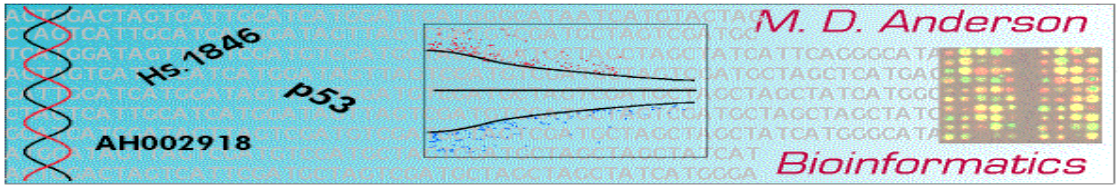
Reproducibile Research and "Omics": Thoughts from an IOM Review

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

SIAM CSE, Feb 28, 2013



Why is Reproducibility Important in Genomics? With "Big Data" in General?

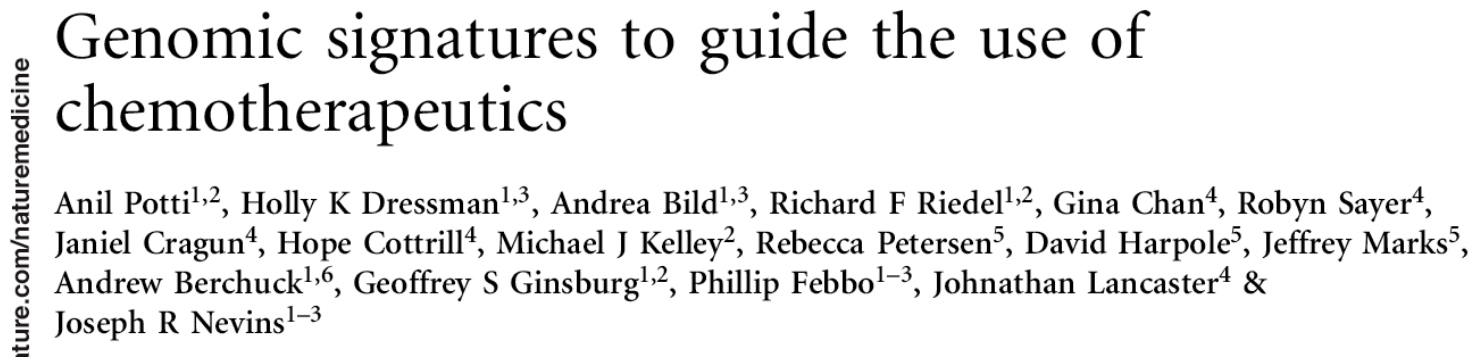
Our intuition about what "makes sense" is very poor in high dimensions.

To use "genomic 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.

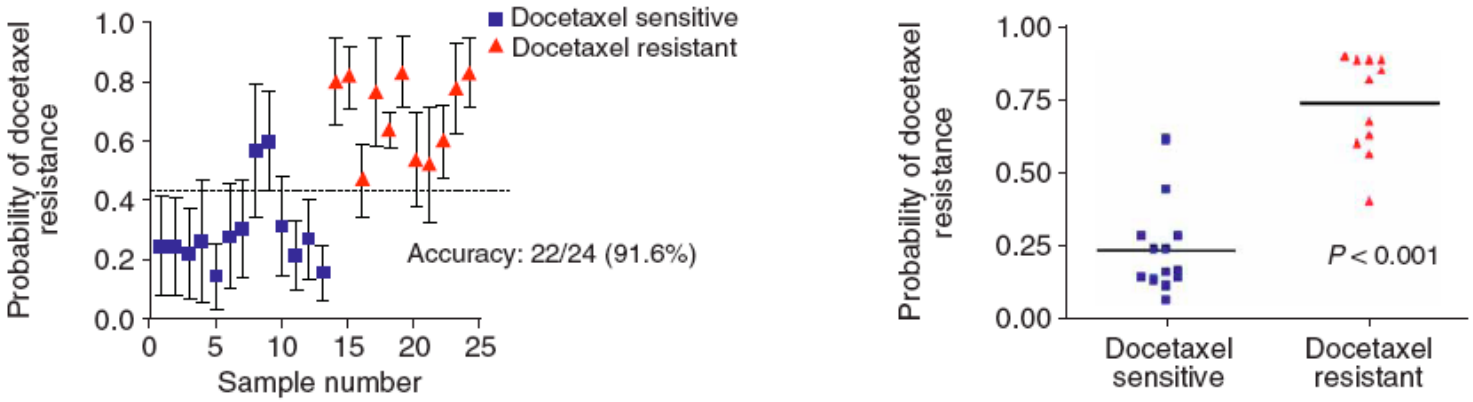
The main conclusion: we can use microarray data from cell lines (the NCI60) to define drug response “signatures”, which can predict whether patients will respond.

