Ethics, Scientific Integrity, and Deontology

Arnaud Legrand

Master 2 MOSIG
December 2024

Toward a Responsible Science

Three pillars

HCERES OFIS, CNRS COMETS, Inria COERLE, CERNA

1. Ethics
   - Deep questions raised by science progress on society; Moral values motivating our acts and their consequences.
   - Regular discussions, connection with society. [COMETS/COERLE/...]
   - Retrospective [law, approval] vs prospective view
2. Scientific Integrity
   - Rules governing research practices. Code of conduct for trustable research. Domain dependant.
   - Universal, *soft law* professional code. [Scientific integrity/whistleblowers referent]
3. Deontology
   - Le Pors 1983 rév. 2016: *"Le fonctionnaire exerce ses fonctions avec dignité, impartialité, intégrité et probité"*
     - Independence, equity, private activities, ... of civil servants
     - Conflicts of interest, connection w. industry, harassment, intellectual property, ...
   - Law, deontology referent

Law evolution (in France)

- **2015**: EU puts forward as a priorities the opening of science, of innovation and openness to the world. Societal science
- **2016**: Doctoral training on ethics and scientific integrity is mandatory (doctoral schools)
  - Related to the Corvol report (Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique)
- **2018**: National Plan for Open science
  - open access (HAL) and open data (FAIR)
  - sustainable and international movement
- **2021**: 2nd National Plan for Open science
  - 100% open access data by 2030
  - open source, open science = new norm
- **2022**: the ~~Hippocratic~~ PhD Oath

  In the presence of my peers. Having completed my doctorate in [discipline], and having thus practiced, in my quest for knowledge, the exercise of a demanding scientific research, cultivating intellectual rigor, ethical reflexivity and respect for the principles of scientific integrity, I commit myself, for what will depend on me, in the continuation of my professional career, whatever the sector or the field of activity, to maintain an honest conduct in my relationship to knowledge, my methods and my results.

Content

Deontology

Deontology (charte de déontologie du CNRS)

Obligation de dignité ~~violent behavior~~, ~~offensive and inappropriate comments, insults~~; ~~hostile/degrading/humiliating environment,~~...

Obligation de probité et d'intégrité ~~using your position to obtain personal benefit~~, ~~using public resources or funds for your own~~, ~~favoring relatives~~, ~~getting someone else's credit~~ (honest and disinterested engagement)

Obligation d'impartialité treat anyone in a fair and objective way (~~bias~~, ~~partisan considerations~~, ...)

Obligation de neutralité behavior and engagement should not depend on political/philosophical/religious opinions.

- No conflict with freedom of speech and critical thoughts
- No discrimination based on origin, sex, identity, age, family situation, physical appearance, religion, ...

Obligation de secret/discretion professionnel confidential data, economic and industrial espionage

Obligation d'exécution des fonctions et d'obéissance hiérarchique

Conflict of Interest

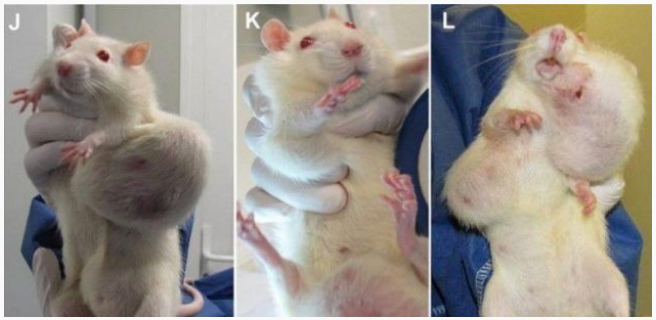
Industry lobby (tobacco, sugar, oil, chemistry) may fund your research so that you give the conclusions they like ...
or study alternative explanations, etc.

- Relationship between Funding Source and Conclusion among Nutrition-Related Scientific Articles
- Known as the Funding bias

Conflict of Interest

Industry lobby (tobacco, sugar, oil, chemistry) may fund your research so that you give the conclusions they like ...
or study alternative explanations, etc.

- Relationship between Funding Source and Conclusion among Nutrition-Related Scientific Articles
- Known as the Funding bias



Greenpeace, anti-GMO, anti-4|5G lobby

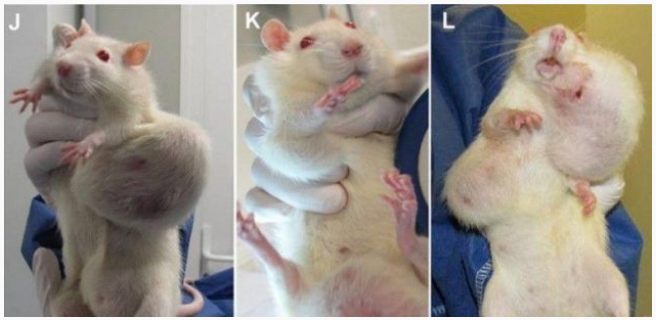
- Long-term toxicity of Roundup-tolerant GMO on rats by Séralini et al.

