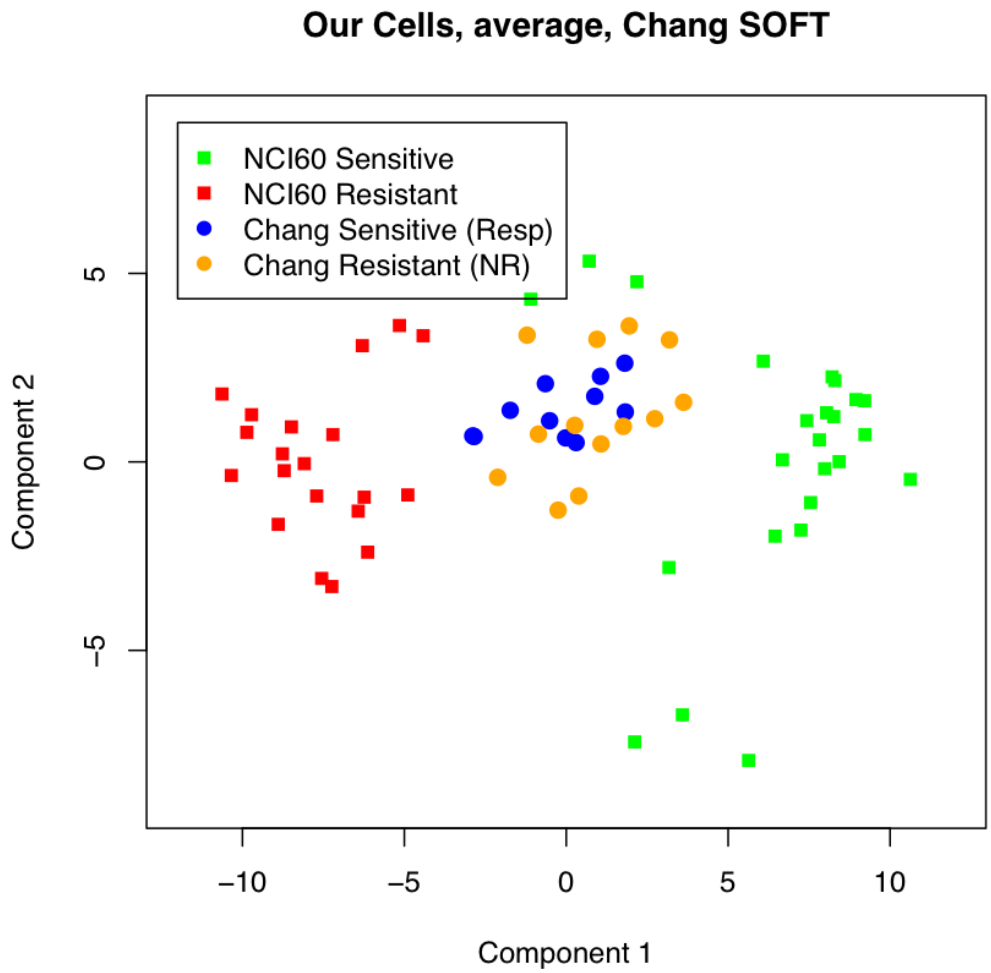


Potti et al, Nat Med 2006, 12:1294-300, Fig 2c



Holleman et al, NEJM 2004, 351:533-42, Fig 1

Trying it Ourselves



When we try it, *it doesn't work.*

Partial Timeline

2006:

* Nov 8: Our first questions to Potti and Nevins.

* Nov 21: Our first report describing errors.

* Nov-Dec: More reports/questions: Nov 27, Dec 4, 13, 27.

2007:

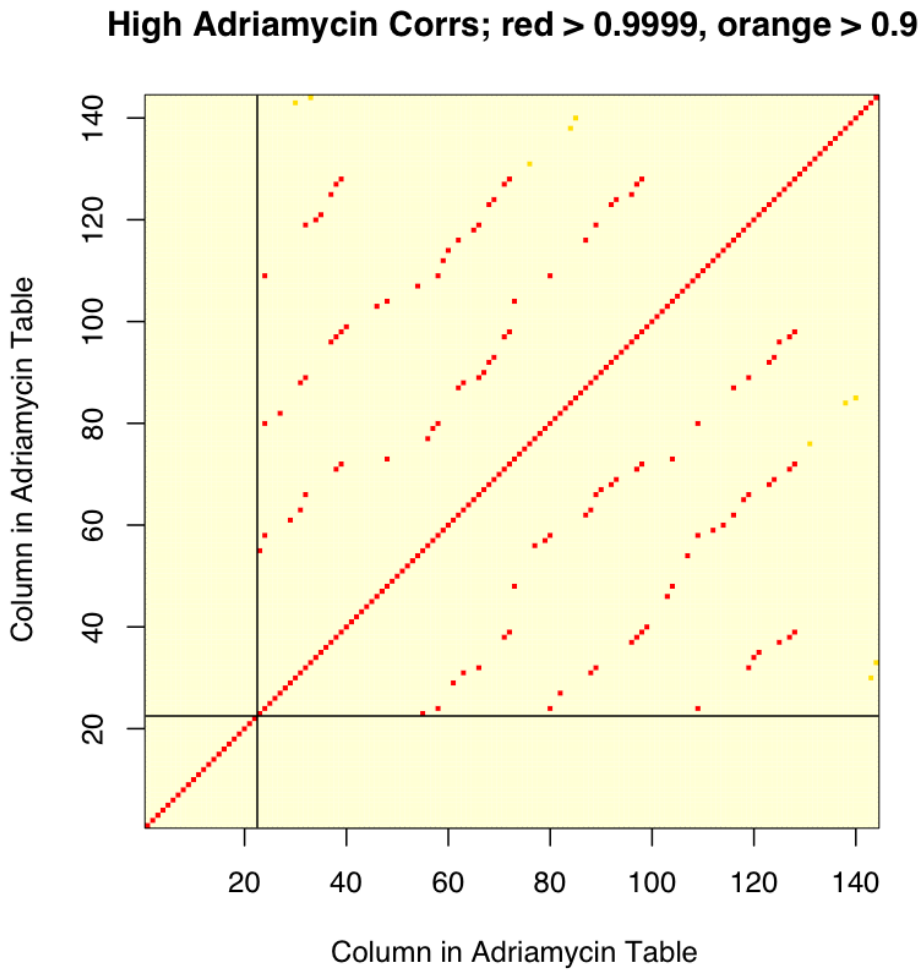
* Jan 24: We meet with Nevins at M.D. Anderson. We urge him to review the data.

* Feb-Apr: New data and code are posted. Some numbers change. We tell them we don't think it works.

* Apr 25: We send Potti and Nevins a draft for comment.

* May: We find problems with outliers. Potti and Nevins continue to insist it works, and want to "bring this to a close".

Adriamycin 0.9999+ Correlations (Reply)



Redone Aug 08, “using ... 95 unique samples”.

Validation 1: Hsu et al

Pharmacogenomic Strategies Provide a Rational Approach to the Treatment of Cisplatin-Resistant Patients With Advanced Cancer

David S. Hsu, Bala S. Balakumaran, Chaitanya R. Acharya, Vanja Vlahovic, Kelli S. Walters, Katherine Garman, Carey Anders, Richard F. Riedel, Johnathan Lancaster, David Harpole, Holly K. Dressman, Joseph R. Nevins, Phillip G. Febbo, and Anil Potti

J Clin Oncol, Oct 1, 2007, 25:4350-7.

Same approach, using Cisplatin and Pemetrexed.

For cisplatin, U133A arrays were used for training. ERCC1, ERCC4 and DNA repair genes are identified as “important”.

With some work, we matched the heatmaps. (Gene lists?)