They provide examples using 7 commonly used agents.

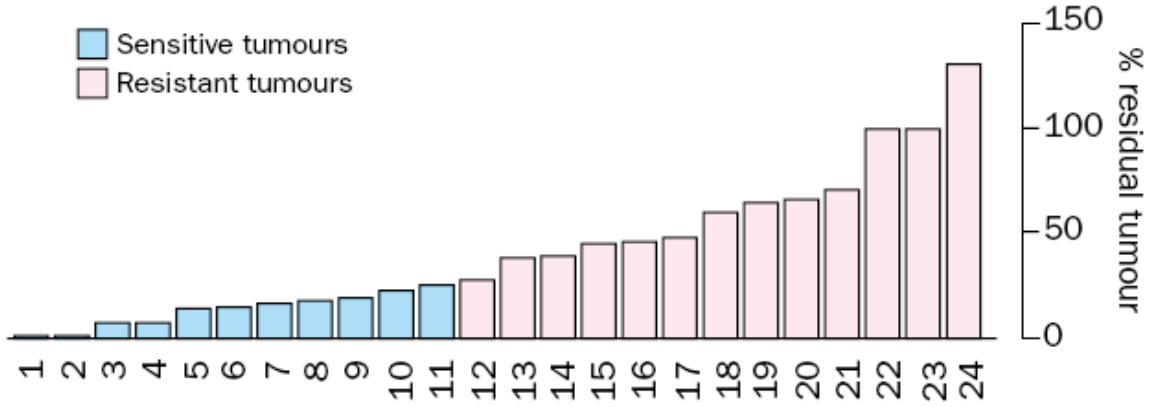
Their Gene List and Ours

```
> 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

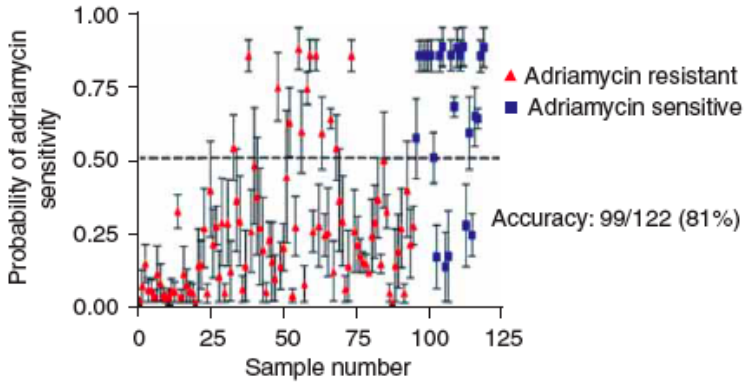


Potti et al, Nat Med 2006, 12:1294-300, Fig 1d

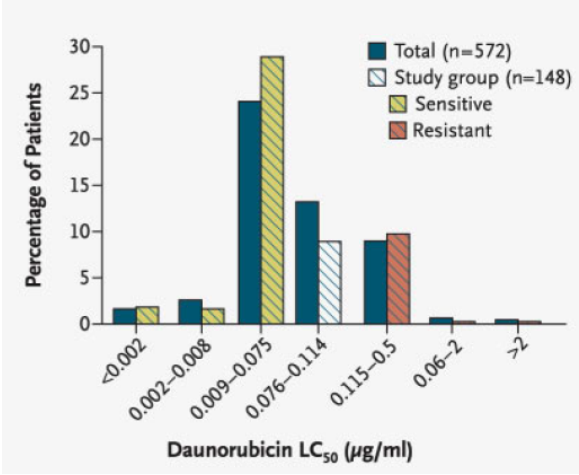


Chang et al, Lancet 2003, 362:362-9, Fig 2 top

Predicting Response: Adriamycin

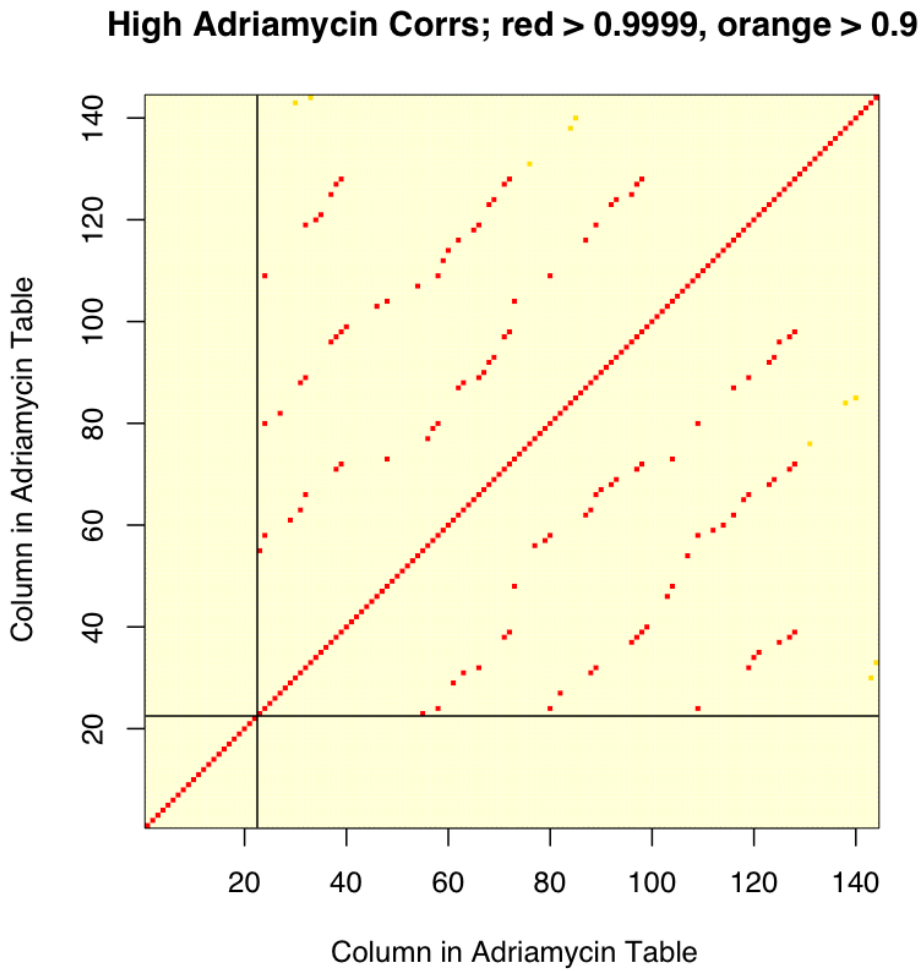


