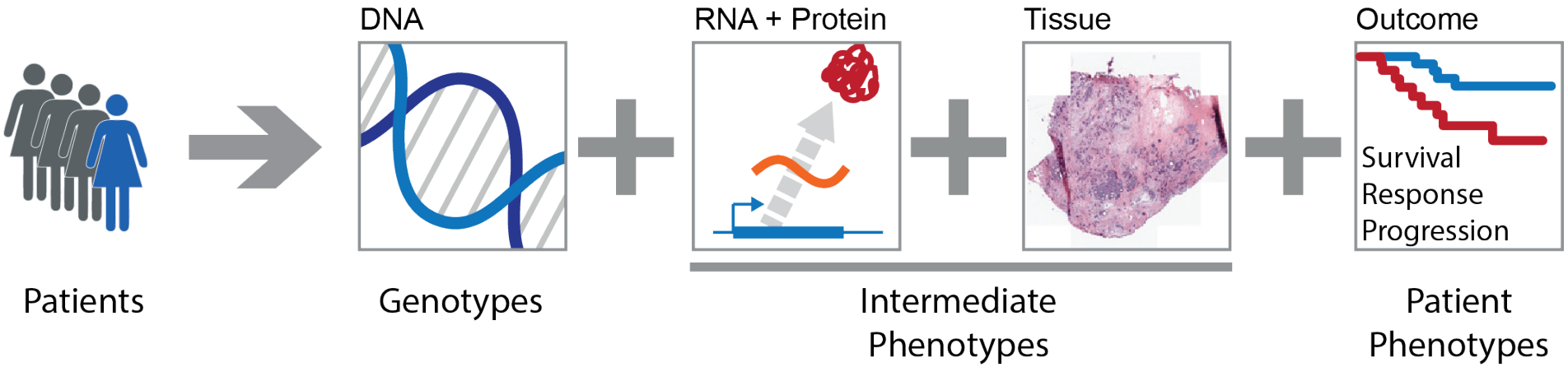
5 selfish reasons to work **reproducibly**

More publications, more grants, more awesome!

Florian Markowetz
CRUK Cambridge Institute
www.markowetzlab.org

Systems Genetics of Cancer



Genetic variation → **Phenotypic variation**

- In people
- In tumours
- In clones

- Tumour subtypes
- Aggressiveness
- Survival

Cancer genome
Function

Andy

Amanda

Hyunjin

Federico

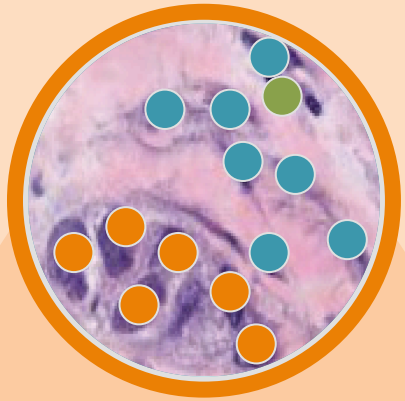
Cancer genome
Evolution

Geoff

Edith

Ruben

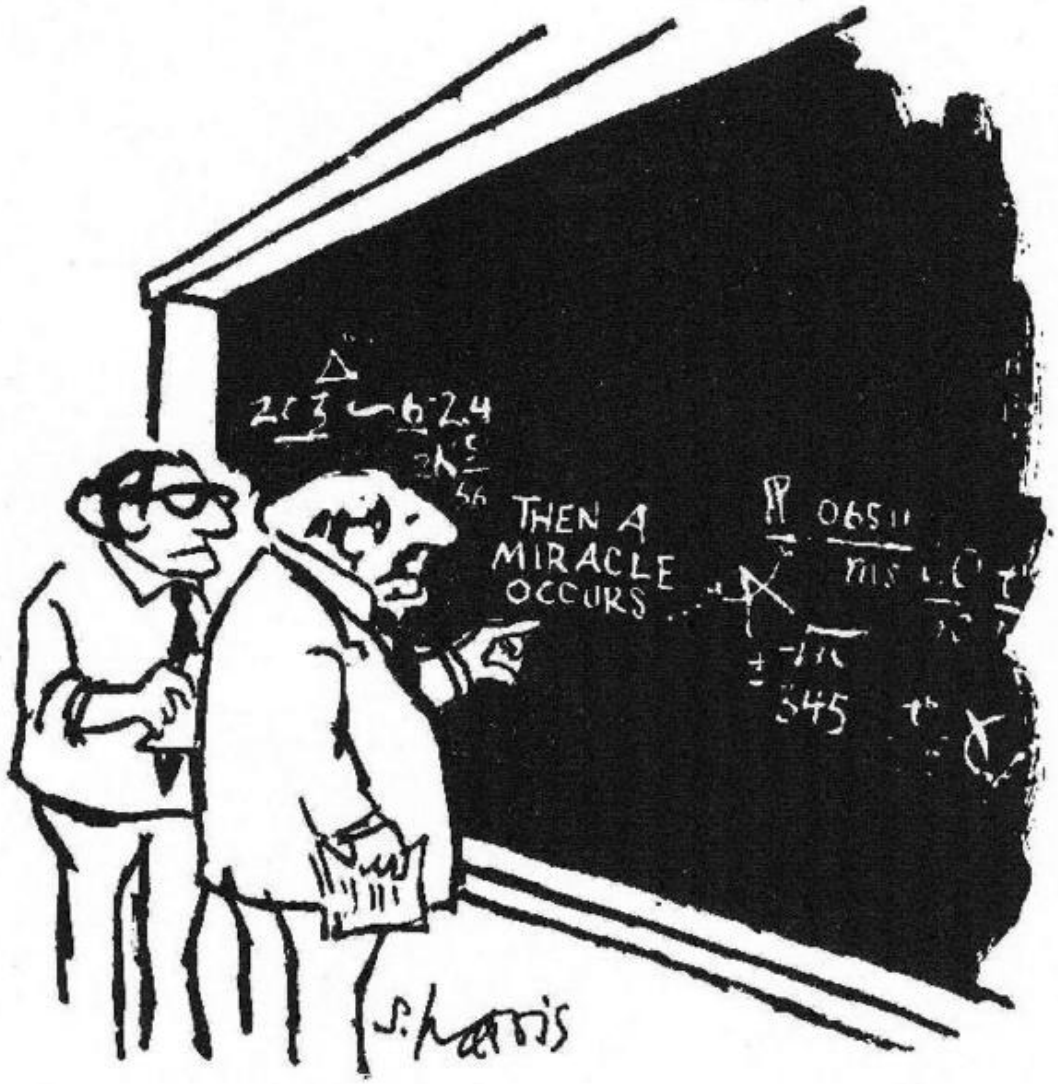
Cancer tissue
Context



Leon

Turid

Science ≠ miracles



"I think you should be more explicit here in step two."

“How Bright Promise in Cancer Testing Fell Apart”

New York Times 2011

The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1309–1334
DOI: 10.1214/09-AOAS291

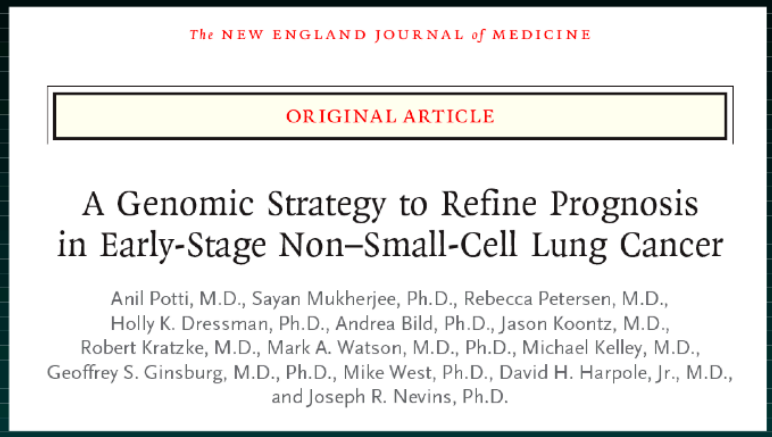
DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

By Keith A. Baggerly[1] and Kevin R. Coombes[2]

University of Texas

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results,

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.

