Responsible Conduct of Research:

Data Acquisition, Ownership, Management, and Sharing

Prescott Deininger, Ph.D.

Tulane Cancer Center

Tulane University

New Orleans, Louisiana

Special thanks to Brian Weimer and Lisa Wurtzel

“The institution of science involves an implicit social contract between scientists so that each can depend on the trustworthiness of the rest…the entire cognitive system of science is rooted in the moral integrity of aggregates of individual scientists.”

The Common Sense of Science
Jacob Bronowski

I. Data Acquisition

II. Objectivity

III. The Laboratory Notebook

IV. Data Errors

V. Data Ownership

I. Data Acquisition

Principles of Data Acquisition

- *Data acquisition* – The process of obtaining and recording primary experimental information
  - Proper data acquisition and record keeping
    - Provides the foundation on which subsequent data analysis and generalizations are based.
    - Without good data collection and record keeping, all subsequent use of the data is tainted.
    - Proper record keeping is of vital importance for patentable inventions.
    - Do not depend on your memory over time, document everything.

II. Objectivity

What Do Scientists Recognize as Data?

- Quantitative
  - Raw data
    - Recorded by hand (“I counted 153 cells”)
    - Instrument output (NMR peaks)
  - Processed data
    - Charts and graphs
- Qualitative
  - Notes (“The mice in group E were all running and doing backflips”)
  - Some instrument output
    - Pictures
    - Slides
    - Movies
- Potential (Unprocessed)
  - Biological specimens

- Impartial, not biased
- Not motivated by personal gain
- Rigorous test of hypothesis

- Avoid becoming personally attached to a hypothesis or concept
  - Be willing to modify concepts or position
  - Accept responsibility for the validity of the report

- Desire to please one’s supervisors or mentors
- Desire for promotions and advancements
- Improve chances for grant funding
- Facilitate acceptance of publication

“These are the data, we cannot change them”

Principles of Data Acquisition
Case Study I – Personal Gain

- A principal investigator (PI) outlines their theory for a certain effect to a graduate student who is involved in acquiring data to confirm or refute the theory. The PI explains their anticipation of finding experimental data having these values. The graduate student generates the experimental data with approximately the anticipated values. 1) The data are published 2) The PI garners recognition 3) her grant is renewed 4) The graduate student receives his Ph.D.

- The next student working on this project, however, has difficulties reproducing the data. After further investigation, it is found that the first graduate student chose, in cases of ambiguity, values that came closer to the PI's predicted values, thereby <u>emphasizing a trend in the data that was not present in general</u>.

Questions to Consider

1. Should the PI have restrained herself from mentioning his/her anticipations?

2. Should the PI have more closely supervised the acquisition of the data to verify the accuracy?

3. Should the PI have insisted on a more thorough check of data reproducibility?

4. What should the PI do now about the situation?

Discussion of *Case Study I*

- The PI must supervise the design of experiments and the processes of acquiring, recording, examining, interpreting, and storing data. Accordingly, the PI must acknowledge negligence on his/her part.
- Scientists have a duty to avoid contaminating the literature with incorrect information. At this point, the PI should supervise the experiments necessary to resolve any uncertainties. If they conclude that the original publication was seriously flawed, they should publish a correction.

Principles of Data Acquisition
Case Study II - Outliers

A PI acquires data that portray a wonderful correlation between administration of a drug and a physiological response. About 10 % of the data, however, lie far removed from the predicted values.

There are explanations as to why these data are different. For example, some experimental parameters have not been well-controlled, and were different for the experiments in which the results deviate. The PI chooses to ignore the outlying data in the publication.

Questions to Consider

1. Should the PI have published the outlying data with an explanation, thereby weakening the conclusions?

Discussion of *Case Study II*

The investigator may or may not have erred in throwing out the "outliers."

Just because a group of data does not fit your hypothesis is not a valid reason to ignore the data.