Potti et al, Nat Med 2006, 12:1294-300, Fig 2c



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

Adriamycin 0.9999+ Correlations (Reply)



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

Extensions to Combination Therapy

Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial

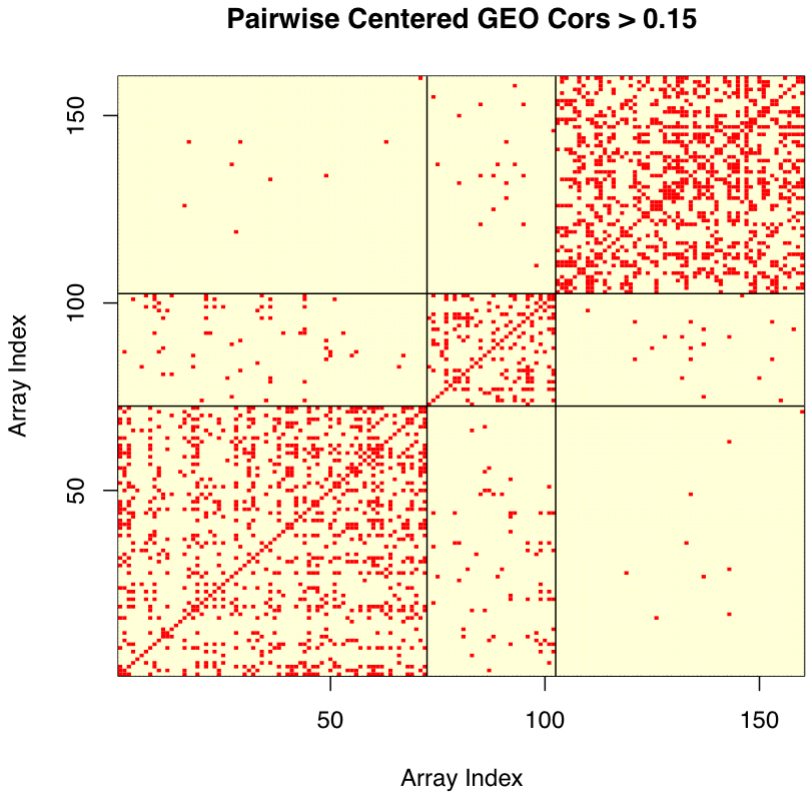
Hervé Bonnefoi, Anil Potti, Mauro Delorenzi, Louis Mauriac, Mario Campone, Michèle Tubiana-Hulin, Thierry Petit, Philippe