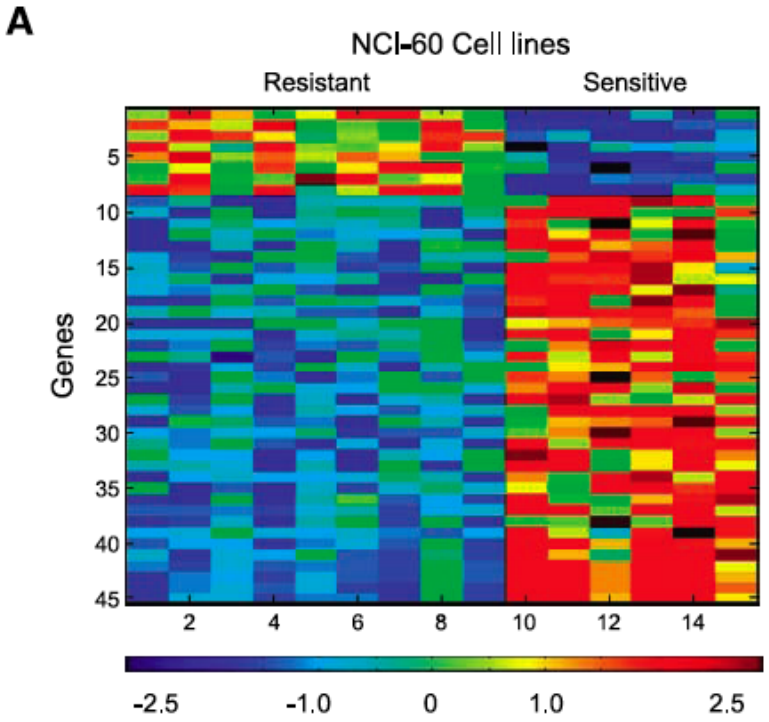
The 4 We Can't Match (Reply)

203719_at, ERCC1,
210158_at, ERCC4,
228131_at, ERCC1, and
231971_at, FANCM (DNA Repair).

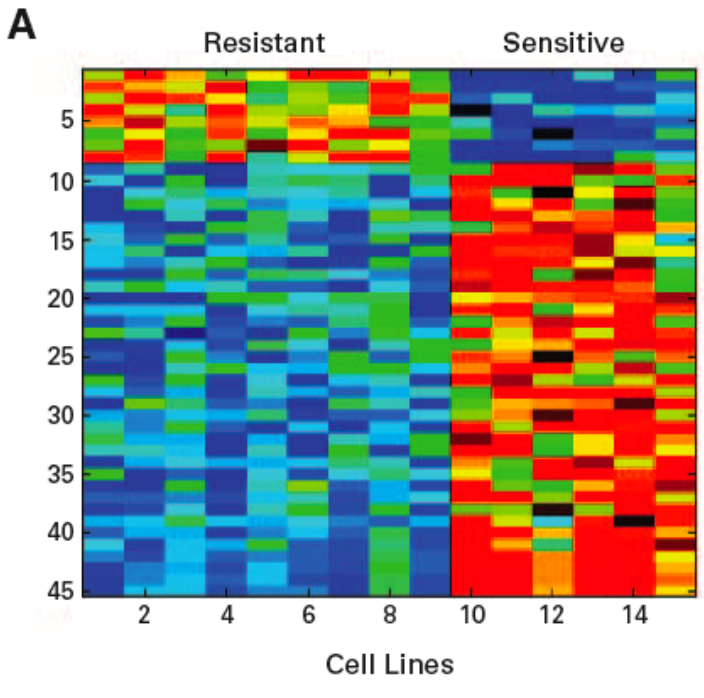
Another problem –

The last two probesets aren't on the U133A arrays that were used. They're on the U133B.

Temozolomide Heatmaps



Augustine et al., 2009, *Clin Can Res*, **15**:502-10, Fig 4A. Temozolomide, NCI-60.



Hsu et al., 2007, *J Clin Oncol*, **25**:4350-7, Fig 1A. Cisplatin, Gyorffy cell lines.

The Reason We Really Care

Jun 2009: we learn clinical trials had begun.
2007: pemetrexed vs cisplatin, pem vs vinorelbine.
2008: docetaxel vs doxorubicin, topotecan vs dox (Moffitt).

Sep 1, 2009: We submit a paper describing case studies to the *Annals of Applied Statistics*.

Sep 14, 2009: Paper accepted and available online at the *Annals of Applied Statistics*.

