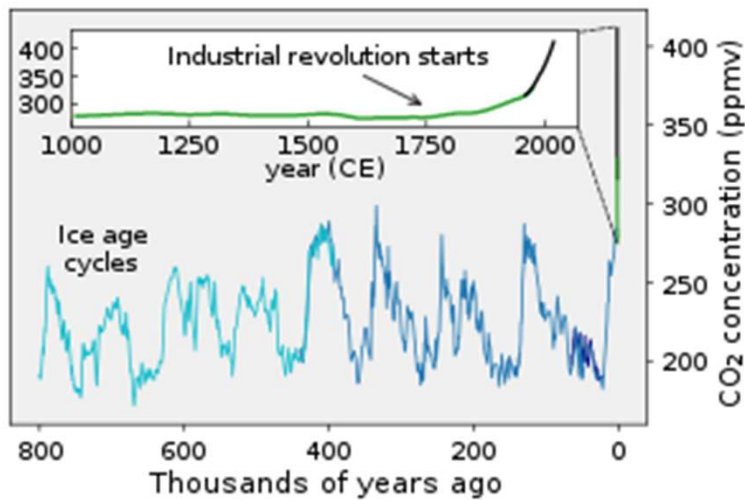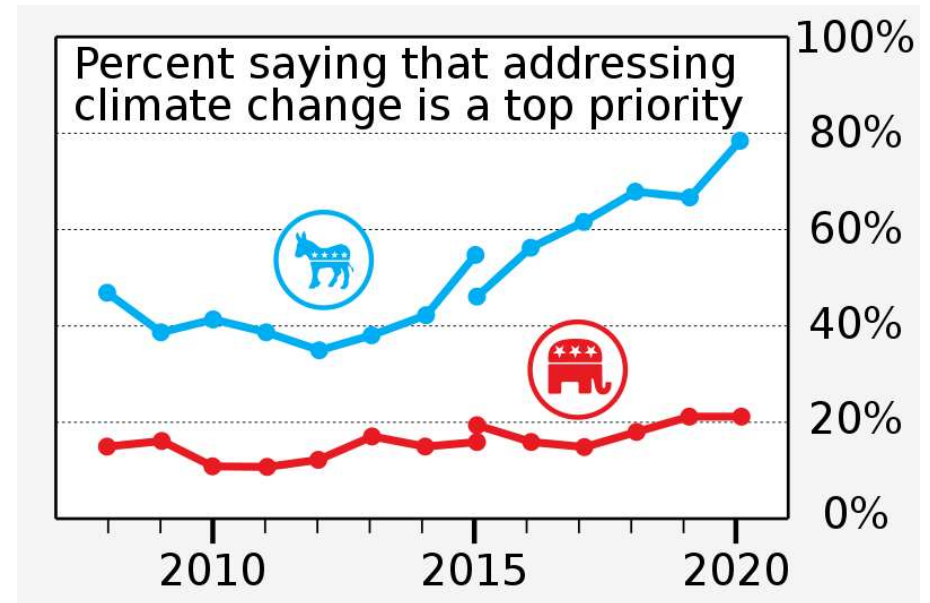# FCCC_21RCR#1_Data irreproducibility/fraud: why, why not, and how not to
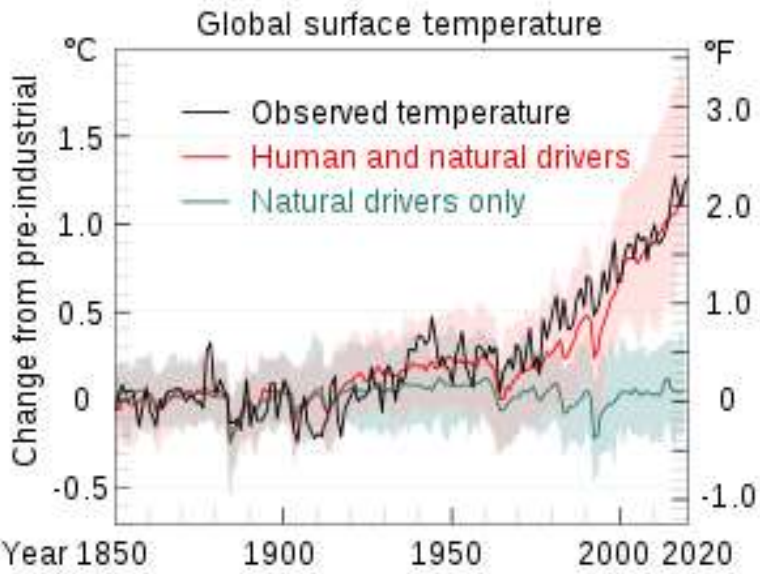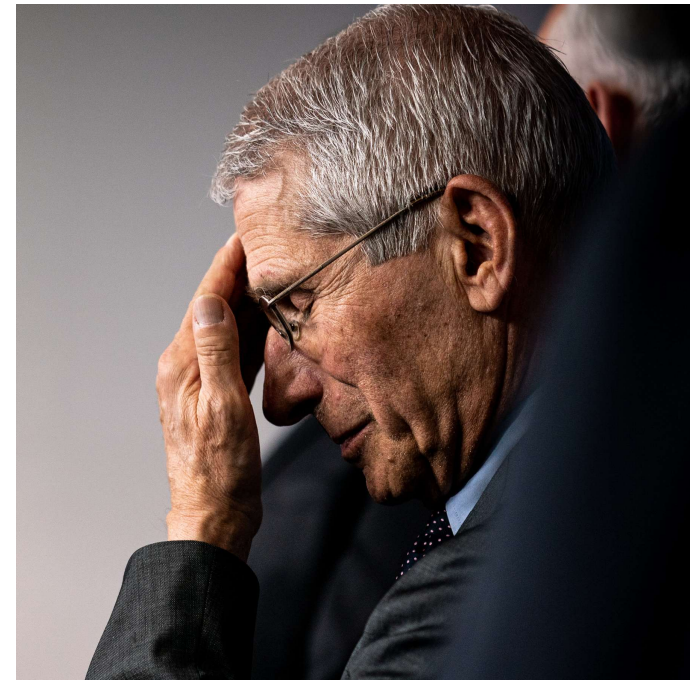
Eroding public trust in science and institutions

