It can happen to you:
Sources and proximity of lack of reproducibility

Ross Hardison

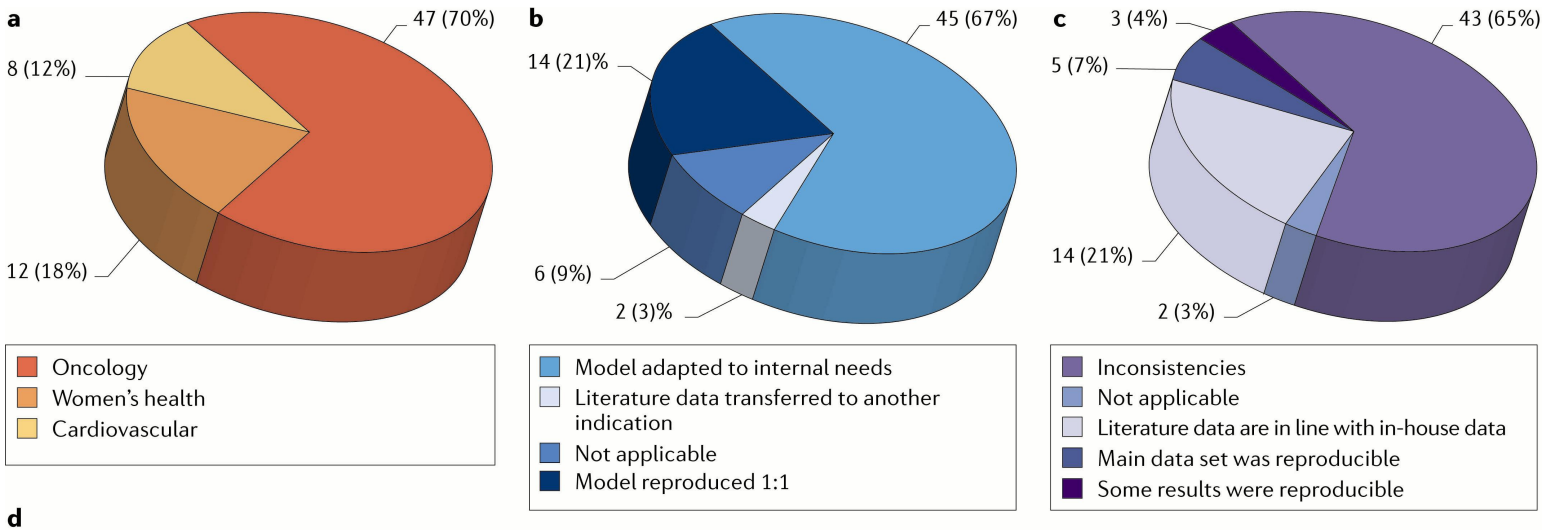
Penn State University

July 10, 2017

There is a problem

- For example, a substantial majority of prominent results in preclinical cancer studies **fail** to be reproduced independently.
- Prinz et al. 2011. Nature Reviews Drug Discovery. Bayer HealthCare, Germany
- Begley and Ellis. 2012. Nature. Amgen, Thousand Oaks, CA

Only 20-25% of published results are consistent with independent tests in-house



d

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

Nature Reviews | Drug Discovery

Prinz et al. 2011. Nature Reviews Drug Discovery.

Similar poor experience at Amgen

- Begley and Ellis. 2012. Nature 483: 531-533.
- Over the past decade, scientists in the hematology and oncology department at Amgen tried to confirm published findings related to potential drug targets of interest.
- 53 papers described “landmark” studies
- Significant findings were confirmed in only 6 cases – 11% of the studies.
- Contact original authors, discuss discrepant findings, exchange reagents, etc: Still low rate of reproducibility

Sources of lack of reproducibility

- A. Fabrication
- B. Inadequate measures for data quality
- C. Inadequate measures for reproducibility
- D. Biased reporting of results
- E. Inappropriate analysis
- F. Incomplete description of methods

A. Fabrication: Infamous examples

- William Summerlin (1974) Memorial Sloan-Kettering Research Institute
  - Transplant research: expected change in coat color; drew patches on mice with a black marker pen
- Eric Poehlman (1992-2002), University of Vermont
  - Fabricated data in 10 research papers on hormone replacement therapy and ageing
- Andrew Wakefield (1998): Lancet paper linking autism with MMR vaccine
  - "highly selective reporting of data"
- Hwang Woo-Suk (2004-2005) : papers in Science on production of human embryonic stem cells by somatic cell nuclear transfer
  - Data fabrication
- Later today: Keith Baggerly video on inability to reproduce results for cancer treatment predictions from transcriptomes. Was Anil Potti guilty of fabrication, or was it all data mix-ups and poor analysis?