Slides by Keith Baggerly

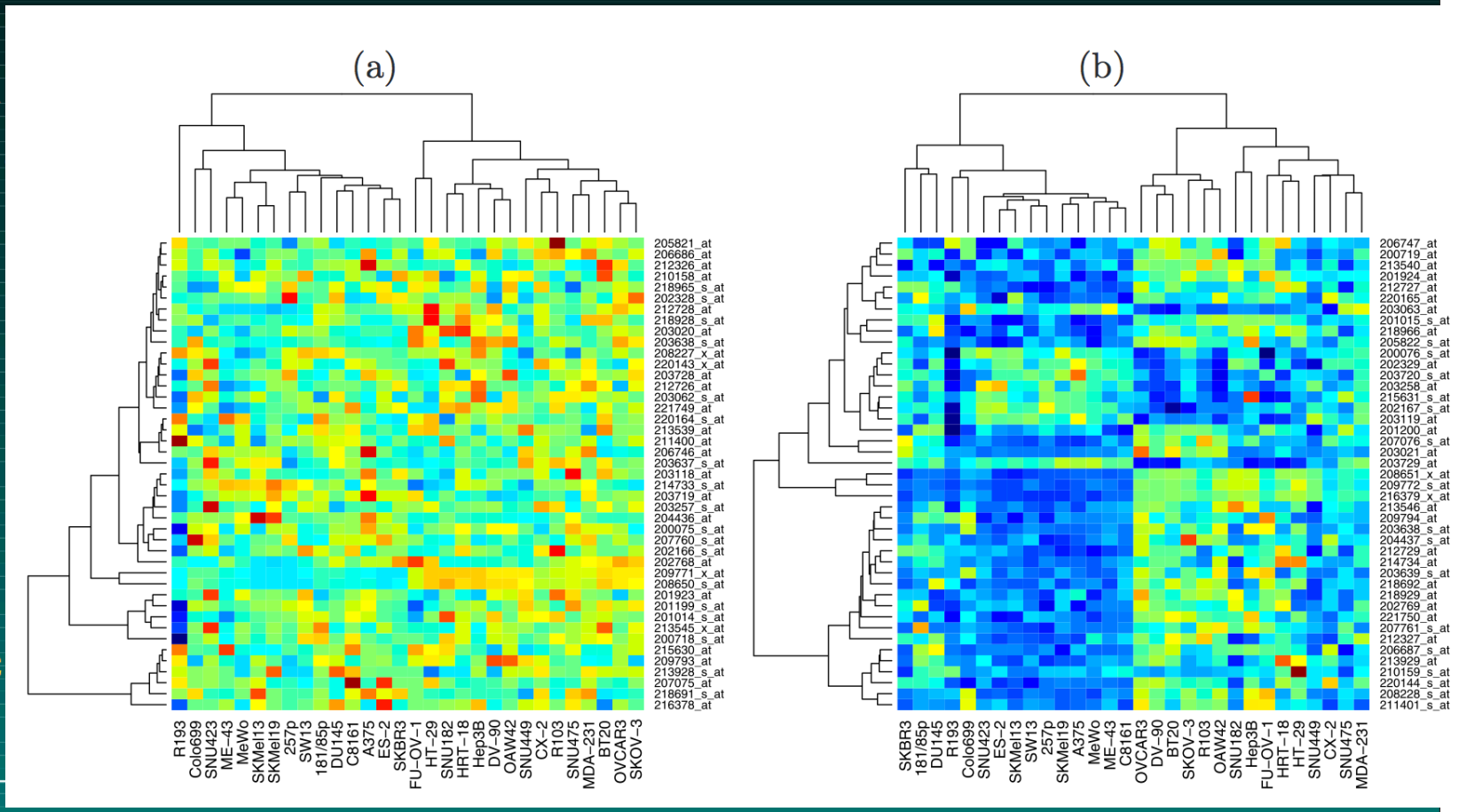
Off by One

```
> 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"
...
```

