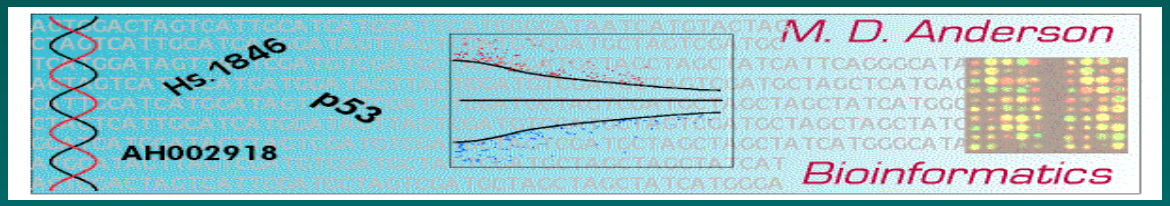
The Importance of Reproducibility in High-Throughput Biology: Case Studies

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

AAAS, February 19, 2011



Why is Reproducibility Important in H-TB?

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 *forensic bioinformatics* to infer what was done to obtain the results.

Let's examine some case studies involving an important clinical problem: *can we predict how a given patient will respond to available chemotherapeutics?*

Using the NCI60 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]

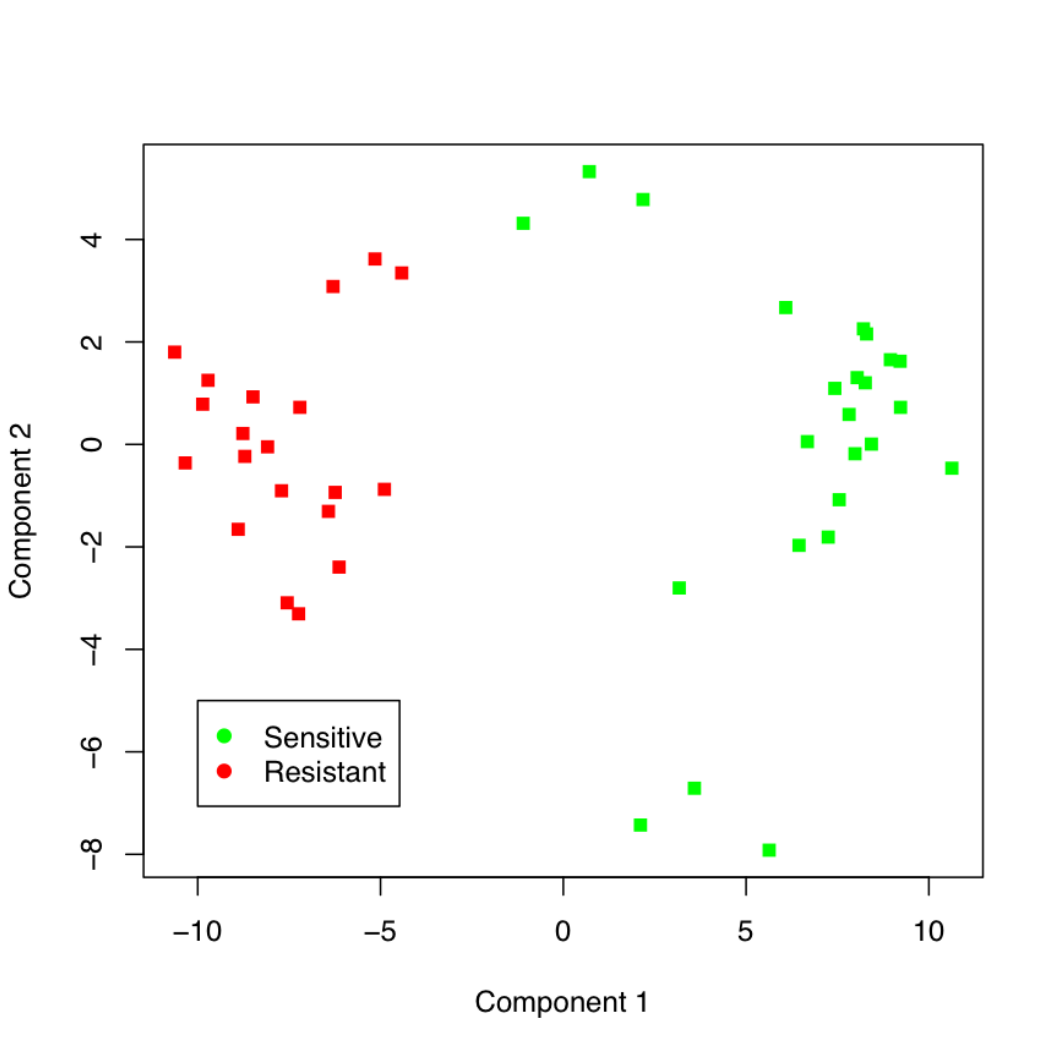
Potti et al (2006), Nature Medicine, 12:1294-1300.

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.