However, if there is an objective reason that could justify discounting the data, notable before any data were collected, then the group may be removed from statistical analyses. For example, all mice in group E escaped from the cage and ate some reagents the night before data were collected. In this case, the corresponding data may be disregarded and left out of the publication.

Remember that the most exciting discoveries often come from unexpected results. You are not smarter than Nature. Maybe the confounding results are reflecting something you have not thought of yet.

III. The Laboratory Notebook

What should a good notebook have?

- Bound and numbered pages (no erasure or page removal)
- Date labeled for each day's expts.
- Description of why an experiment is done.
- Description of expt (tables, procedures, ref. to standard protocols etc..)
- Raw data….images (labeled), data fastened, location of original electronic files.
- Description of conclusions or problems
- Description of what's next.
- ELECTRONIC NOTEBOOKS? (https://mynotebook.labarchives.com/login)

Principles of Data Acquisition
Case Study III – The Notebook

Smith, a chemistry graduate student, begins a laboratory research project. At the start, Smith discusses the project thoroughly with his thesis advisor, who also provides relevant references for the student to read. The advisor, however, does not mention laboratory notebook practices.

Smith begins laboratory work and soon begins to obtain interesting results. The student and advisor discuss the results periodically, and excitement for the project increases. After about six months, Smith is asked to write up the results for a publication. In Smith's draft, the raw data have been processed into graphs, tables, and text.

Upon studying the draft, the advisor has a number of questions about the raw data and asks to see the Smith's notebook. To his dismay, the advisor finds that no notebook exists; Smith has been keeping records on loose pieces of paper. The records are undated, and many can not be found at all.

Questions to Consider

1. What should be done?

2. Who is responsible?

Discussion of *Case Study III*

1. What can be done?

   The data don't exist and must be repeated to be used in publication, grant or patent.

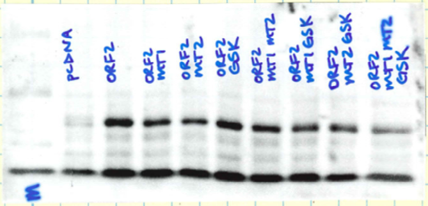
2. Who is responsible?

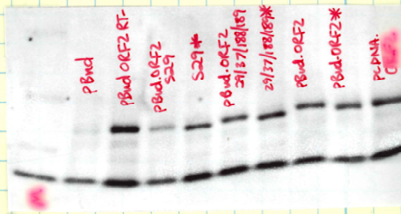
   It was the advisor's responsibility to make sure that Smith had a notebook and understood how to use it from the outset.

   While the advisor may be primarily responsible for the problem, Smith will bear the brunt of it. The relevant experiments must be repeated.

Weds 3/11/15

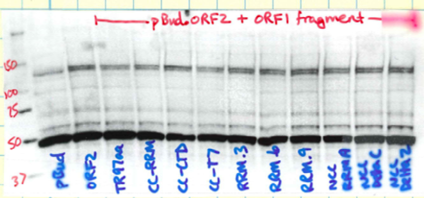
- 2° anti rabbit for westerns → 11ml buffer + 4ml 3T3 media, 3% milk





ORF2 960 on tris glycine
~30 sec exp
old MT1/MT2/GSK western on p28

wash
store
Tris



→ ORF2 960 on tris acetate
~21 sec exp

→ reprobe w/GAPDH
- Block RT normal buffer 1hr

- Try transforming ~~XXXX~~ pBud.ORF2 ligations w/company cells

- change HYG media on Hela EN1 chronic tox - D7, starting to se colonies by eye now. The EN-- #s look super high! Hope the p post CO_2 changes don't differ by much. May need to do retro a if so.

- Hela Zeo dose curve ~ D8 → finally having more aggressive die-offs, still a lot left on the flask. No colonies yet.
- Hela ORF2/Zeo - D2 - nothing yet.