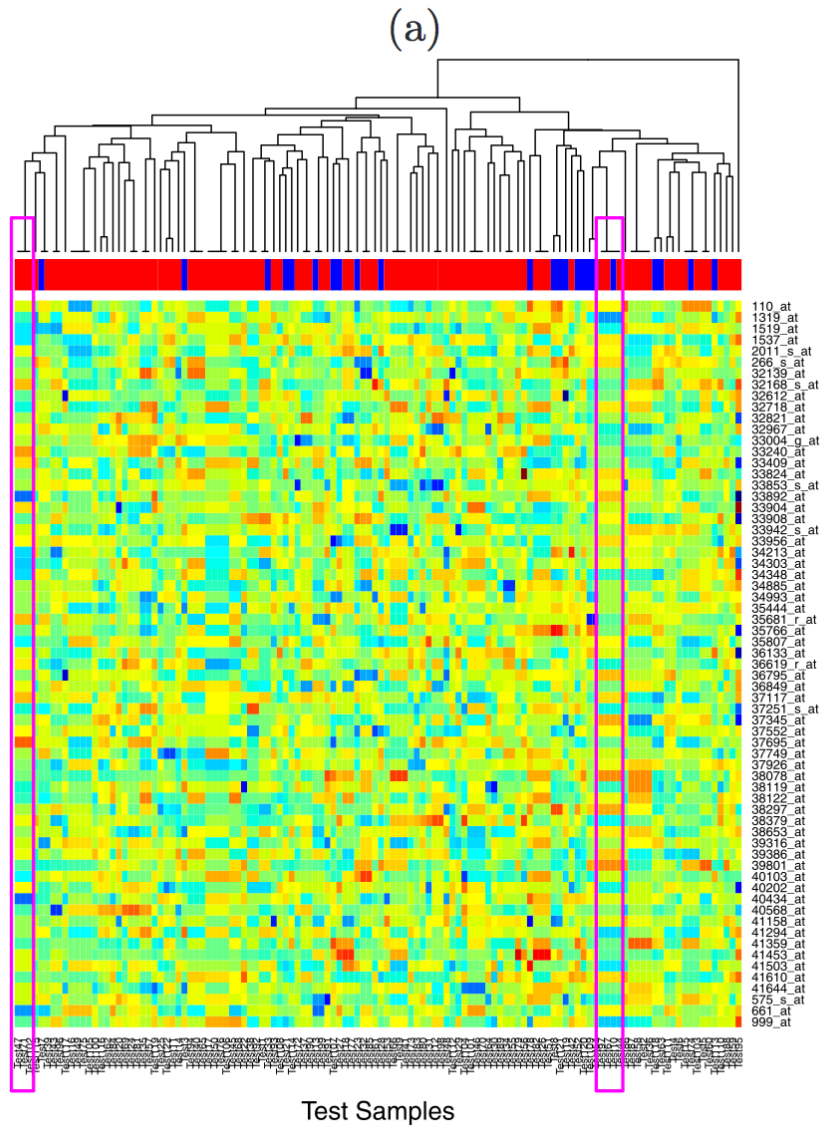
Dec 4

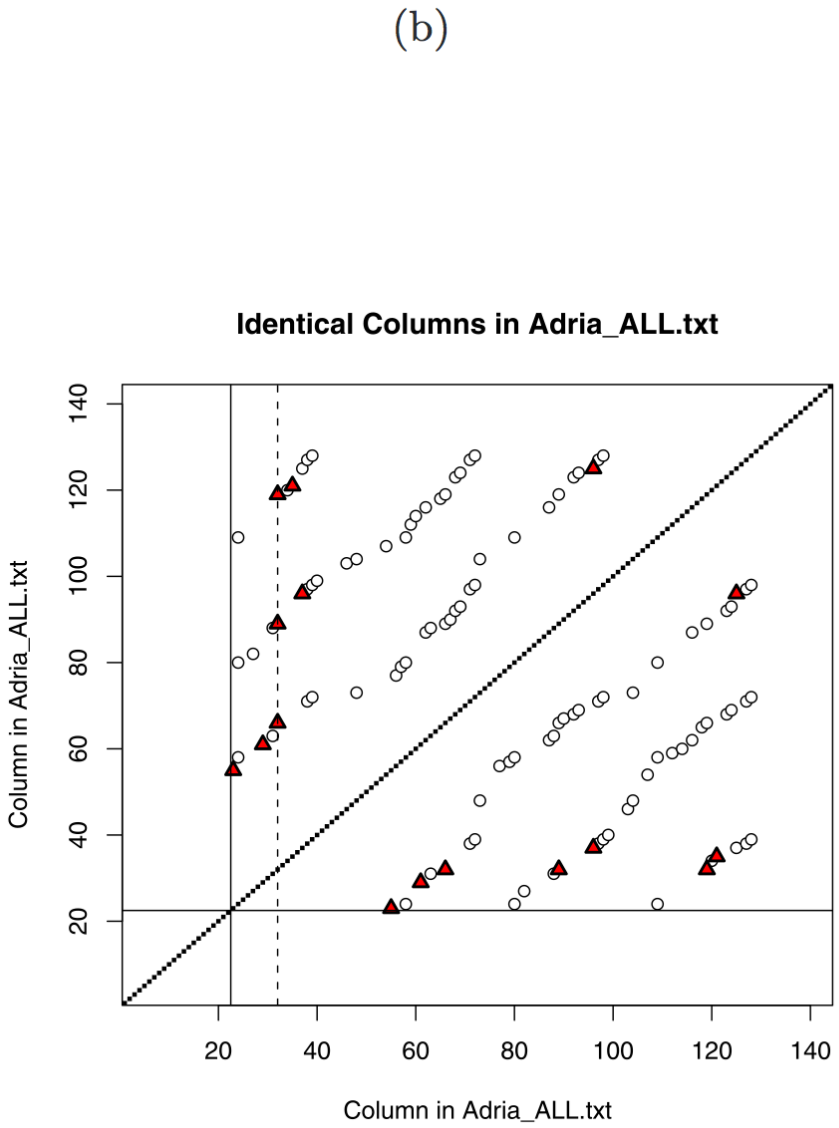
Off by One

```
> temp <- cbind(
    sort(rownames(pottiUpdated)[fuRows]),
    sort(rownames(pottiUpdated)[
> c
> t
...
[3,
[4,
[5,
...
Dec
```



Slides by Keith Baggerly





Baggerly & Coombes, AOAS 2009

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.

Switch off the lights

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

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

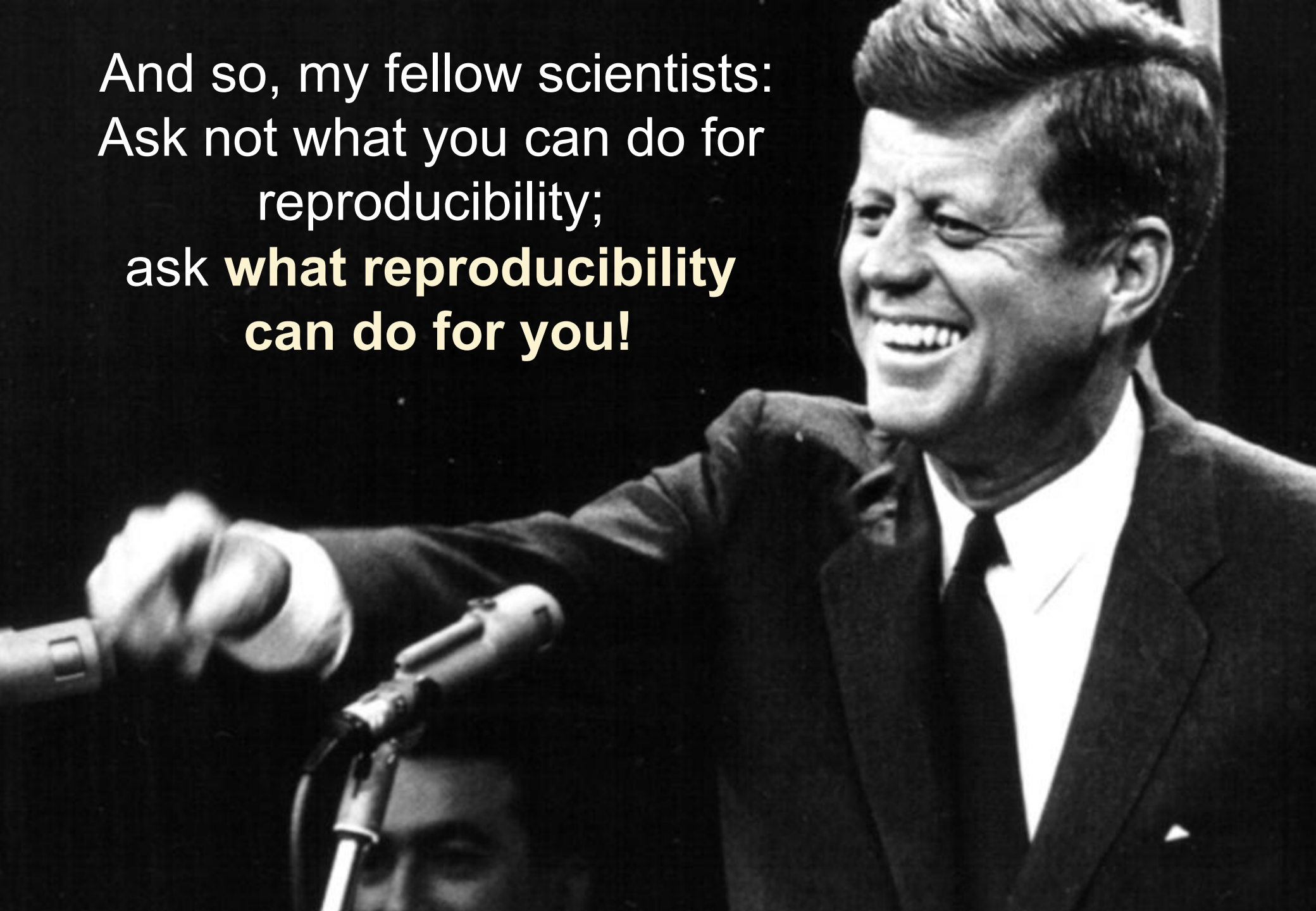
Cambridge, 4 September 2010

Slides

http://videolectures.net/cancerbioinformatics2010_baggerly_irrh/

Reproducible Research

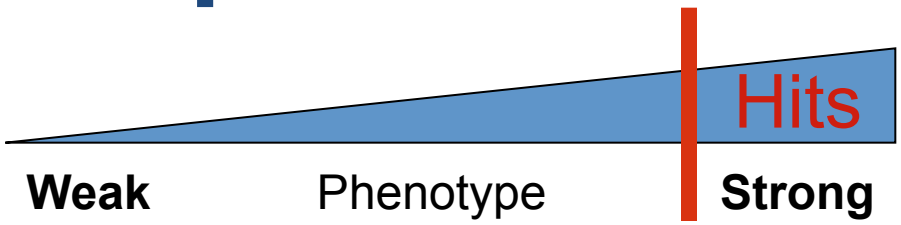
- It's the **right thing** to do!
- The world would be a **better place** if everyone did it!
- It's the **foundation** of Science!
- It's the **honourable** thing to do!