Rouanet, Jacek Jassem, Emmanuel Blot, Véronique Becette, Pierre Farmer, Sylvie André, Chaitanya R Acharya, Sayan Mukherjee, David Cameron, Jonas Bergh, Joseph R Nevins, Richard D Iggo

Lancet Oncology, Dec 2007, 8:1071-8. (early access Nov 14)

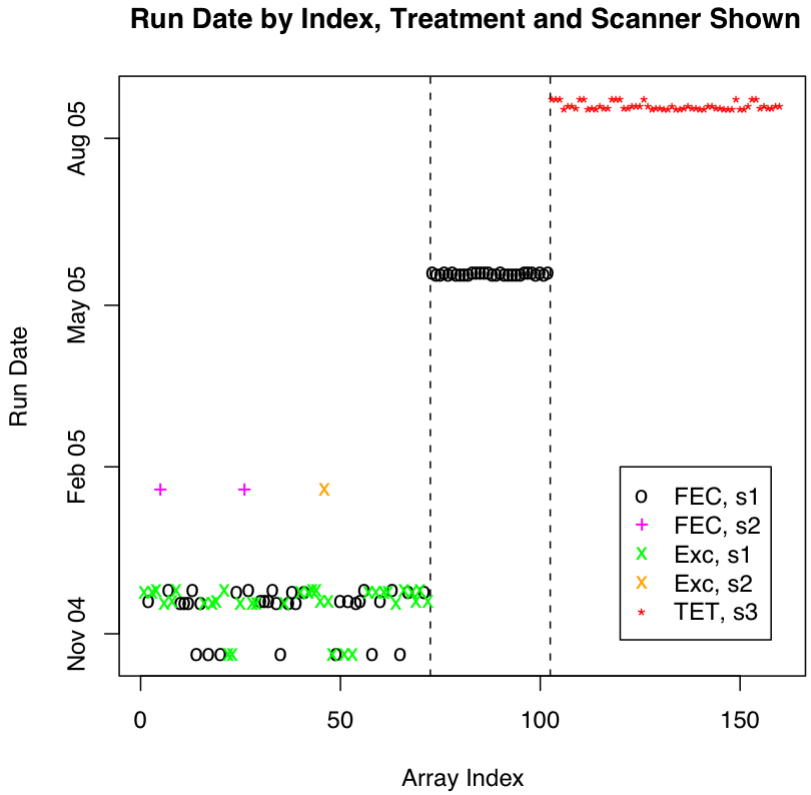
Similar approach, using signatures for Fluorouracil, Epirubicin, Cyclophosphamide, and Taxotere to predict response to combination therapies: FEC and TET.