Eroding public trust in science and institutions

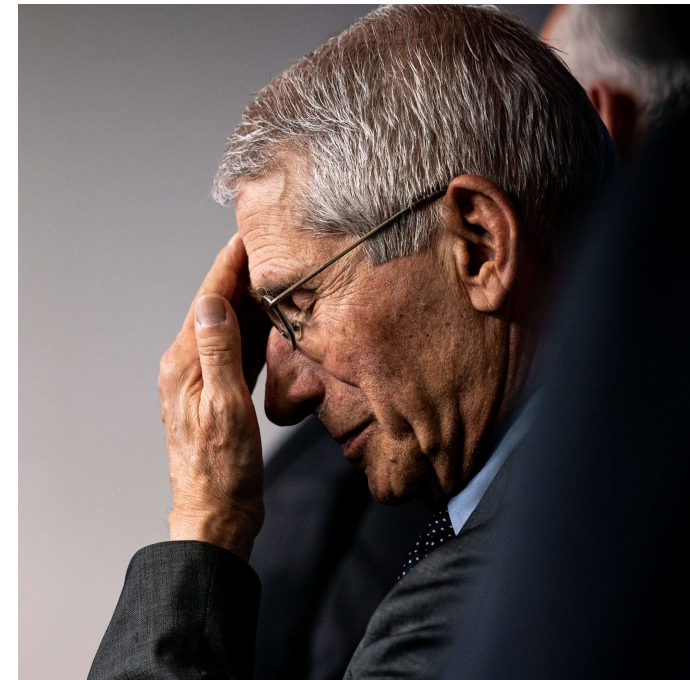# Eroding public trust in science and institutions

Science's own-goals - **irreproducibility and fraud**

Misinformation, echo-chambers, internet cults are here to stay

Meanwhile... What the narrative of science is blemished by ... Unforced errors

Erosion, alternative facts

Tougher funding climates

differentiating quality competing with bio-tech industry

Science has a reproducibility crisis and a systematic fraud issue.

We need to talk about systematic fraud

Software that uncovers suspicious papers will do little for a community that does not confront organized research fraud, says Jennifer Byrne.

nature > news feature > article

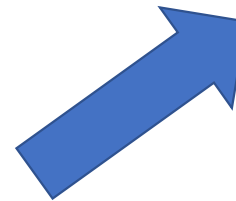## 1,500 scientists lift the lid on reproducibility

Monya Baker

- 90% agrees there is a reproducibility crisis

- 70% failed to reproduce other's data

- 50% could not reproduce data from the same lab

Science has a reproducibility crisis and a systematic fraud issue.

OPEN ACCESS | Perspective | 07 June 2016

The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications

Authors: Elisabeth M. Bik, Arturo Casadevall, and Ferric C. Fang | AUTHORS INFO & AFFILIATIONS
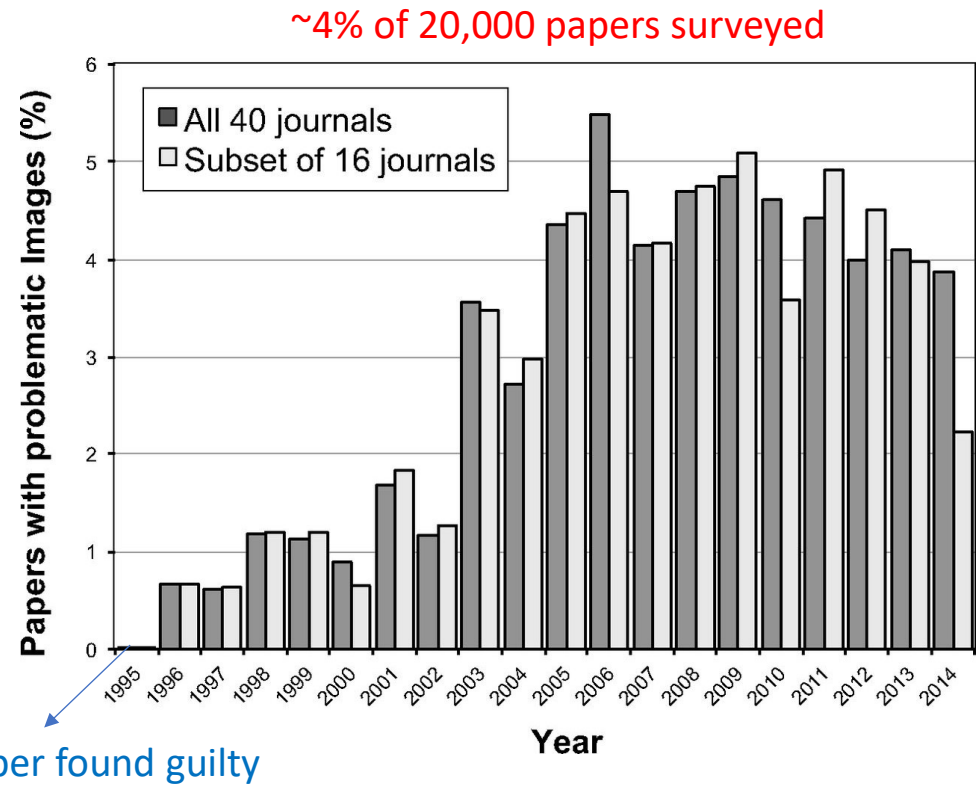
~4% of 20,000 papers surveyed

Zero paper found guilty

Science has a reproducibility crisis and a systematic fraud issue.
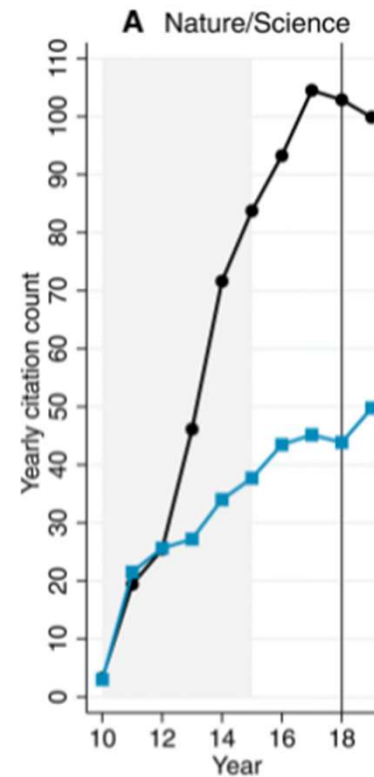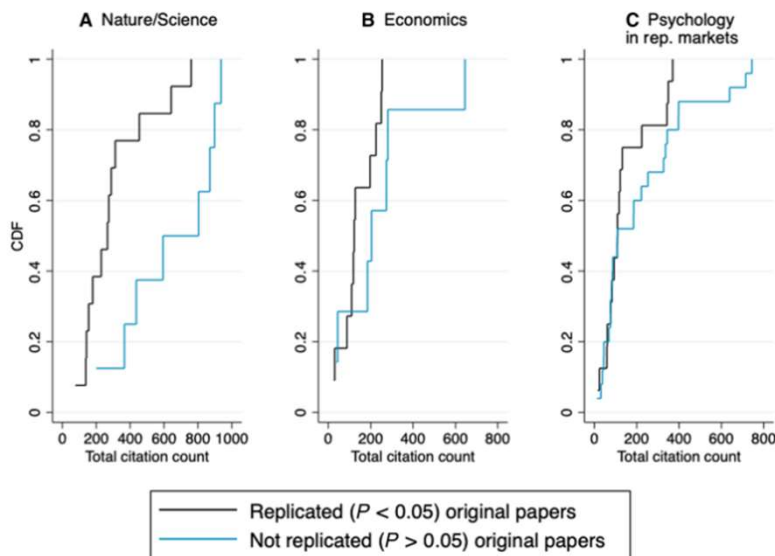
- Reviewer/editor tradeoff: Irreproducible papers in Nature/Science are cited more often than reproducible ones.

