Office of Research Integrity

- **Research Data Management and Integrity Practices**

William C Trenkle, Ph.D.

Scientist-Investigator, Division of Investigative Oversight

Office of Research Integrity, HHS

Presented at the National Data Integrity Conference, 2015

Topics

- Office of Research Integrity (ORI) mission and process.
- Examples of Research Data and Forensic Analysis.
  - The slippery slope from beautification to Research Misconduct
- Discussion of developing Challenges facing the Research Community.

ORI's Mission

Mission: To promote the integrity of Public Health Service supported extramural and intramural research programs

- Respond effectively to allegations of research misconduct
- Promote research integrity
- Deter research misconduct through public disclosure of findings and penalties.

Definition of Research Misconduct

Title 42: Public Health

PART 93—PUBLIC HEALTH SERVICE POLICIES ON RESEARCH MISCONDUCT

Subpart A—General

§ 93.103 Research misconduct.

Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.

(a) *Fabrication* is making up data or results and recording or reporting them.

(b) *Falsification* is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

(c) *Plagiarism* is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.

(d) *Research misconduct* does not include honest error or differences of opinion.

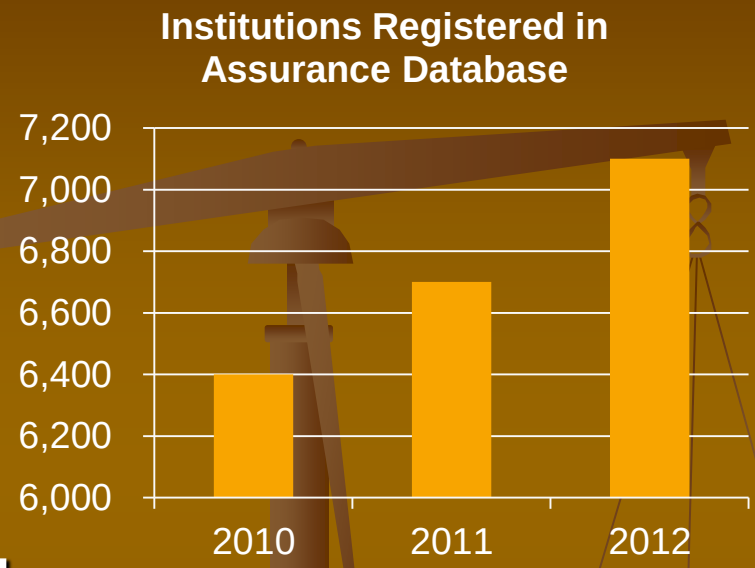
Proof of Research Misconduct

A finding of misconduct requires -

- That there be a significant departure from accepted practices of the relevant research community, and
- The misconduct be committed intentionally, knowingly, or recklessly; and
- The allegation be proven by a preponderance of the evidence, (42 CFR Part 93.104)

Additional ORI Activities

- Administer the Assurance program, a database of all institutions eligible to receive PHS funds
- Correct or retract research publications to protect the integrity of the scientific literature
- Protect the confidentiality of respondents, complainants, and witnesses
- Protect witnesses from retaliation (42 CFR 93.300 (d))



ORI Activities (cont)

- Provide education in RCR
- Collaborate with the research community to improve biomedical research
- Exclude dishonest investigators from PHS and Federal agency funded research
- Make public findings of misconduct so that institutions and individuals will be aware of wrongdoing

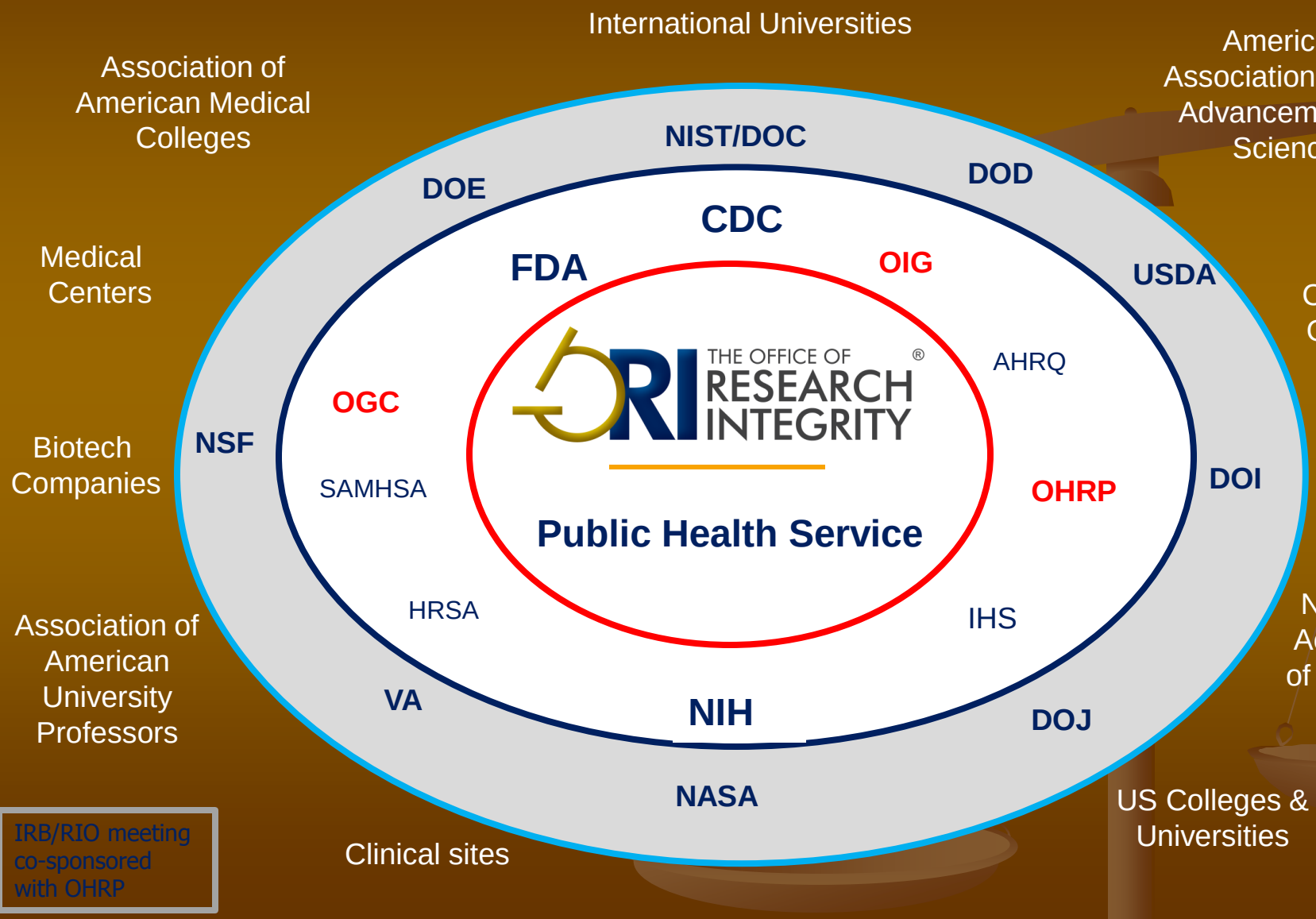
ORI lacks jurisdictions for many types of inappropriate behavior: some are referred to other agencies

- Misuse of human or animal subjects
- Misconduct and other complaints involving FDA-regulated research
- Financial mismanagement
- Radiation or biosafety hazards
- Conflict of interest

Other issues not within ORI's jurisdiction:

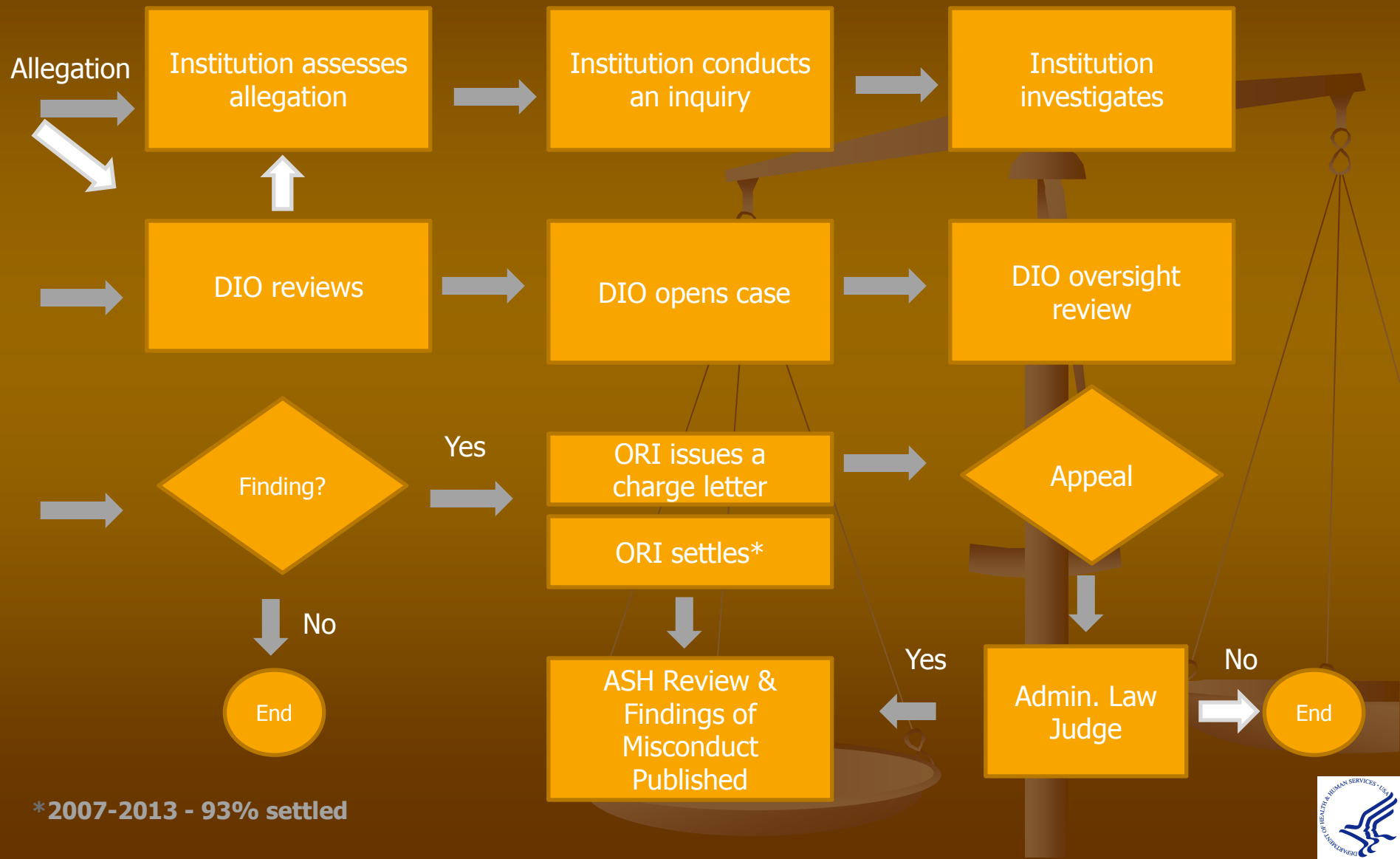
- Honest error or honest differences in interpretations or judgments of data
- Authorship or credit disputes
- Duplicate publication
- Collaboration agreements or research-related disputes among collaborators
- Intellectual property

Supporting Coordination & Collaboration



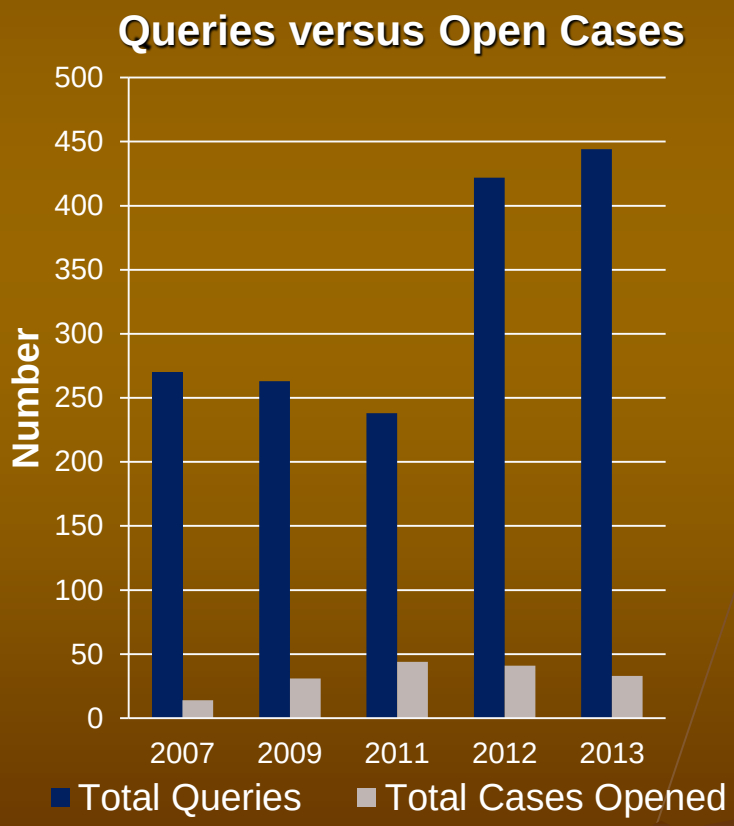
IRB/RIO meeting co-sponsored with OHRP

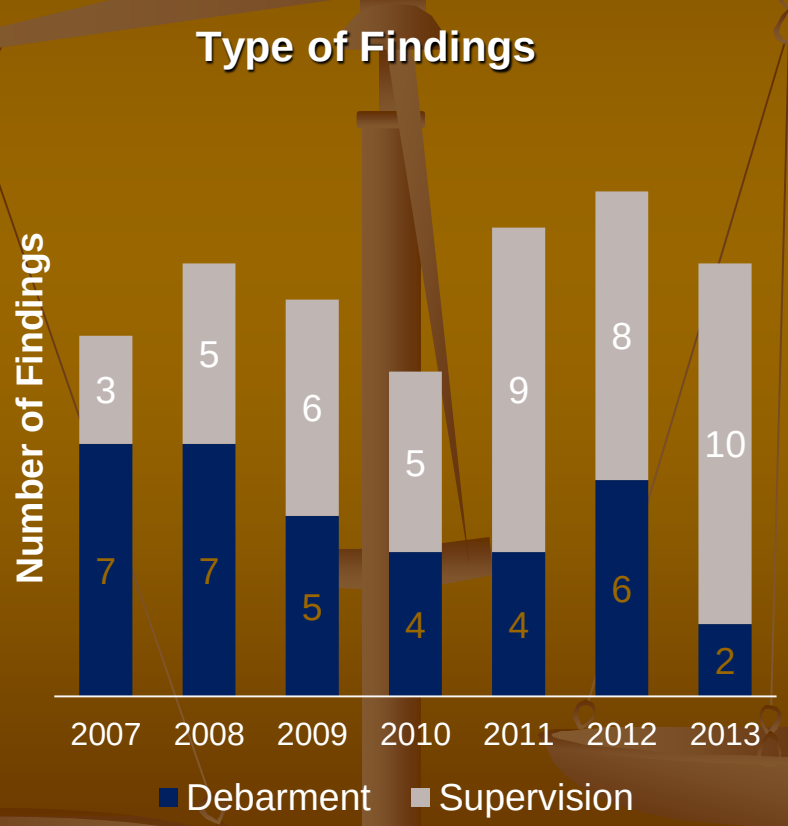
Handling Cases of Research Misconduct



*2007-2013 - 93% settled

Key Metrics – Division of Investigative Oversight





A few key issues that ORI has found contribute most significantly to allowing misconduct

1. Inadequate record keeping and lack of guidance from mentors and Institutions on how to record and retain research data;
2. Failure of mentors to regularly review raw data; overreliance on derivative data (PowerPoint presentations) at lab meetings
3. Unquestioning acceptance of data that others consider "too good to be true"