- Selected examples from presentation by Chris Willmott (University of Leicester)
- http://www.slideshare.net/cjrw2/infamous-cases-of-research-misconduct

Is it just someone else's problem?

- When I was training, I thought you'd have to be crazy to think you'd get away with fabrication
- Seriously – using your marker pen to paint mice???

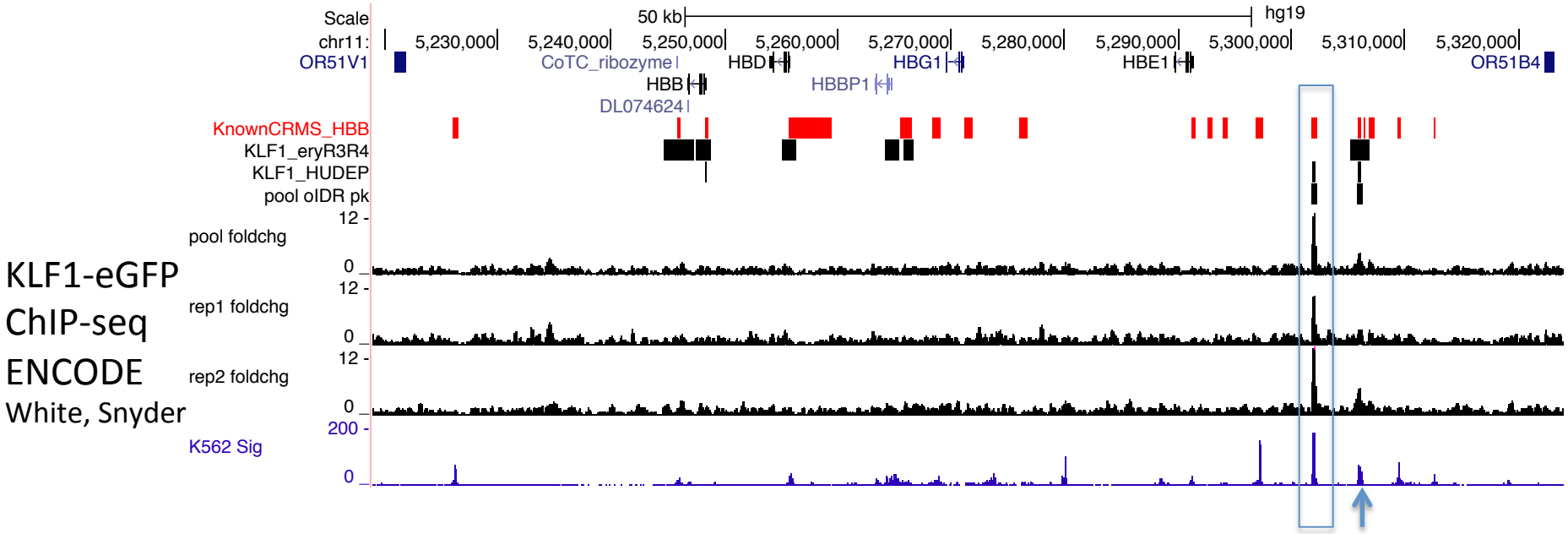
Is fabrication rare or common?

- I've reviewed cases of fabrication *at this University*, under the direction of the Office for Research Protections
  - If you suspect data fabrication or other research misconduct, contact Candice A. "Candy" Yekel, Associate Vice President for Research, Director, Office for Research Protections
- Two years ago, a Ph.D. thesis and degree were withdrawn because of plagiarism
- *It does happen*!

B. Inadequate measures for data quality

- Ideal situation: high quality measurements fall within a range of a quality metric score agreed upon by the community of experimentalists
- Unfortunately, this is not always achieved
- The issue is more challenging with large volumes of data from massively parallel assays, such as hybridizations to microarrays and second generation sequencing machines (e.g. Illumina)