Potentially improves ER- response from 44% to 70%.

Experimental Design Matters

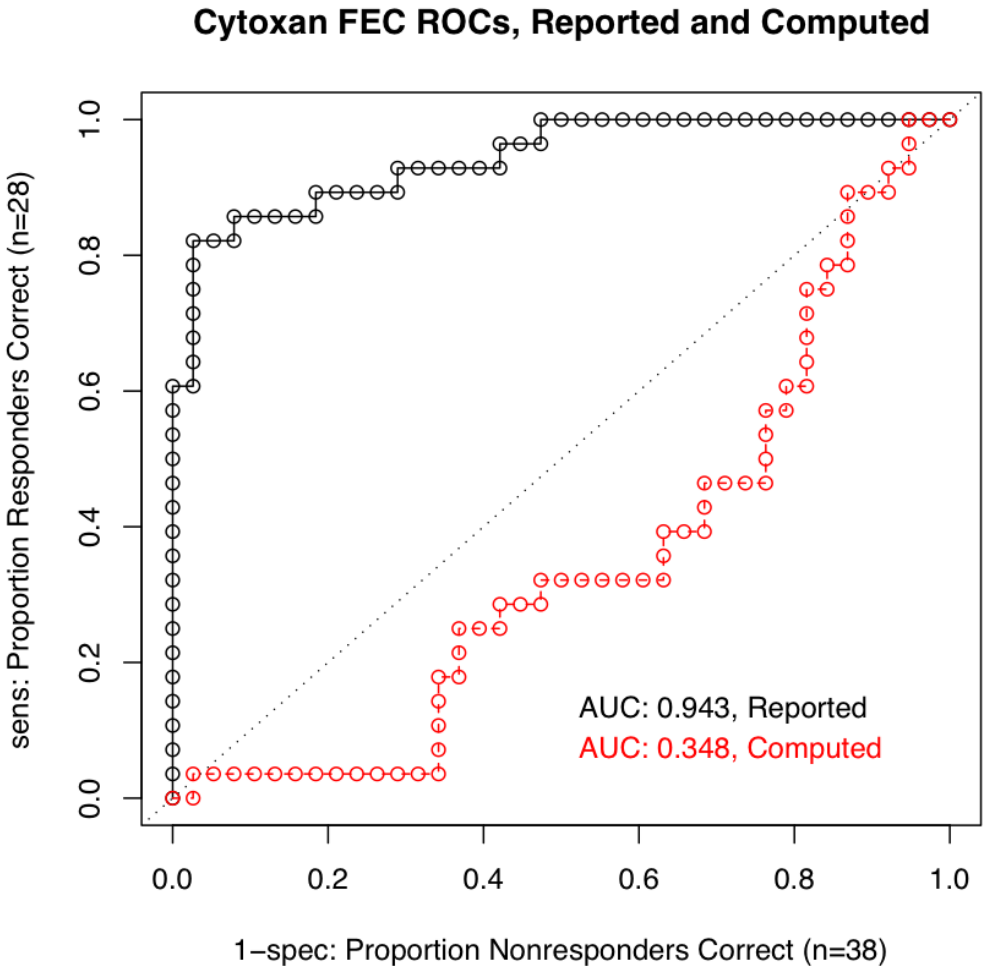


High Sample Correlations after Centering by Gene



Array Run Dates

Predictions for Individual Drugs?