Super hard and sensitive topics. I wouldn't take position.

Conflict of Interest

Industry lobby (tobacco, sugar, oil, chemistry) may fund your research so that you give the conclusions they like ...
or study alternative explanations, etc.

- Relationship between Funding Source and Conclusion among Nutrition-Related Scientific Articles
- Known as the Funding bias



Greenpeace, anti-GMO, anti-4|5G lobby

- Long-term toxicity of Roundup-tolerant GMO on rats by Séralini et al.

Super hard and sensitive topics. I wouldn't take position.

An other completely different example:

Collusion Rings Threaten the Integrity of Computer Science Research

Bullying and Sexual Harassment

Surprisingly present despite a high level of education ☹

2016 Jane Willenbring filed a Title IX complaint of sexual harassment against her adviser. Released the day after the New York Times story of Harvey Weinstein's alleged sexual assault and harassment. #MeToo

https://en.wikipedia.org/wiki/Picture_a_Scientist

Recent public harassment example: D. Raoult vs. E. Bik

Bik pointed many flaws on PubPeer in Raoult's articles

Raoult and his supporters answer with intimidation and threats on social media, lawsuits

> *She is a fake expert, a charlatan, and she is paid off. A bunch of nonsense, tantamount to murder. Now she is trying to stop Parkinson's, What a disgusting person*

Raoult accuses the PubPeer admin, Boris Barbour, of complicity

COMETS finally takes position

> *Ces procédés relèvent de stratégies d'intimidation inadmissibles. Le COMETS tient à faire part de son inquiétude face à de telles pratiques qu'il convient de condamner fermement. Au-delà, le COMETS déplore la judiciarisation progressive des questions d'intégrité dans la recherche qui relèvent avant tout d'une expertise et d'une autorité scientifiques.*

Ethic Regulation

Credits & Acknowledgments

This whole part is stolen from Carole Peyrin's presentation:

- https://pod.univ-lille.fr/video/16631-s12-histoire-de-lethique-de-la-recherche/
- https://pod.univ-lille.fr/video/16633-s14-les-fondamentaux-de-lethique-de-la-recherchemp4/

Tuskegee Syphilis study

1932 – 1972 US Public Health Service (PHS) and the Centers for Disease Control and Prevention (CDC) conduct a study on a group of nearly 400 African Americans with syphilis.

Observe the effects of the disease when untreated

Tuskegee Syphilis study

1932 enroll 600 impoverished African-American sharecroppers

- 399 with latent syphilis, 201 uninfected for the control group
- promised medical care
- never informed of their disease, placebo disguised as bad blood treatment

1943-1947 penicilin has become the standard treatment for syphilis

- free physical examinations at Tuskegee University,
- free rides to and from the clinic, hot meals on examination days
- autopsy agreements in return for funeral benefits

1972 leak to the press

- 28 died directly from syphilis, 100 from complications, 40 wives infected, 19 children born with congenital syphilis

1997 Public apology from President Clinton

Unethical clinical studies

- No free and informed consent
  - Concealment: hide information on disease and risks
  - Deception: fraudulent medical care
  - Influence: money incitation, friendly nurse, military exemption
  - Concealment: hide the cure

- 1966 : Henri Beacher, *Ethics and Clinical Research*, New England Journal of Medicine.
  - 22 examples of unethical clinical research that risked patients' lives

Ethical regulation

1946-1947 Nazi doctor's trial and Nuremberg Code on Human Research

- free and informed consent should be the norm

1964 Helsinki declaration (code of conduct)

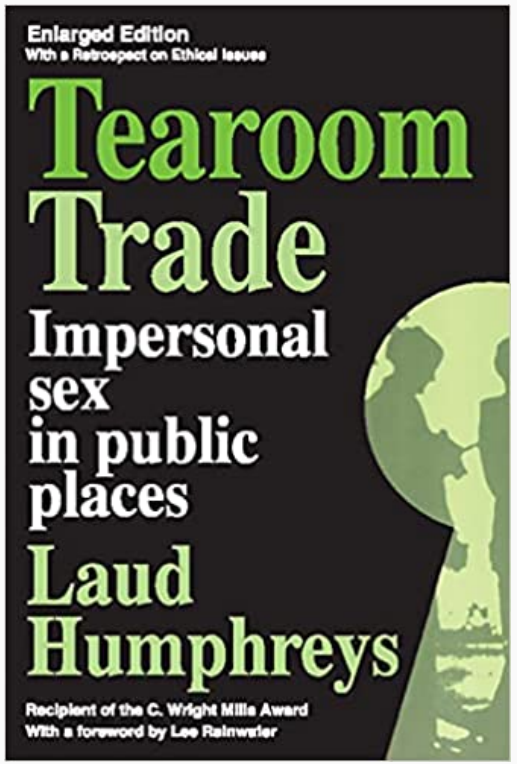
- Need for free and informed consent

1975 Helsinki II

- Mandatory free and informed consent
- Ethical board pre-approval
- Should the research be unethical, no publication