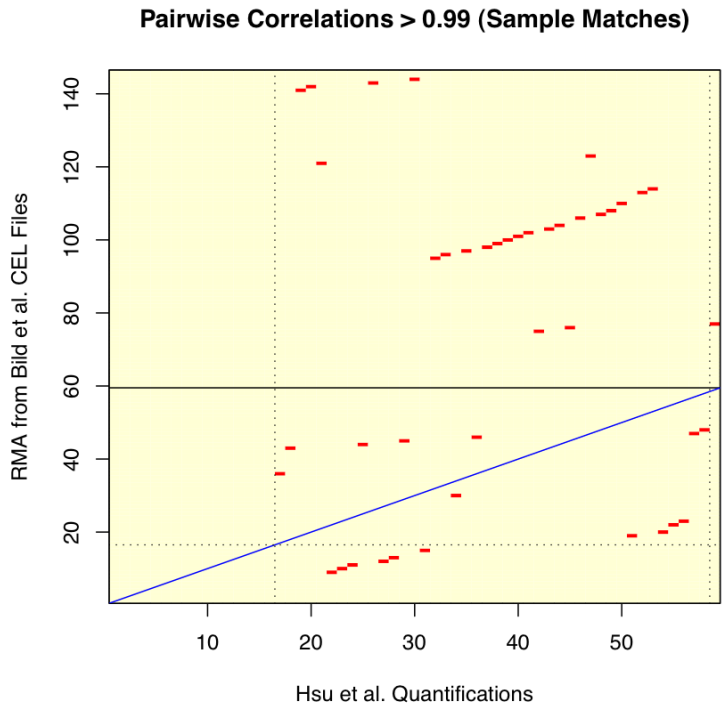
Sep-Oct 2009: Story covered by *The Cancer Letter*.
NCI raises concerns with Duke's IRB behind the scenes.
Duke starts internal investigation, suspends trials.

New Data

Early-Nov ’09 (mid-investigation), the Duke team posted new data for cisplatin and pemetrexed (in lung trials since ’07).

These included quantifications for the 59 ovarian cancer test samples (from GSE3149, which has 153 samples) they used to validate their predictor.

We Tried Matching The Samples



43 samples are mislabeled.

16 samples don't match because the genes are mislabeled.

All of the validation data are wrong.

We reported this to Duke and to the NCI in mid-November.

Jan 29, 2010

THE CANCER LETTER

PO Box 9905 Washington DC 20016 Telephone 202-362-1809

Duke In Process To Restart Three Trials Using Microarray Analysis Of Tumors

By Paul Goldberg

Duke University said it is in the process of restarting three clinical trials using microarray analysis of patient tumors to predict their response to chemotherapy.

Their investigation's results *"strengthen ... confidence in this evolving approach to personalized cancer treatment."*

We Asked for the Data

"While the reviewers approved of our sharing the report with the NCI, *we consider it a confidential document*" (Duke). A *future paper* will explain the methods.

This did give us one more option...

In May 2010, we obtained a copy of the reviewers' report from the NCI under FOIA.

In our assessment, it did not justify restarting trials.

There was no mention of our Nov 2009 report.

A Catalyzing Event: July 16, 2010

THE CANCER LETTER

PO Box 9905 Washington DC 20016 Telephone 202-362-1809

Prominent Duke Scientist Claimed Prizes He Didn't Win, Including Rhodes Scholarship

By Paul Goldberg

Jul 19/20: Letter to Varmus; Duke resuspends trials.

Oct 22/9: First call for paper retraction.

Nov 9: Duke terminates trials.

Nov 19: call for Nat Med retraction, Potti resigns

Other Developments

117 patients were enrolled in the trials.
Sep, 2011: Patient lawsuits filed (11+ settlements).