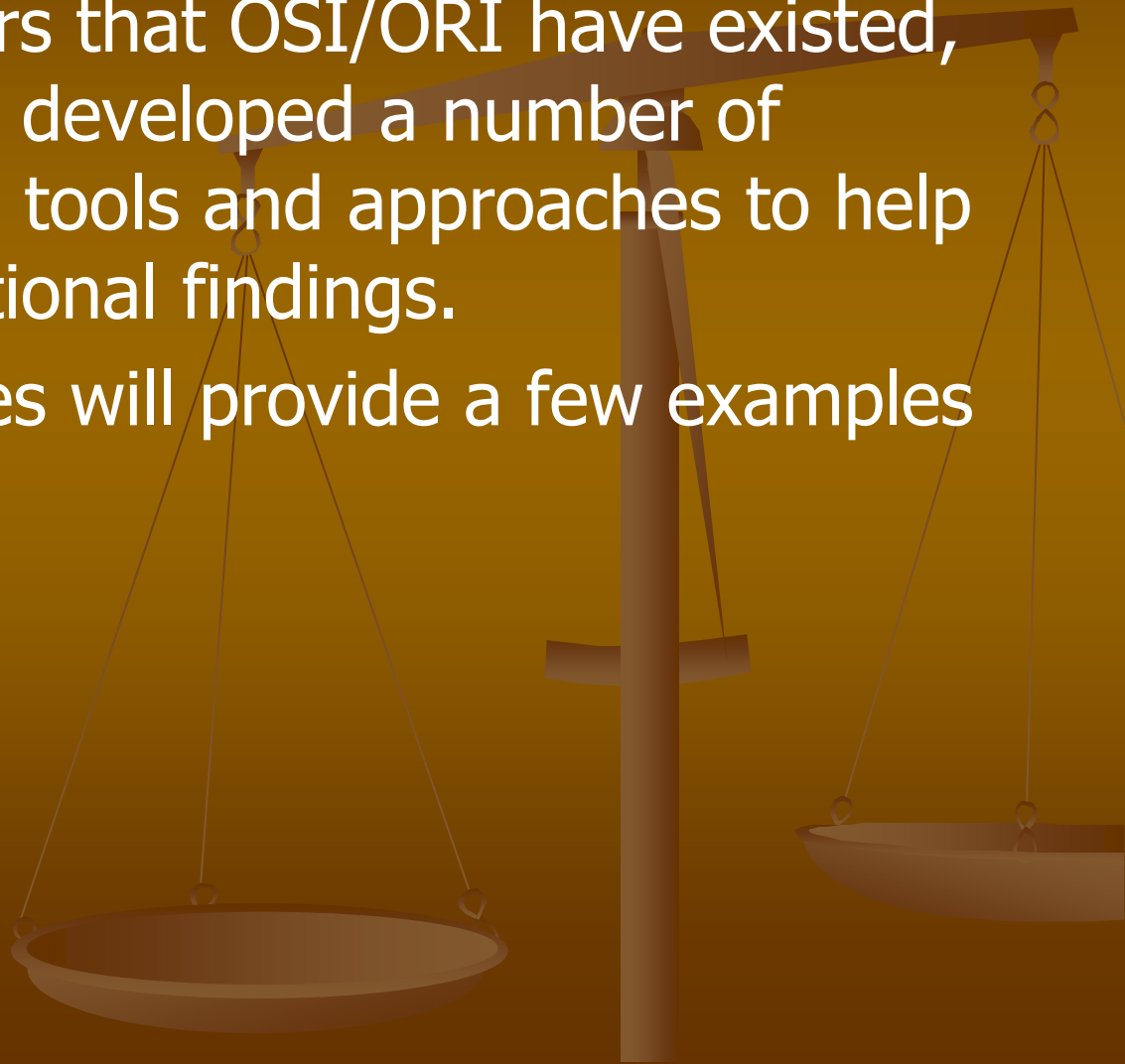
More issues that facilitate misconduct

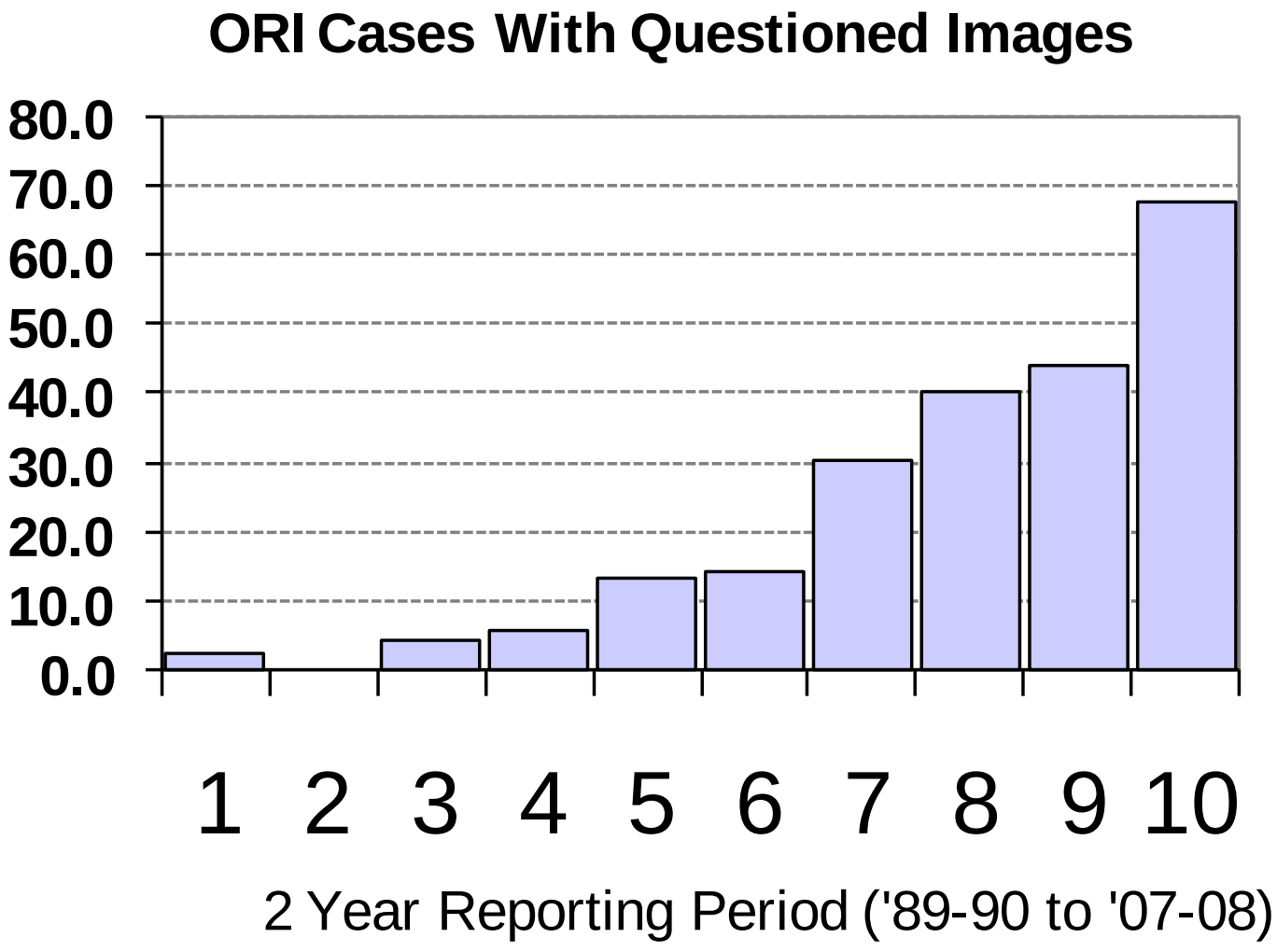
- 4. Lack of transparency within the laboratory and among the staff
- 5. Labs so large that authority becomes diffuse
- 6. P.I.s are spread too thin, and do not provide adequate training and guidance to students

The bottom line – good mentorship and the consistent review of raw data can profoundly reduce the likelihood of research misconduct.

DIO Oversight: Forensics

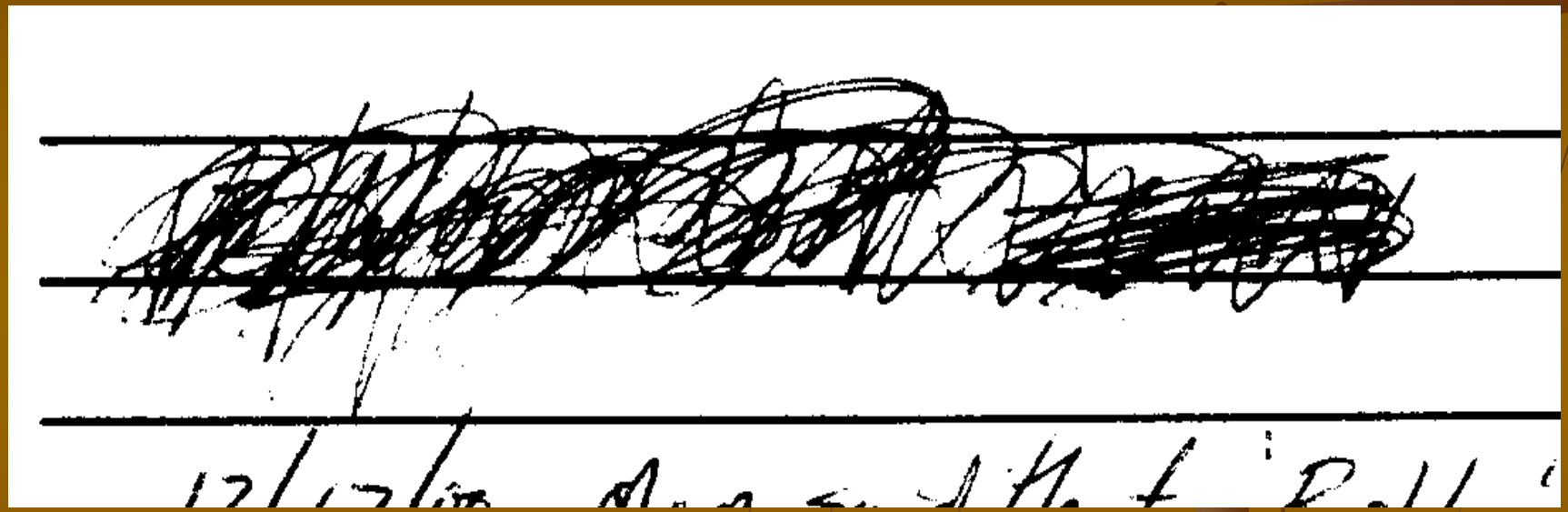
- During the 20 years that OSI/ORI have existed, investigators have developed a number of computer-assisted tools and approaches to help strengthen institutional findings.
- The following slides will provide a few examples of this.



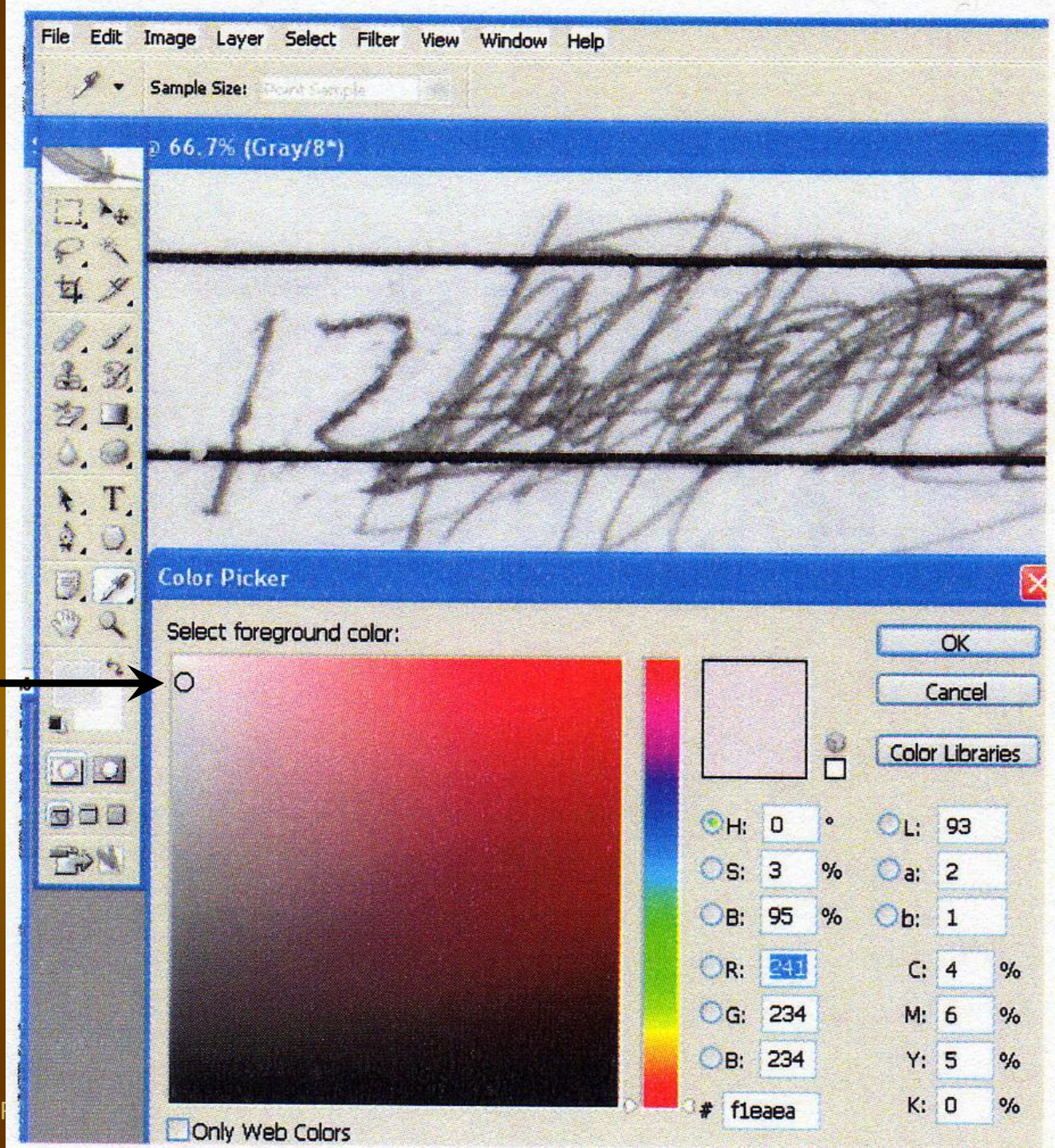

ORI Cases With Questioned Images
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10
2 Year Reporting Period ('89-90 to '07-08)

Examples of analyzing images

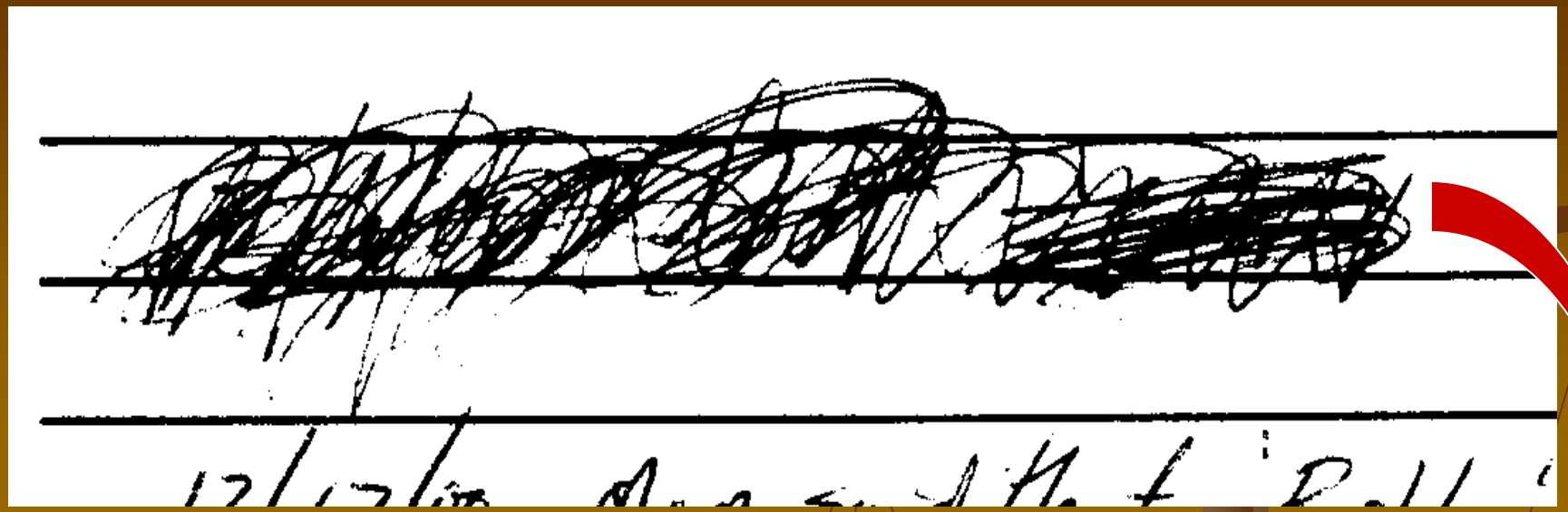
- Several examples follow which illustrate how ORI can examine images provided by institutions during their investigation.
- Many of ORI's cases involve images that are duplicated from paper to paper or paper to grant application. This may be duplicate publication, but when such images are said to be the result of different experiments, one of the images, at minimum, has been potentially falsified.
- The first example, however, is a little different.