Does cytoxan make sense?

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*.

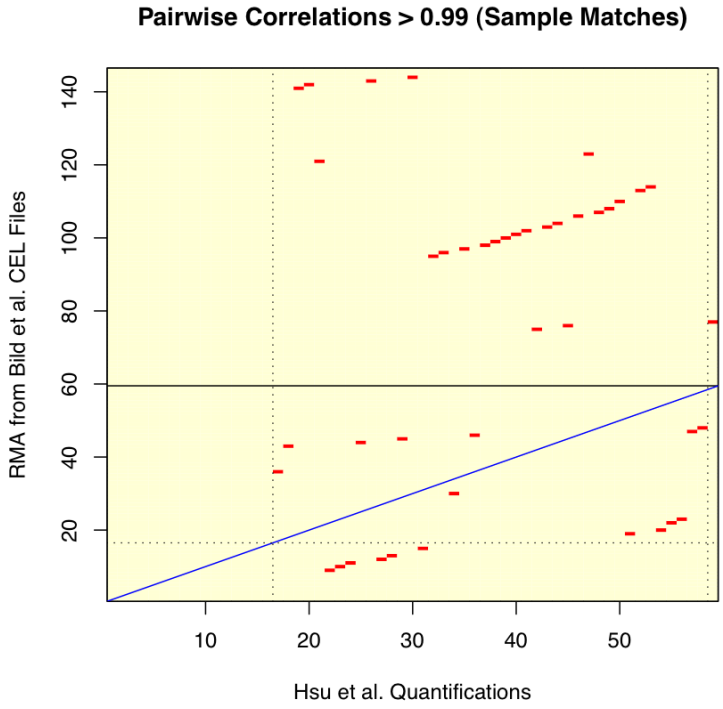
Sep-Oct 2009: Story covered by *The Cancer Letter*.
NCI raises concerns with Duke's IRB behind the scenes.
Duke starts internal investigation, suspends trials.

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."*

A Catalyzing Event: July 16, 2010

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

The IOM Review

Dec 20, 2010: NCI, FDA Presentations.

Mar 30-1, 2011: Case Studies. Joe Nevins presents.
I present. Duke historical document supplied.
Details clarify what happened with our Nov 2009 report.

Jun 30, 2011: NCI Presentation.

Aug 22, 2011: Duke Institutional Response.

Nov 4, 2011: Moffitt trial in *The Cancer Letter*.

Links to MP3 audio, documents, our annotations:

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

Other Developments

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

Misconduct investigation (ongoing).
10 retractions, 6 corrections/partial retractions to date.

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

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.

Subject of an NCI focus meeting in Sep 2012.

One Lesson: What Should the Norm Be?

In our group we've prepared reports in *Sweave* since 2007. (That's changing to knitr/Markdown now.)

For papers? (Baggerly + lots, *Nature*, Sep 22, 2010)

Things we look for:

1. Data (often mentioned, given MIAME)
2. Provenance
3. Code
4. Descriptions of Nonscriptable Steps
5. Descriptions of Planned Design, if Used.

For clinical trials?