- start G418 on Hela acute EN1 tox - spike 80μl
- H2^s → feed
- H1^s → count → 465,000   500K = 1.075ml
    400K = 860μl
  ↳ seed EN1 T25×19
  EN1.5 T25×6 } 9PM

- sequences, trying to figure out L1Neo+/- ; L1GFP+/- ; L1Blast+/

Thurs 3/12/15

Date: 03/12/15
Cell type: HeLas, kept at 20% O2
Media: MEM + FBS +++
Seed: 400K per T25; 03/11/15 ~9PM

- cells confluent but not super packed

#	Plasmid 1	ng/ul	ug per rxn	DNA for 1 rxn	Plus 4ul per rxn	DMEM w/out serum
1	pcDNA	51.6	2	38.76	4.0	157
2	EN1--	47	2	42.55	4.0	153
3	EN1.5--	48.35	2	41.37	4.0	155
4	EN1.5 - CK1 mut	38.83	2	51.51	4.0	144
5	EN1.5 - GSK3B	36.75	2	54.42	4.0	142
6	EN1.5 - PDZ	36.93	2	54.16	4.0	142
7	EN1.5+	51.85	2	38.57	4.0	157
8	EN1 - S29A	44.7	2	44.74	4.0	151
9	EN1 - S33A	44.5	2	44.94	4.0	151
10	EN1 - S37A	50.35	2	39.72	4.0	156
11	EN1 - S79A	48.9	2	40.90	4.0	155
12	EN1 - S188A	52.5	2	38.10	4.0	158
13	EN1 - S228A	48.1	2	41.58	4.0	154
14	EN1 - S29A/S37A	51.45	2	38.87	4.0	157
15	EN1 - S79A/T82A	49.6	2	40.32	4.0	156
16	EN1 - S188A/T189A	48.55	2	41.19	4.0	155
17	EN1 - S29A/S37A/S188A	40.05	2	49.94	4.0	146
18	EN1 - S29A/S37A/S228A	54.1	2	36.97	4.0	159
19	EN1 - S33A/T47A/T140A	42	2	47.62	4.0	148
20	EN1 - S29A/S37A/S188A/T189A	46.75	2	42.78	4.0	153
21	EN1 - CK1 mut	46.4	2	43.10	4.0	153
22	EN1 - GSK3B	46.3	2	43.20	4.0	153
23	EN1 - PDZ	45.7	2	43.76	4.0	152
24	EN1+	49.4	2	40.49	4.0	156

Incubate for 15 min at RT

	25
Lipofectamine 8ul	200
DMEM 92ul	2300

Add 100ul Lipo mix to DNA mix per reaction
Incubate for 10 min at RT ~12:40
***Add 2ml DMEM w/out serum to flasks**
***300ul of DNA/Lipo mix to each flask**

***3 hours later, remove DNA/Lipo mix and add complete media**

- 2° anti rabbit for GAPDH, reuse 2ND time from 1/31

- change Zeo on dose curve & Hela ORF2 transf, hasn't started dying yet on transfected group.
- H1^s → good
- H2^s → split 1:20 × 5
- lab meeting w/ Victoria

- Block RT 1 hr for yesterday's TrisGlycine gel. Try probing w/ p84 4° O/N. Reuse from 1/25/13 1:50K, 2ND use

Aug 5 2008

1. Western blot

[illegible]

A.

M	293 PES1	293 PEG	293 P21 S1	293 P21 G	293 PE mah/3 S1	293 PE mah/3 GFP	293 P21 mah/3 S1	293 P21 mah/3 GFP

SIRT2 Rb (180kD) — 1 hr 88
pTEN Rb (54kD) — 1 hr 88
Rad51 (ms) (37kD) — 1 hr 88

B.

M	Cos	C 81	C 4	C PES1	C 07	C 85	C 86	C 2	C G

SIRT2 Rb (180kD) — 1 hr 88
NF-kB Ms (65kD) — 1 hr 107
Rad51 (ms) (37kD) — 1 hr 107

2. DAB (IHC: for mpGTES1, 15MPGD)

Immediately before use, prepare the substrate solution as follows.