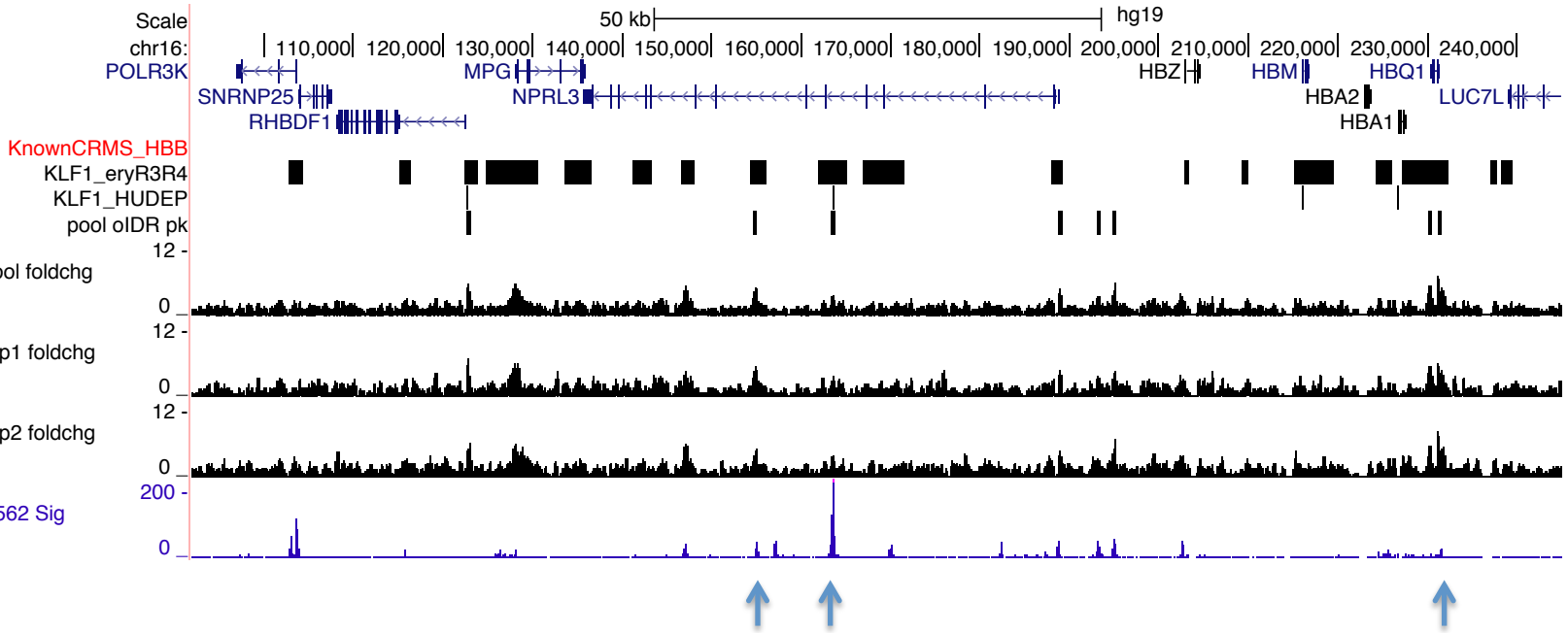
Quality inference, KLF1 at *HBB* locus



- KLF1 is an important transcription factor, but notoriously hard to ChIP
- Earlier data was very noisy (eryR3R4)
- Tagged protein KLF1-eGFP has an obvious peak, but how robust is evidence for 2nd peak?

Quality inference, KLF1 at *HBA* locus

KLF1-eGFP ChIP-seq ENCODE White, Snyder



- Some of these peak calls are surprising

Progress on quality metrics

- With respect to massively parallel, sequencing data, the ENCODE consortium has set some standards
- Landt et al. 2012. Genome Research 22: 1813-1831: ChIP-seq
  - E.g. fraction of reads in peaks (FRiP)
- RNA-seq, others: ENCODE data portal
- https://www.encodeproject.org
- Work on standards continues.

C. Inadequate measures for data reproducibility

- First of all: Always replicate the experiments!
  - Major issue stressed by Begley and Ellis
- When replicates are done, can still have errors arising from not knowing what is reproducible
- More later from Dr. Qunhua Li
- Back when we had 3 determinations in an assay for each condition, and tens of experiments, reproducibility or not was pretty obvious
- Data space now is enormous
- When you have hundreds of millions of observations (e.g. mapped sequencing reads), how do you assess reproducibility in an objective manner?

