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

Ross Hardison

Penn State University

June 6, 2016

Sources of lack of reproducibility

- A. Fabrication
- B. Inadequate measures for data quality and reproducibility
- C. Biased reporting of results
- D. Inappropriate analysis
- E. 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: 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
- Last year, a Ph.D. thesis and degree were withdrawn because of plagiarism
- *It does happen*!

B. Inadequate measures for data quality and reproducibility

- 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, 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?

C. 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
- More from Dr. Broach on this topic (Begley papers)
- One remedy: Have a different investigator replicate the result

D. Inappropriate analysis

- Analysis is misleading
- Usually inadvertant
- 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

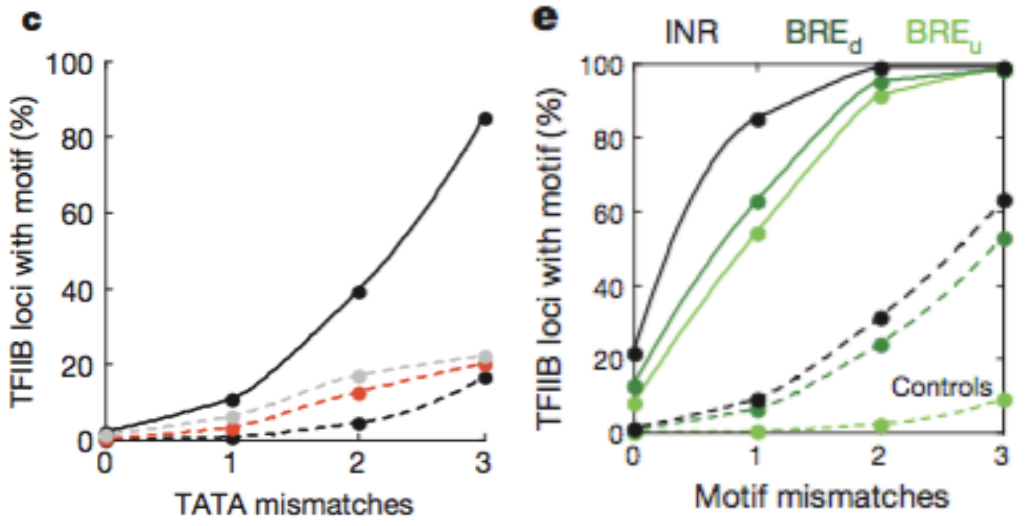
- Genomes of most eukaryotes are large, complex, and highly heterogeneous
- The sequences are not random
- What do we mean by the “null expectation” when we calculate enrichment?
- This question does not have one common answer for all applications
  - Random sequences of the same base composition as the targets of interest
  - Randomly chosen DNA segments in the vicinity of the targets of interest (Genome structure correction)
  - Others

Specific example: Motif enrichment

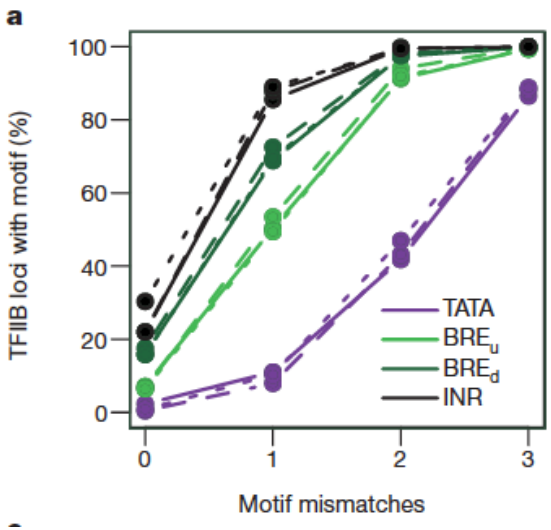
- Paper in Nature with high resolution ChIP-exo data on binding sites for TFIIB and POL2 in a mammalian cell line
  - Venters and Pugh (2013) Nature 502:53-58
- Valuable data; new, comprehensive insights into promoters genome-wide, defined by the key proteins that act at promoters
- Found that these biochemically defined promoters have matches to Core Promoter Elements
- Also found less enrichment in negative controls

But are the motifs enriched in the promoters?

- Brief Communication Arising paper says that the motifs are **not** enriched
  - Siebert and Söding (2014) Nature 511: E11-E12
- Used a different method for calculating frequency in negative controls
- Venters and Pugh retracted the paper, despite the fact that the basic data are sound.



Venters and Pugh



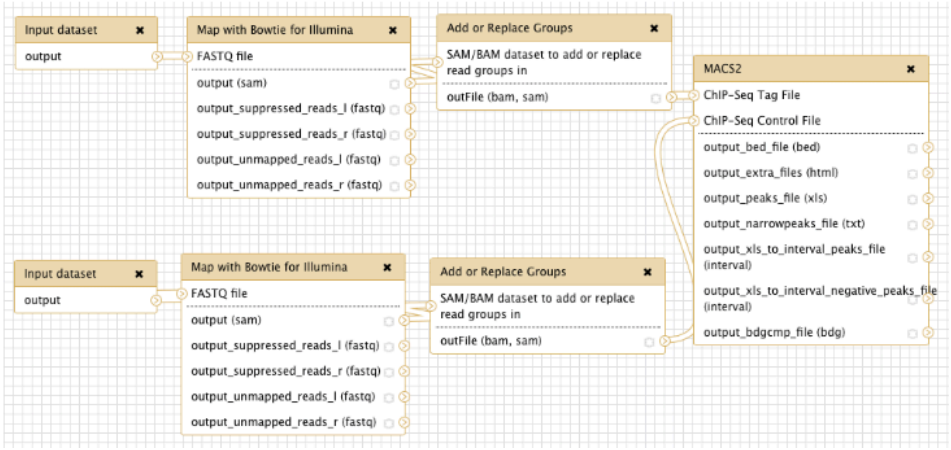
Siebert and Söding

D. 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

E. 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



Sources of lack of reproducibility

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