① 5 ml. ddH2O 2 drop buffer stock solution
② 4 drop DAB mix
③ 2 drop Hydrogen peroxide solution
④ 2 drop Nickel Solution

mix

room temperature
2–10 mins
wash 5 min water

3. IFS CCD

08	85	87
GFP	hs4	S1

β-Catenin (mouse)

4°C overnight

66

Aug 6. 2008

1. western blot

[illegible]

M	L08	S87	L9	S1	S9	S85	86	S 19	S 29

EGR1 (M) ... 1 Day 1:1000

pEGFR (R) ... 1:1000

β actin (M) ... 1:1000

p-44 (m)

M	L08	S87	S2	S1	S9	S85	S86	S	L29

Stat3 (M) ... 1 Day ... 87 14

pPTEN (R) ... 88 W

pCNA (36 KD) (M) ... 89 7

Survivin (R) ... 89 V

67

2. gene transfection

HEK293	15 HPlus	[illegible]
HEK293	15 HPlus	[illegible]
HEK293	15 HPlus	[illegible]
HEK293	15 HPlus	[illegible] 26 [illegible]

A: 20ul plasmid + [illegible] 75 ul [illegible]

B: 30ul lipofectamine + 5ul optimem

→ 15 min RT → Add 15 min → incubate

→ Change medium after 3 hours..

7. IFS

	08	80	87		08	85	82	
Hep3B	Con	L2	S1		GFP	L2	S1	Huh7

anti-β-Catenine

68

Aug 7 2008

1. Western blot

M	C 68	C 87	C	C	C	C 07	C 87	C 86	C	C 12

SARA
TLR4
PCNA

LPS

3.

M	C 07	C 87	C	C	C	C 07	C 87	C 86	C	C

Control

SARA
TLR4
PCNA

2. IFS

07	87	86
GFP		

anti-PTEN

the top panel

2. IP wash buffer

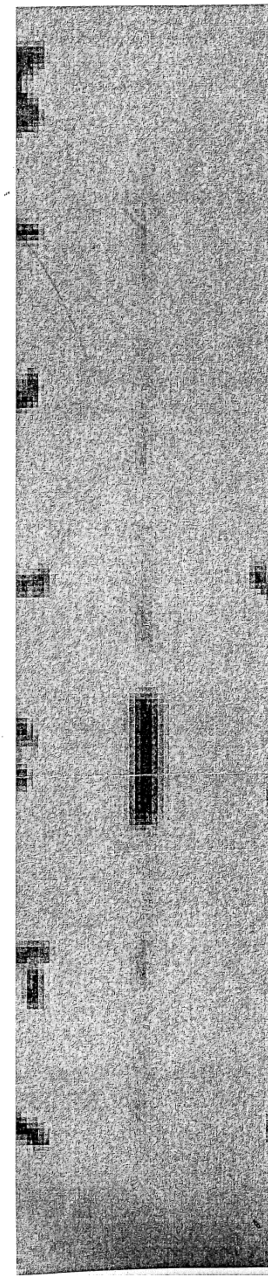
50mM pH8.0 Tris
150mM NaCl
1% NP40
1mM EDTA
ddw

3. extract protein

HEK293
HEK293 15H
HEK293 15H
HEK293 15H

4. HEK293 for luc

5. rotating ... PPAR
p53 TP



No. 22133087

LABARCHIVES

https://libguides.tulane.edu/labarchives/about

What is LabArchives?

LabArchives is a secure, cloud-based electronic lab notebook (ELN) designed for research. This ELN system enables researchers to capture, manage, store, and share information and data. LabArchives facilitates collaboration amongst researchers within a lab, institution, and with external stakeholders.

Key features:

- **Upload and store files** – including text, tables, images, spreadsheets, and attachments – in their original format
- **Create standard notebook templates** for your research group
- **Store data securely** on LabArchives servers: multiple redundancy ensures 24/7 data availability
- **Share information seamlessly** within your lab and invite collaborators from outside of Tulane to join your notebook
- **Maintain all revisions** of notebook entries
- **Provide compatibility across multiple platforms**, including mobile devices
- **Access and connect seamlessly** to lab teams
- **Navigate data easily and quickly** with simple search functions

Made possible through the support of Tulane University Office of Research.