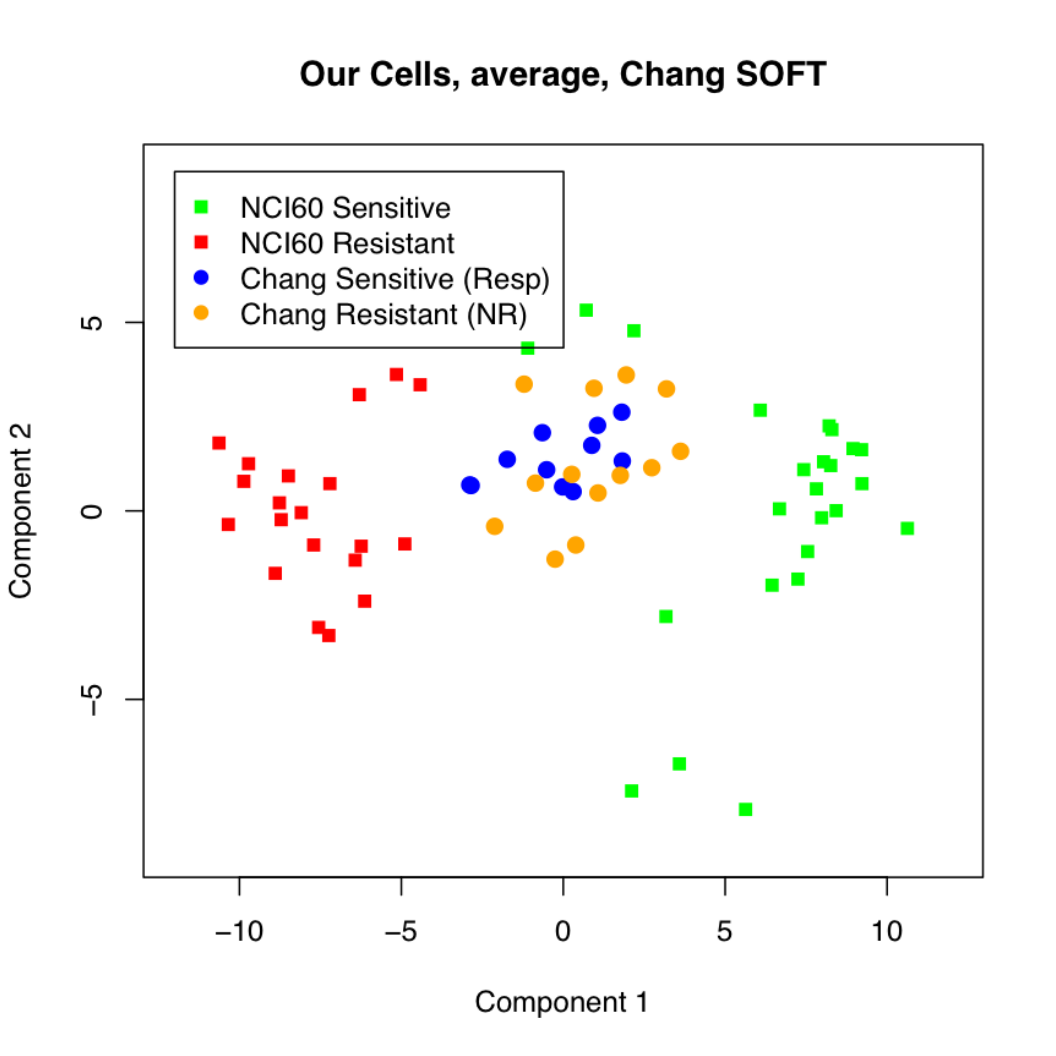
This got people at MDA very excited.

Fit Training Data



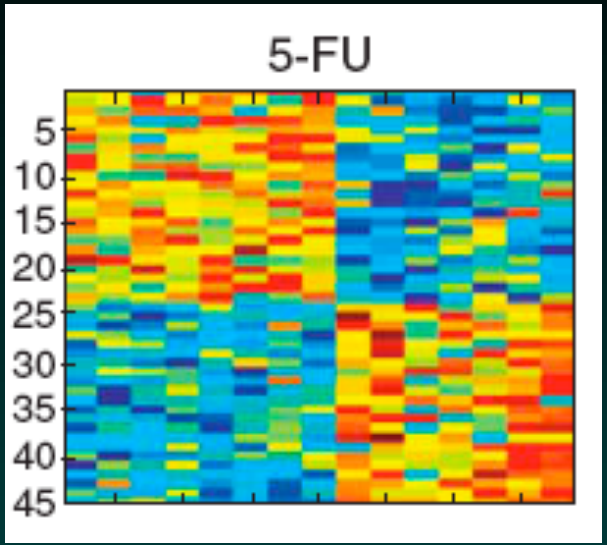
We want the test data to split like this...

Fit Testing Data



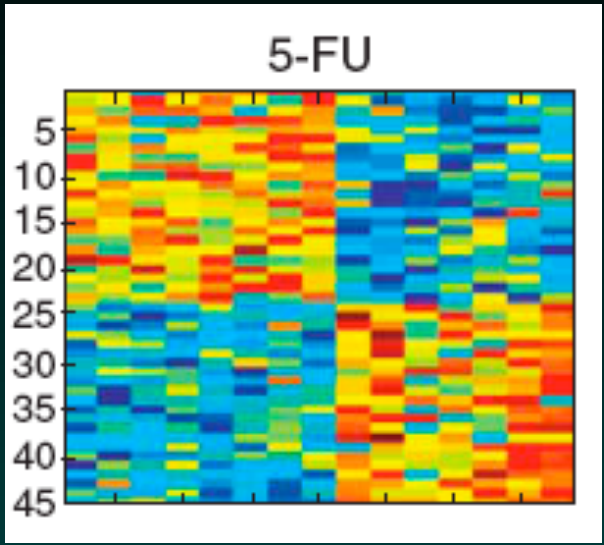
But it *doesn't.* Did we do something wrong?