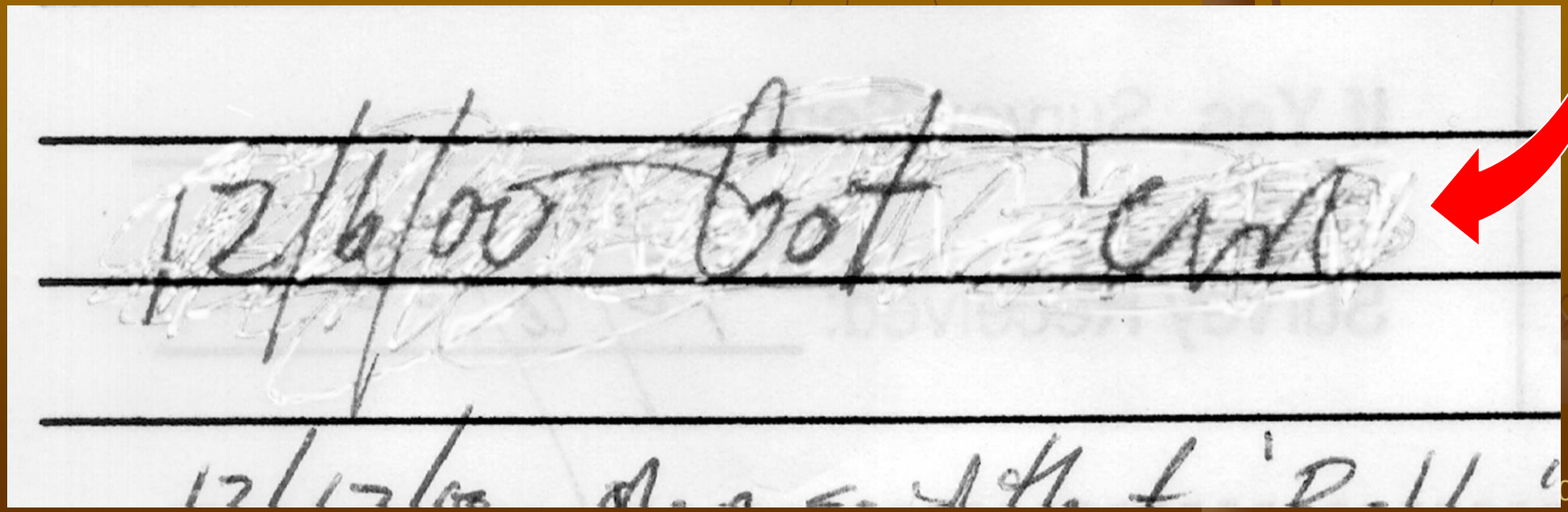
Screen shot from Photoshop showing analysis under way – the small circle in the Color Picker is the brush size moved to a color approximately matching the image's background.



An obscured data entry; two slightly different colored inks.



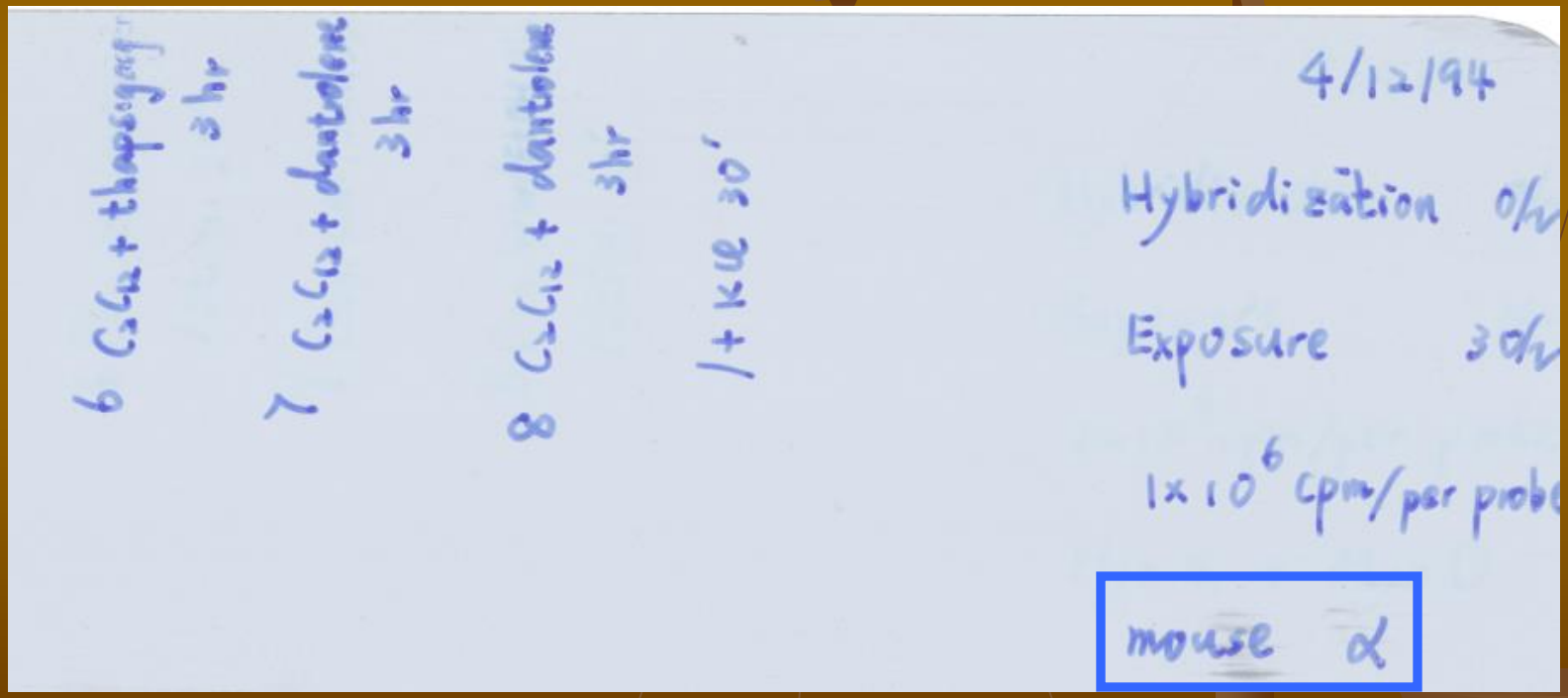
The result of "removing" most of the "scribbling."



This is why the RIO sent the previous sample to ORI for Review

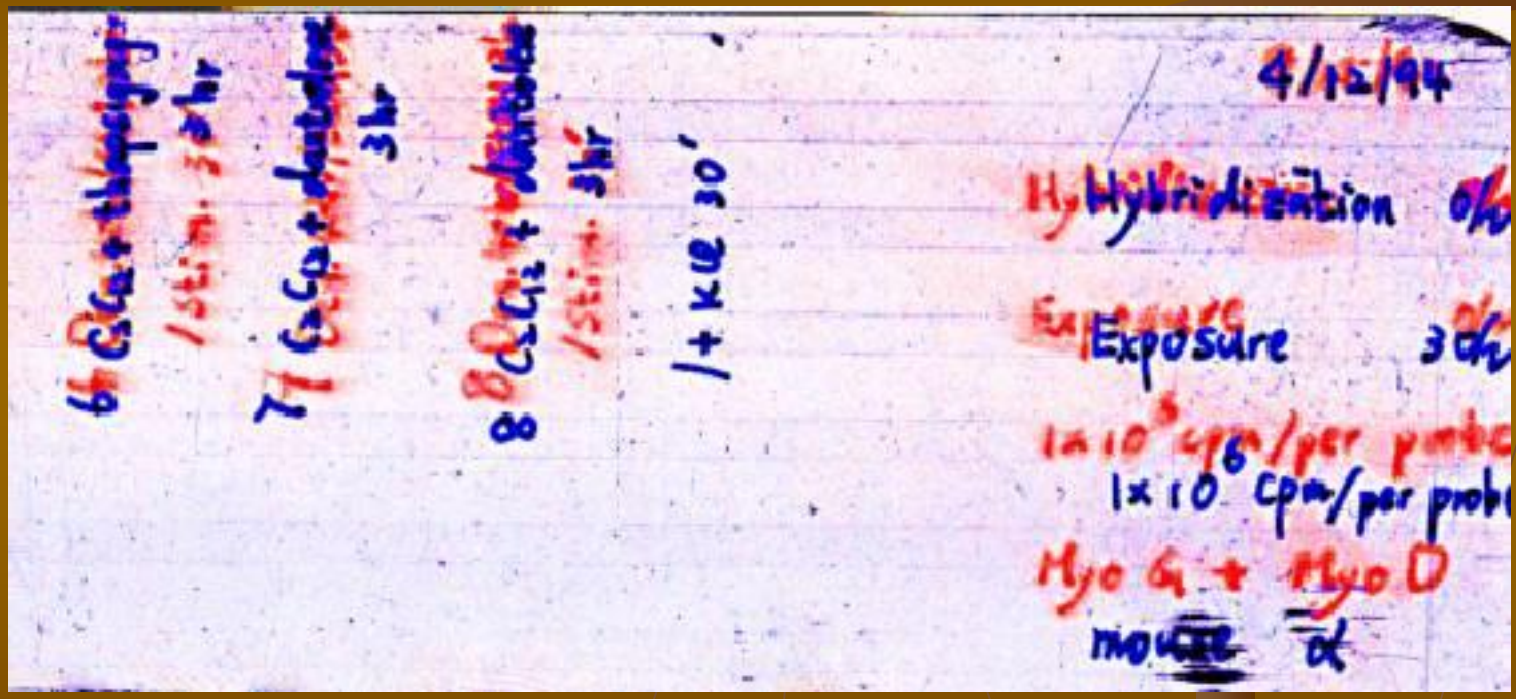
In this case, 1 film was used to represent 2 experiments

Corner of Film



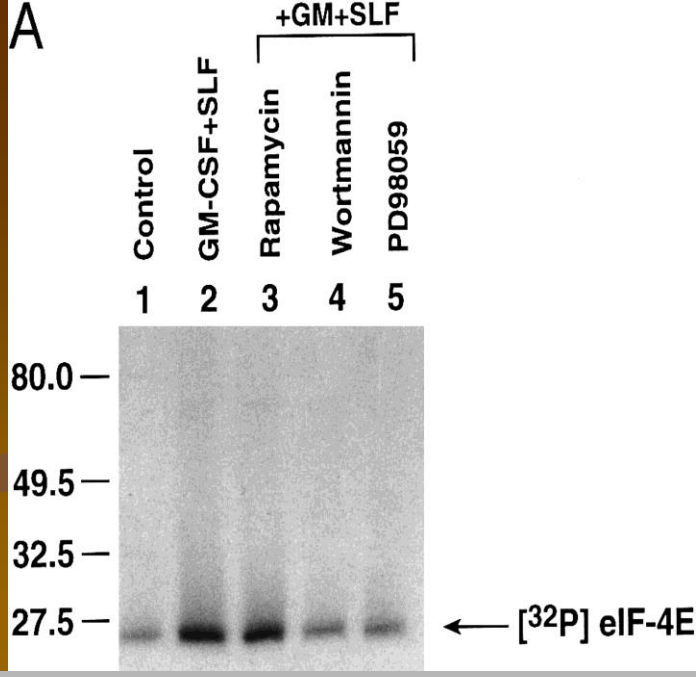
the same film used for mouse a and mouse Myo D and Myo G

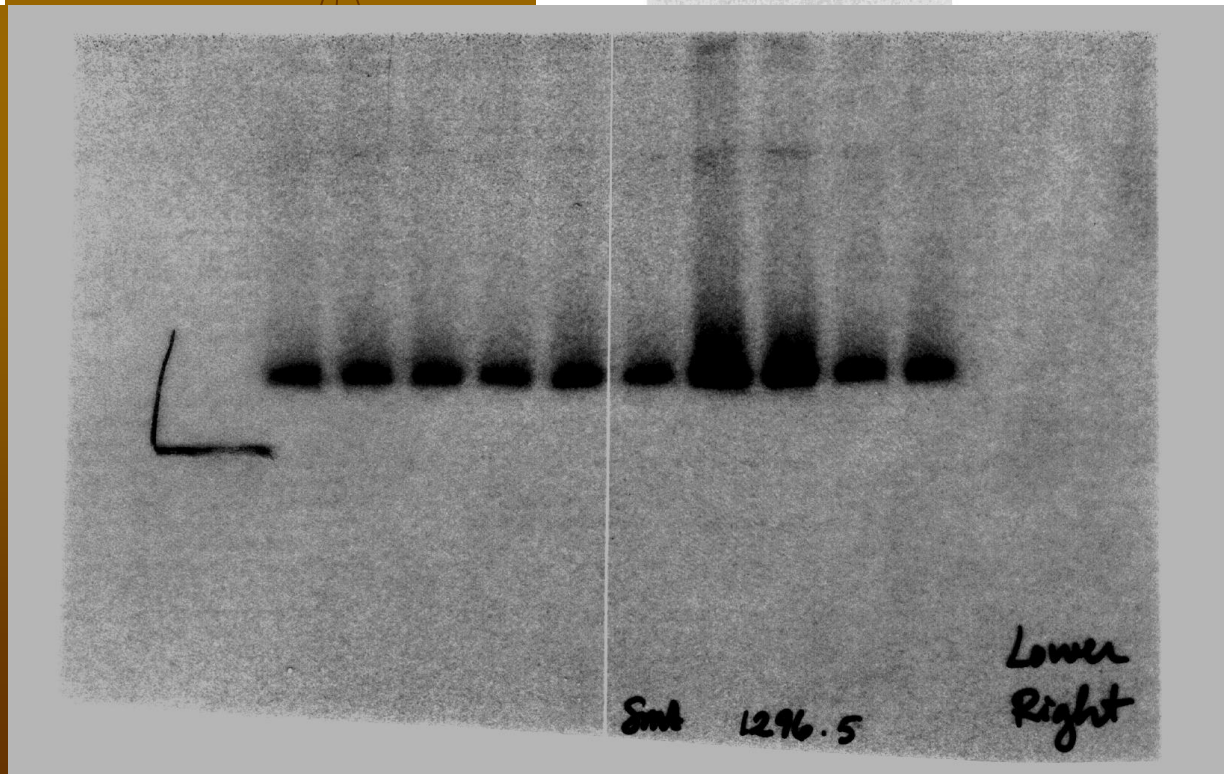
Scanned film separated by hue



Result: writing in red erased from film then re-labeled

In this example, the respondent published a figure (shown to the right) and claimed that the blot had been stripped and re-probed to provide a loading control (not shown). ORI's review of the notebook showed that she had cut a film into two fragments and claimed that one-half was the loading control. However, forensic examination clearly established that the two films were cut from a single exposure of a blot.





When the films were aligned and scanned in reflection mode:

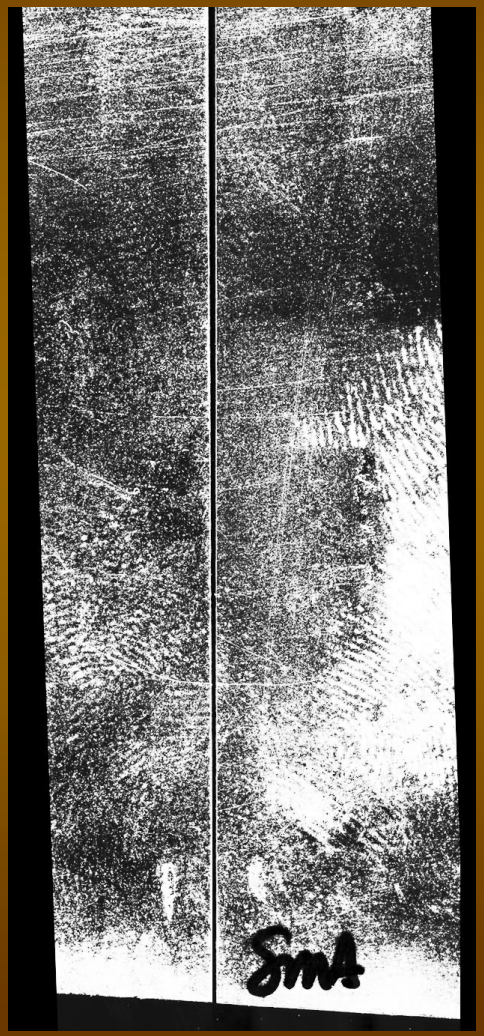


Image processing tools bring out hidden features:

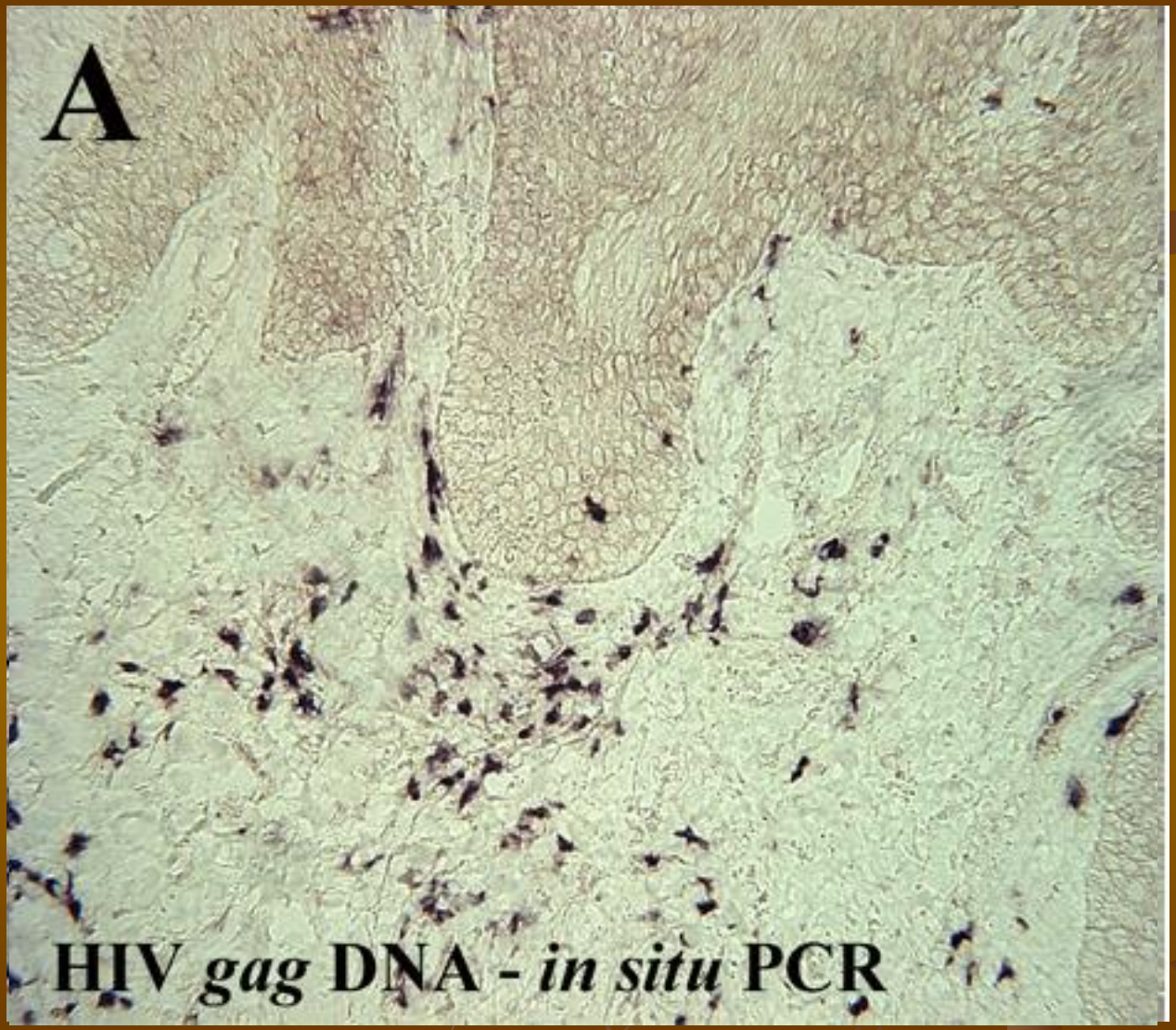
- common edge
- scratches and prints crossing edge

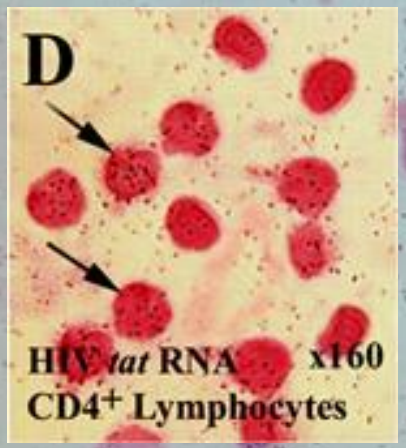
Moving to the digital age

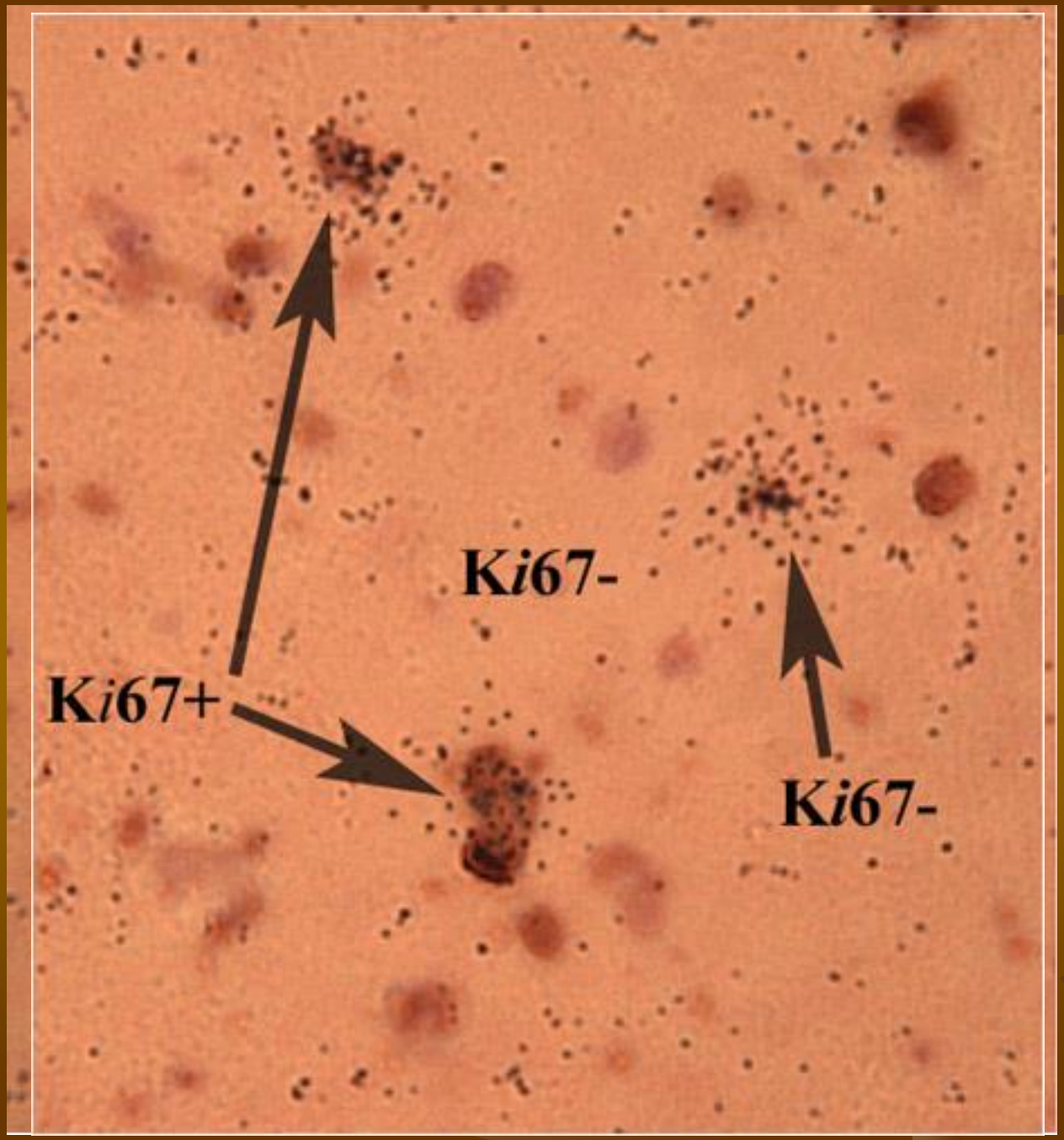
- The next examples illustrate the importance of the eye to detect evidence of inappropriate image manipulation.
- In addition, many ORI cases rely on prompt sequestration of evidence, including hard drives and portable storage media, to ensure that manipulated images can be shown to have originated with a particular individual.
- Time-date stamps are often probative with respect to how and when the manipulations occurred.

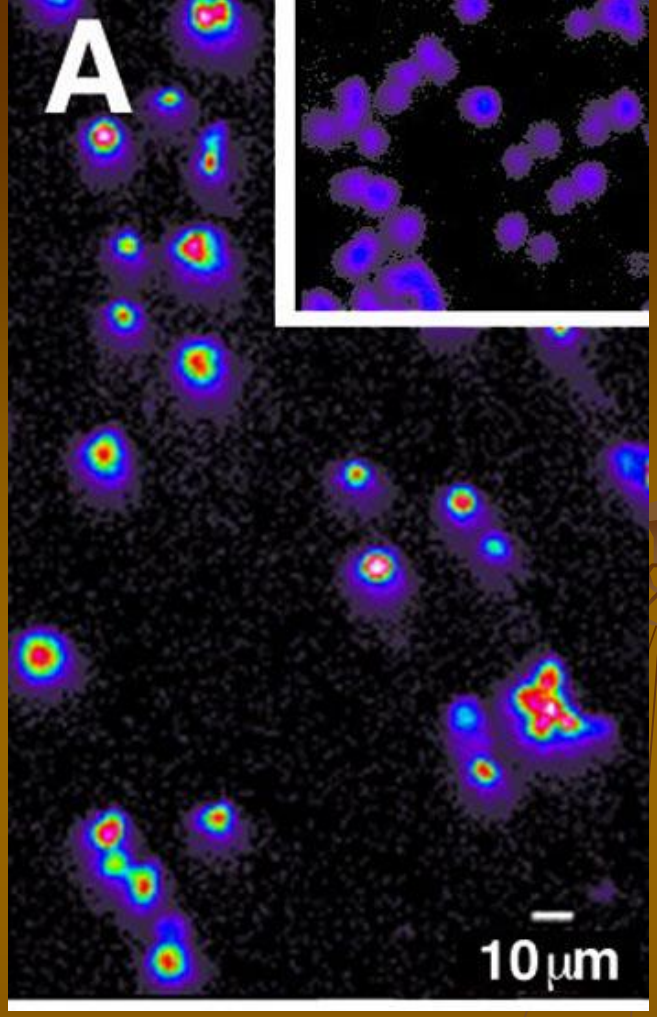
A Slippery Slope

- The next few slides show how difficult it can be to determine if a manipulation is appropriate, possibly inappropriate, or obviously fraudulent.
- Generally, ORI is reluctant to make findings of misconduct when an image has been "beautified" by altering background, or by reuse of loading controls, when the actual data verifies the factual findings claimed in the grant or paper.
- However, adding or removing important elements of a figure can often be considered evidence for intentional falsification.

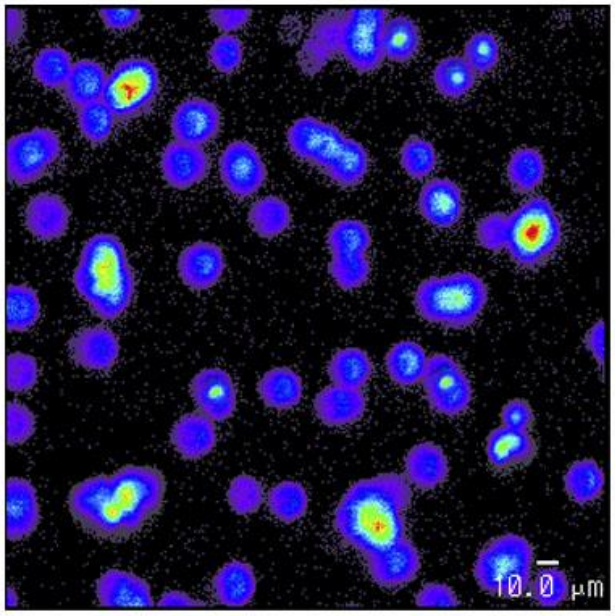

A
HIV gag DNA - in situ PCR


D
HIV *tat* RNA x160
CD4+ Lymphocytes

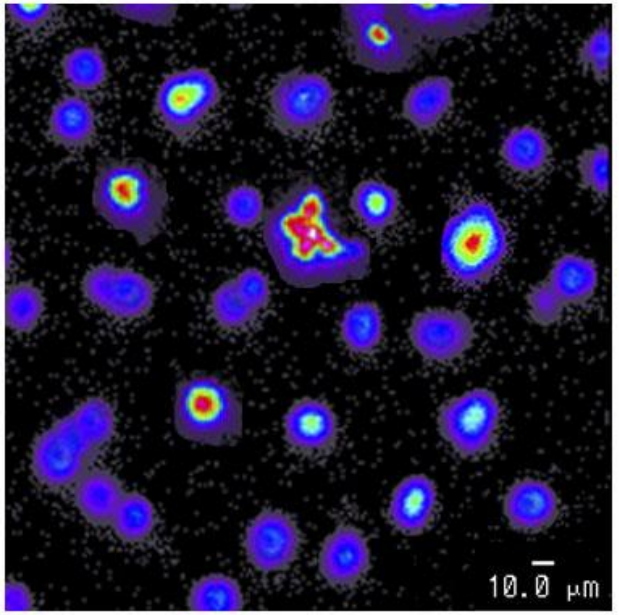

Ki67-
Ki67+
Ki67-



Cells expressing HIV RNA



Cells harboring HIV DNA



An example of unique images that alerted the institution to apparent falsification. The original images were of all positive or all negative cells positive for a gene different from HIV DNA or RNA.

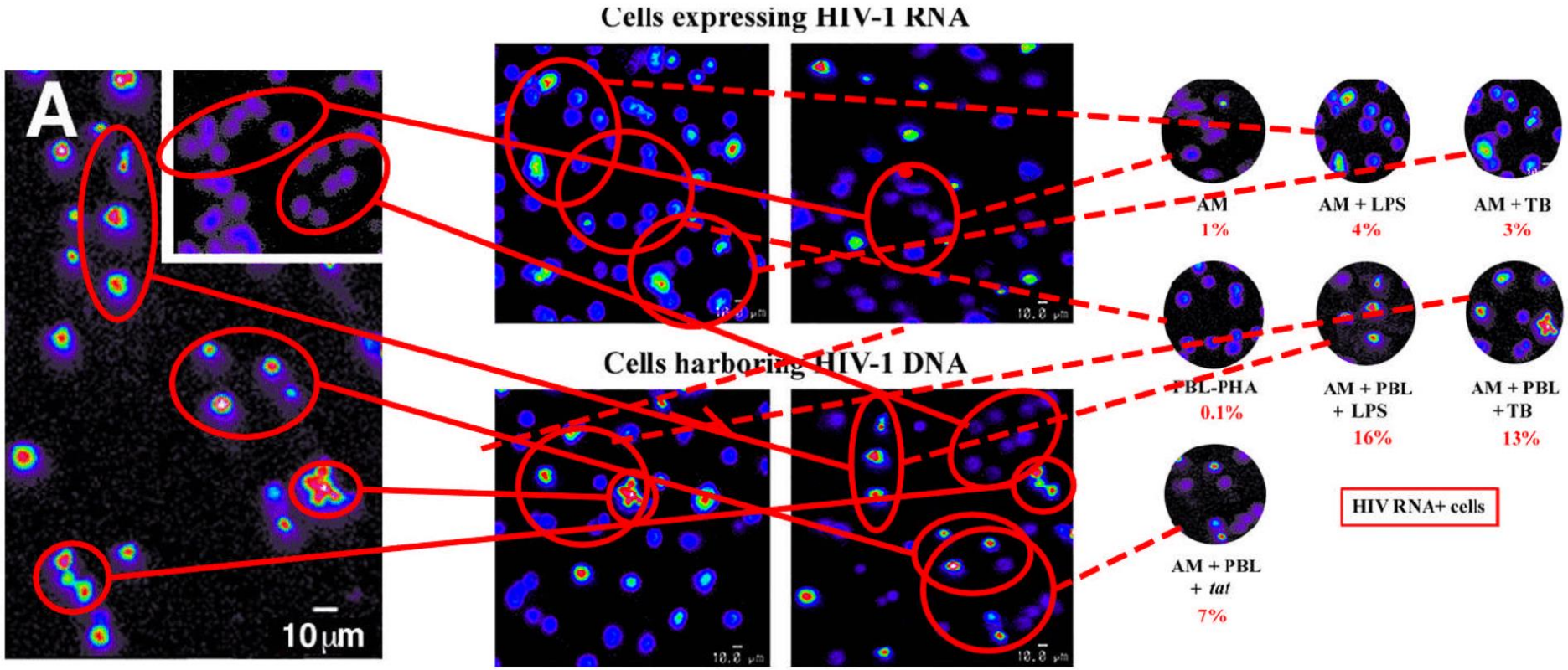
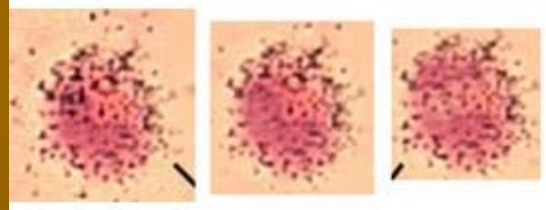
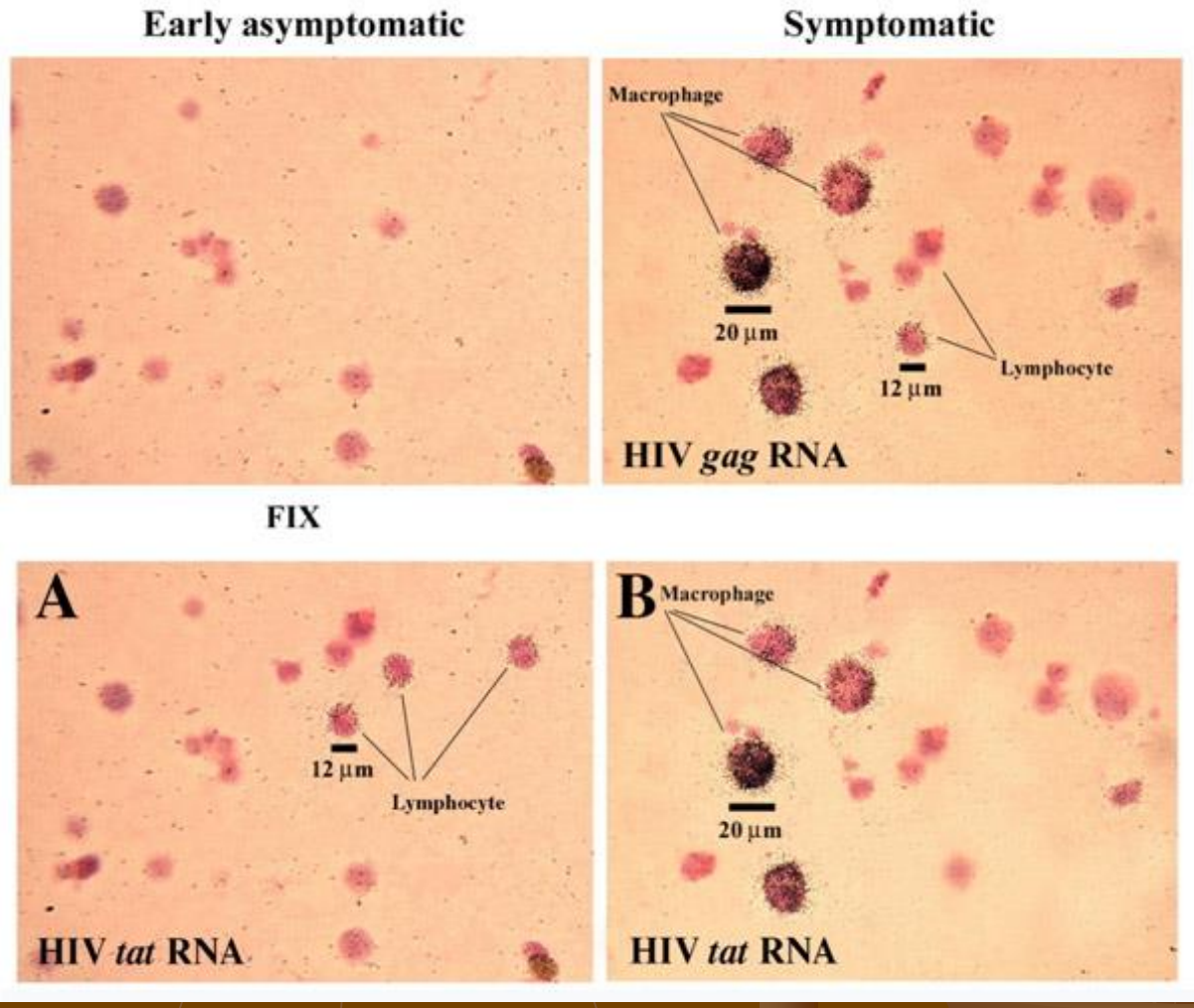