And so, my fellow scientists:
Ask not what you can do for reproducibility;
ask **what reproducibility can do for you!**

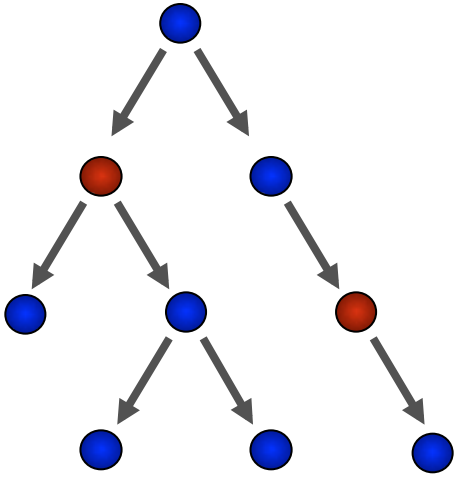
Reproducibility helps to **avoid disaster**

Anatomy of the NFκB pathway

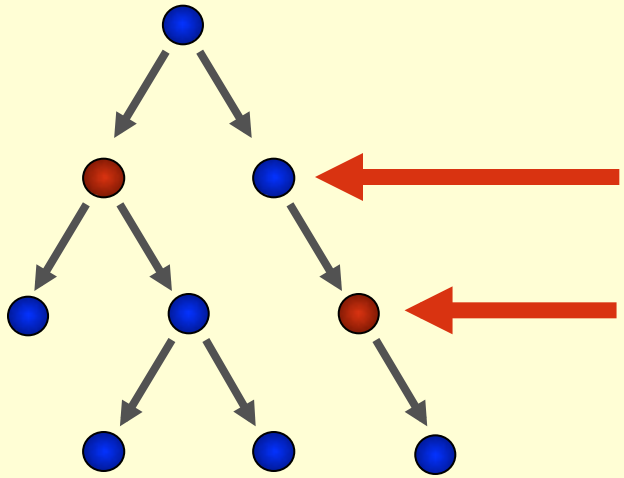
Step 1



?