5-FU Heatmaps

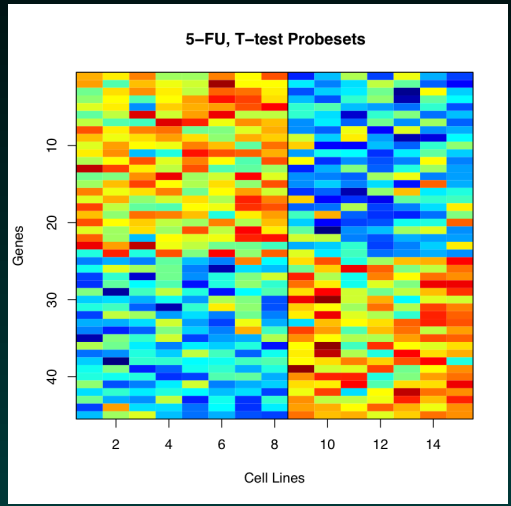


Nat Med Paper

5-FU Heatmaps

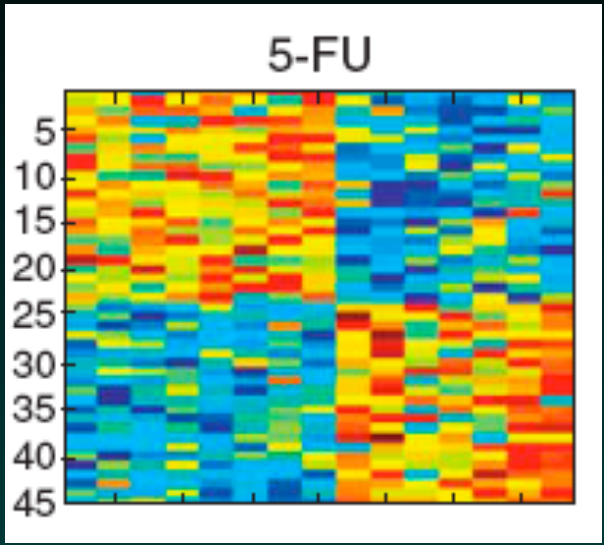


Nat Med Paper

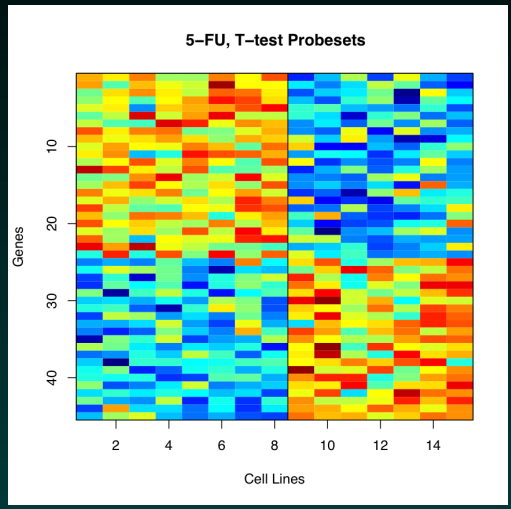


Our t-tests

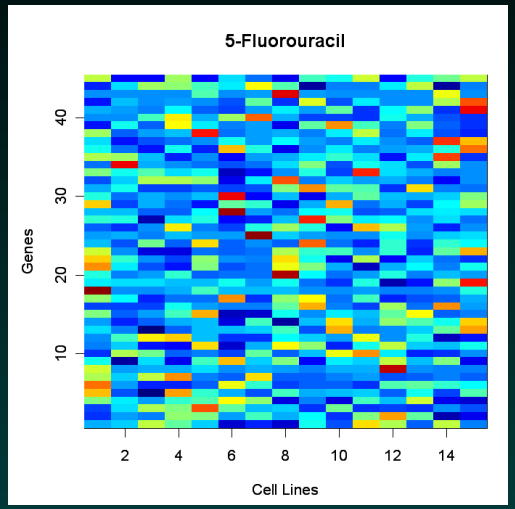
5-FU Heatmaps



Nat Med Paper



Our t-tests

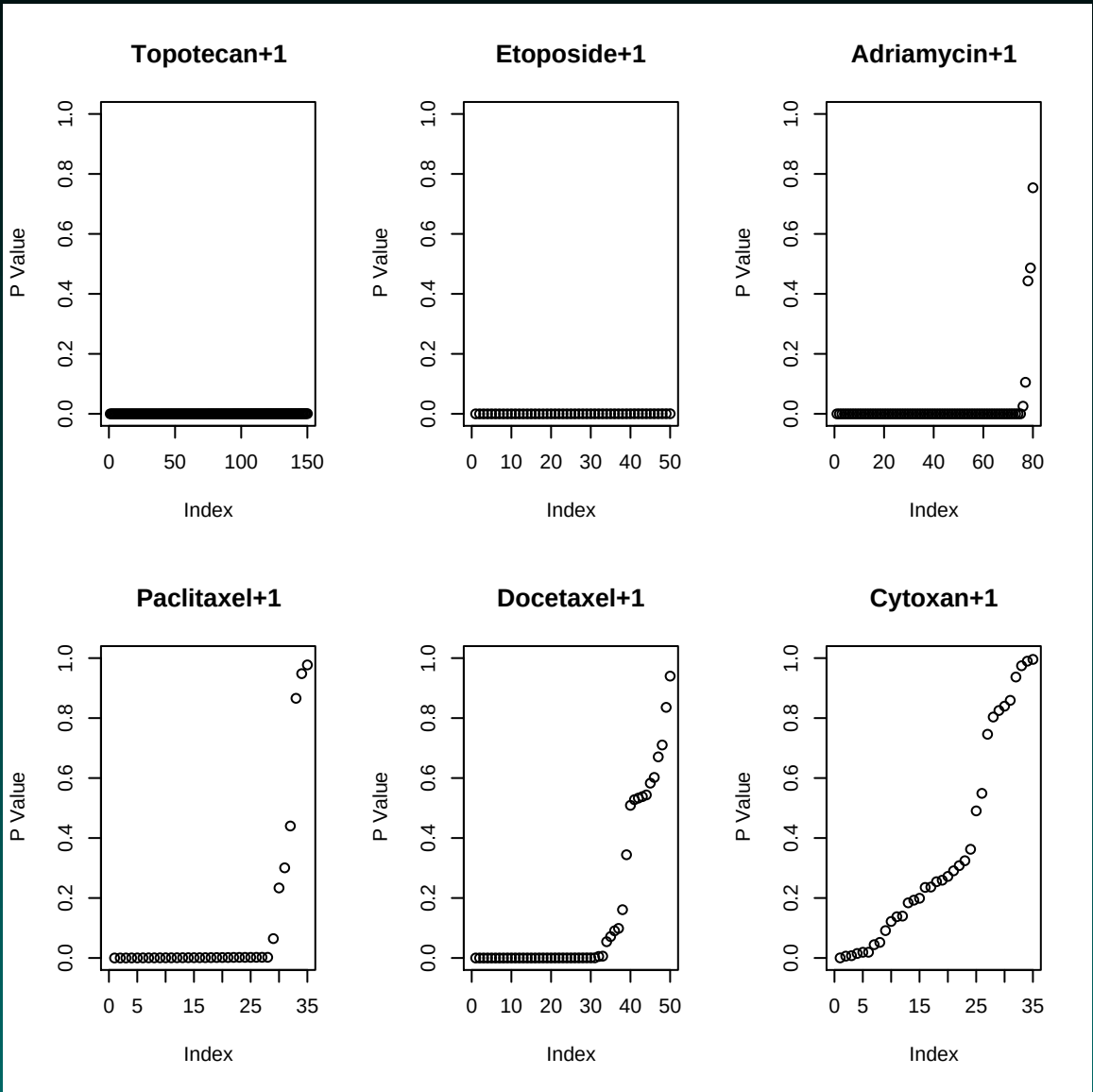


Reported Genes

Their 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"
...
```

Offset P-Values: Other Drugs



Using Their Software

Their software requires two input files:

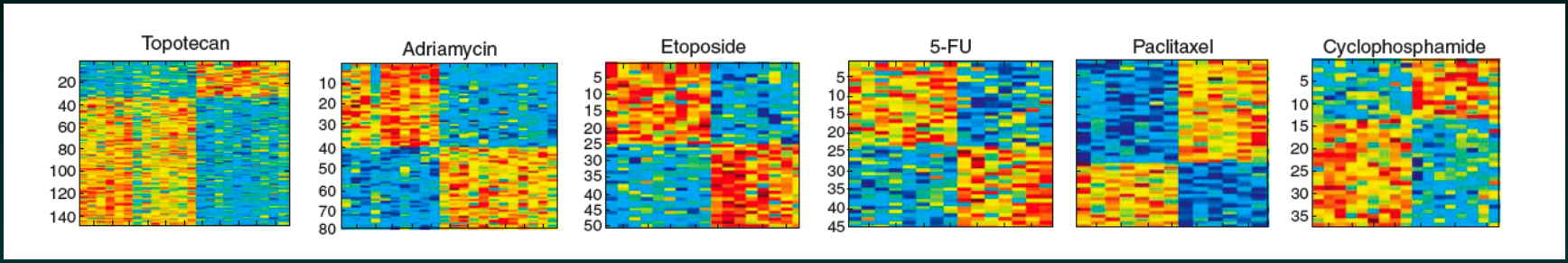
1. *a quantification matrix*, genes by samples, with a header giving classifications (0 = Resistant, 1 = Sensitive, 2 = Test)

2. *a list of probeset ids* in the same order as the quantification matrix. *This list must not have a header row.*

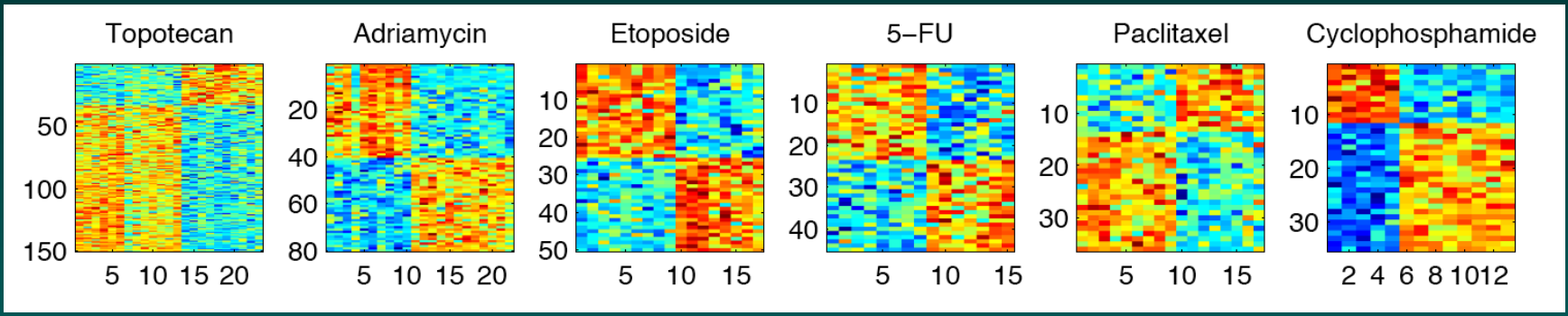
What do we get?

Heatmaps Match Exactly for Most Drugs!

From the paper:

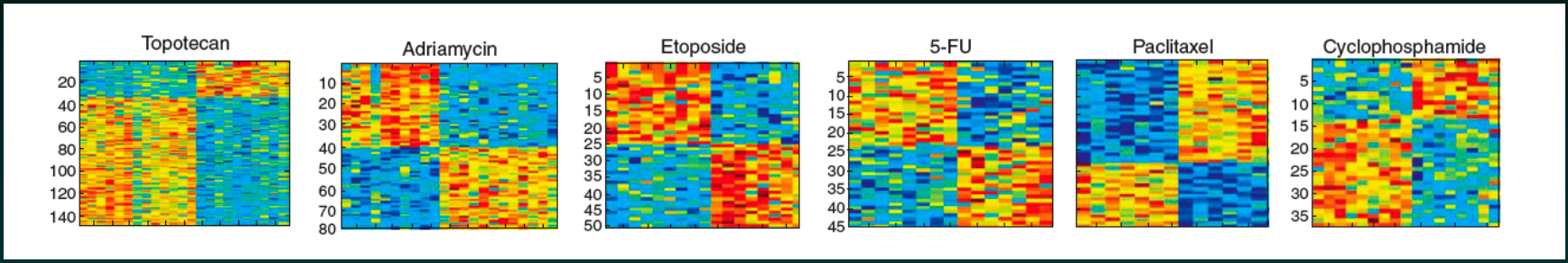


From the software:

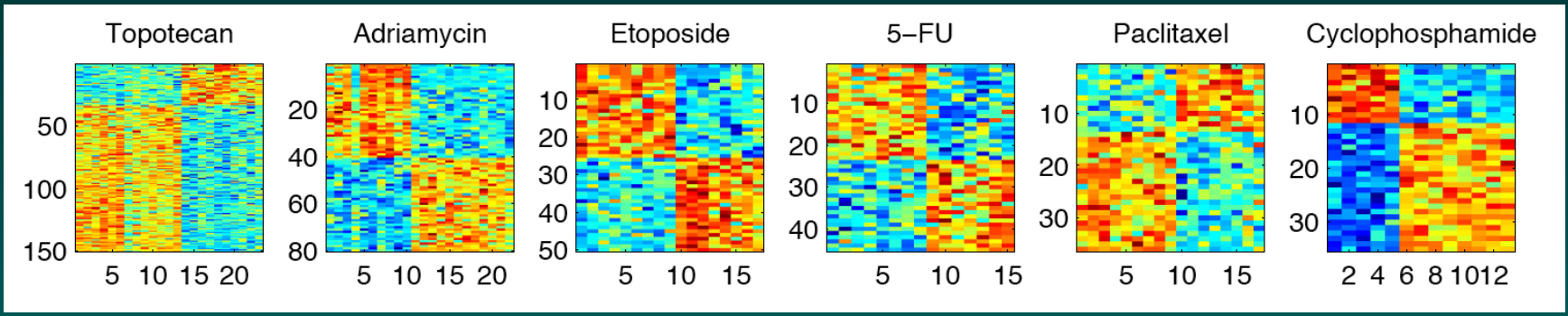