Figure 5a, *Journal of Clinical Investigation*, 105:1407-17, 2000

Fig. 2D5, *Jour. Exp. Med.* manuscript; Fig. 3.5, Mullins grant; Fig. 11, Greenberg grant

Figure C.2.3, Mittler grant



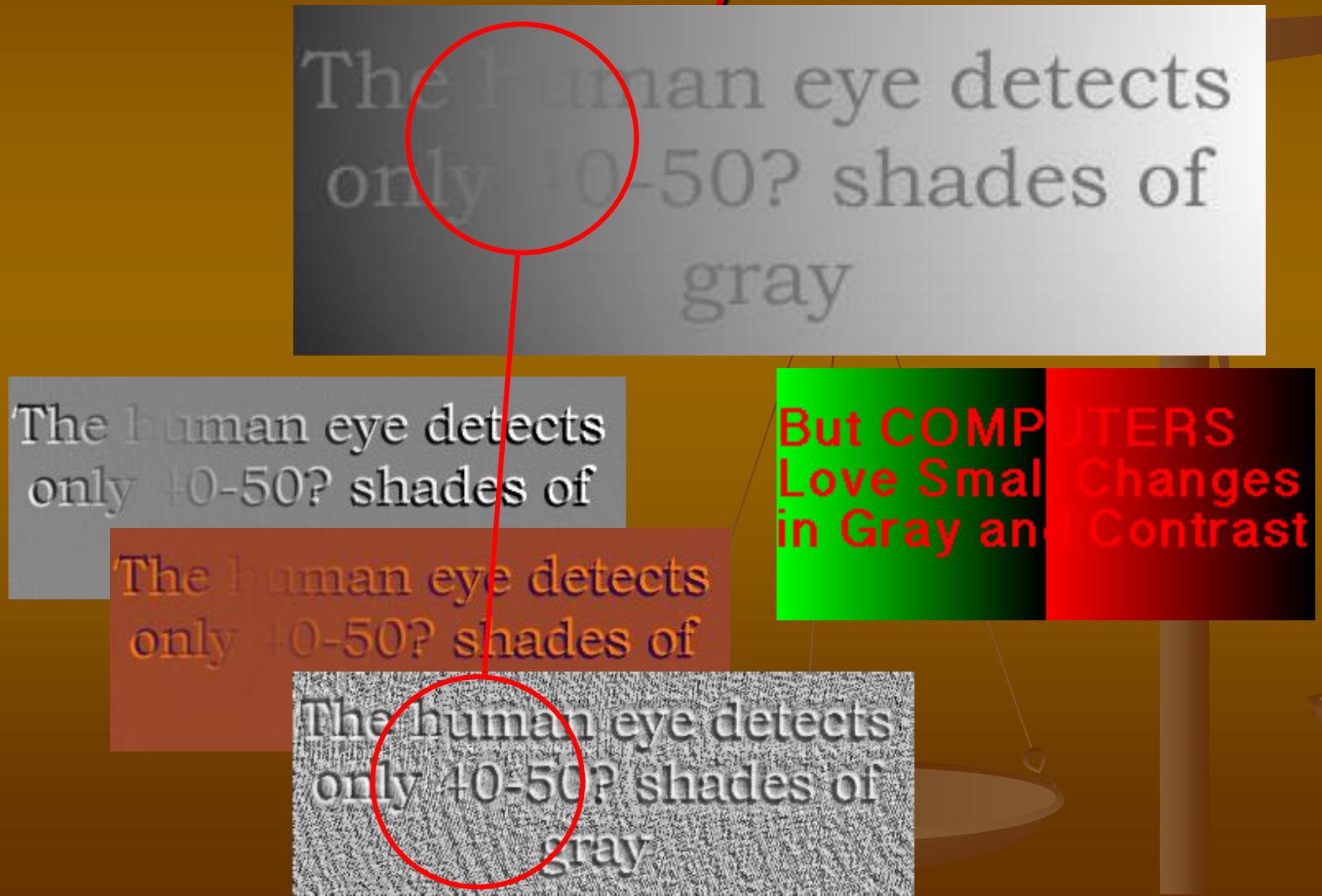

Early asymptomatic
Symptomatic
Macrophage
20 µm
12 µm
Lymphocyte
HIV gag RNA
FIX
A
12 µm
Lymphocyte
HIV tat RNA
B
Macrophage
20 µm
HIV tat RNA

How to detect non-obvious changes

- Some of the following slides will illustrate how Photoshop can be used to help our eyes visualize alterations to images, and verify suspected duplications, through the use of specific tools such as the gradient map, contours, and various enhancements such as contrast and intensity.

DETECTION

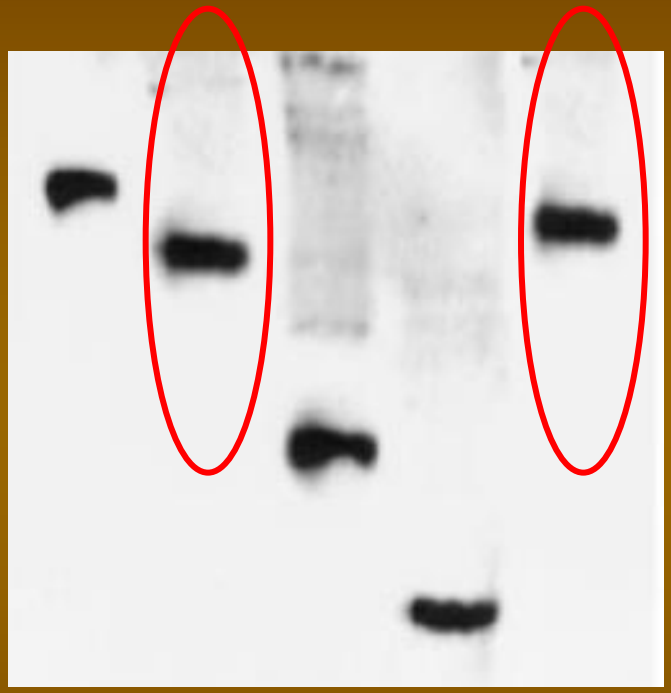
Increase Visibility of "Hidden" Details



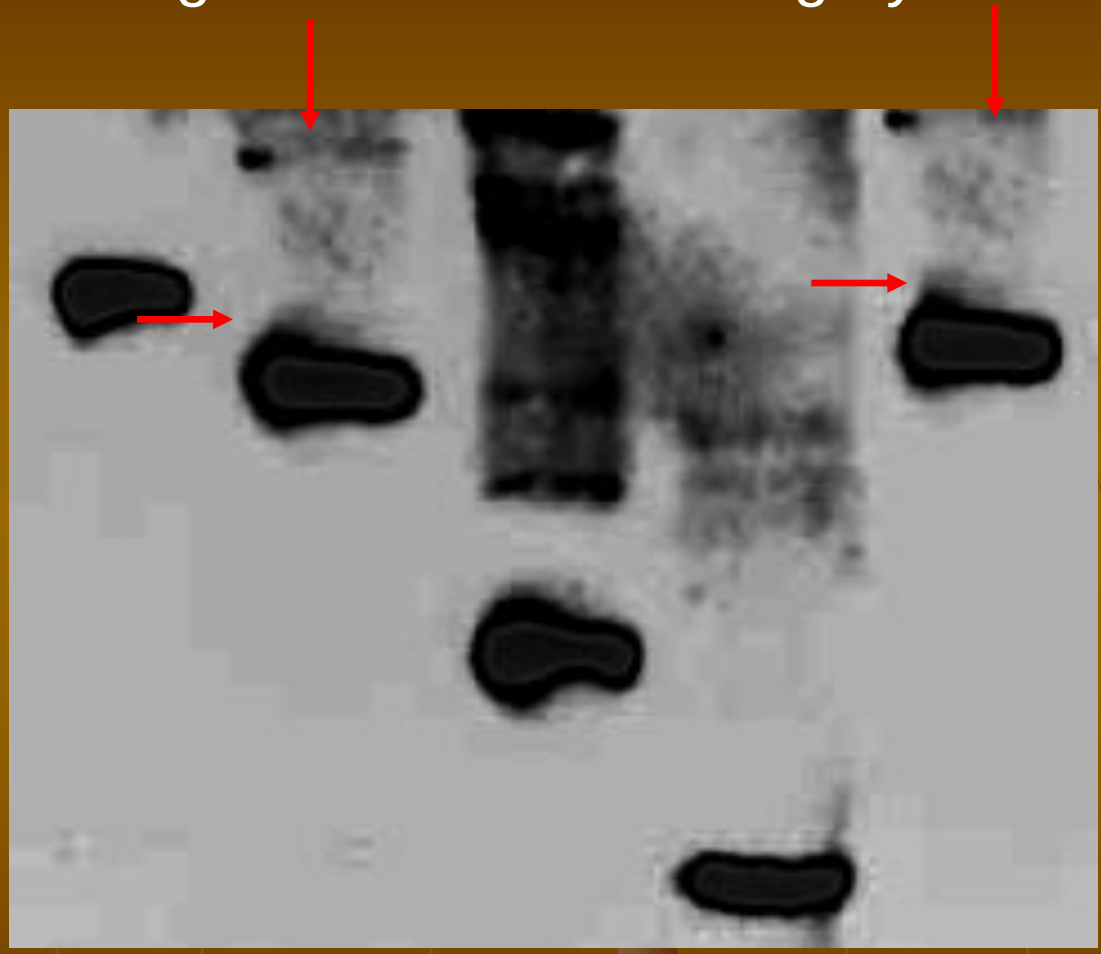
Forensic Examination of Scientific Images

1. Contrast Enhancement ("Curves") - human eye is not very good at detecting small differences in gray scale
2. Texture, Variance – examination for erasures
3. Histogram Equalization – quick look for background inconsistencies
4. Gradient Map – powerful tool to reveal many similarities in background and band morphologies
5. Embossing – shadowing makes the image slightly dimensional to reveal borders in background or edges
6. Overlay of Images – shows similarities of images

Contrast Enhancement – detecting small differences in gray scale

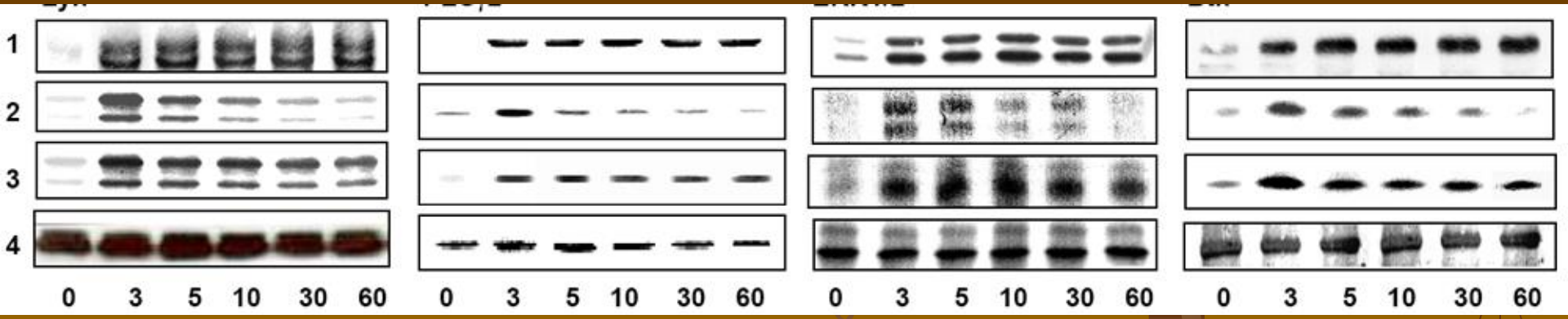


Original Data



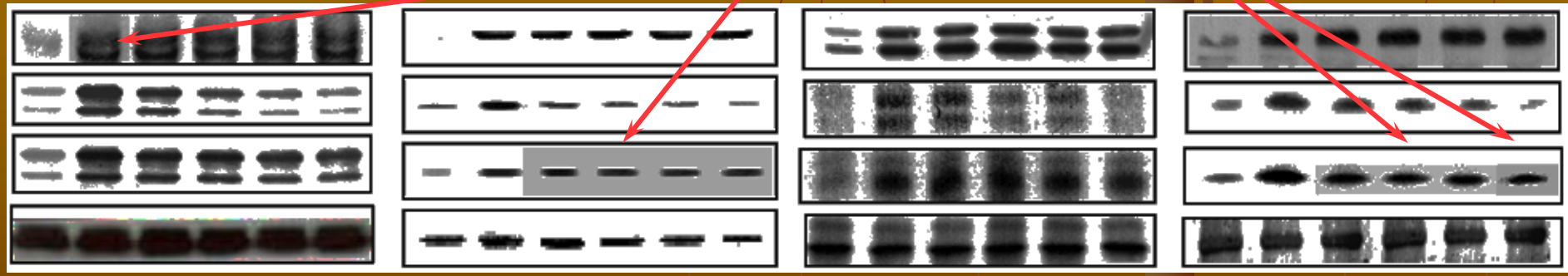
Contrast Enhanced

Effect of Histogram Equalization: background inconsistencies



Effect of Histogram Equalization:

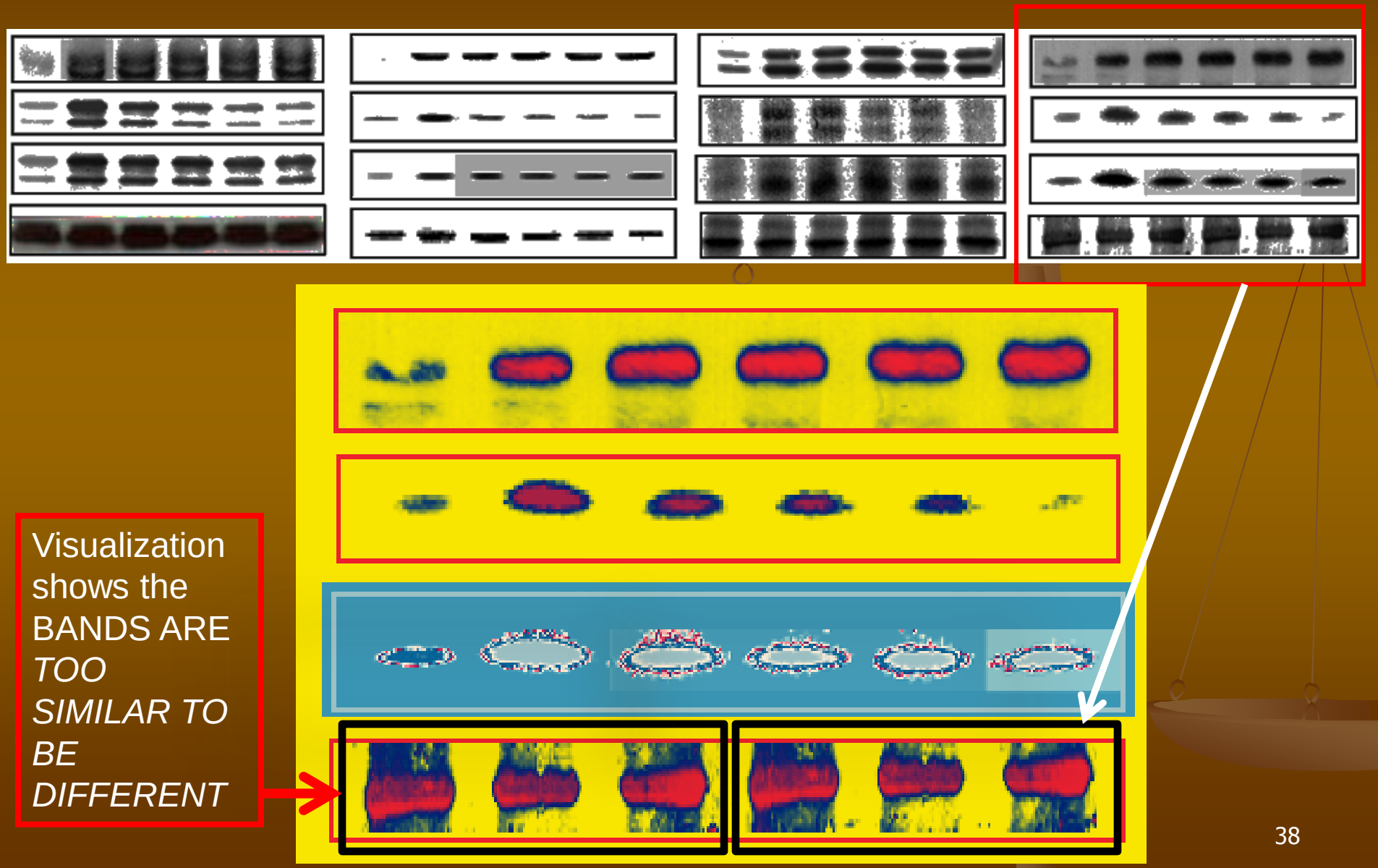
"SPLICED" DATA?



manipulation DOES NOT mean research misconduct

(shown with permission)

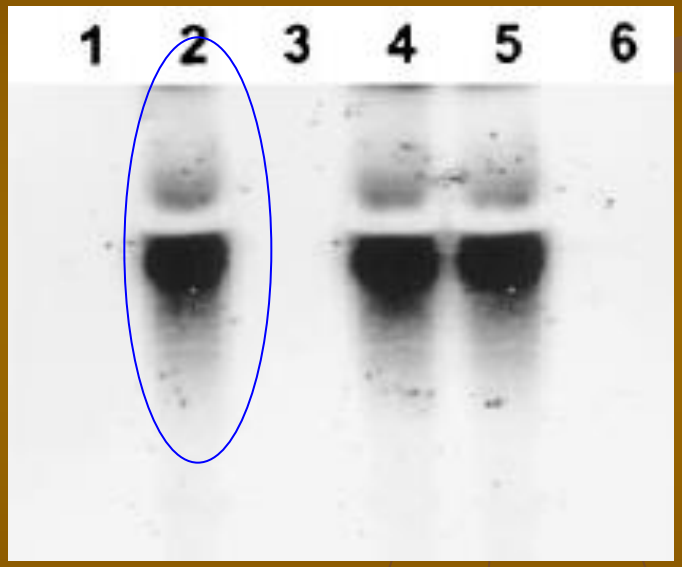
When a is Problem Found: Extend Visualization



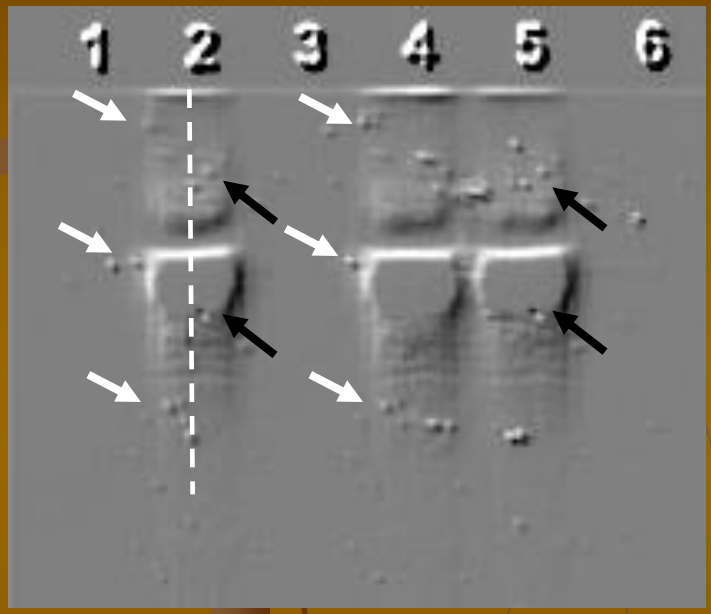
Visualization shows the BANDS ARE *TOO SIMILAR TO BE DIFFERENT*

Embossing – reveals borders in background or edges

Original Data



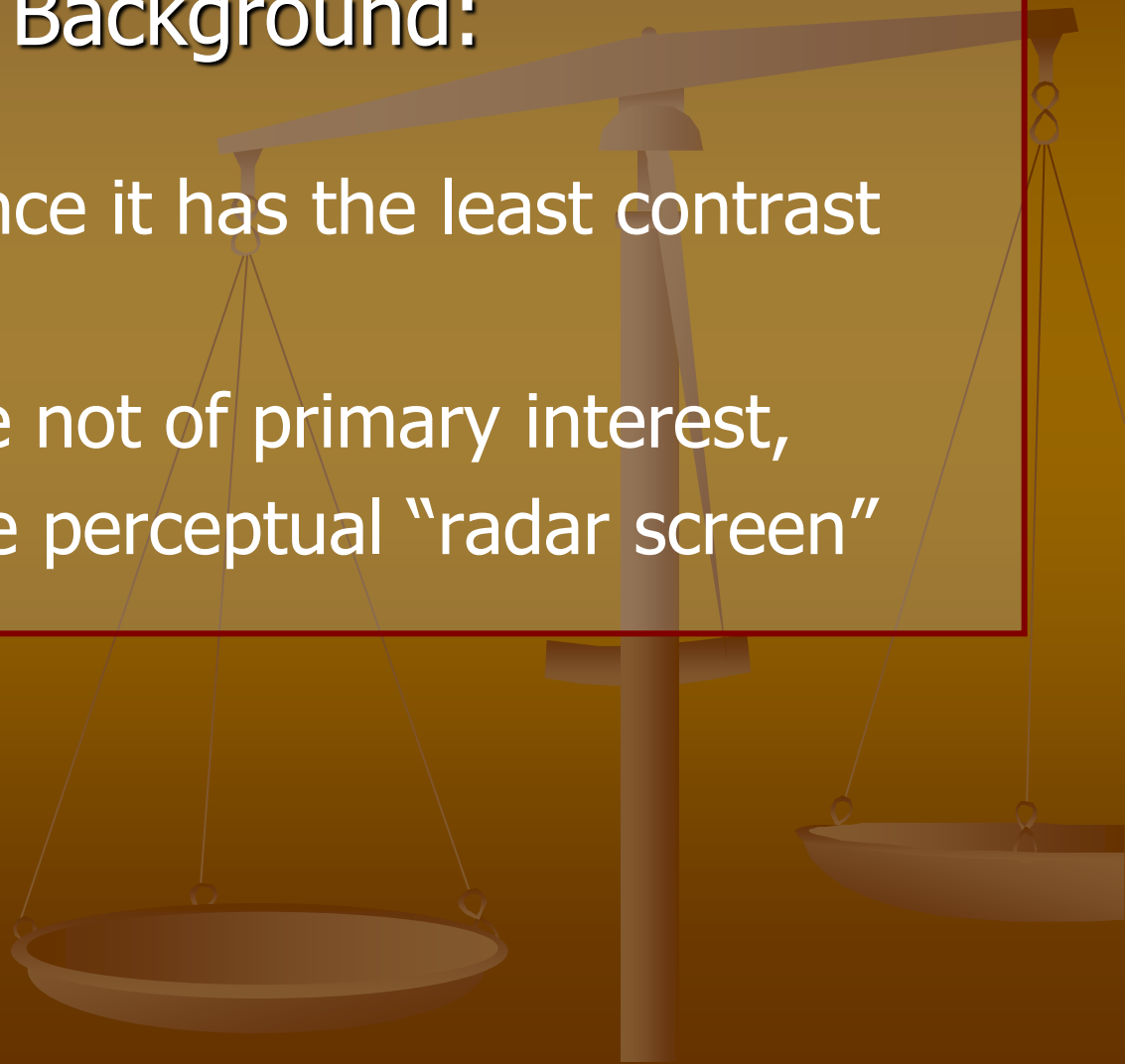
Published Data



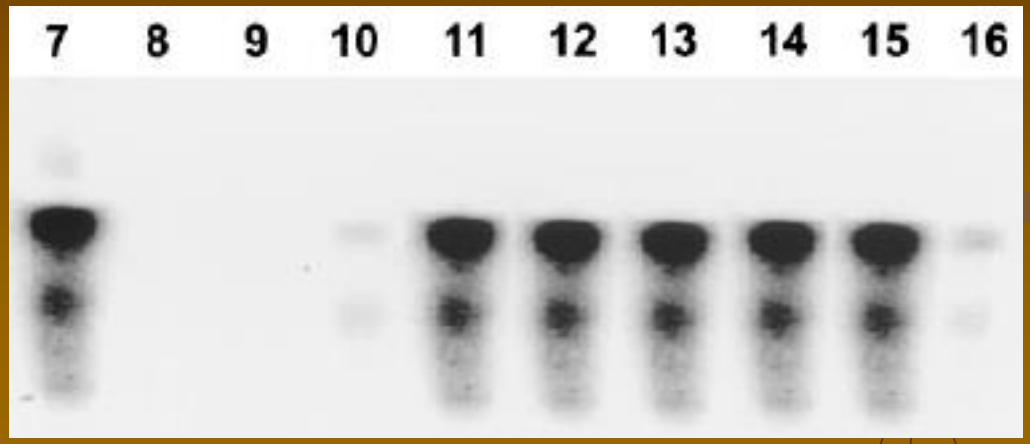
Embossed

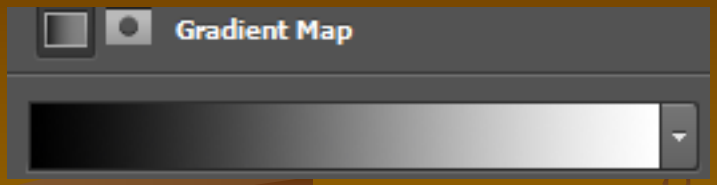
Forensic Value of Background:

- Harder to see since it has the least contrast
- Overlooked since not of primary interest,
  i.e., below the perceptual "radar screen"