- "The hype does die" : After the irreproducible results came out, only 12% of post-exposure citation acknowledge

- Significant citation drop only seen after an irreproducible paper is retracted.

Science has a reproducibility Crisis! – because of science

- Science is driven by human nature: passion, creativity, limitation
- Science is reiterative: breaking old rules, establish new paradigms
  - "Scientists are wrong all the time"
  - Two type of research: textbook science, vs. frontier science
- There is no right or predetermined formula for desirable science

Sloppy vs. fraud – the boundary often blurry

○Example of "honest" mistake
- Antibody specificity issue, e.g. cross-reactivity
- RNAi, CRISPR, morpholino off-target effect
- Phenotypic drift due to transgenic silencing
- Correlation vs. causation: "Starbuck" case

○Example of sloppy science - missing controls
- no Tamoxifen control,
- no primary Ab ctrl for staining
- loading ctrl for western blot

○Example of Fabrication & Falsification
- Reuse published data
- Cherry-picking, selectively omit contradictory results to expedite publication
- Create data without doing it or altering
- Try different methods of analysis until one yields **statistically** significant results
- Putting untruthful statement in grant application
- Super-spotter test level 1-4

Fraud-spotter level 1

36
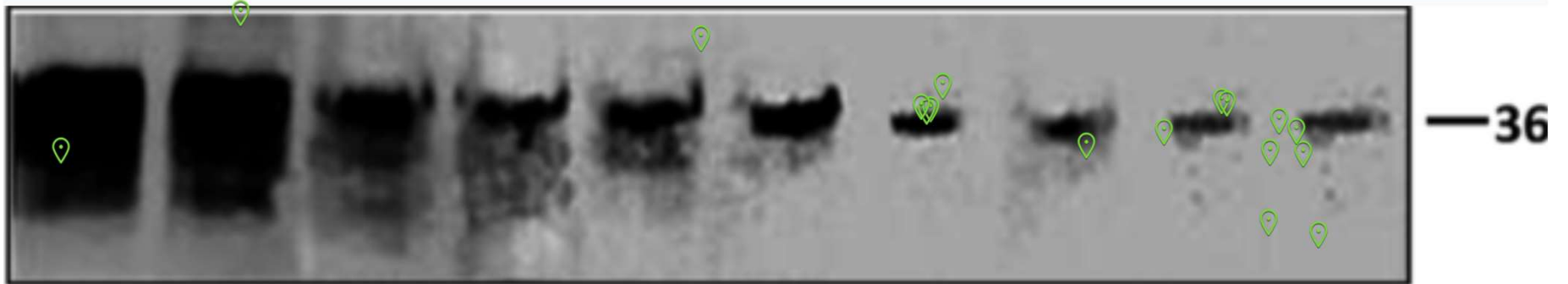
Fraud-spotter level 1
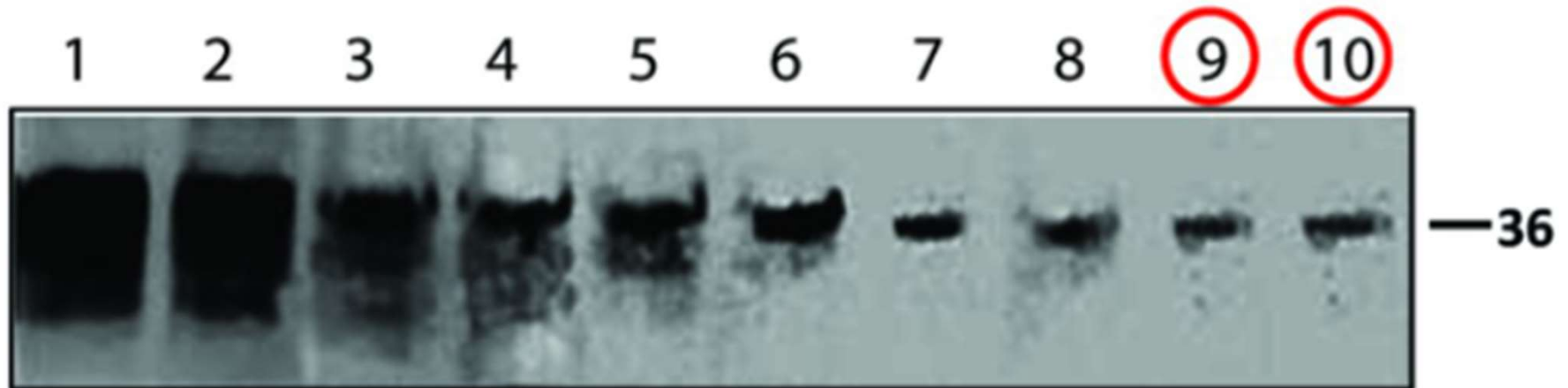
The figure shown here displays Western blotting results for 10 different protein fractions isolated from a density gradient. The figure appears to show a single blot, but the last two lanes (red circles) appear to contain an identical band. Exposure was altered to bring out details in reference 37; the figure was corrected in reference 38.