D. Biased reporting of results

- Examples
  - Reporting only some of the results – the ones that support the major conclusion
  - Showing only portions of a Western blot
  - Not using fully validated reagents
- This issue is emphasized by Begley and Ellis and by Prinz et al.
- Remedies:
  - Analysts blind to experimental and control groups
  - Have a different investigator replicate the result
  - Journals publish negative results

E. Inappropriate analysis

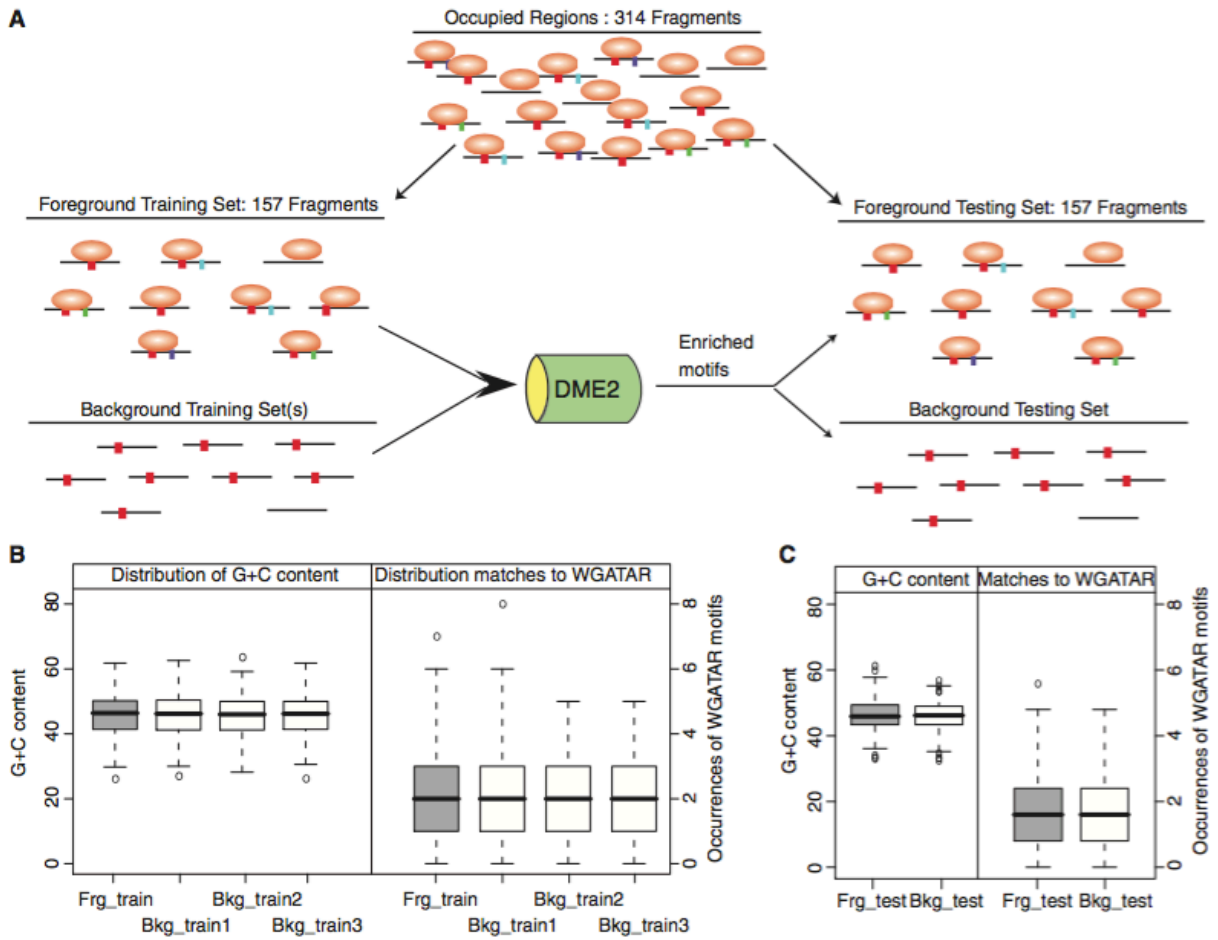
- Analysis is wrong
  - Striking examples in video lecture from Keith Baggerly, UTHSC Houston (this afternoon)
  - Reports of gene expression signatures that distinguish drug-sensitive from resistant cancers, Anil Potti and colleagues at Duke University
  - Errors in gene ids and mix-ups of labels (resistant vs sensitive)
  - Suspect that “the most simple mistakes are common”
- Analysis is misleading
  - Usually inadvertent
  - Have plenty of high quality data in a well-designed experiment
  - But are the results of your analyses robust and biologically meaningful?