Step 2



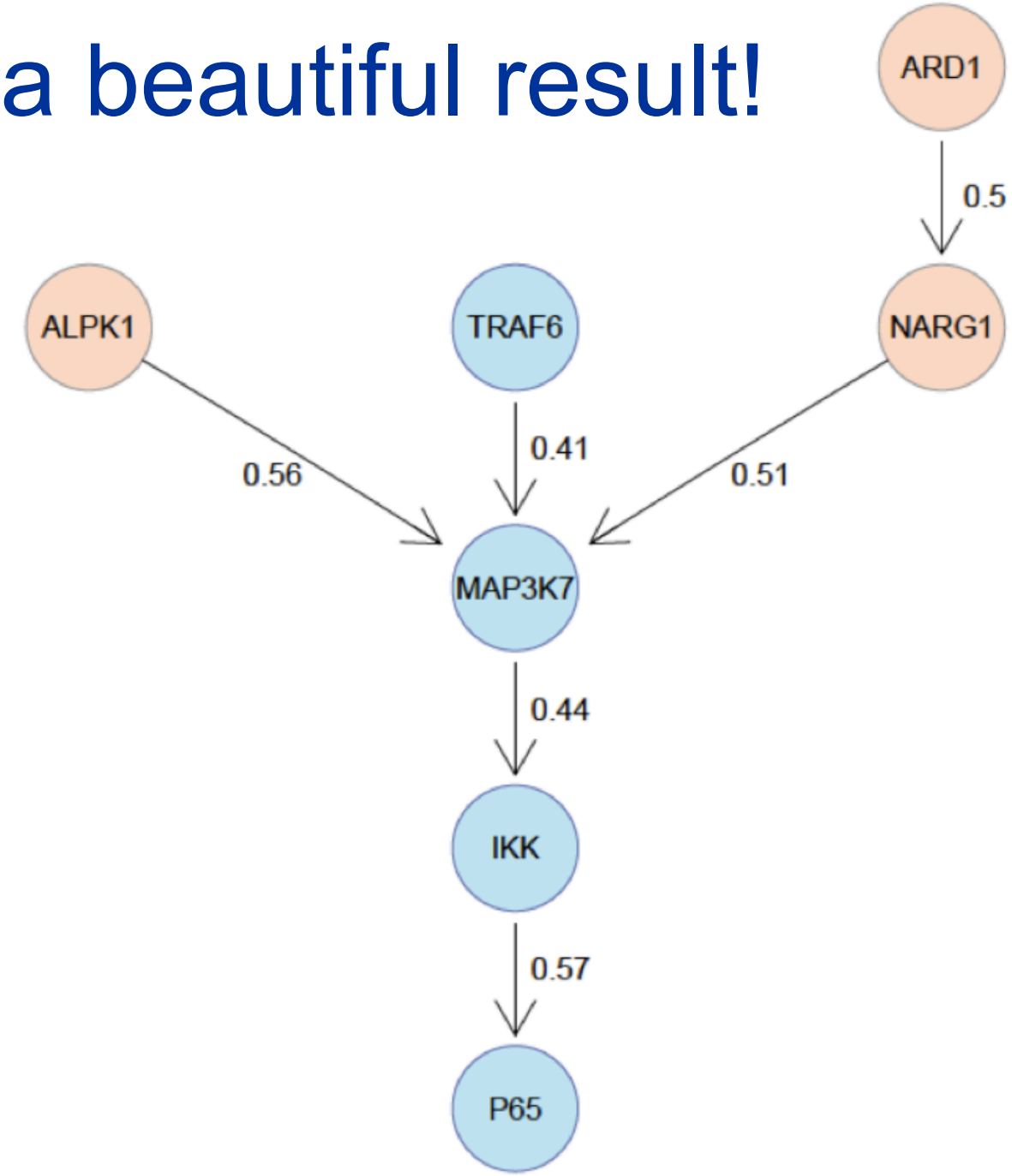
Knock-down

Known pathway members

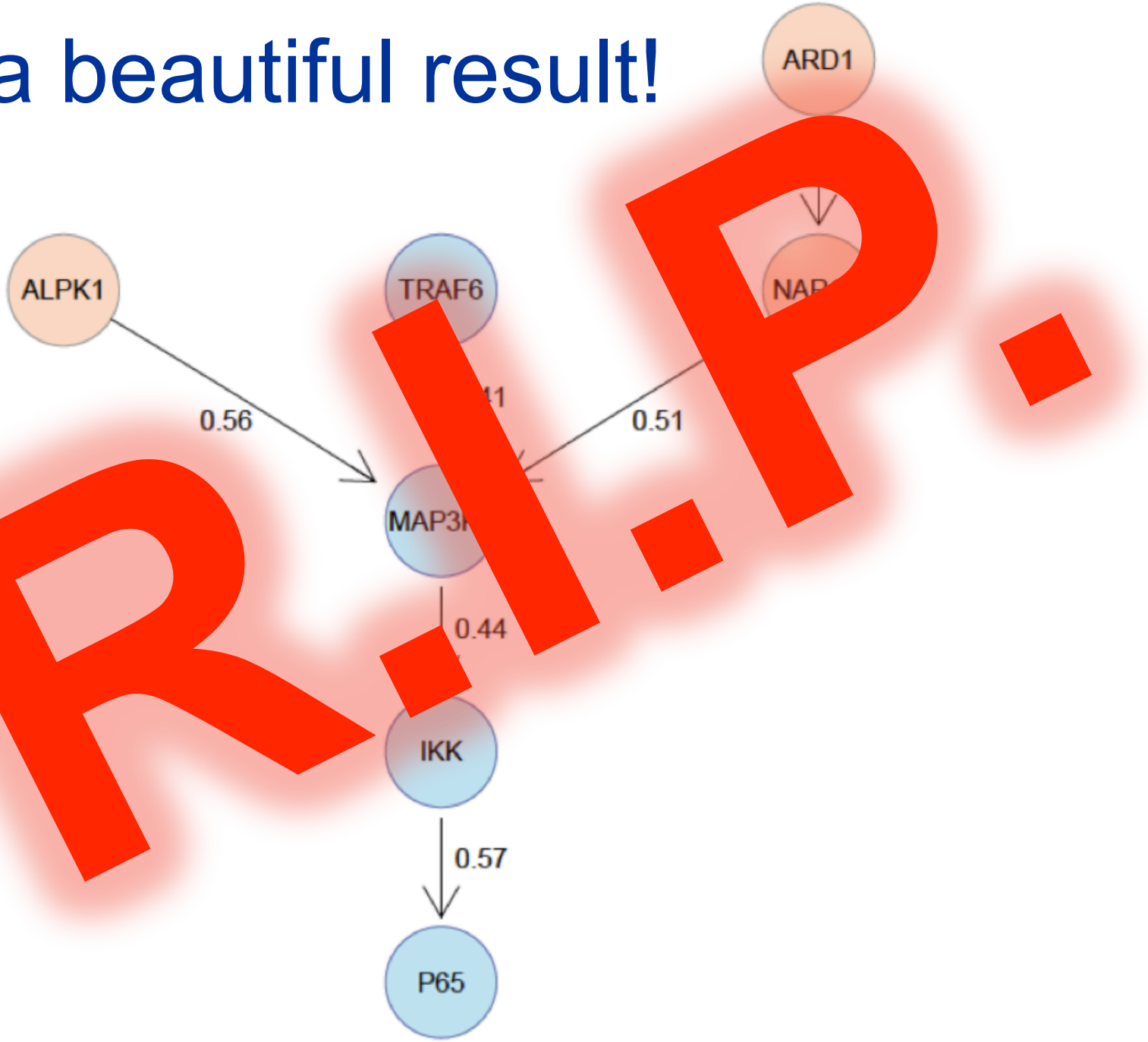
New RNAi Hits

Compare expression phenotypes by **NEMs**

What a beautiful result!



What a beautiful result!



A project is more than a beautiful result!

Starting with reproducibility *early* helps saving time later

Reproducibility helps **writing papers**

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis

Roland F. Schwarz, Charlotte K. Y. Ng, Susanna L. Cooke, Scott Newman, Jillian Temple, Anna M. Piskorz, Davina Gale, Karen Sayal, Muhammed Murtaza, Peter J. Baldwin, Nitzan Rosenfeld, Helena M. Earl, Evis Sala, [...], James D. Brenton
[view all]

Published: February 24, 2015 • DOI: 10.1371/journal.pmed.1001789 • Featured in PLOS Collections

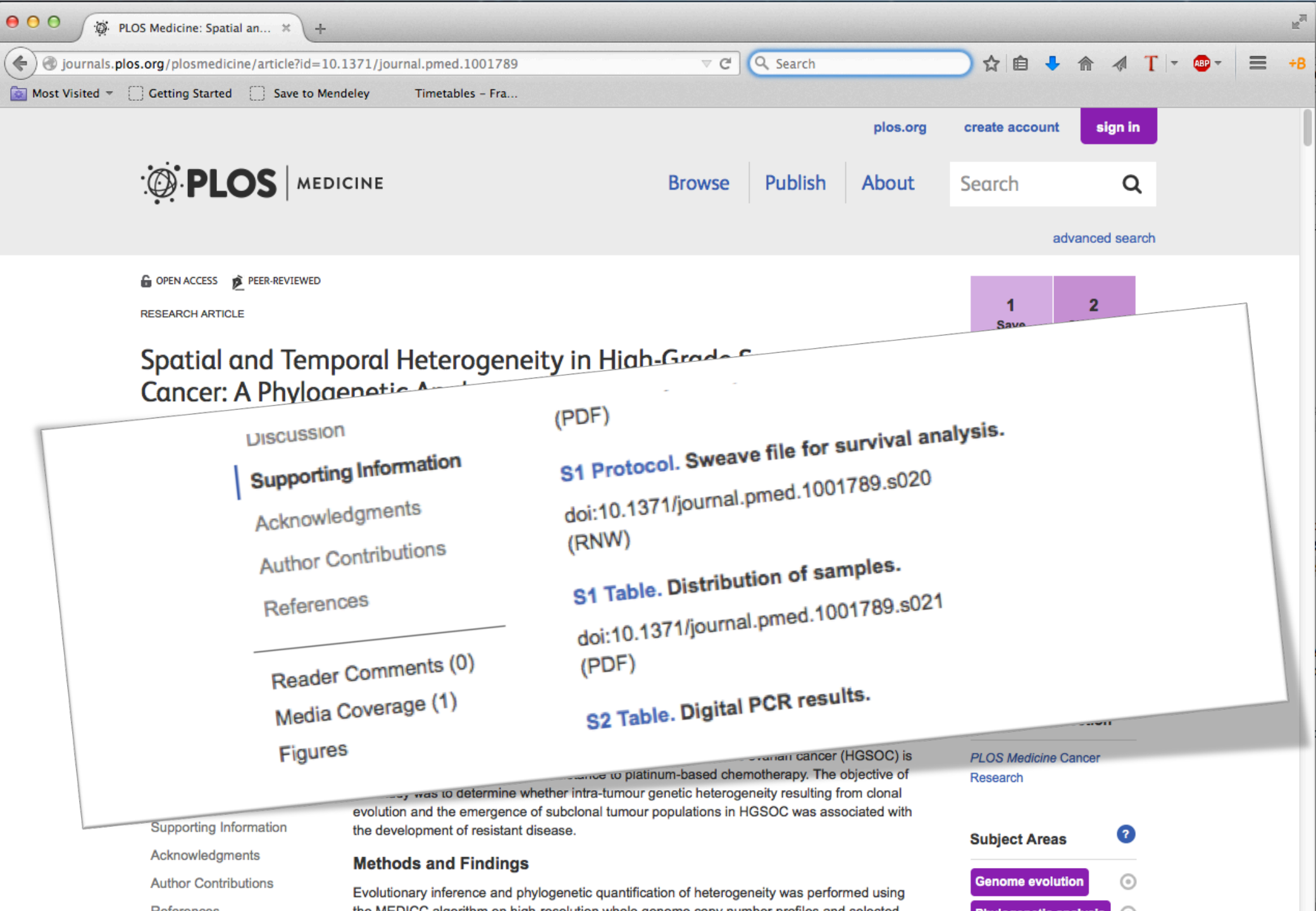
Abstract

Background

The major clinical challenge in the treatment of high-grade serous ovarian cancer (HGSOC) is the development of progressive resistance to platinum-based chemotherapy. The objective of this study was to determine whether intra-tumour genetic heterogeneity resulting from clonal evolution and the emergence of subclonal tumour populations in HGSOC was associated with the development of resistant disease.