Fraud-spotter level 2
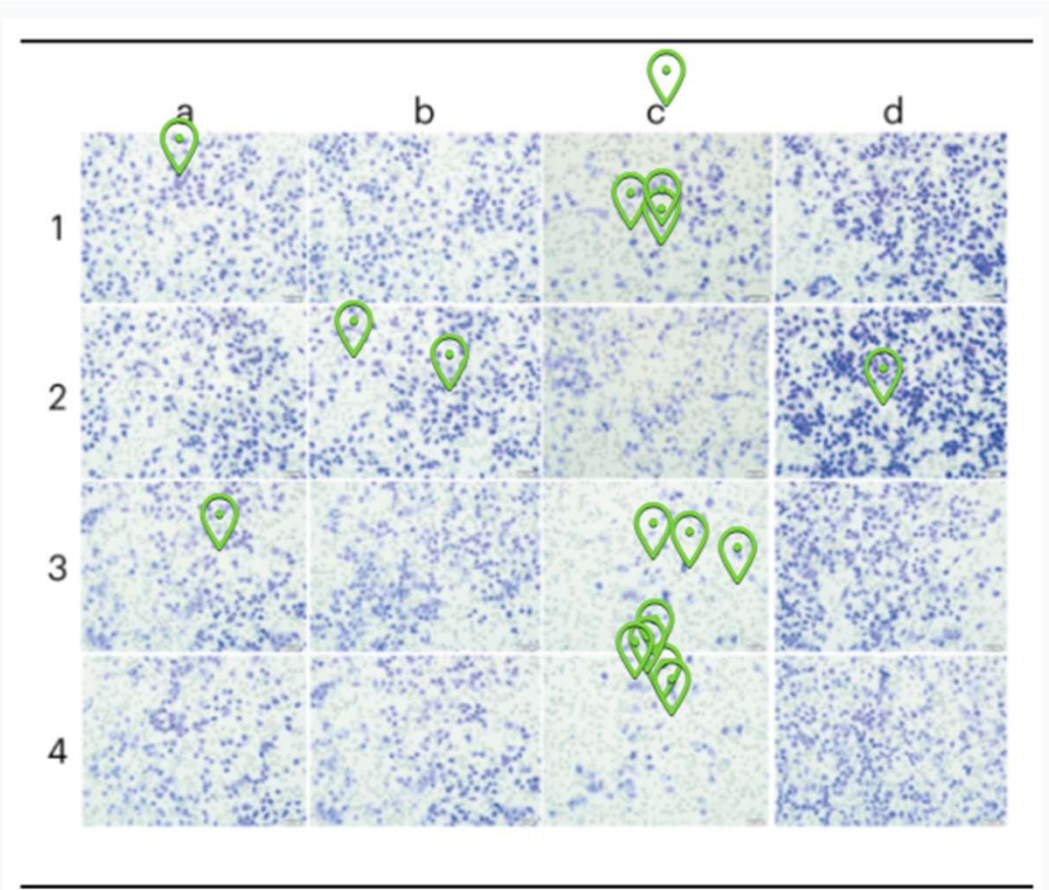
Fraud-spotter level 2
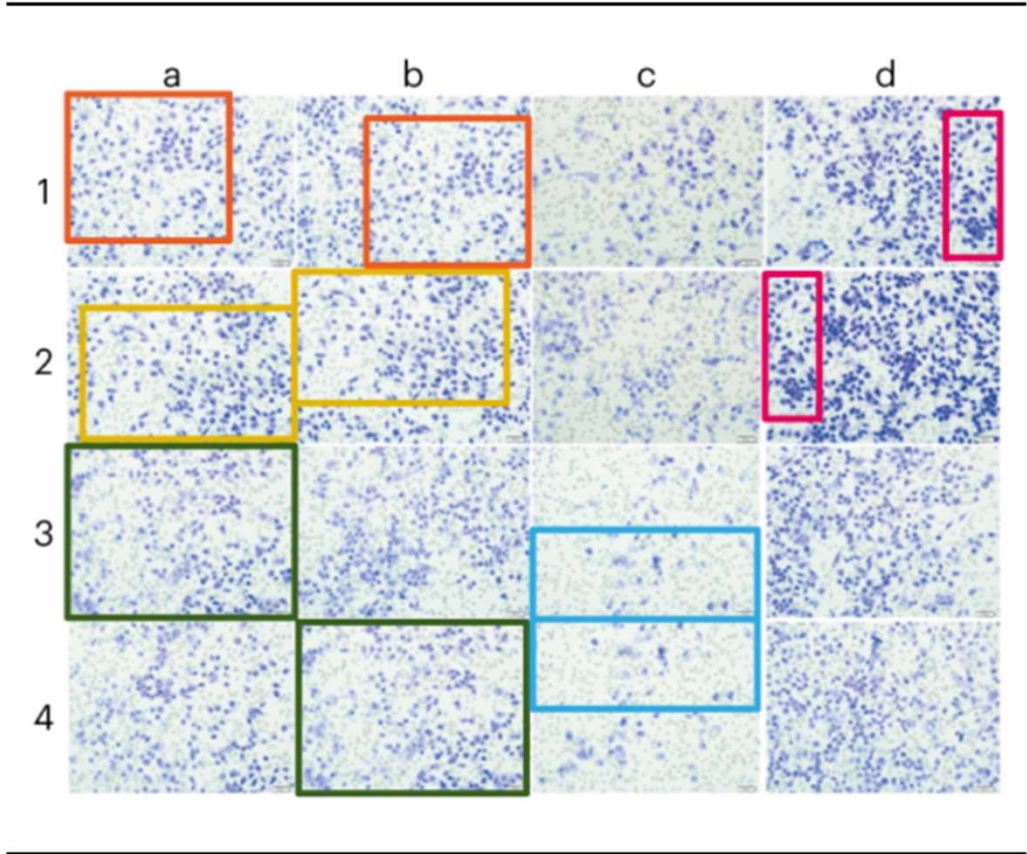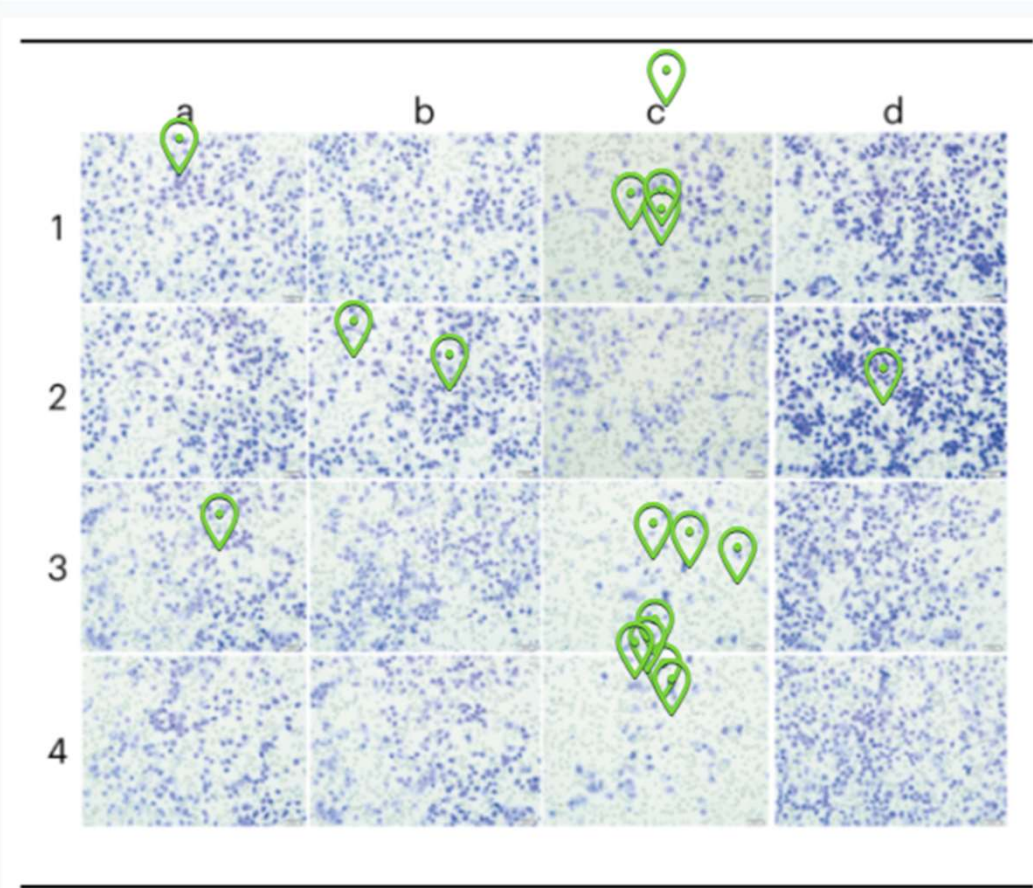