Heatmaps Match Exactly for Most Drugs!

From the paper:

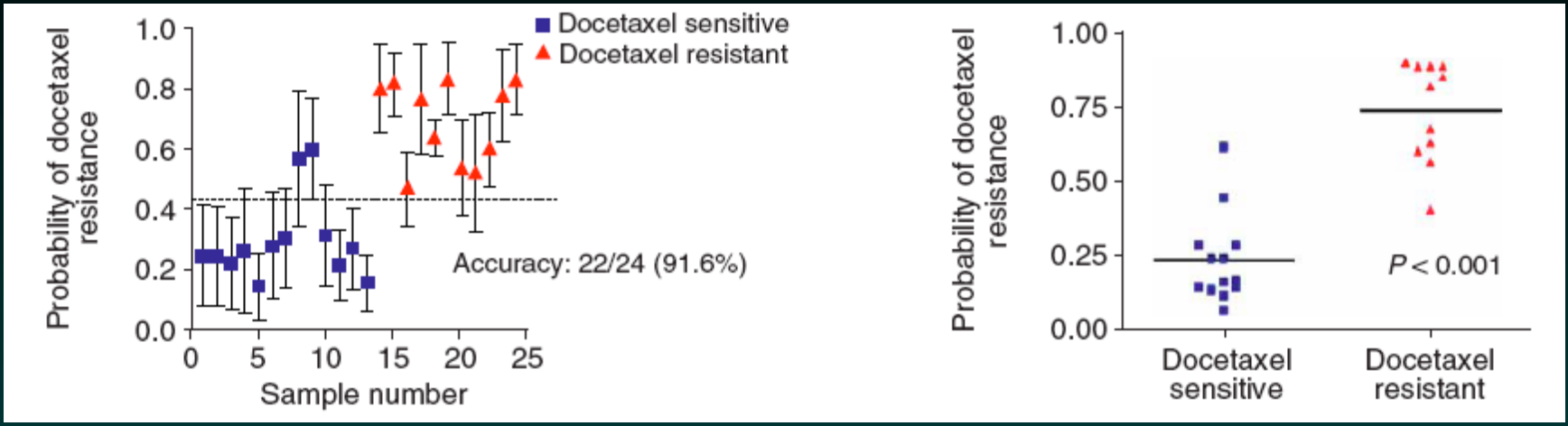


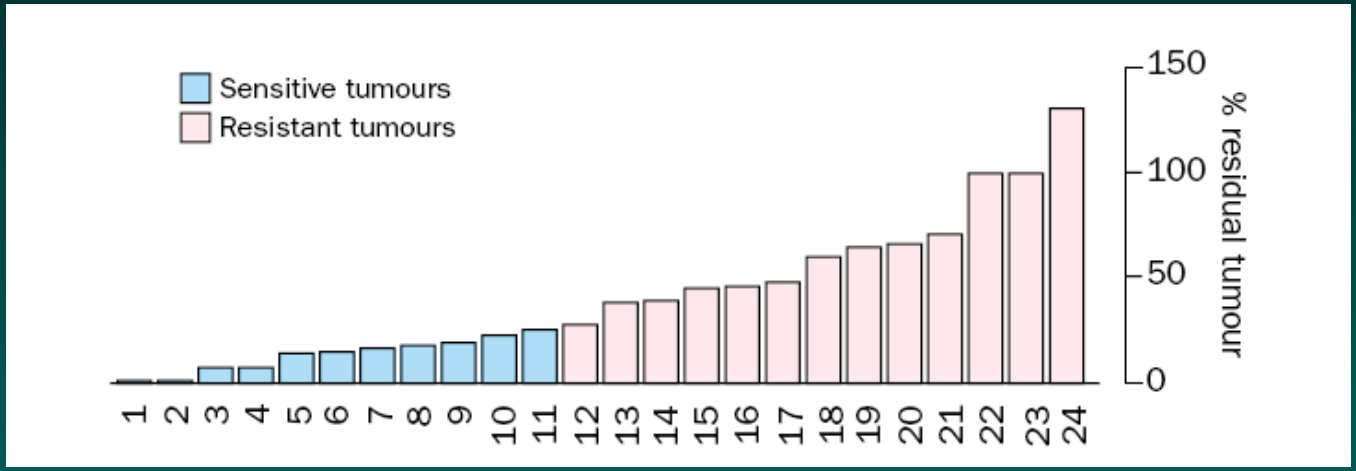
From the software:



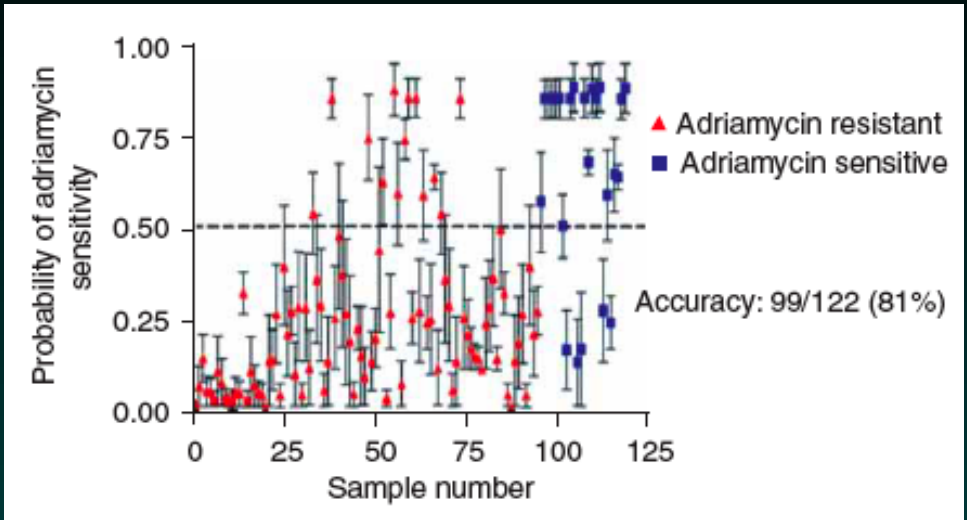
We match heatmaps but not gene lists? We'll come back to this, because their software also gives *predictions*.

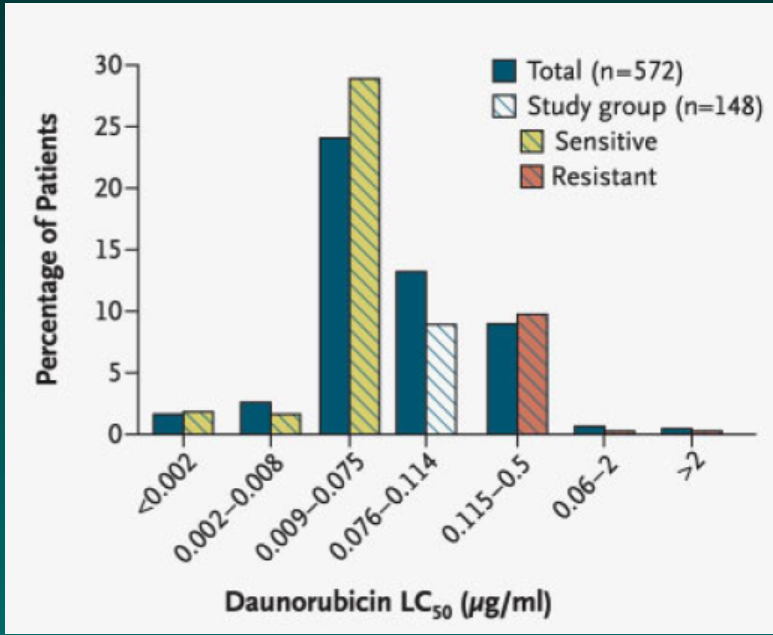
Predicting Docetaxel (Chang 03)





Predicting Adriamycin (Holleman 04)





There Were Other Genes...

The 50-gene list for docetaxel has 19 “outliers”.

The initial paper on the test data (Chang et al) gave a list of 92 genes that separated responders from nonresponders.

Entries 7-20 in Chang et al’s list comprise 14/19 outliers.

The others: ERCC1, ERCC4, ERBB2, BCL2L11, TUBA3.
These are the genes named to explain the biology.

A Repro Theme: Don't Take My Word For It!

Read the paper! Coombes, Wang & Baggerly, Nat Med, Nov 6, 2007, 13:1276-7, author reply 1277-8.

Try it yourselves! All of the raw data, documentation*, and code* is available from our web site (*and from Nat Med):