General example: Choice of negative controls

- Genomes of most eukaryotes are large, complex, and highly heterogeneous
- The sequences are not random
- What do we mean by the "null expectation" when we calculate enrichment?
- This question does not have one common answer for all applications, e.g.
  - Random sequences of the same base composition as the targets of interest
  - Randomly chosen DNA segments in the vicinity of the targets of interest
  - Randomly chosen DNA segments with a similar distribution of distances from gene features (start site, exons, etc) as those in the targets of interest

E.g. Search for features that distinguish TF-bound from unbound DNA segments

7028 Nucleic Acids Research, 2009, Vol. 37, No. 21



- Match foreground (TF bound) and background (not bound) for potentially confounding features
- GC content
- Matches to primary motif

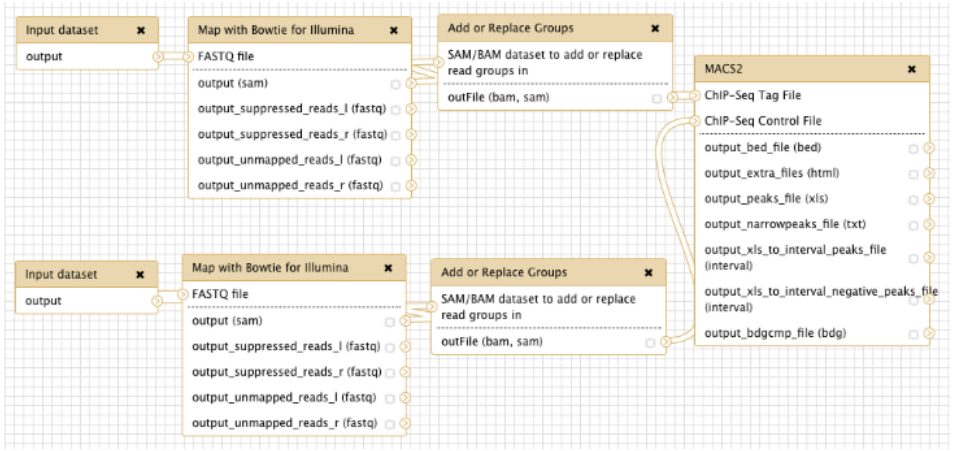
Ying Zhang et al. 2009. NAR

E. Inappropriate analysis: summary

- This may be one of the most pervasive problems in scientific research
- It is certainly a huge issue in Big Data
- In genomics, the negative dataset is not always obvious
  - Use more than one!
- Even with appropriate negative datasets, care is needed in applying the analysis
  - Vetting by independent analysts

F. Incomplete description of methods

- If you don't tell people what you did, how can they reproduce it?
- This is a common problem, but it is not acceptable
- Supplementary material rarely has a limit, you can explain it there
- Have another researcher read the methods and ask them – Can you do this procedure following these methods?
- Use workflows that you make public, e.g. via Galaxy
- More later in the Boot Camp



Reproducibility studies are being funded and published

- Shan et al. 2017. eLife; 6:e25306
- "The Reproducibility Project: Cancer Biology (RP:CB) … seeks to address concerns about reproducibility in scientific research by conducting replications of selected experiments from a number of high-profile papers in the field of cancer biology (Errington et al., 2014).
- For each of these papers a Registered Report detailing the proposed experimental designs and protocols for the replications was peer reviewed and published prior to data collection.
- The present paper is a Replication Study that reports the results of the replication experiments detailed in the Registered Report (Fung et al., 2015), for a paper by Dawson et al.
- Collaboration between the Center for Open Science and Science Exchange"

Replication study: Inhibition of BET recruitment …

eLIFE
elifesciences.org

REPLICATION STUDY

Replication Study: Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia

Xiaochuan Shan[1], Juan Jose Fung[2], Alan Kosaka[2], Gwenn Danet-Desnoyers[1], Reproducibility Project: Cancer Biology*

[1]University of Pennsylvania, Perelman School of Medicine, Stem Cell and Xenograft Core, Philadelphia, United States; [2]ProNovus Bioscience, LLC, Mountain View, United States

- "We found treatment of MLL-fusion leukemia cells … with the BET bromodomain inhibitor I-BET151 resulted in selective growth inhibition, ..., this is similar to the findings reported in the original study"

Another replication study

eLIFE elifesciences.org

REPLICATION STUDY

Replication Study: The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate

Megan Reed Showalter[1†], Jason Hatakeyama[2,3†], Tomas Cajka[1†], Kacey VanderVorst[2,3], Kermit L Carraway III[2,3], Oliver Fiehn[1], Reproducibility Project: Cancer Biology*

[1]West Coast Metabolomics Center, University of California, Davis, United States; [2]Department of Biochemistry and Molecular Medicine, University of California, California, United States; [3]University of California Davis Comprehensive Cancer Center, University of California, California, United States

- "These results are similar to those reported in the original study"

Sources of lack of reproducibility

- A. Fabrication
- B. Inadequate measures for data quality
- C. Inadequate measures for reproducibility
- D. Biased reporting of results
- E. Inappropriate analysis
- F. Incomplete description of methods