IV. Data Errors

- Design level
  - Skew: experimental design that favors certain results
- Experimental Level
  - Undesirable or negative results are disregarded
- Analysis Level
  - Statistical treatment is not appropriate
  - Grouping is forced
- Interpretation Level
  - Personal bias leads to erroneous interpretation
- Fraud
  - Deliberate error with intent to deceive

- Sampling Error
  - Due to chance variation in sample selection
    - Sample size may be too small
- Selection Bias
  - Distortion resulting from manner in which subjects were selected
    - Taking a poll in front of the Democrat National Headquarters

- Information Bias
  - Measurement error
    - Machine not calibrated
  - Misclassification of subjects

- Confounding Error
  - Influence of uncontrolled variables that are linked with the independent and dependent variables under study
    - Incubator ran dry, cells responded differently
    - Mice were affected by researcher turning on the lights in the middle of the night for another experiment
    - Phase of the moon?

There are many different reasons why erroneous data may result or incorrect interpretations may be made. Mistakes of this nature can be costly and may have serious consequences … but they are NOT FRAUD

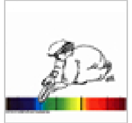
Fraud involves the <u>deliberate intent to deceive</u>

- Saying you ran experiments that you did not
- Saying you ran more experiments than you did
- Changing the data to fit your bias
- Only counting results you like
- Intentional misinterpretation of data

•https://ori.hhs.gov/education/products/RIandImages/guidelines/list.ht

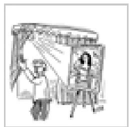
Guidelines for Best Practices in Image Processing

Please click on each guideline for further details. Also, see the guidelines demonstrated in **Photoshop Videos**.



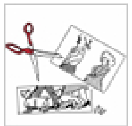
Treating Images as Data: Digital scientific images should be treated as data

Saving the Original: Manipulations of digital images should always be done on a copy of the raw image data. The original must be retained.

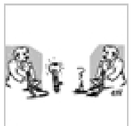


Making Simple Adjustments: Simple adjustments to the entire image are usually acceptable.

•https://or

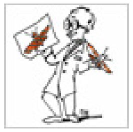


Cropping is usually OK: Cropping an image is usually acceptable.

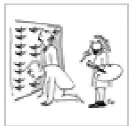


Comparing Images: Digital images that will be compared to one another should be acquired under identical conditions.

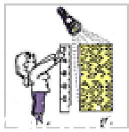
Manipulating the Entire Image: Manipulations that are specific to one area of an image and are not performed on other areas are questionable.



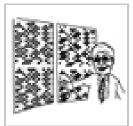
Filters Degrade Data: Use of software filters to improve image quality is usually not recommended for biological images.



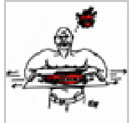
Cloning Degrades Data: Cloning objects into an image or from other parts of the image is very questionable.



Making Intensity Measurements: Intensity measurements of digital images should be performed on raw data and the data should be calibrated to a known standard.



Lossy Compression Degrades Data: Avoid the use of lossy compression.