http://bioinformatics.mdanderson.org/Supplements/ReproRsch-Chemo.

Potti/Nevins Reply (Nat Med 13:1277-8)

Labels for Adria are correct – details on their web page.

They've gotten the approach to work again. (Twice!)

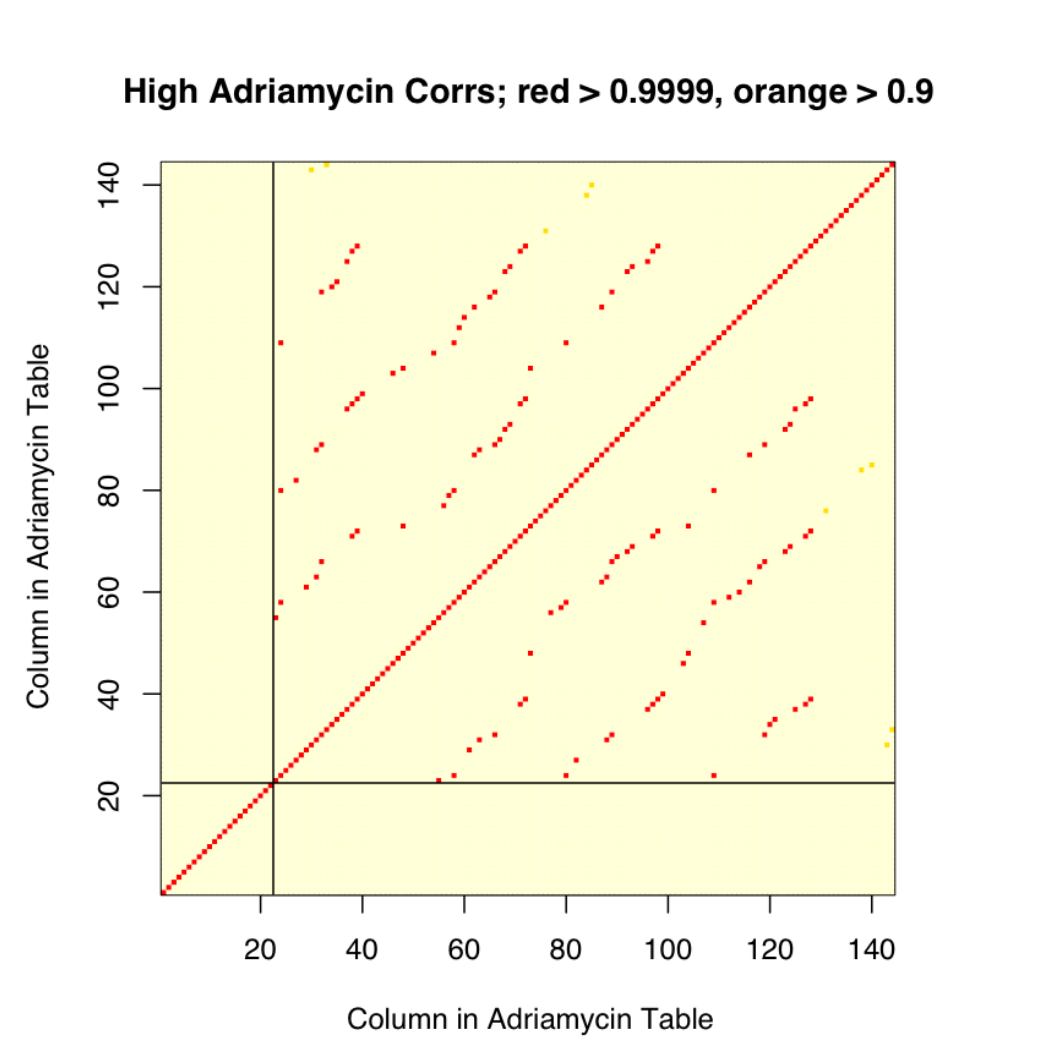
Pharmacogenomic Strategies Provide a Rational Approach to the Treatment of Cisplatin-Resistant Patients With Advanced Cancer

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

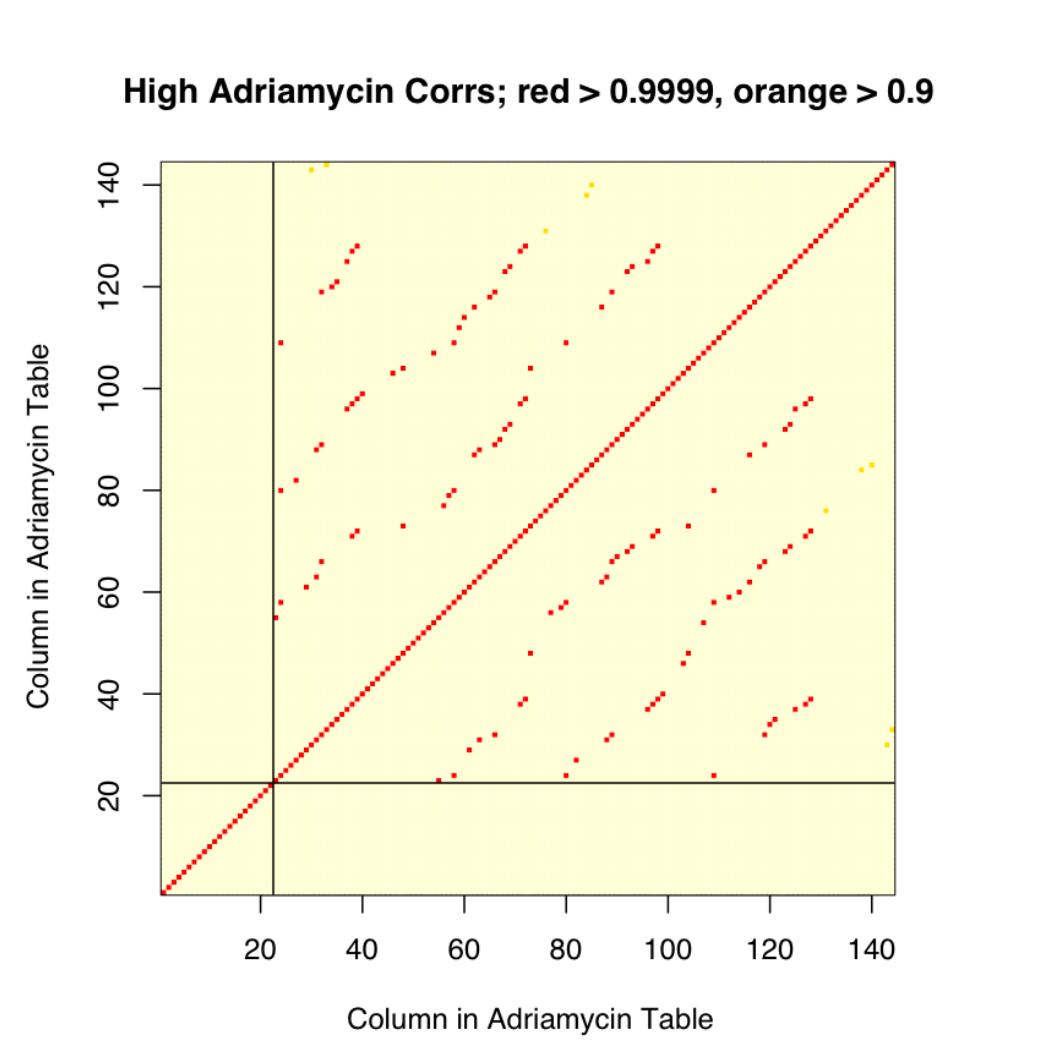
Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial

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

Adriamycin 0.9999+ Correlations (Reply)



Adriamycin 0.9999+ Correlations (Reply)



Redone Aug 08, “using ... 95 unique samples” (also wrong)

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).

The last two probesets are special.

These probesets aren't on the U133A arrays that were used. They're on the U133B.

Some Timeline Here...

Nat Med Nov 06*, Nov 07*, Aug 08. JCO Feb 07*, Oct 07*. Lancet Oncology Dec 07*. PLoS One Apr 08. CCR Jan 09*. (* errors reported)