Fraud-spotter level 3
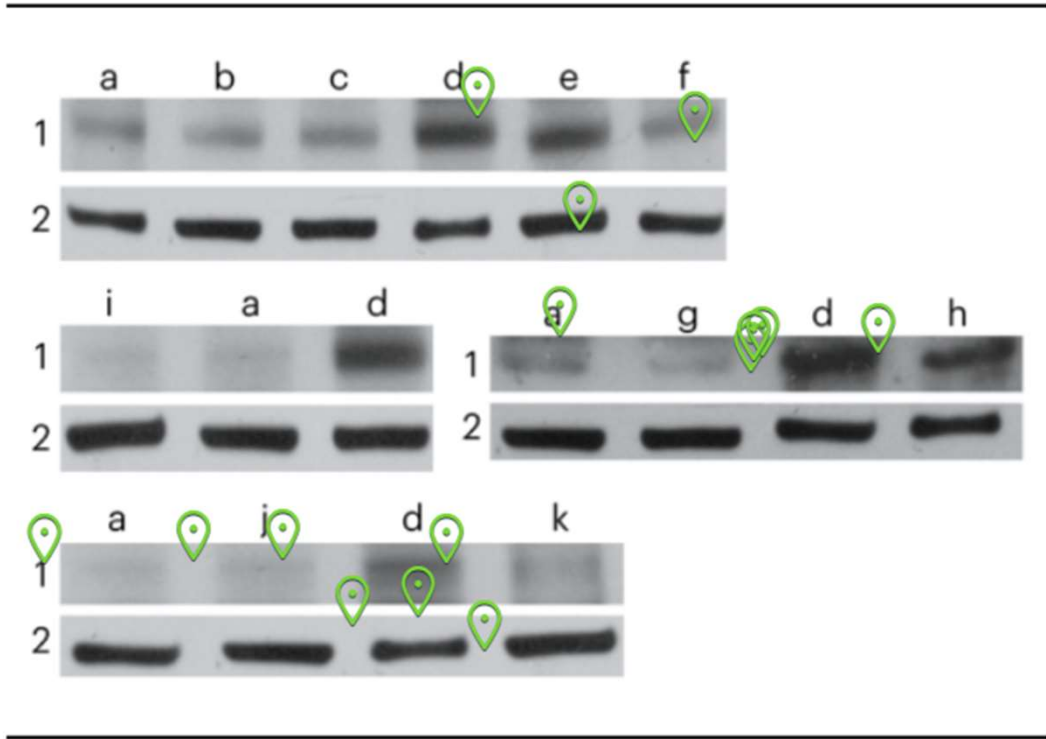
Fraud-spotter level 3
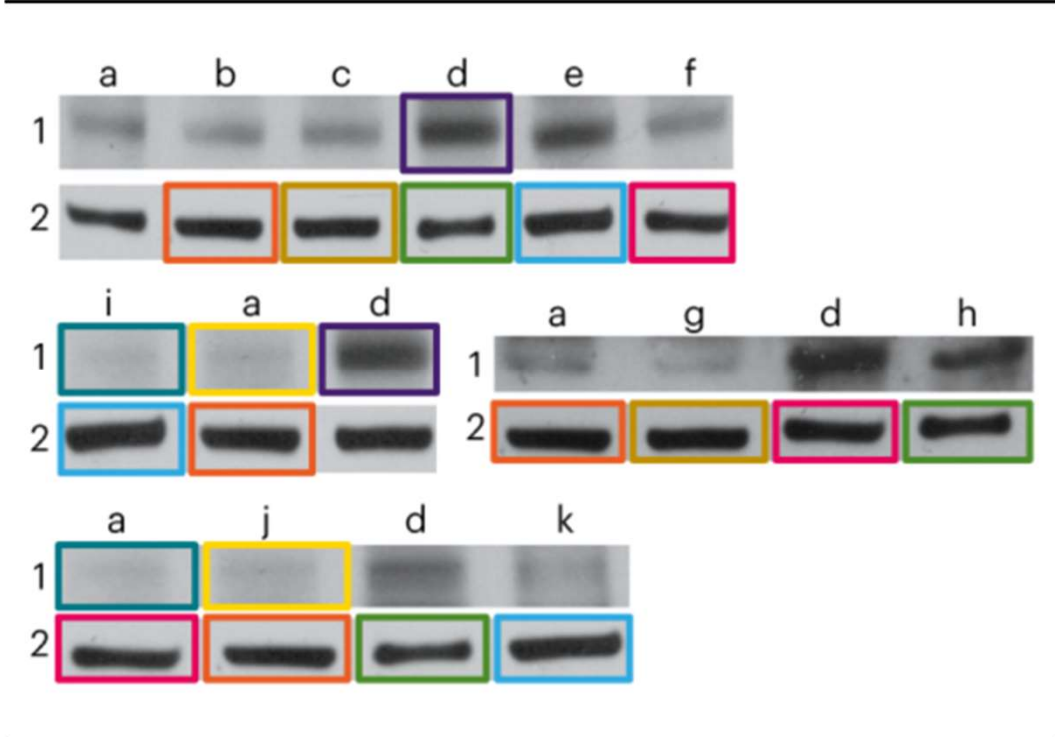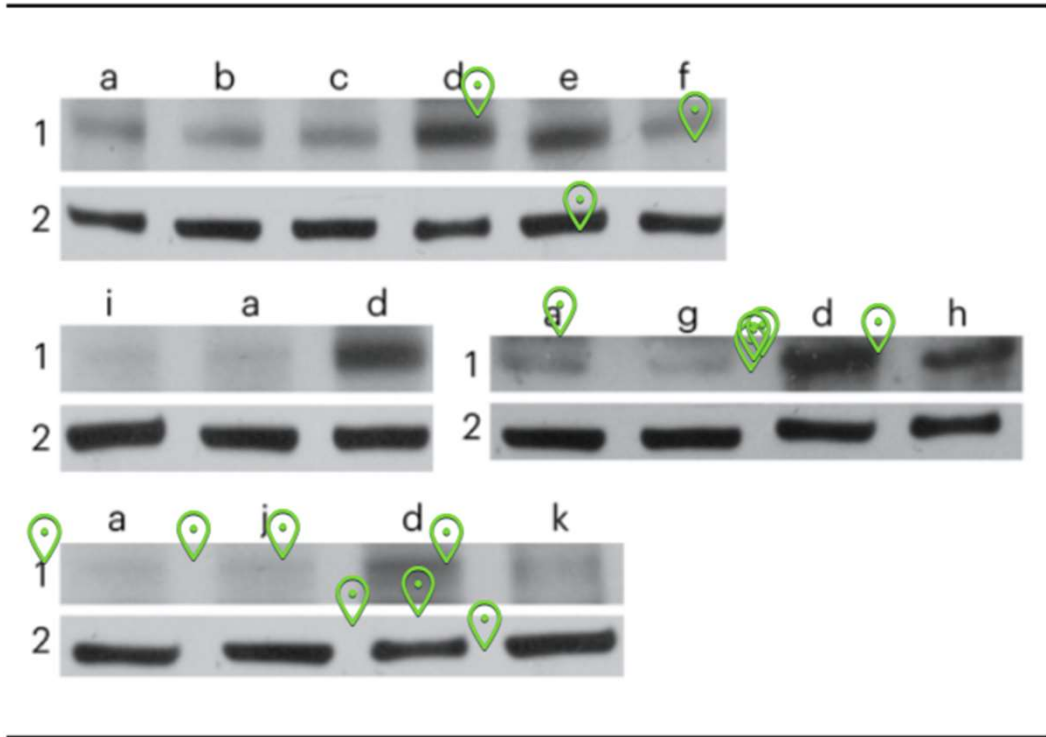