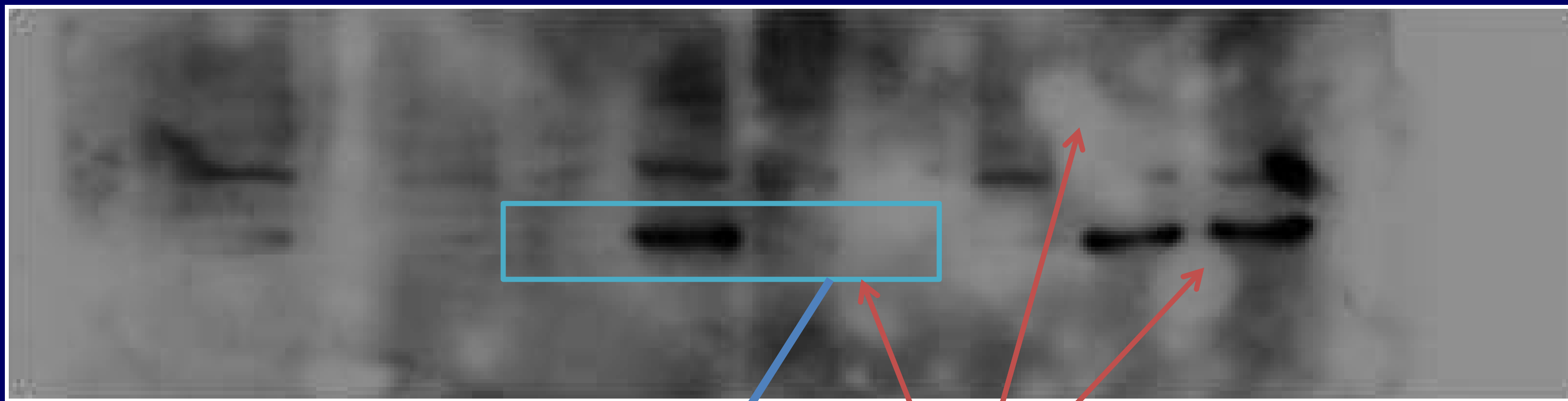
Issues With Magnification: Magnification and resolution issues are important.

Issues With Pixels: Be careful when changing the size (in pixels) of a digital image.

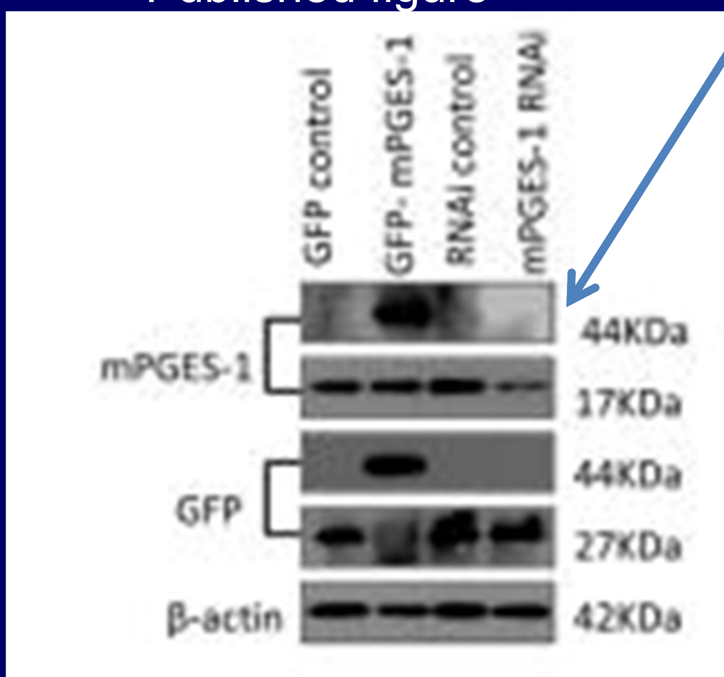
https://ori.hhs.gov/infographics

- Look at the first two segments of this web site for:
- 1. real-world examples of fraud detected.
- 2. general suggestions on appropriate image manipulation.