Some Timeline Here...

Nat Med Nov 06*, Nov 07*, Aug 08. JCO Feb 07*, Oct 07*. Lancet Oncology Dec 07*. PLoS One Apr 08. CCR Jan 09*. (* errors reported)

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

Some Timeline Here...

Nat Med Nov 06*, Nov 07*, Aug 08. JCO Feb 07*, Oct 07*. Lancet Oncology Dec 07*. PLoS One Apr 08. CCR Jan 09*. (* errors reported)

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

Sep 1. Paper submitted to *Annals of Applied Statistics*.
Sep 14. Paper online at *Annals of Applied Statistics*.
Sep-Oct: Story covered by *The Cancer Letter*, Duke starts internal investigation, suspends trials.

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."*

Why We're Unhappy...

"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.

Why We're Unhappy...

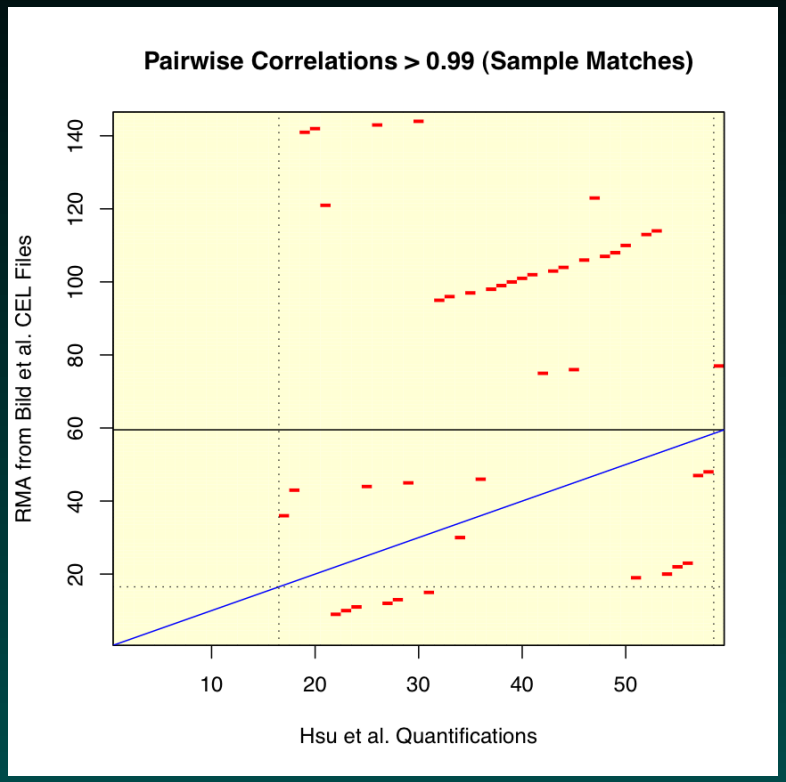
"While the reviewers approved of our sharing the report with the NCI, *we consider it a confidential document*" (Duke). A *future paper* will explain the methods.