Fraud-spotter level 4
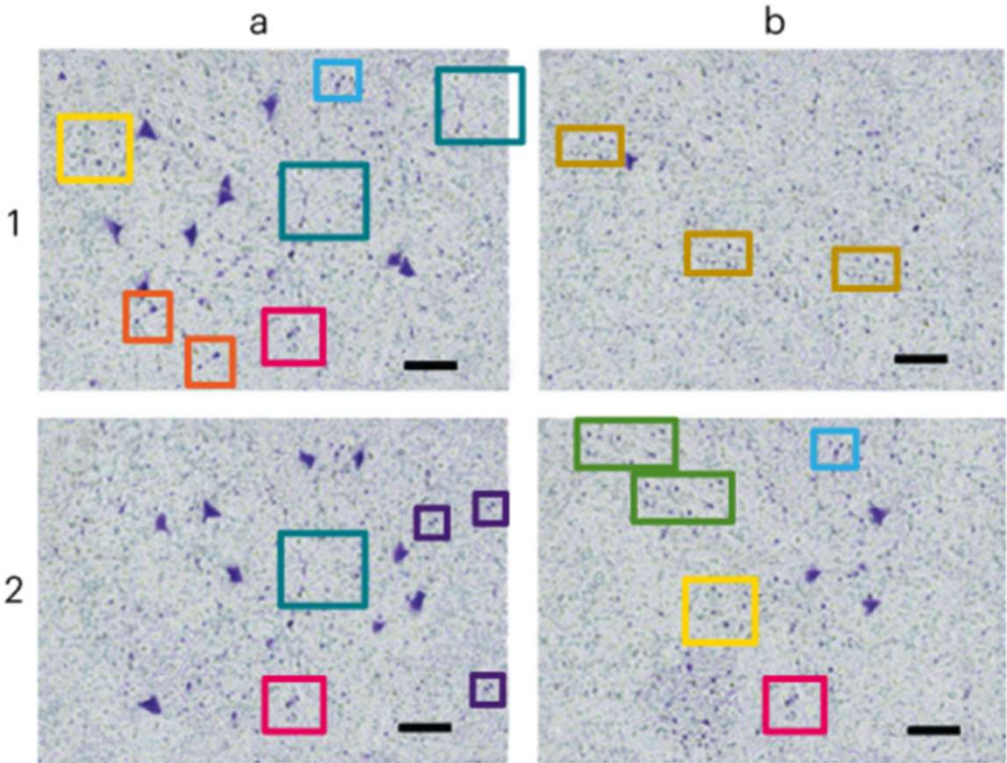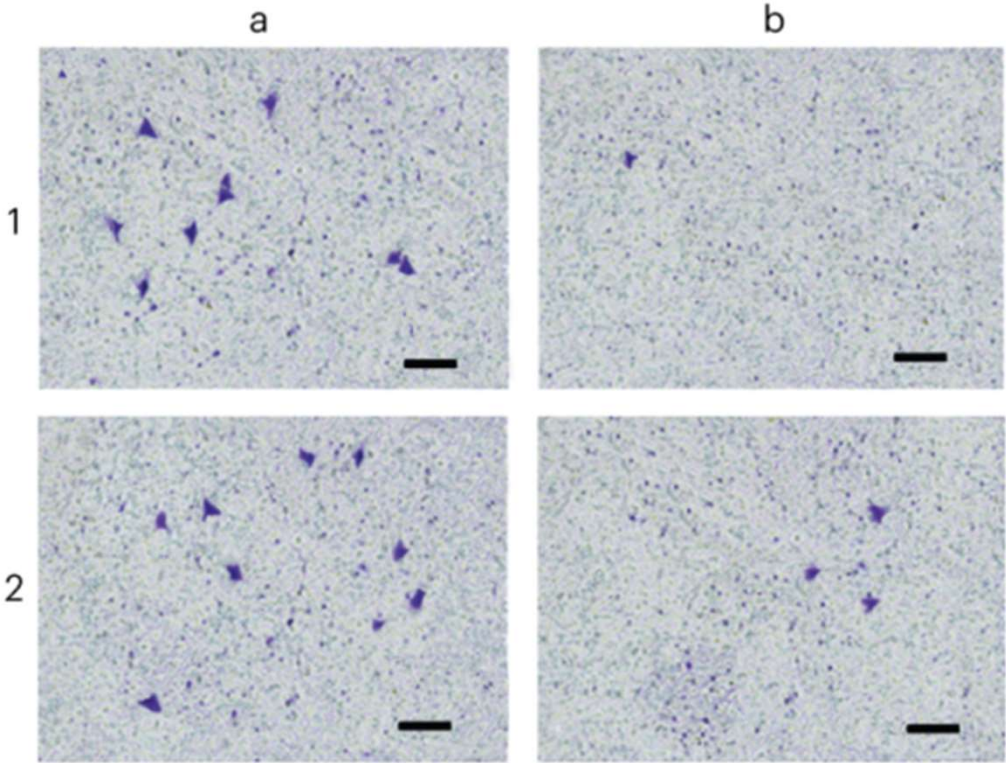
Why one should not commit frauds