- Original Image



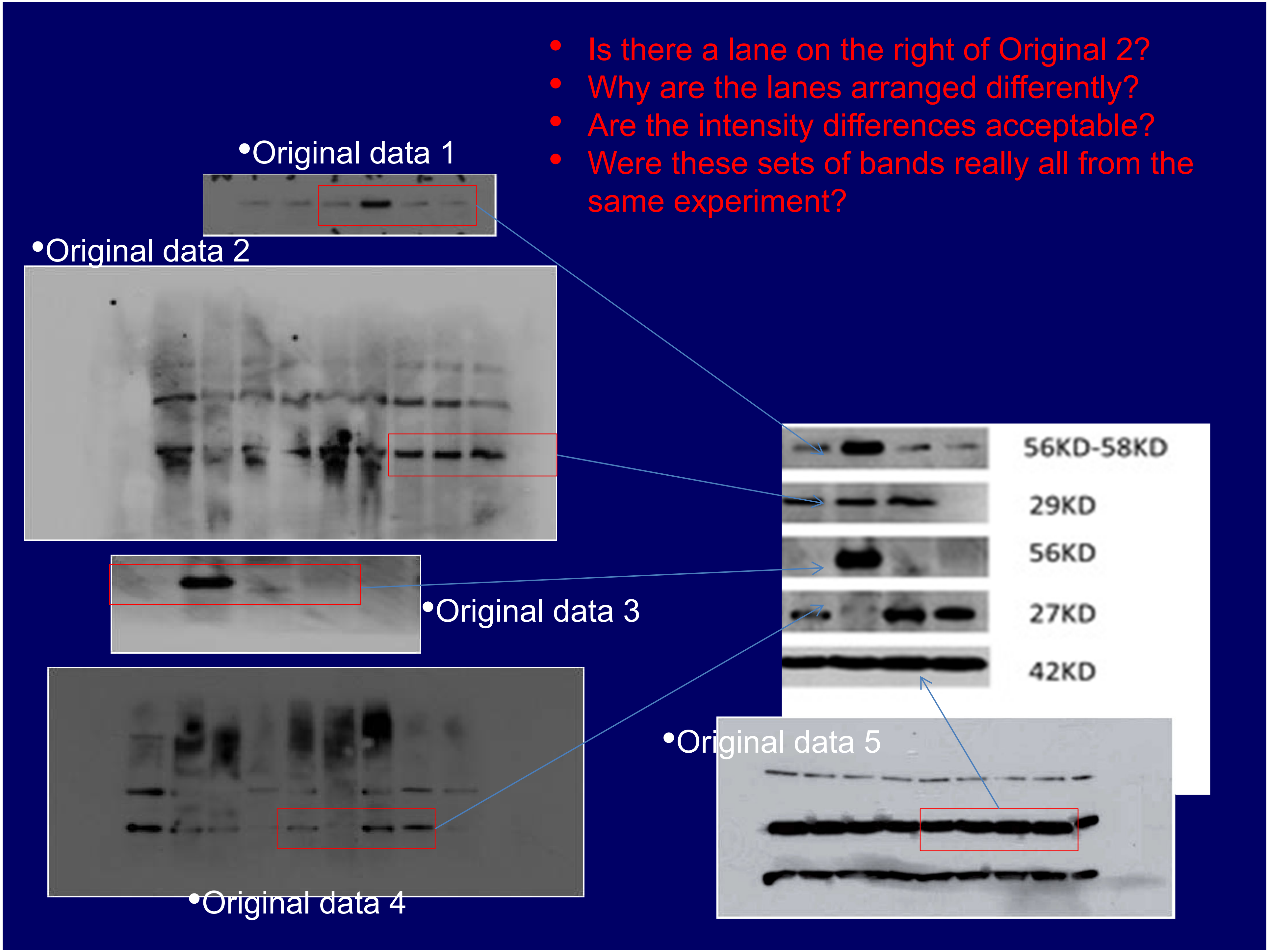
- Published figure



- Bubbles during transfer????

- The top panel does seem to authentically represent the gel.
- However, it looks like the gel has significant defects that likely hide potential bands.

The student prepared the figure using the above data. 1: Who is responsible for any interpretation mistakes?
2: Is it fraud?


Is there a lane on the right of Original 2?
Why are the lanes arranged differently?
Are the intensity differences acceptable?
Were these sets of bands really all from the same experiment?
Original data 1
Original data 2
Original data 3
Original data 4
Original data 5
56KD-58KD
29KD
56KD
27KD
42KD

V. Data Ownership

Ownership and Sharing of Data from Federally Funded Research

- The Bayh-Dole Act (A.K.A. The Patent and Trademark Amendment)
  - Established a national policy encouraging government, universities, and industry to work together to commercialize new technologies.
  - Removed obstacles that previously blocked transfer of technologies developed with federal funding.
  - Before the Bayh-Dole Act, funding agencies owned the intellectual property developed with their support, and fewer than 5% of the 28,000 patents held by the U.S. government were licensed to industry.
  - Now, the universities usually retain title to intellectual property and are free to market it.
  - However, journals and funding agencies often impose rules about data and reagent sharing.