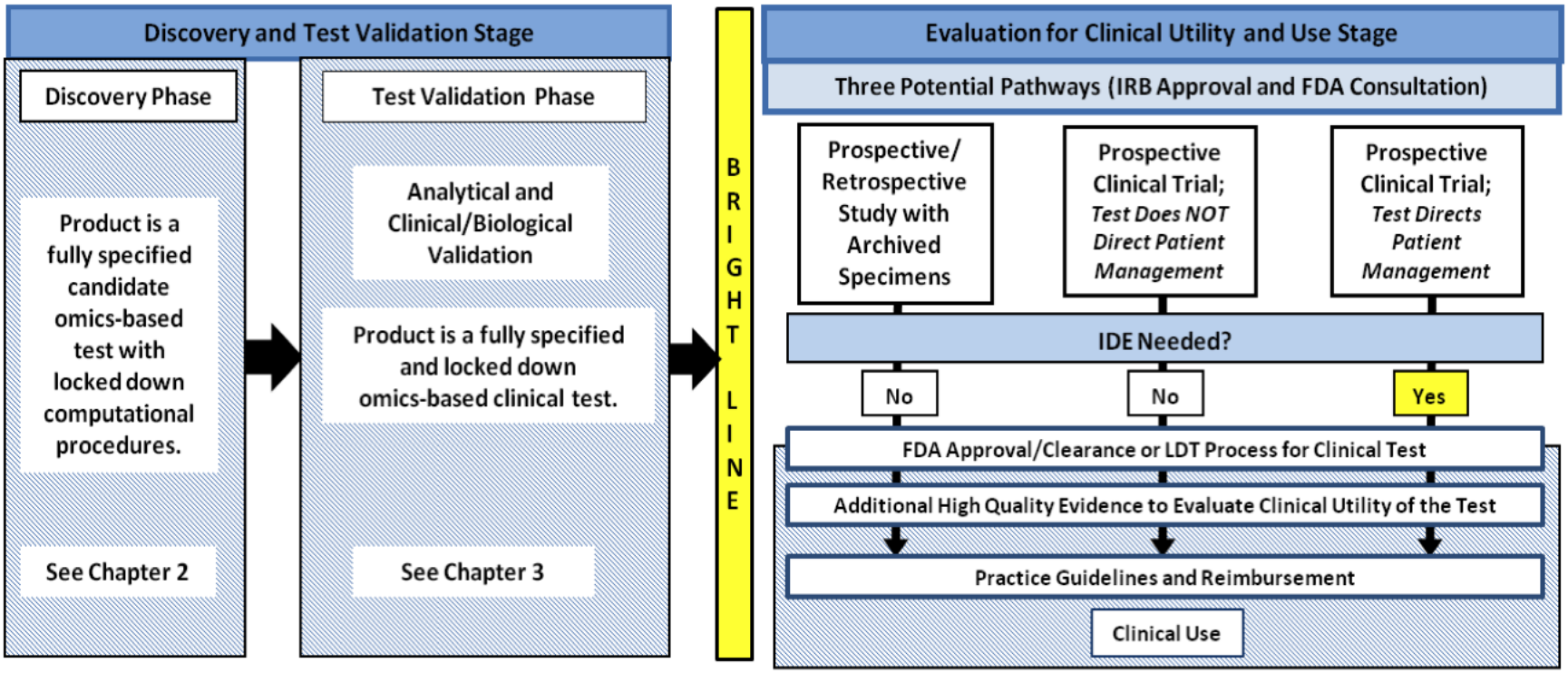
Can We Get the Data?

Ochsner et al. (2008), *Nat Meth*, **5(12)**:991. Deposition rates for raw data in array studies are $< 50\%$.

Witwer (2013), *Clin Chem*, **59(2)**:392-400. Deposition rates for raw data in miRNA studies are $< 50\%$.

Vines et al. (2013), *FASEB J*, Epub ahead of print. Checking at editorial level improves deposition rates.

The IOM Report: A Snapshot



274 pages, outlining best practices and lessons learned

Recs 1-3: Discovery, Validation, Evaluation

Recs 4-7: Institutions, Funders, FDA, Journals

Science, Nature, AACR, ASCO, NCI, AACC

Reasons for Hope

1. Our Own (Evolving!) Experience
2. Better tools (knitr, Markdown, Galaxy, GenePattern/Firehose)
3. Journals, Code and Data
4. The IOM, the FDA, and IDEs*
5. The NCI and Trials it Funds

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

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`.

Index