- Why frauds happen?
  - Pressure? Money? Fame?

- "Paper cannot hold fire" – truth will prevail
  - Findings will be checked, cheating will inevitably be caught
  - The self-correcting nature will not let frauds unnoticed
  - Not worth the money and time - better to try to be a novelist, or a film director, a con-artist, a thief - may have better investment-return ratio

Why one should not commit frauds

- Why frauds happen?
  - Pressure? Money? Fame?

- Paper cannot hold fire – truth will prevail
  - Important findings will be checked and scrutinized, cheating will inevitably be caught
  - The self-correcting nature will not let frauds unnoticed
  - Not worth the money and time – better try a novelist, a film director, or if immoral, a con-artist, or a thief - may have better investment-return ratio

- The cops are getting more and more sophisticated
  - Pubpeers
  - Retraction watch
  - Super-spotters hired by journals and institutions
  - AI, machine learning, and automation, image-checker, hired by journals and institutions

SEEING DOUBLE

Elisabeth Bik quit her job to spot errors in research papers – and has become the public face of image sleuthing. By Helen Shen

What if you encounter research frauds in Fox Chase?

- To be or not to be (a whistle blower)

- Internal Due Process

- Anonymous and confidentiality

- Potential Mechanisms

What if you encounter research frauds in Fox Chase?

- To be or not to be (a whistle blower)

- Internal Due Process

- Anonymous and confidentiality

- Potential actions

CAO- Glenn Rall

CSO- Jon Chernoff

Program Leaders

Other faculty members you trust

Outside approaches:

How to better insulate us from frauds

Rank factors contribute to data irreproducibility/fraud

- pressure to publish — 1st
- selective reporting — 2nd
- insufficient oversight/mentoring — 3rd
- not replicated enough in original lab — 4th
- raw data not available from original lab — 5th
- insufficient statistic power — 6th
- methods, code insufficient sharing — 7th
- insufficient peer review — 8th
- reagent variability — 9th
- the blind eyes by journals — 10th

How to better insulate us from frauds

Journals and the Peer-review

When should be a problematic paper be retracted?

If multiple papers have been retracted or been flagged from the same group, should there be accountability and necessary repercussion on future submissions?

Hype train to where? – should there be lower standard on trendy topics and ideas? (tradeoff for higher citability – worth it?)

Demanding raw data and detailed methodology

How to better insulate us from frauds

Journals and the Peer-review

Demanding raw data and detailed methodology

Supplementary information

Supplementary Information

This file contains Supplementary Text, Su Table 1.

Reporting Summary

Supplementary Data 1

Pymol session showing the top 10 hTR m core.

Supplementary Data 2

Supplementary Data 3

Supplementary Data 4

Peer Review File

Source data

Source Data Extended Data Fig. 2

Extended Data Figure 7g immunoblot

α-Strep   α-H2A   α-H2B   α-H3   α-H4

Extended Data Figure 7h immunoblot

α-H2A   α-H2B/α-Dyskerin   α-H3   α-H4

- α-Dyskerin
- α-H2B

Extended Data Figure 7f left
SDS-PAGE H2A:H2B

Extended Data Figure 7f right
urea PAGE CR4/CR5

Fig. 3f
H2A:H2B EMSA replicate 1
autoradiogram

Fig. 3f
H2A:H2B EMSA replicate 2
autoradiogram

Figure 3f
H2A:H2B EMSA replicate 3
autoradiogram
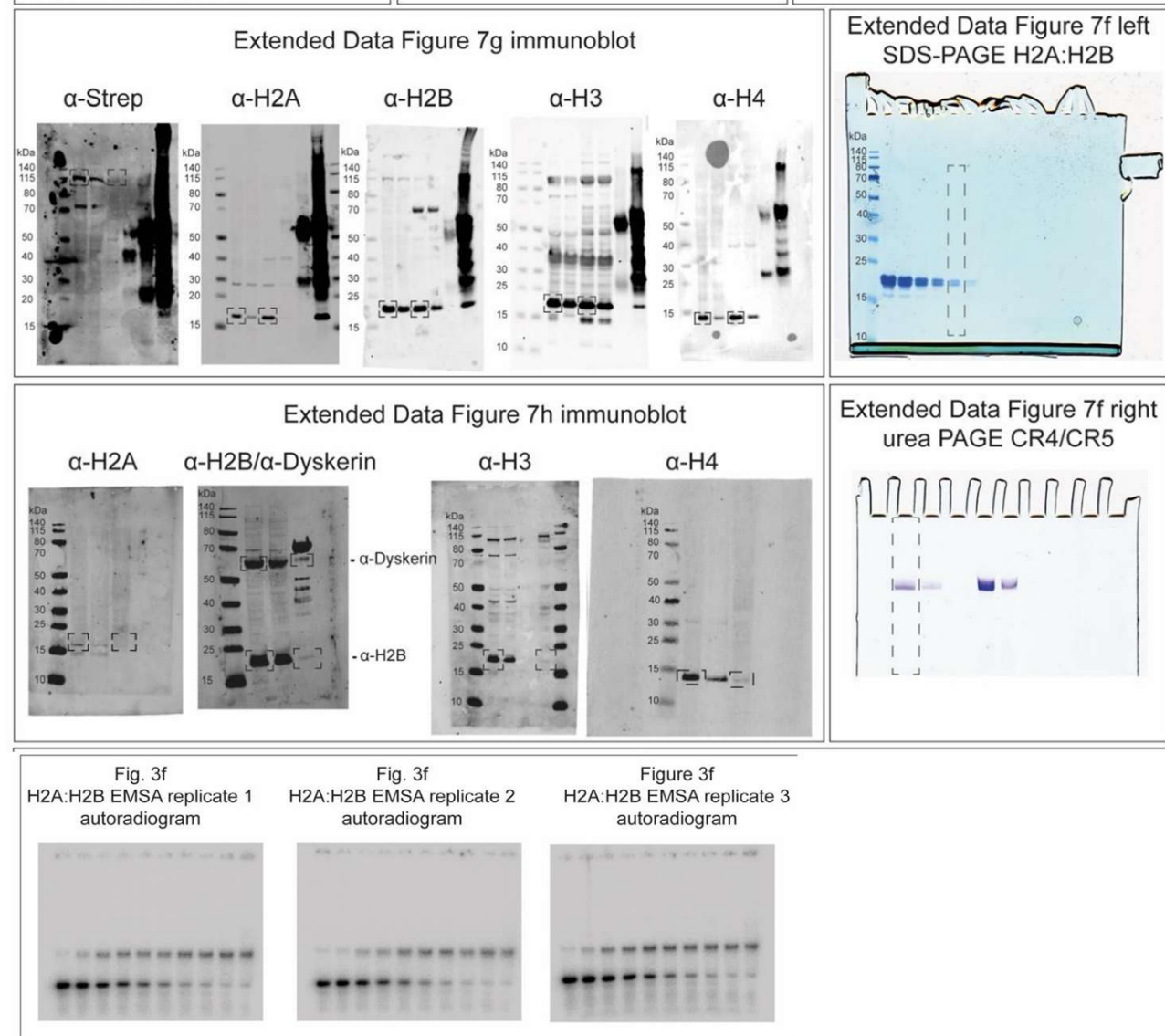
How to better insulate us from frauds