Tulane Regulations and Policies

- Human Research Protection Program / Institutional Review Board (http://tulane.edu/asvpr/irb/upload/Tulane-HRPP-SOPs.pdf) (http://www2.tulane.edu/asvpr/irb/)
- Institutional Animal Care and Use Committee (IACUC) (http://tulane.edu/asvpr/iacuc/hsc/sops.cfm)
- Research Compliance and Research Integrity (http://www2.tulane.edu/asvpr/research-compliance.cfm)

Who Owns Research Done at TU?

Tulane!*

- Raw data (including laboratory notebooks)
- Processed data

*subject to conditions established by granting agencies or contracts with sponsors.

- It is the Principal Investigator who manages the data
  - The Administrator of the unit in which (s)he works may also manage data.
  - All data and notebooks stay with the university!!!! Copies may be taken away, but only used in conjunction with the university.

Lawsuit Alleges Professor Stole Student's Research and Defrauded University of Millions

By Zipporah Osei | MARCH 01, 2019

U. of Missouri at Kansas City

U. of Missouri at Kansas City

A University of Missouri at Kansas City professor who resigned in January amid allegations that he had exploited graduate students stole a student's research and secretly sold it to a pharmaceutical company, according to a lawsuit filed on Tuesday by the university system.

The suit, filed in federal district court in Kansas City, Mo., alleges that the research has already earned the professor, Ashim Mitra, more than $1.5 million and could earn him $10 million more in royalties over the next five years.

• Bill Priestap Assistant Director, Counterintelligence Division Federal Bureau of Investigation Statement Before the Senate Judiciary Committee Washington, D.C. December 12, 2018

China's Non-Traditional Espionage Against the United States: The Threat and Potential Policy Responses

Moffitt Cancer Center details links of fired scientists to Chinese talent programs

By Jeffrey Mervis | Jan. 19, 2020 , 10:30 AM

Six Florida cancer researchers who were **dismissed last month** for hiding their ties to a Chinese medical university appear to have been motivated by simple greed and a disregard for both institutional and federal rules.

Improper Influence

It did not violate Moffitt policies, for these individuals to have participated in the Talents programs, or to have had other academic positions, consulting positions, or research collaborations with Chinese colleagues or Chinese institutions. However, under Moffitt policies and NIH regulations, such activities must be timely disclosed and approved in advance after they have been analyzed for possible conflicts of interest or other compliance risks. Problems also arise when the participation in Chinese activities specifically conflicts with a Moffitt leader's or faculty member's duties to Moffitt and/or to U.S. government agencies like the NIH, or when a Moffitt official accepts undisclosed personal compensation from an entity (TMUCIH) with which Moffitt does business, which would represent a conflict of interest. Other problems arise if a full-time Moffitt leader or faculty member agrees to spend significant professional time and effort on non-Moffitt activities, without permission, which would represent a conflict of commitment

HOW TO STAY OUT OF TROUBLE

1. Disclose any new potential funding for your research (It should always be pre-routed through grants and contracts.
2. Disclose any significant personal income that is not routed through Tulane and is professionally related (This should include expensive trips etc. paid by other entities)
3. Disclose investments, income, leadership roles in any company that is affected by your actions.
4. Evaluate collaborative relationships, particularly with visiting scientists to protect Tulane's ownership of discoveries.
5. Remember that foreign entities aren't the only source of conflict. More traditionally it applies to relationships with companies, such as drug companies.

SHARING

- Once data are published, all related resources must be shared.
- Once data are public in any way, there is one year before patent potential is lost.
- Making data available informally, but publicly, can limit publication options.