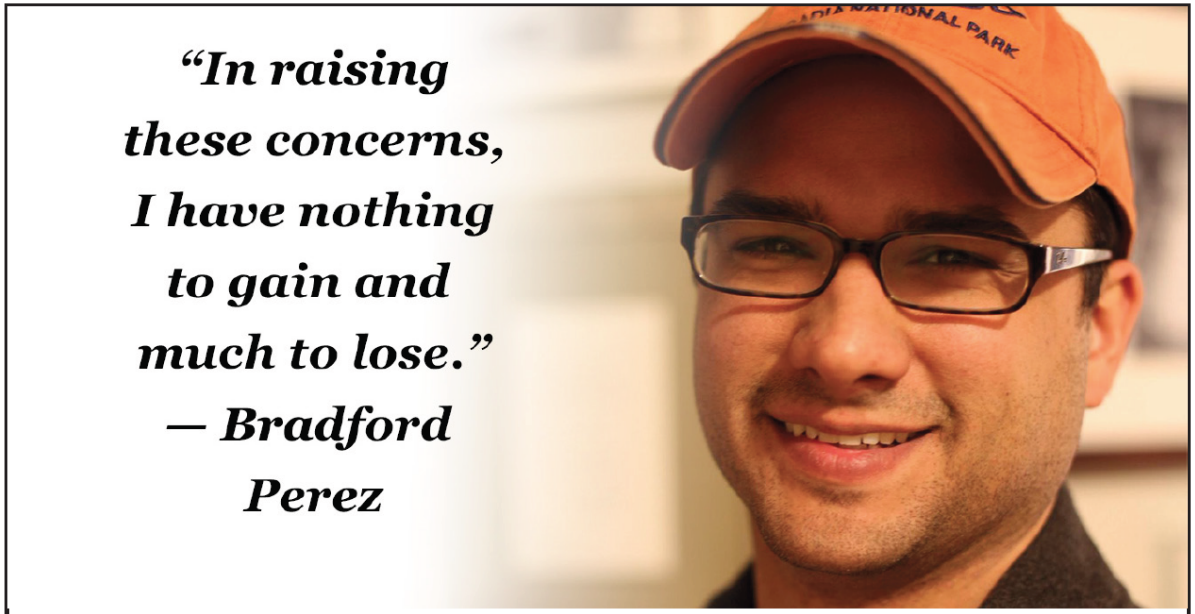
Misconduct investigation (ongoing).
FDA Review, Discussions with Duke IRB

Jul 8, 2011: Front Page, NY Times.
Feb 12, 2012: 60 Minutes.
http://www.cbsnews.com/8301-18560_
162-57376073/deception-at-duke/

Mar 23, 2012: IOM Report Released.
http://www.iom.edu/Reports/2012/
Evolution-of-Translational-Omics.aspx

What Would You Do?

Jan. 9, 2015 • www.cancerletter.com •



Internal Emails Raise New Questions

Duke Officials Silenced Med Student Who Reported Trouble in Anil Potti's Lab

By Paul Goldberg

The patient lawsuits settled at the end of April.

Some Cautions/Observations

This case is pathological.

But we've seen similar problems before.

The most common mistakes are simple.

Confounding in the Experimental Design
Mixing up the sample labels
Mixing up the gene labels
Mixing up the group labels
(Most mixups involve simple switches or offsets)

This simplicity is often hidden.

Incomplete documentation

This is not an Isolated Problem

Ioannidis et al. (2009), *Nat. Gen.*, **41**:149-55. Tested reproducibility of microarray papers. Could reproduce 2/18.

Begley and Ellis (2012), *Nature*, **483**:531-3. Amgen attempted validation of clinical “breakthroughs” prior to further study. Validated 6/53.

NCI focus meeting Sep 2012.

Tabak and Collins, Nature 2014.

NAS meeting Feb 26-7, 2015.

What Did We Suggest?

Exploiting a Teachable Moment...

Baggerly et al *Nature* (2010)

Give us your data, your code, your huddled masses

Records of data provenance

Checking existence as a task for journals and reviewers (are there links? are they live?)

What Did Others Suggest?

IOM/NCI/FDA -

Locking down the rules you used (can someone else use it)?

Discussion of biological rationale

What will be measured, and how?

McShane at IOM; Gene Lists as Predictors; AACC; MTEC

NCI Guidelines in *Nature* Oct 2013

How Do We Teach The Importance of RR?

If you give me an hour, this way ;)

Expand on the ethics of the situation - what are the downstream effects of getting this wrong?

See the May 22, 2015 *Cancer Letter* for an interview with the last surviving trial participant who was party to the lawsuits

Expand on how to use the tools (this needs a workshop)

Reasons for Hope

1. Our Own (Evolving!) Experience
2. Better tools (knitr, Markdown, GitHub)
3. Journals, Code and Data
4. The IOM, the FDA, and IDEs*
5. The NCI and Trials it Funds
6. OSTP, Congress, Science, Nature

Reasons to Do It

TCGA

TARGET

CCLE

CTD2

Achilles

cBIO

and more!

Acknowledgments

Kevin Coombes

Yang Zhao, Ying Wang, Shelley Herbrich
Shannon Neeley, Jing Wang
David Ransohoff, Gordon Mills
Jane Fridlyand, Lajos Pusztai, Zoltan Szallasi

M.D. Anderson Ovarian, Lung and Breast SPOREs

Baggerly and Coombes (2009), *Annals of Applied Statistics*, **3(4)**:1309-34.

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

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

Thanks!