Methods and Findings

Evolutionary inference and phylogenetic quantification of heterogeneity was performed using


PLOS Medicine: Spatial an...
journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001789
Search
Most Visited
Getting Started
Save to Mendeley
Timetables – Fra...
plos.org
create account
sign in
PLOS | MEDICINE
Browse
Publish
About
Search
advanced search
OPEN ACCESS
PEER-REVIEWED
RESEARCH ARTICLE
1
2
Spatial and Temporal Heterogeneity in High-Grade S
Cancer: A Phylogenetic
Discussion
Supporting Information
Acknowledgments
Author Contributions
References
Reader Comments (0)
Media Coverage (1)
Figures
(PDF)
S1 Protocol. Sweave file for survival analysis.
doi:10.1371/journal.pmed.1001789.s020
(RNW)
S1 Table. Distribution of samples.
doi:10.1371/journal.pmed.1001789.s021
(PDF)
S2 Table. Digital PCR results.
ovarian cancer (HGSOC) is
platinum-based chemotherapy. The objective of
was to determine whether intra-tumour genetic heterogeneity resulting from clonal
evolution and the emergence of subclonal tumour populations in HGSOC was associated with
the development of resistant disease.
Methods and Findings
Evolutionary inference and phylogenetic quantification of heterogeneity was performed using
PLOS Medicine Cancer
Research
Subject Areas
Genome evolution

journal.pmed.1001789.s020.RNW

Format | Compile PDF | Run | Chunks

First we load the data and R packages. The data file is part of the paper supplement, and we have also made a copy available online.

```
<<message=FALSE,tidy=FALSE>>=
library(survival)
library(kernlab)
library(rms)
library(spatstat)
library(RColorBrewer)
library(gplots)

## load data file from local copy or from URL
if (file.exists("Schwarz2015-supplement.Rdata")){
  load("Schwarz2015-supplement.Rdata")
  cat("Data loaded from local copy")
} else {
  load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
  cat("Data loaded from URL") }
@
```

The first object in the .Rdata file is a table \texttt{D} with patient information:

```
<<message=FALSE>>=
D
attach(D)
@
```

```
<<echo=FALSE,message=FALSE>>=
## Print the data table in LaTeX format for inclusion into main mansucript
library(xtable)
print(xtable(D),file="TableOverview.tex")
@
```

Rownames correspond to sample identifiers. Columns indicate the patient numbers used in the manuscript (\texttt{Nr}), as well as values for temporal heterogeneity (\texttt{TH}), clonal heterogeneity index (\texttt{CE}), overall survival in days (\texttt{OS}), progression-free survival in days (\texttt{PFS}) and indicators for survival (\texttt{dead}) and progression (\texttt{prog}) as well as covariates for age, stage (ordered factor), residual disease after debulking surgery (\texttt{residual}, ordered factor) and the number of samples per patient (\texttt{N}).

153:13 | Chunk 3 | R Sweave

journal.pmed.1001789.s020.RNW

Format ▾ Compile PDF

```
148
149 First we load the data and R packages. The data file is part of the pap
    also made a copy available online.
150 <<message=FALSE,tidy=FALSE>>=
151 library(survival)
152 library(kernlab)
153 library(rms)
154 library(spatstat)
155 library(RColorBrewer)
156 library(gplots)
157
158 ## load data file from local copy or from URL
159 if (file.exists("Schwarz2015-supplement.Rdata")){
160   load("Schwarz2015-supplement.Rdata")
161   cat("Data loaded from local copy")
162 } else {
163   load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplem
164   cat("Data loaded from URL") }
165 @
166
167 The first object in the .Rdata file is a table \texttt{D} with patient
168 <<message=FALSE>>=
169 D
170 attach(D)
171 @
172
173 <<echo=FALSE,message=FALSE>>=
174 ## Print the data table in LaTeX format for inclusion into main mansucr
175 library(xtable)
176 print(xtable(D),file="TableOverview.tex")
177 @
178
179 Rownames correspond to sample identifiers. Columns indicate the patient
    t (\texttt{Nr}), as well as values for temporal heterogeneity (\texttt{
    index (\texttt{CE}), overall survival in days (\texttt{OS}), progressio
    (\texttt{PFS}) and indicators for survival (\texttt{dead}) and progress
    covariates for age, stage (ordered factor), residual disease after debu
    l}, ordered factor) and the number of samples per patient (\texttt{N}).
180
```

153:13 Chunk 3 R Sweave

1 Clinical data

1.1 Data overview

First we load the data and R packages. The data file is part of the paper supplement, and we have also made a copy available online.

```
library(survival)
library(kernlab)
library(rms)
library(spatstat)
library(RColorBrewer)
library(gplots)

## load data file from local copy or from URL
if (file.exists("Schwarz2015-supplement.Rdata")){
        load("Schwarz2015-supplement.Rdata")
        cat("Data loaded from local copy")
} else {
        load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
        cat("Data loaded from URL") }

## Data loaded from URL
```

The first object in the .Rdata file is a table D with patient information:

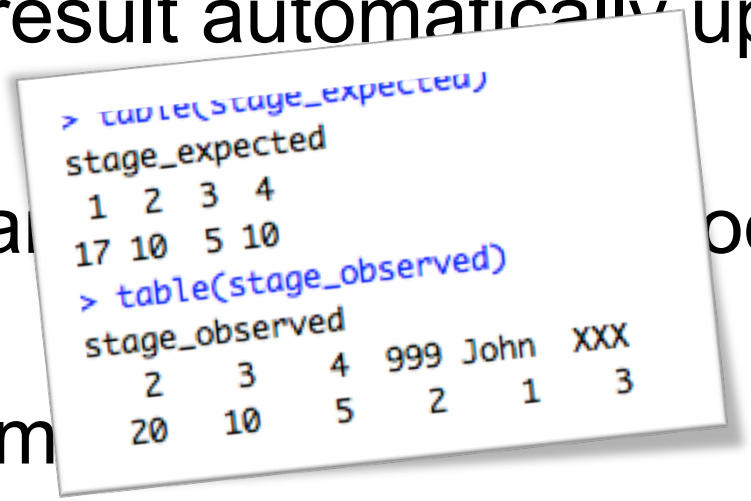
```
D

##          Nr       TH        CE   OS  PFS dead prog  Hist Age Stage residual
## OV03-01   1 4.730231 1.2605274  511  271    1    1 HGSOC  47    IV      <1cm
## OV03-02   2       NA 0.7105901  977  363    1    1 HGSOC  62    IV      <1cm
## OV03-04   3 3.735366 1.2432629  209  153    1    1   HGS  69    IV      >1cm
## OV03-07   4       NA        NA  625  616    1    1 HGSOC  48  IIIC       Nil
## OV03-08   5 3.801712 1.4705531  547  303    1    1 HGSOC  63    IV      <1cm
## OV03-10   6 6.588895 0.7298828  744  298    1    1 HGSOC  59    IV      <NA>
## OV03-13   7 3.000290 0.6836961 1587  358    1    1 HGSOC  61    IV      >1cm
## OV03-17   8 3.423112 2.2357817  889  373    1    1 HGSOC  51  IIIC      <1cm
## OV03-20   9 4.487828 0.6494353 1278  563    1    1 HGSOC  71    IV      >1cm
## OV03-21  10 4.719848 0.8686309 1139  303    1    1 HGSOC  60  IIIC      >1cm
## OV03-22  11 5.702720 0.4834086 1556  382    1    1 HGSOC  58  IIIC      <1cm
## OV03-23  12       NA        NA 1565  534    1    1 HGSOC  60  IIIC       Nil
## OV03-24  NA       NA        NA  376  375    1    1 HGSOC  53  IIIC      >1cm
## OV03-25  13       NA 0.6215297 1166  776    1    1 HGSOC  57  IIIC      >1cm
## OV04-20  14 4.621984 0.6083119 1513  601    0    1 HGSOC  63  IIIC       Nil
## OV04-21  15       NA 0.7412773  706  332    1    1 HGSOC  54    IV       Nil
## OV04-27  16       NA        NA 1408 1408    0    0 HGSOC  58  IIIC       Nil
## OV04-30  17       NA 0.8591205  849  293    1    1 HGSOC  60  IIIC      >1cm
##           N
```

Why is well-documented and easily accessible code+data useful?