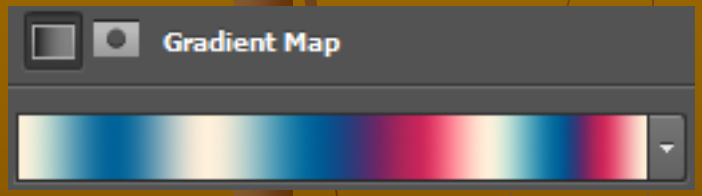


Gradient Map – reveals similarities in background and bands



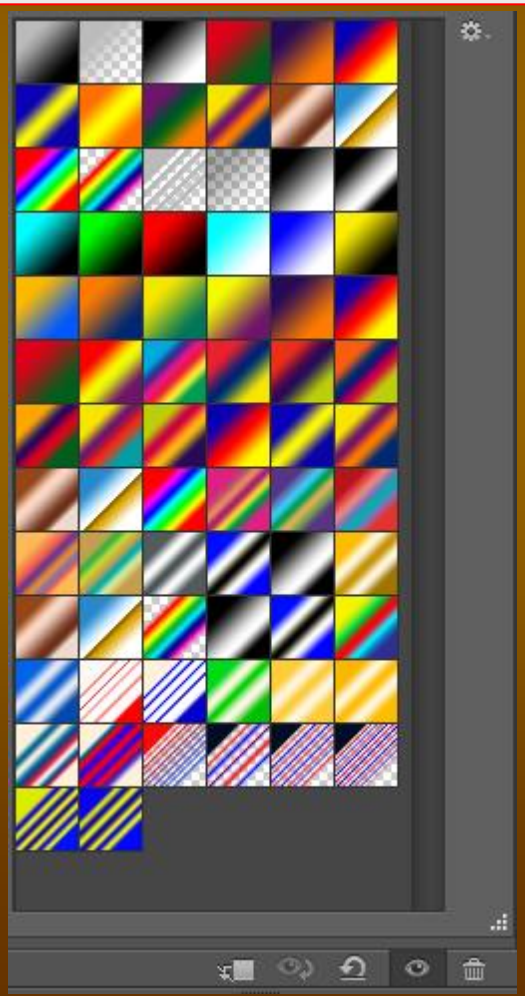


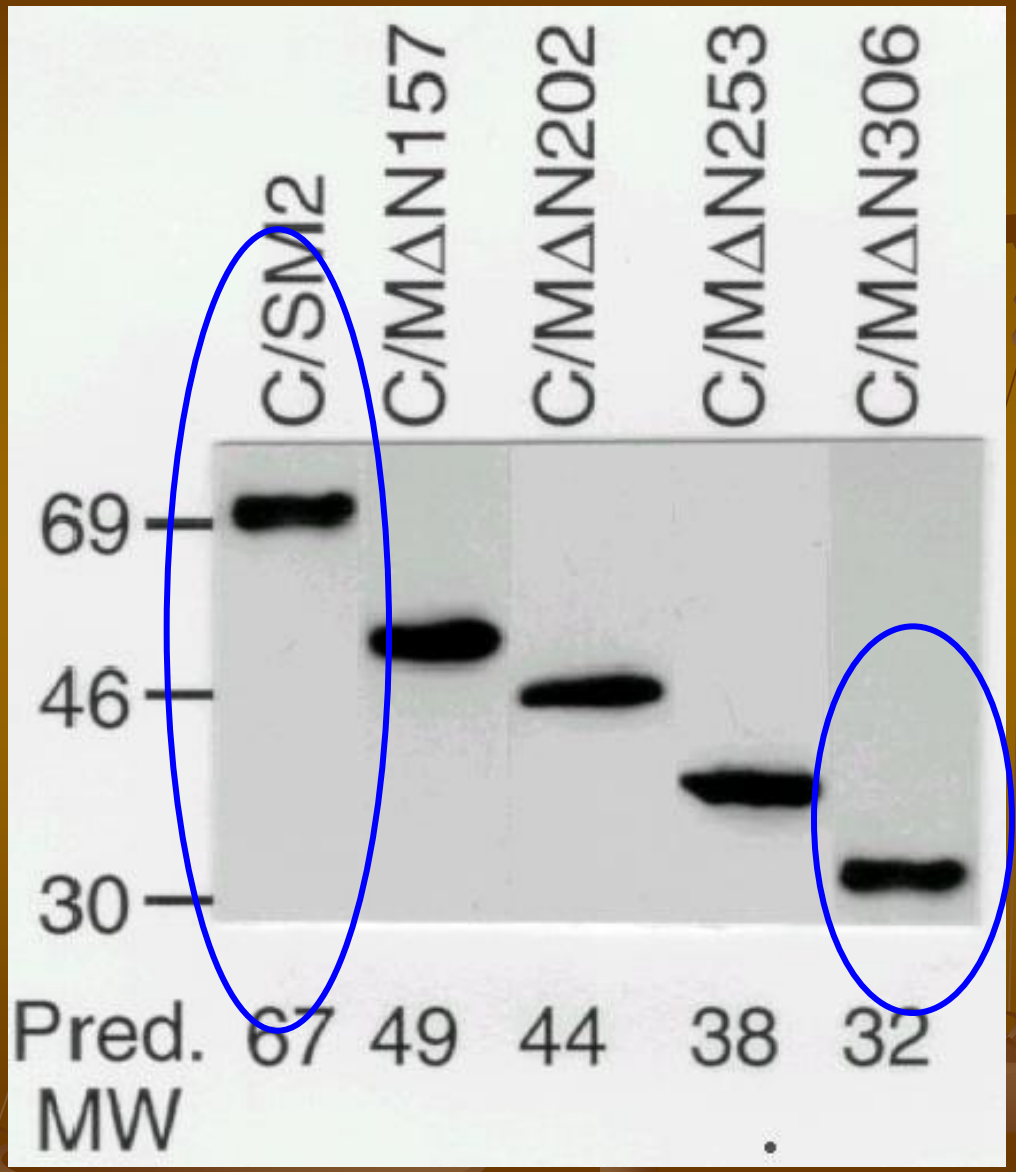
Original Data



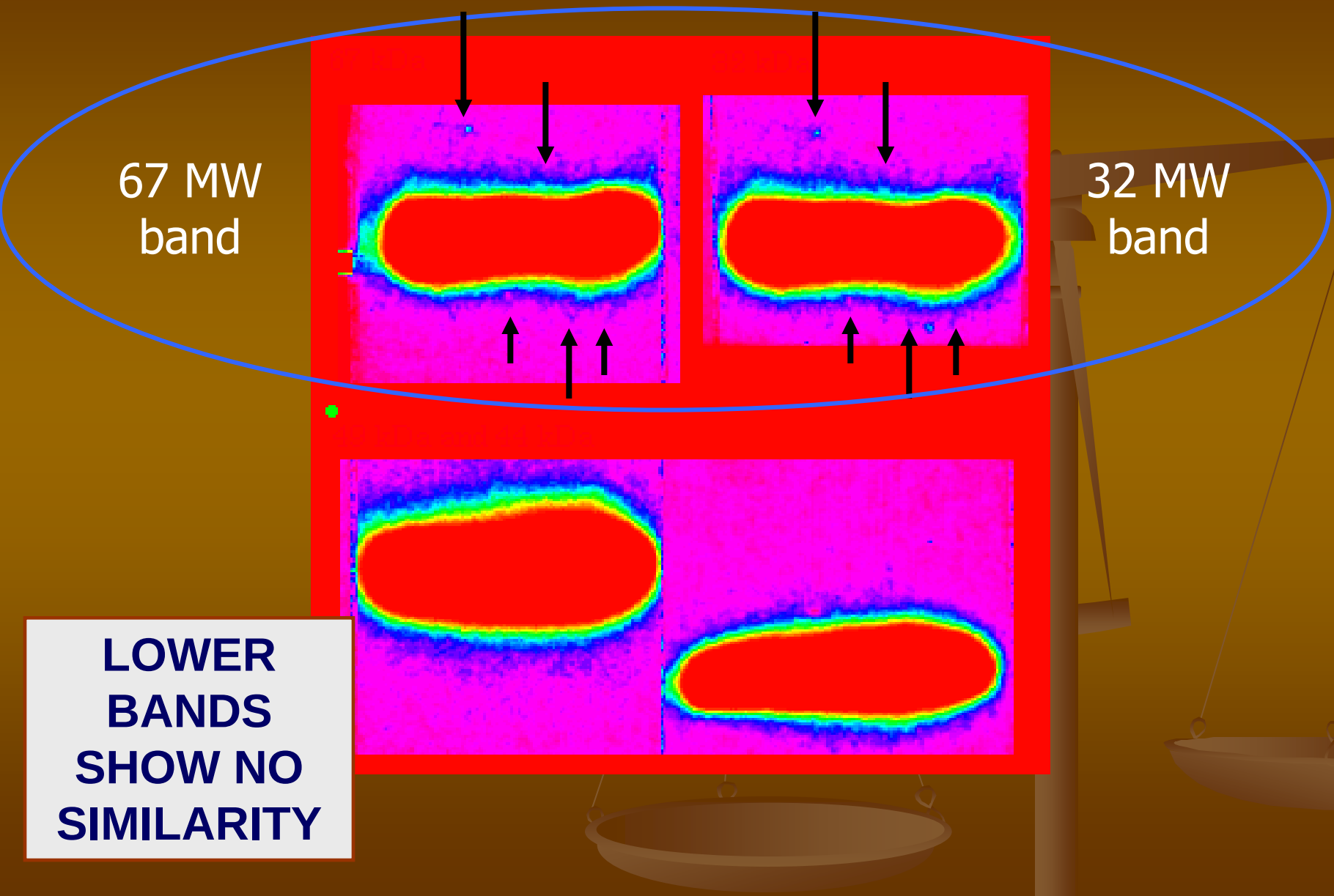
Colorized Data

MORPHOLOGICAL FEATURES OF BANDS





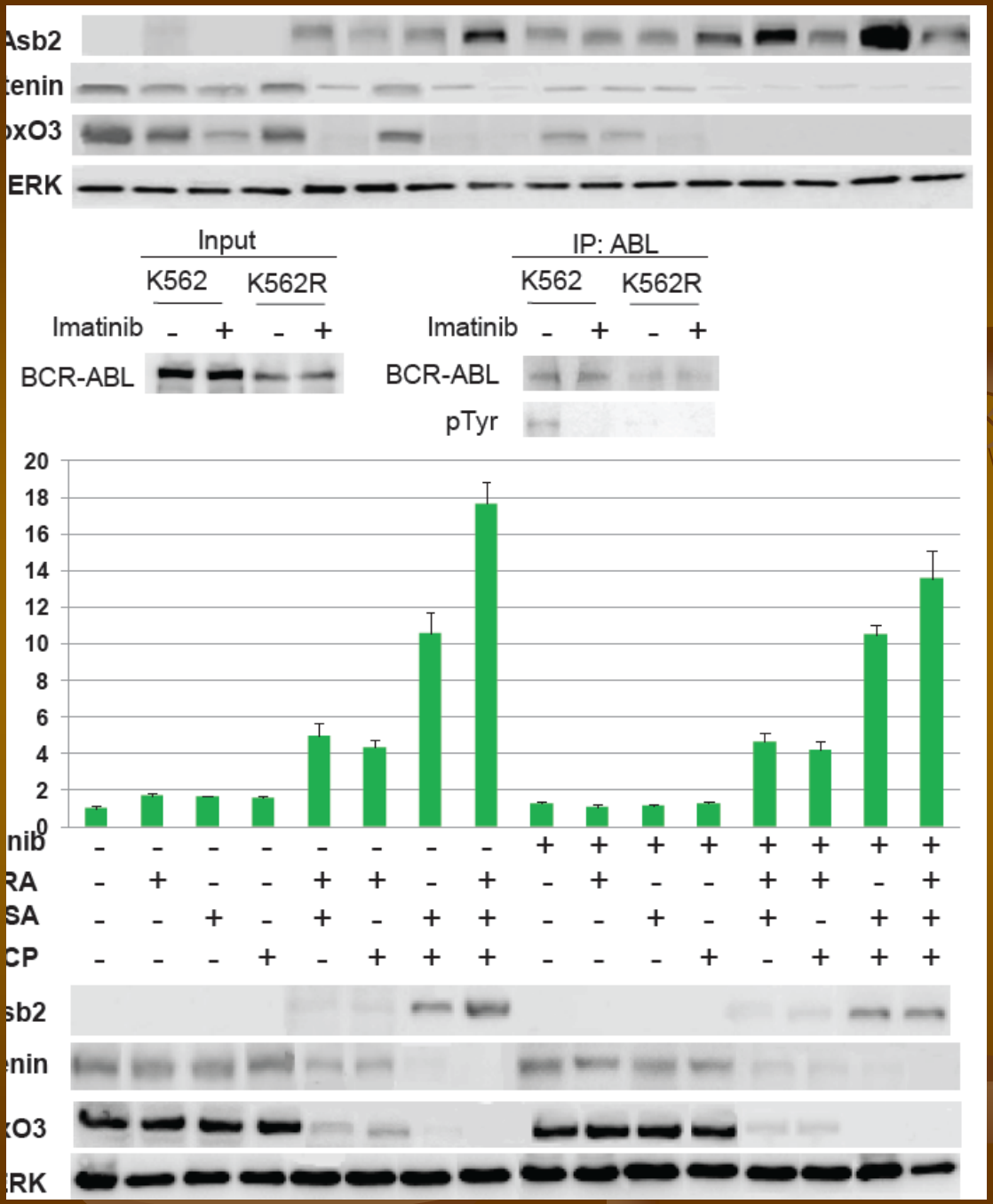
False Colorization "Gradient Map" Reveals Mini-features

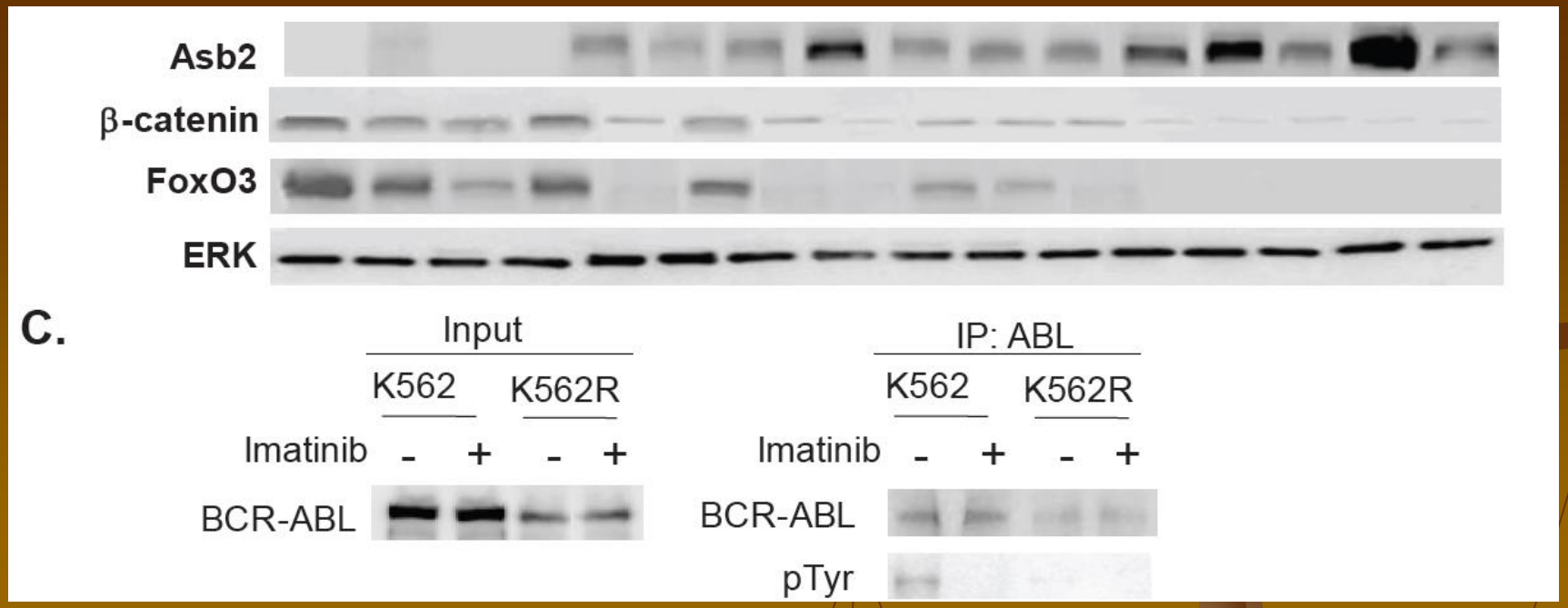


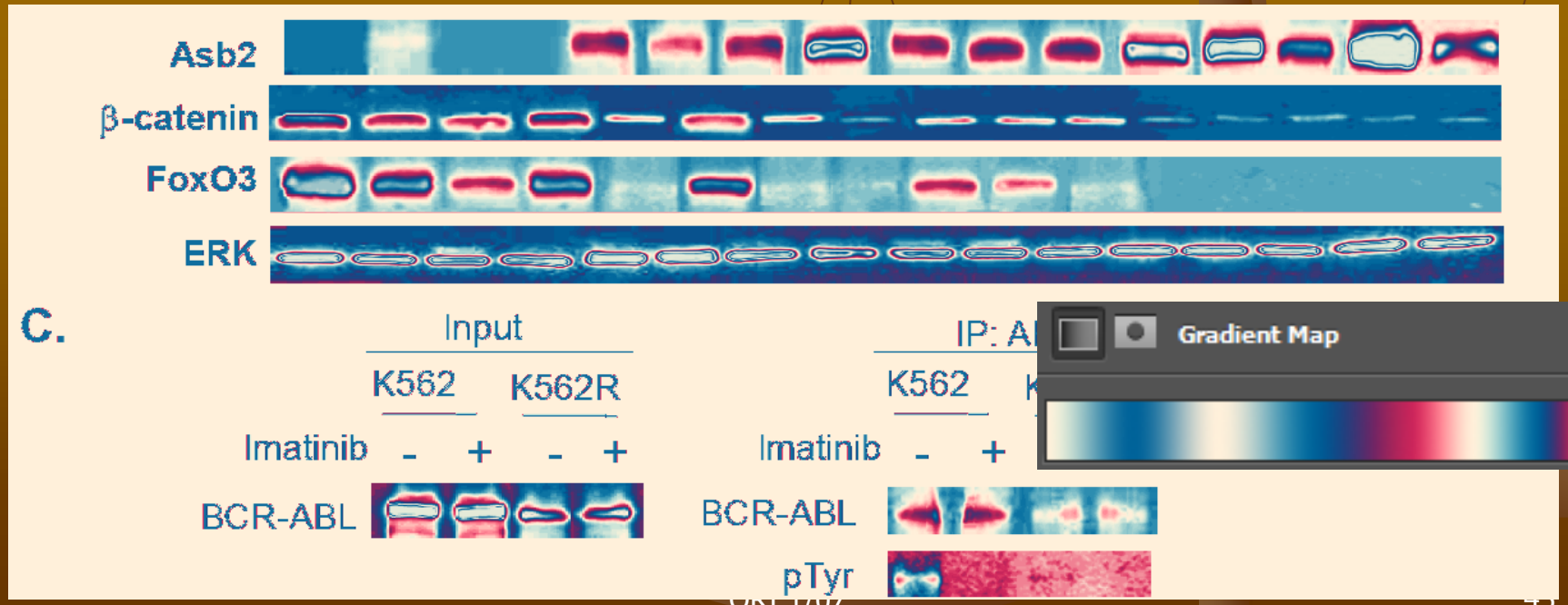
LOWER BANDS SHOW NO SIMILARITY

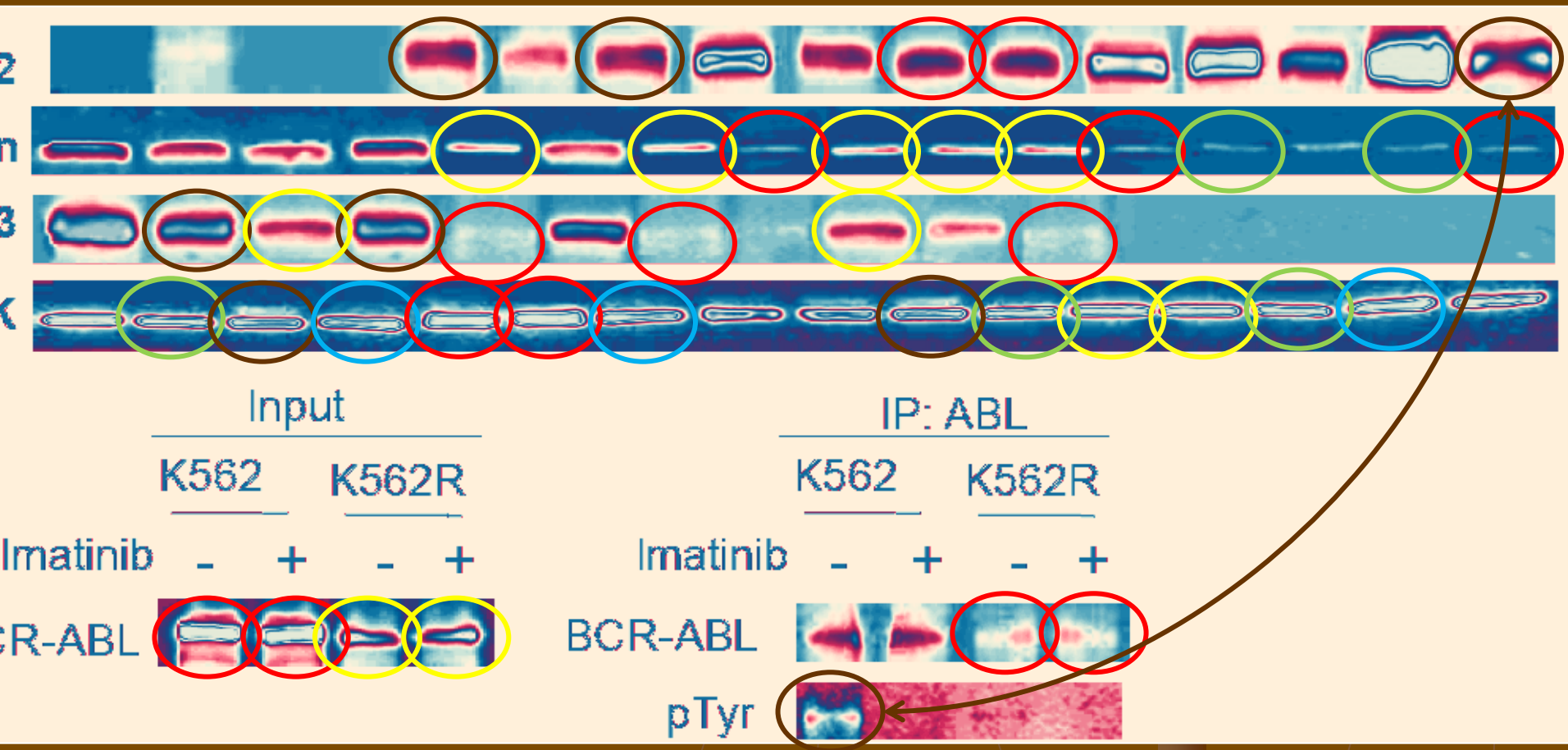
Duplication Example

- Grant applicant accepted data from a postdoctoral fellow without reviewing the raw data.
- Visual inspection and forensic comparison of gel images suggested duplication/reuse/relabeling had occurred.
- Postdoctoral fellow admitted to extensive falsification and/or fabrication through reuse and/or resizing and/or alteration of images.
- To the right is a figure that purportedly showed research results, but none of the experiments were ever run.