There was also a major new development that the restart announcement didn't mention.

In mid-Nov (mid-investigation), the Duke team posted new data for cisplatin and pemetrexed (in trials since '07).

These included quantifications for 59 ovarian cancer test samples (from GSE3149) used for predictor validation.

We Tried Matching The Samples



43 samples are mislabeled; 16 don't match at all.
The first 16 don't match because the genes are mislabeled.
We reported this to Duke and to the NCI in mid-November.
All data was stripped from the websites within the week.

More Timeline

April, 2010. Review report sought from NCI under FOIA.
May, 2010. Redacted report supplied; gaps noted.
May, 2010. NCI and CALGB pull lung metagene signature from an ongoing phase III trial.

Duke trials continue.

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

Subsequent Events

July 19/20: letter to Varmus; Duke resuspends trials
Oct 22/29: call to retract JCO paper
Nov 9: Duke announces trials terminated
Nov 19: call to retract Nat Med paper, Potti resigns

Jan 11, 2011: *Nature* talks to Duke
The Duke deans overseeing the investigation, in consultation with the acting head of Duke's IRB, decided not to forward our report to the reviewers.

Dec '10/Jan '11: The IOM Meets, the NCI Speaks
Jan 28, 2011: The FDA visits

Some Cautions/Observations

We've seen problems like these 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

Unfortunately, we suspect
The most simple mistakes are common.

Is Our Work Reproducible?

For the past three years, we have required reports to be prepared in *Sweave*.

What should the norm be?

For papers? (Baggerly, *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?

Some Acknowledgements

Kevin Coombes

Shannon Neeley, Jing Wang

David Ransohoff, Gordon Mills

Jane Fridlyand, Lajos Pusztai, Zoltan Szallasi

MDACC Ovarian SPORE, Lung SPORE, Breast SPORE

Now in the *Annals of Applied Statistics!* Baggerly and Coombes (2009), 3(4):1309-34.
http://bioinformatics.mdanderson.org/Supplements/ReproRsch-All

Index