- Easy to look up numbers and put them in manuscript
- Be confident your figures and tables are up-to-date
- Numbers and result automatically update when data change.
- It is engaging and more eyes can look over the analysis.
- Easier to spot mistakes.

Why is well-documented and easily accessible code+data useful?

- Easy to look up numbers and put them in manuscript
- Be confident your figures and tables are up-to-date
- Numbers and result automatically update when data change.
- It is engaging a ook over the analysis.
- Easier to spot m

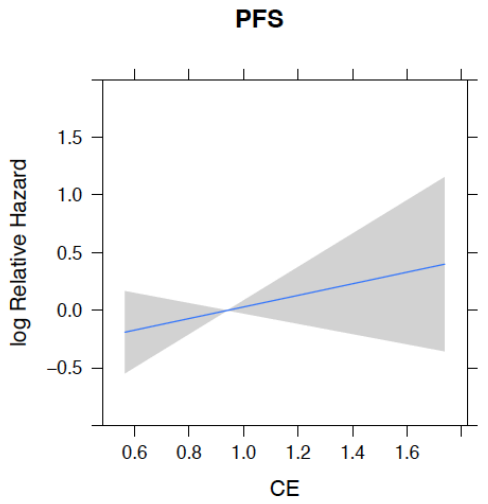


3

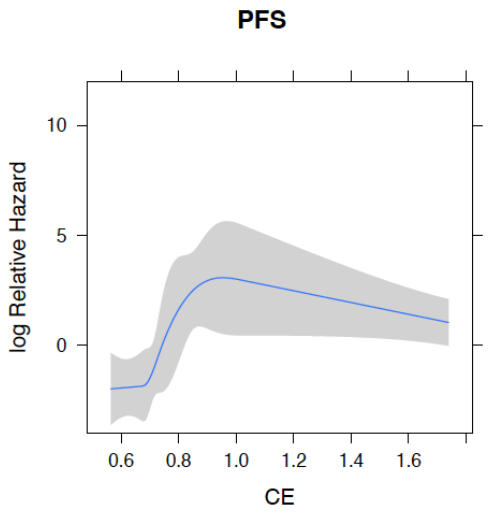
Reproducibility helps **reviewers see it your way**

A very engaged reviewer

- **Reviewer:** “I downloaded the authors’ data and tried out a variation of their analysis which gave an insignificant result”



- **We**: “Thank you, the reason is XXX and if you do YYY everything is fine.”




4
Reproducibility
enables
continuity

“I am so busy,
I can’t remember all
the details of all my
projects”

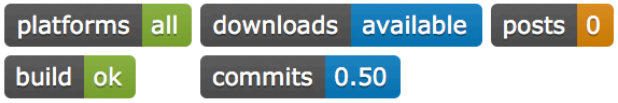
“I’m sorry,
I did this analysis
6 months ago.”

“My PI said I should continue the project of a previous postdoc.

But that postdoc is long gone and hasn’t saved any scripts or data.”


5
Reproducibility helps to build your reputation

Mulder2012



Predicting functional networks and modules of chromatin factors controlling adult stem cell fate from RNA interference screens

Bioconductor version: Release (3.1)

This package provides functions to reproduce results and figures in Mulder K. et. al. published in Nature Cell Biology 2012 and Wang X. et. al. published in PLoS Computational Biology 2012.

Author: Xin Wang <Xin.Wang at cancer.org.uk>, Florian Markowetz <Florian.Markowetz at cancer.org.uk>

Maintainer: Xin Wang <Xin.Wang at cancer.org.uk>

Citation (from within R, enter `citation("Mulder2012")`):

> Wang X and Markowetz F (2012). *Mulder2012: Predicting functional networks and modules of chromatin factors controlling adult stem cell fate from RNA interference screens*. R package version 0.7.1.

Installation

To install this package, start R and enter:

Workflows »

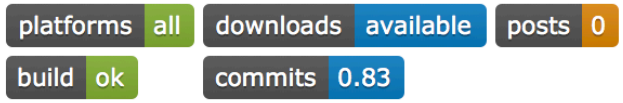
Common Bioconductor workflows include:

- Oligonucleotide Arrays
- High-throughput Sequencing
- Counting Reads for Differential Expression (parathyroideSE vignette)
- Annotation
- Annotating Variants
- Annotating Ranges
- Flow Cytometry and other assays
- Candidate Binding Sites for Known Transcription Factors
- Cloud-enabled cis-eQTL search and annotation
- RNA-Seq workflow: gene-level exploratory analysis and differential expression
- Changing genomic coordinate systems with rtracklayer::liftOver
- Mass spectrometry and proteomics data analysis

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the posting guide before posting!

Fletcher2013a



Gene expression data from breast cancer cells under FGFR2 signalling perturbation.

Bioconductor version: Release (3.1)

The package Fletcher2013a contains time-course gene expression data from MCF-7 cells treated under different experimental systems in order to perturb FGFR2 signalling. The data comes from Fletcher et al. (Nature Comms 4:2464, 2013) where further details about the background and the experimental design of the study can be found.

Author: Mauro Castro, Michael Fletcher, Florian Markowetz and Kerstin Meyer.

Maintainer: Mauro Castro <mauro.a.castro at gmail.com>

Citation (from within R, enter `citation("Fletcher2013a")`):

> Fletcher M, Castro M, Wang X, Santiago Id, O'Reilly M and al. e (2013). "Master regulators of FGFR2 signalling and breast cancer risk." *Nature Communications*, **4**, pp. 2464.

Installation

Workflows »

Common Bioconductor workflows include:

- Oligonucleotide Arrays
- High-throughput Sequencing
- Counting Reads for Differential Expression (parathyroideSE vignette)
- Annotation
- Annotating Variants
- Annotating Ranges
- Flow Cytometry and other assays
- Candidate Binding Sites for Known Transcription Factors
- Cloud-enabled cis-eQTL search and annotation
- RNA-Seq workflow: gene-level exploratory analysis and differential expression
- Changing genomic coordinate systems with rtracklayer::liftOver
- Mass spectrometry and proteomics data analysis

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the posting guide before posting!

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni**

Transparency, openness, and reproducibility are readily recognized as vital features of science (*1*, *2*). When asked, most scientists embrace these features as disciplinary norms and values (*3*). Therefore, one might expect that these valued features would be routine in daily practice. Yet, a growing body of evidence suggests that this is not the case (*4*–*6*).

Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.


So What?
You and 82 others don't give a fuck.

“It’s only the **result** that matters!”

“I’d rather do
real science
than tidy up
my data”

“**Mind your own business**!
I document my data the way I want!”