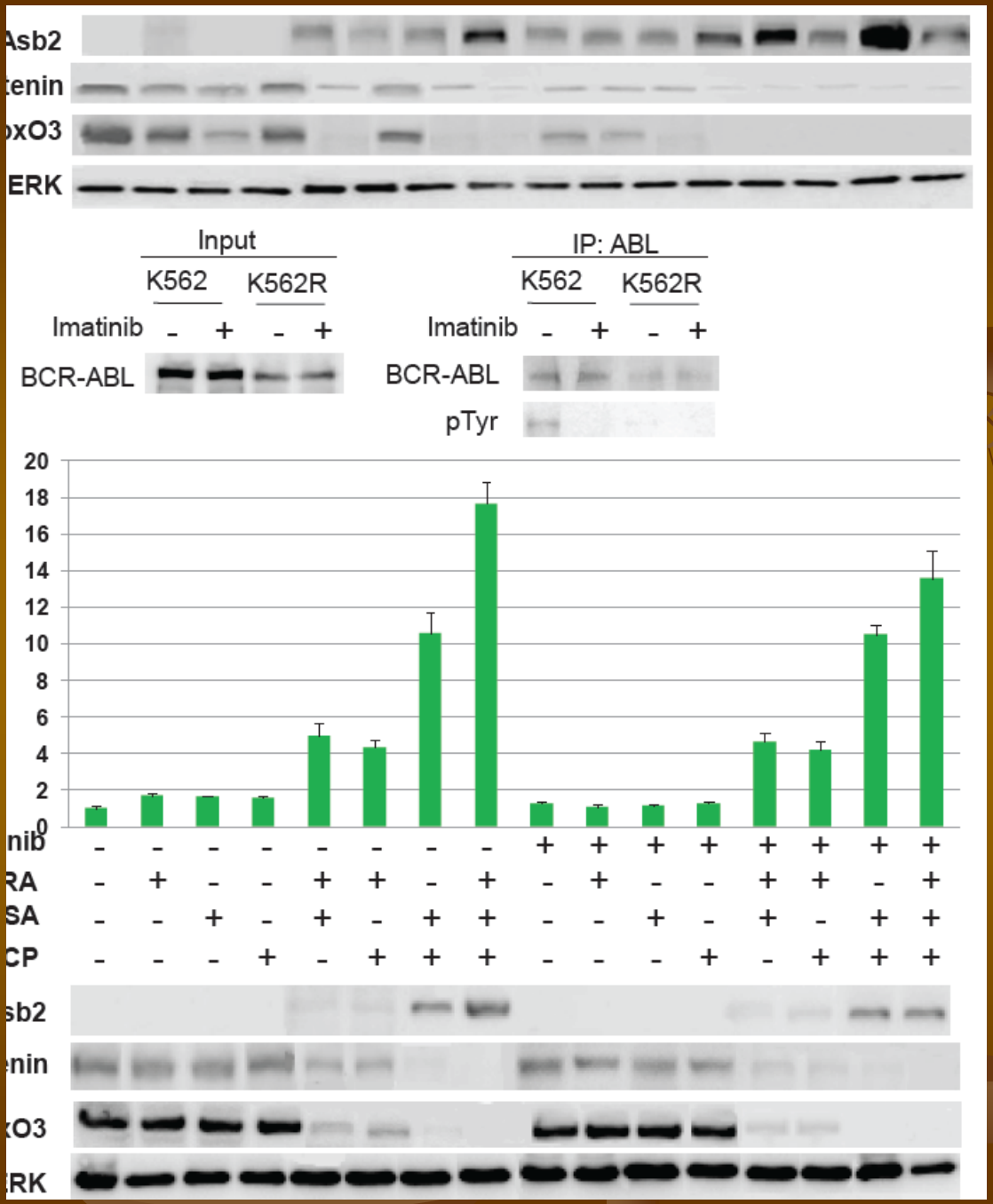

Asb2
β-catenin
FoxO3
ERK
C.
Input
K562
K562R
Imatinib - + - +
BCR-ABL
IP: ABL
K562
K562R
Imatinib - + - +
BCR-ABL
pTyr

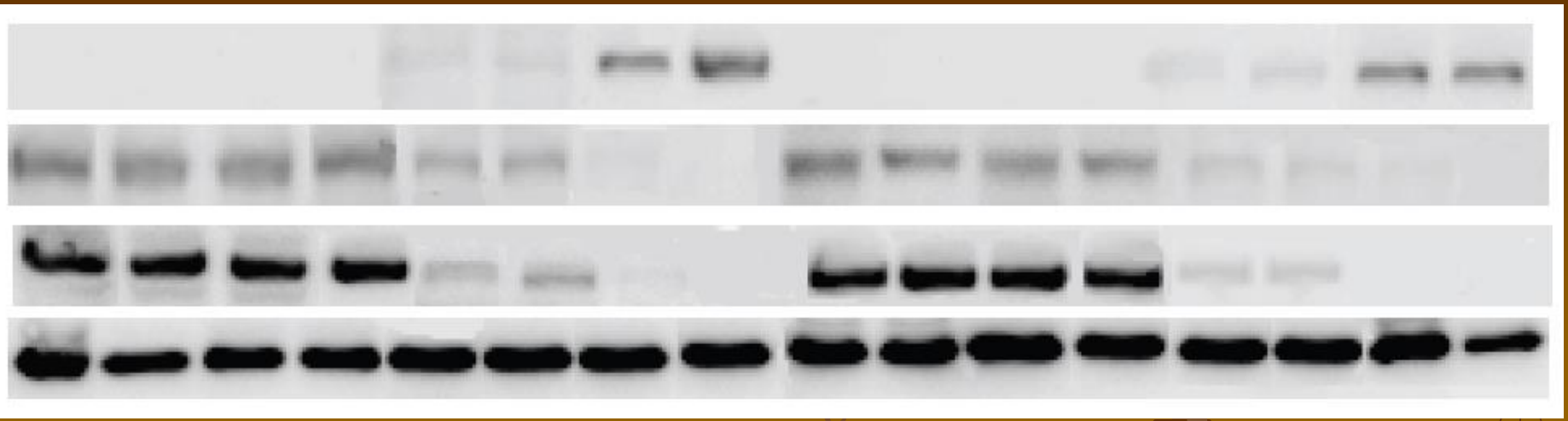

Asb2
β-catenin
FoxO3
ERK
C.
Input
K562
K562R
Imatinib - + - +
BCR-ABL
IP: A
K562
Imatinib - +
BCR-ABL
pTyr
Gradient Map

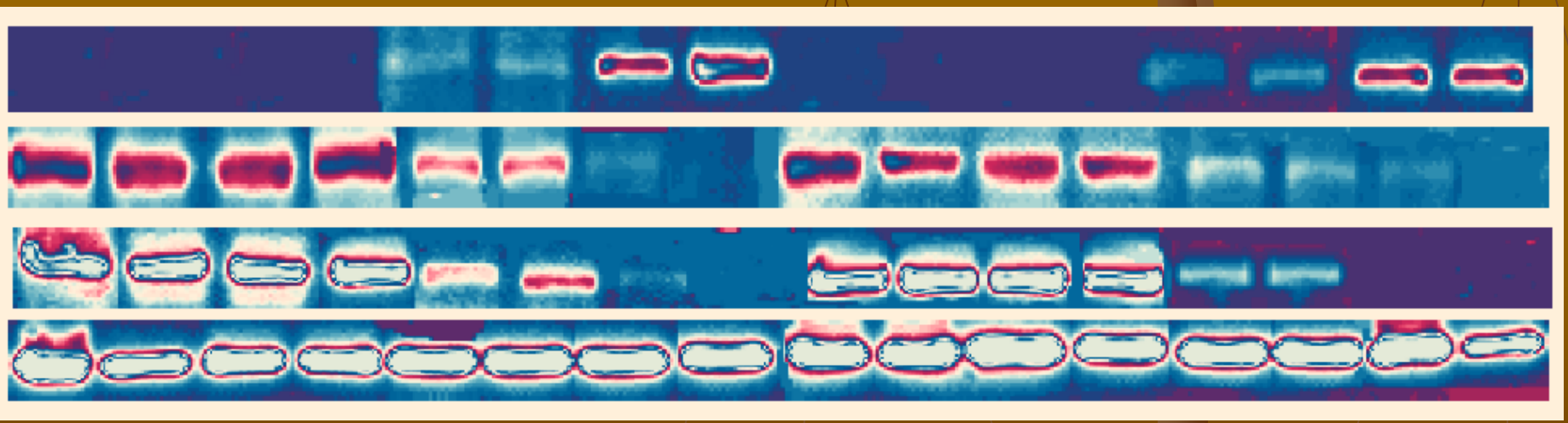

Input
K562
K562R
Imatinib - + - +
BCR-ABL
IP: ABL
K562
K562R
Imatinib - + - +
BCR-ABL
pTyr

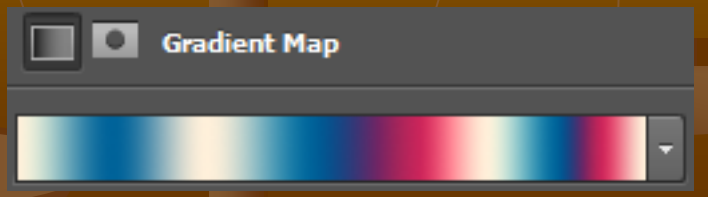
Duplication Example

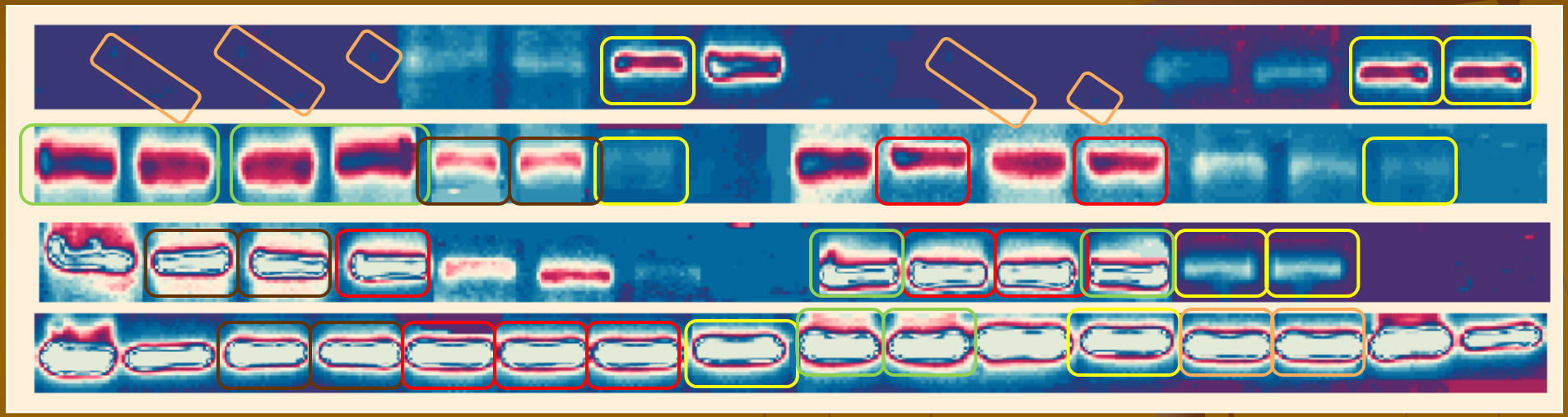
- Grant applicant accepted data from a postdoctoral fellow without reviewing the raw data.
- Visual inspection and forensic comparison of gel images suggested duplication/reuse/relabeling had occurred.
- Postdoctoral fellow admitted to extensive falsification and/or fabrication through reuse and/or resizing and/or alteration of images.
- To the right is a figure that purportedly showed research results, but none of the experiments were ever run.





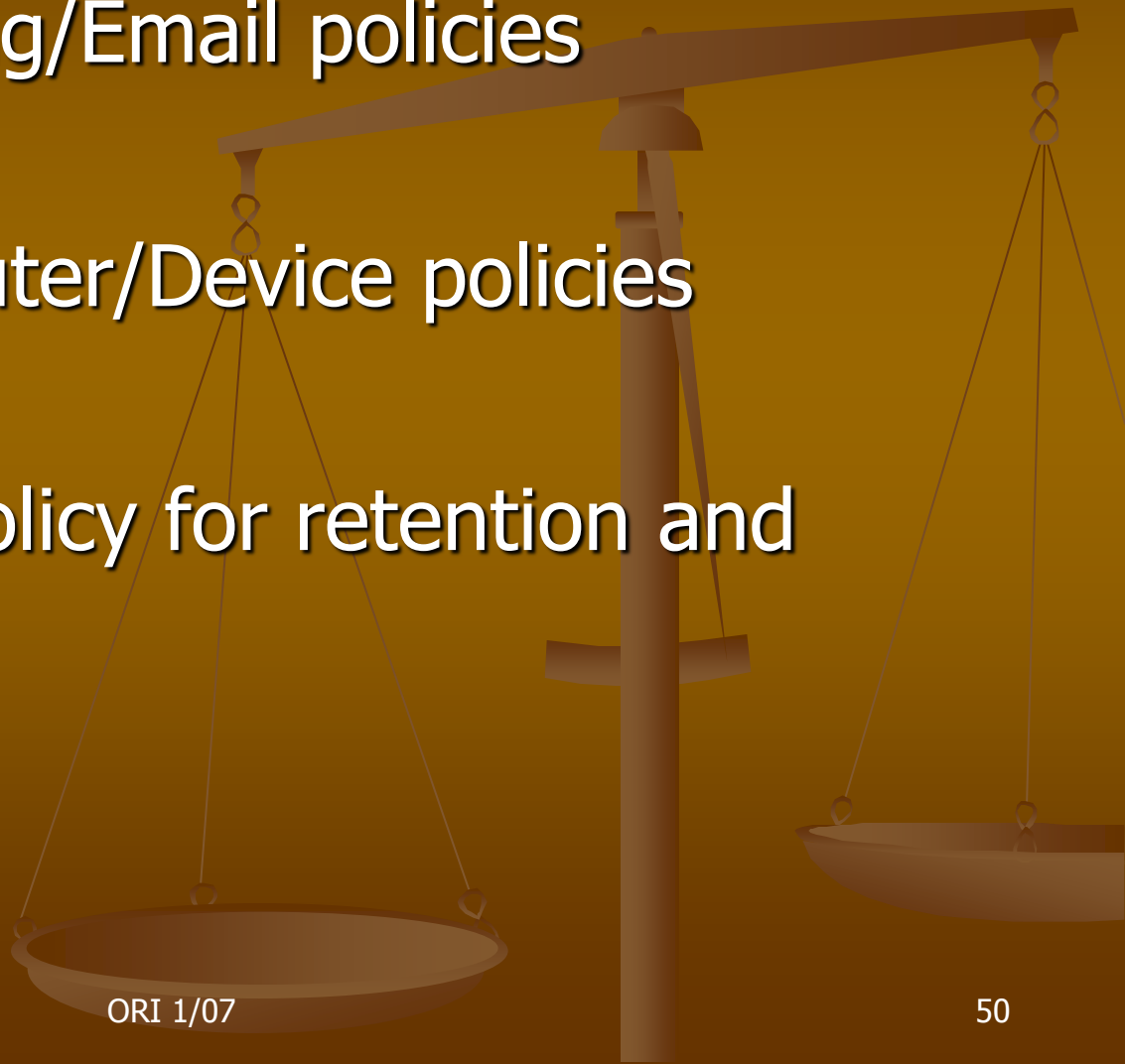



Gradient Map



Developing Challenges

- Cloud Computing/Email policies
- Personal Computer/Device policies
- Original Data-policy for retention and storage



Developing Challenges

- Cloud Computing/Email policies
  - One institution had a senior professor who allowed his Post-doctoral fellows to store raw and processed data on personal DropBox accounts.
  - This data was unrecoverable by the institution without a court order.
  - The Institution determined the best course of action was to ban personal accounts and provided Professional accounts owned by the Institution.

Developing Challenges

- Personal Computer/Device policies
  - Many cases of Research Misconduct involve personal computers.
  - Failure of the institution to obtain these computers has resulted in loss of primary data and inhibited ability to complete investigations.
  - Some institutions have gone to court to recover university owned data residing on personal computers.
  - Some institutions have policies that require the use of institution owned computers.

Developing Challenges

- Original Data-policy for retention and storage
  - The NIH grants policy guide requires retention of data for the term of grant plus three years.
  - A number of recent cases of alleged R.M. have revealed groups with essentially no original data retention.
  - Institutions are developing policies for the retention of primary data and research results to protect their Intellectual Property.

Conclusions

- Data retention and evaluation is important for evaluation of allegations of misconduct.
- Evaluation of the raw data is critical for early detection of problems.
- ORI can provide advice confidentially regarding potential Research Misconduct questions.

ORI can provide assistance 240 453 8800; AskORI@hhs.gov

- Telephone or on site assistance available
- Allegation assessment
- Advice on policies and procedures, for example :
  - Sequestration of evidence
  - Acquisition of digital information (forensic imaging of hard drives)
- Properly getting an inquiry or investigation under way
- Analysis of the evidence, such as assisting with analysis of questioned images
- Investigative strategy and legal problems