Labs: Data recording and notebooks

Archive lab notebooks periodically

Link raw data file – analyses – polished data: naming system, storage and preservation of raw data- especially published ones

Lab notebook:
aims, plan, design, result interpretation, next step.

PI – more emphasis on raw data, be closer to the bench, reproducibility and interval validation

What system(s) do you use for your lab notebook?

physical lab notebooks with routine digital scanning/archiving	33%
physical lab notebooks without any archiving	67%
Benchling (compatible with molecular cloning/construct management)	
Evernote	
OneNote	
Others electronic notebooks	

How to better insulate us from frauds

Labs/mentoring: a culture that focuses on data

!!! Use orthologous approaches, multiple readouts

PI – more emphasis on raw data, be closer to the bench, reproducibility and interval validation – testing alternative hypothesis before publishing

make sure ample backup reagents are available for follow-up or data reproducing

Gel images: lane-cropping; exposures between images (ctrl vs experimental)

Statistics: consult with Eric Ross and co. for optimal statistic tools and model of choice

How to better insulate us from frauds

Labs/mentoring: a culture that focuses on data

!!! Use orthologous approaches, multiple readouts

PI – more emphasis on raw data, be closer to the bench, reproducibility and interval validation – testing alternative hypothesis before publishing

make sure ample backup reagents are available for follow-up or data reproducing

Gel images: lane-cropping; exposures between images (ctrl vs experimental)

Statistics: consult with Eric Ross and co. for optimal statistic tools and model of choice

How to better insulate us from frauds

Institution:

Have an in-house Integrity Officer

Original Data Depository: Stowers Institute

Culture of sharing of unpublished data, unpolished stories

CAO- Glenn Rall

CSO- Jon Chernoff

Program Leaders

Other faculty members you trust

Outside approaches:

Publications

Year	Title	Author/Journal	Accession ID		
2021 ⌄				Search	Reset

Original Research Papers

MOCS2 links nucleotide metabolism to nucleoli function
Sugamuna T, Swanson SK, Gogol M, Garrett TJ, Florens L, Workman JL. [published ahead of print October 26 2021]. *J Mol Cell Biol*. 2021-09-14.
Original Data

A distinct inner nuclear membrane proteome in *Saccharomyces cerevisiae* gametes
Shelton SN, Smith SE, Unruh JR, Jaspersen SL. *G3: Genes, Genomes, Genetics*. 2021:jkab345. doi: 310.1093/g1093journal/jkab1345.
Original Data

Elongin functions as a loading factor for Mediator at ATF6alpha-regulated ER stress response genes
He Y, Sato S, Tomomori-Sato C, Chen S, Goode ZH, Conaway JW, Conaway RC. *Proc Natl Acad Sci U S A*. 2021;118:e2108751118. doi: 10.1073/pnas.2108751118.
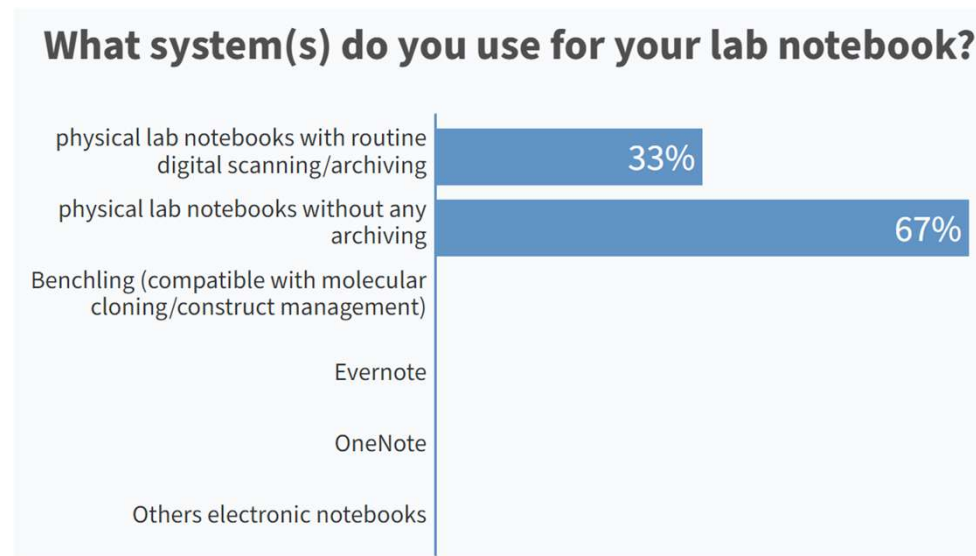
Thanks

How to better insulate us from frauds

○ Lab notebook taking
- ○ Like a police investigator/detective wearing body camera during work
- ○ Not just annotating the data, but also the detailed aim, goal, design, result interpretation, and plan for the next step
- ○ Importantly preserve the raw data, and link the raw data with analysis, and with the information

What system(s) do you use for your lab notebook?

System	Percentage
physical lab notebooks with routine digital scanning/archiving	33%
physical lab notebooks without any archiving	67%
Benchling (compatible with molecular cloning/construct management)	
Evernote	
OneNote	
Others electronic notebooks	

How to better insulate us from frauds

○Original data depository
  ○ Example from Stowers.org
○Publishing raw data for journals
  ○ Nature
  ○ Repercussion for fraud and retraction
○Better mentoring/supervision
  ○More emphasis on raw data, not title, abstract-only
  ○Make sure the repeats, and validations are done
○!!! Orthologous methods!!!
○Image processing
  ○ Cropping lanes, REQUIRED to add a visible landmark
  ○ Exposure between experimental vs. ctrl groups

synopsis

- Why we need to have the integrity discussion?
  - Public trust in science is eroding
  - The reproducibility crisis of Science
    - Intrinsic nature of science and who does it (human)
    - Blurry line between sloppy science vs. fraud
      - Examples of both
- Reasons not to cheat
  - Impossible to cover up
  - Super spotters, whistle-blower
- Better day-to-day practices to avoid the darkside