“**Excel works just fine**.

I don’t need any fancy R or Python or whatever.”

“Sounds alright, but my code and data are **spread over so many hard drives and directories** that it would just be too much work to collect them all in one place”

“We can always sort out the code and data **after submission**”

“My field is very **competitive** and I can’t risk **wasting time**”

Reproducibility is important for

- Phd students
- Postdocs

Learn tools and apply in daily work!

- PIs

Create a '**culture of reproducibility**' in your lab!

In case of fire

1. git commit

2. git push

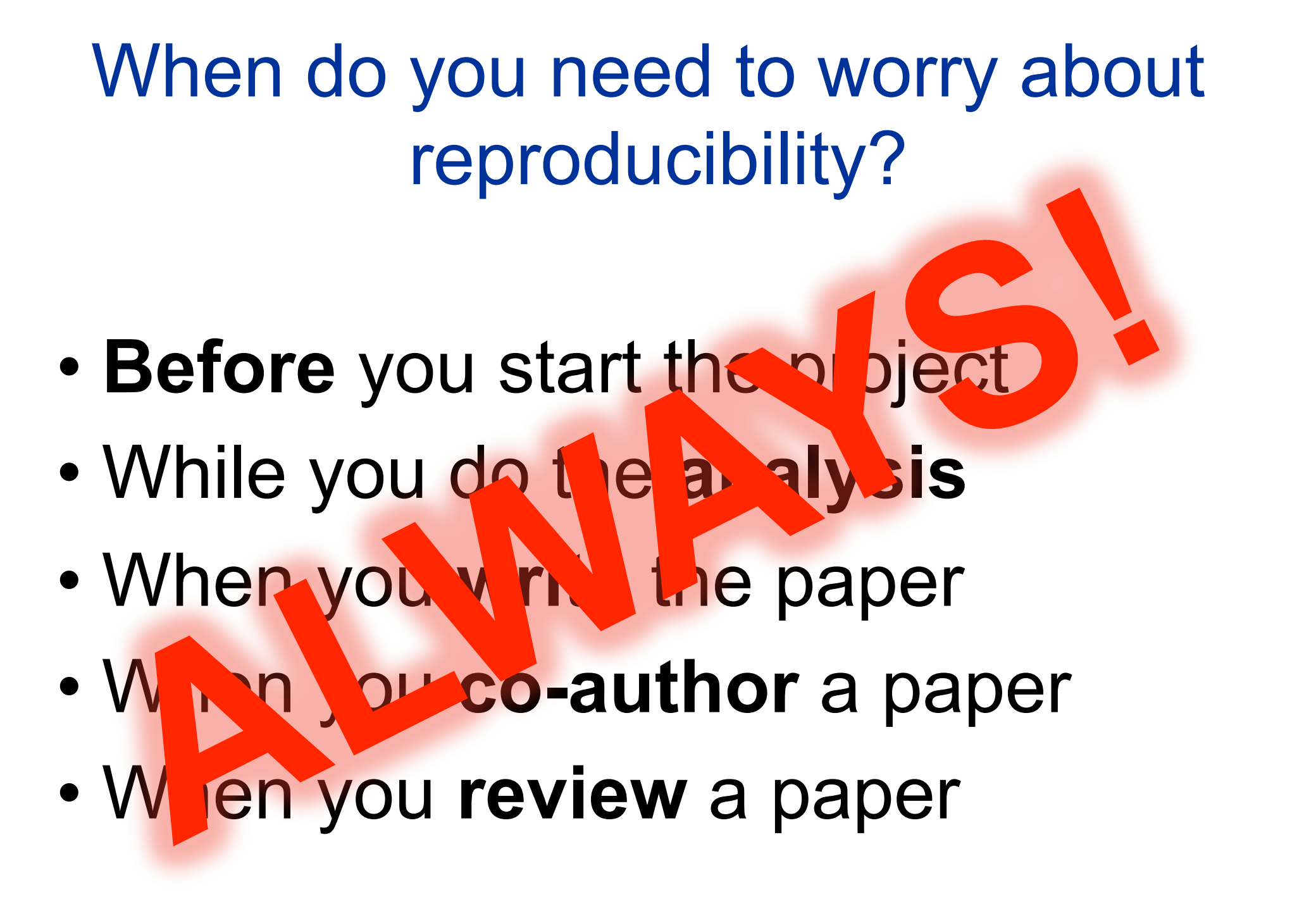
3. leave building

When do you need to worry about reproducibility?

- **Before** you start the project
- While you do the **analysis**
- When you **write** the paper
- When you **co-author** a paper
- When you **review** a paper

When do you need to worry about reproducibility?

- **Before** you start the project
- While you do the **analysis**
- When you **write** the paper
- When you **co-author** a paper
- When you **review** a paper



Scientific SOFT SKILLS

- Organization of project
- Tidy data
- Tidy code
- Control over tools
- Documentation
- Reproducibility

```
\project
      \data
      \code
      \analysis
      \paper
```

Less clicking and pasting,
more scripting and coding

The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1309–1334
DOI: 10.1214/09-AOAS291

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

By Keith A. Baggerly[1] and Kevin R. Coombes[2]

University of Texas

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results,

The Annals of Applied Statistics
2009, V
DOI: 1

Markowetz *Genome Biology* (2015) 16:274
DOI 10.1186/s13059-015-0850-7

Genome Biology

COMMENT Open Access

Five selfish reasons to work reproducibly

Florian Markowetz

Abstract

And so, my fellow scientists: ask not what you can do for reproducibility; ask what reproducibility can do for you! Here, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Keywords: Reproducibility, Scientific career

A complex equation on the left half of a black board, an even more complex equation on the right half. A short sentence links the two equations: "Here a miracle occurs". Two mathematicians in deep thought. "I think you should be more explicit in this step", says one to the other.

This is exactly how it seems when you try to figure how science actually is. And, whether you like it or not, science is all about more publications, more impact factor, more money and more career. More, more, more... so how does working reproducibly help me achieve more as a scientist.

Reproducibility: what's in it for me?

In this article, I present five reasons why working reproducibly pays off in the long run and is in the self-interest of every ambitious, career-oriented scientist.

Reason number 1: reproducibility helps to avoid disaster

"How bright promise in cancer testing fell apart" titled a *The New York Times* article published in summer 2011 [1] highlighting the work of Keith Baggerly and Kevin Coombes, two biostatisticians at M.D. Anderson Cancer

The Annals of Applied Statistics
2009, V
DOI: 1
© Inst

Markowetz *Genome Biology* (2015) 16:274
DOI 10.1186/s13059-015-0850-7

Genome Biology

COMM

Five

Florian Ma

Abstract

And so, r
for repro
for you! H
reproduc
self-inter
scientist.

Keyword

A complex
even more
sentence l
curs". Two
should be
other.

Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOU

Search

New Results

Tools and techniques for computational reproducibility

Stephen R Piccolo, Adam B Lee, Michael B Frampton

doi: http://dx.doi.org/10.1101/022707

Abstract | Info/History | Metrics | Preview PDF

Abstract

When reporting research findings, scientists document the steps they followed so that others can verify and build upon the research. When those steps have been described in sufficient detail that others can retrace the steps and obtain similar results, the research is said to be reproducible. Computers play a vital role in many research

5 selfish reasons
to work reproducibly

1. Avoid disaster
2. Easier to write papers
3. Easier to talk to reviewers
4. Continuity of your work/in the lab
5. Reputation

Publishing Better Science through Better Data 2016 (**#scidata16**)

Cartoon by
Royston Robertson
www.roystoncartoons.com/