Social sciences: gay anonymous relations

- Laud Humphreys, sociology, wants to study/explain such motivation
  - *Tearoom Trade: Impersonal Sex in Public Places* (1970)
  - Huge risk for such persons (fee, reputation, etc.)
- Masquerades as a voyeur, Topological/temporal study
  - Detailed diagrams and maps of tearoom
- No consent, tracked down names and addresses through license plate numbers
- Interviewed later the men in their homes in disguise and under false pretenses



Ethical regulation 2

1946-1947 Nuremberg Code

1964 Helsinki declaration

1975 Helsinki II declaration

1981 US Common rule (biomedical and behavioral research involving human subjects), for federally funded research

1988 Loi Huriet in France (medical research)

- Ethical board pre-approval required for any biomedical research
- What about SHS and ICST ?

2009-2012 Loi Jardé *recherches impliquant la personne humaine*

2016-2018 Revision of the Common rule

- Big Data Analytics and revision of the commun rule

All this is mostly related the risk for human subjects

Do Artifacts have Politics? (Langon Winer)

UC Davis, 1945: Plant breeder Jack Hanna and engineer Coby Lorenzen, team up to invent a machine that could mechanically harvest tomatoes, no one thought they could do it.

- When they started, it was cheap and efficient to pay (Mexican) farm laborers
- Huge 50,000$ machines but reduction by 5-7$ per ton compared to hand-harvesting

Consequences: increase in tons of tomatoes but

- Required ”hard” tomatoes (de-stemmed hybrid vf-145
- Lead to economic concentration (no more small farms, from 4,000 growers in 1960 to 600 in 1973)
- Supression of 2/3 of the jobs in the tomato industry over 30 years

1979-1989: Law suits against UC Davis arguing that tax money is spent on projects favoring a handful of private interests to the detriment of small farmers, consumers, etc. University denied arguing *it would require elimination of all research with any potential practical application.*

What about *consequences* of Science on Society ?

Technologies have political consequences that transcend the simple categories of ”intended” and ”unintended”.

Is this research/technology worth developing?

- Nuclear power, genetic engineering and GMOs, animal testing, spaceships, ...

In computer science:

- Social Networks and recommendation systems
- Facial/voice recognition
- 5G/6G, autonomous cars, ...
- Recent AI developments
  - DALL-E (OpenAI), StableDiffusion, MidJourney (08/22)
  - 11/22 Cicero (Meta) and ChatGPT

This is cool and exciting, but can scientific curiosity justify any research? Let’s keep in mind that it is common that a technological deck has been stacked long in advance to favor certain social interests, and that some people were bound to receive a better hand than others.

Scientific Integrity (Data)

Scientific Misconduct

From https://en.wikipedia.org/wiki/Scientific_misconduct

Danish definition *Intention* or *gross negligence* leading to fabrication of the scientific message or a false credit or emphasis given to a scientist

Swedish definition *Intentional* distortion of the research process by fabrication of data, text, hypothesis, or methods from another researcher's manuscript form or publication; or distortion of the research process in other ways

US NSF definition Scientific misconduct revolves around the three following issues:

1. fabrication (completely made up results).
2. falsification (you obtain result and change them)
3. plagiarism (stealing someone else's sentence or idea without giving credit)

Data fabrication Examples

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received $\approx \$19$ million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $\$7.2$ million!

Data fabrication Examples

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received $\approx \$19$ million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $\$7.2$ million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [..] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Data fabrication Examples

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received $\approx \$19$ million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $7.2 million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [..] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Brian Wansink Professor, Psychological Nutrition, Cornell, 2016

I gave her a data set of a self-funded, failed study which had null results. I said "This cost us a lot of time and our own money to collect. There's got to be something here we can salvage because it's a cool (rich & unique) data set." I told her what the analyses should be. [..] Every day she came back with puzzling new results, and every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses

17 retracted publications

Data falsification

Table 2. Particle Sizes of IOCMCS/DNA Complexes Prepared in Water

charge ratio (N/P)	IOCMCS20 (nm)	IOCMCS 100 (nm)	IOCMCS200 (nm)
1	382.4 ± 12.1	392.1 ± 16.5	392.8 ± 15.4
10	394.1 ± 19.1	383.2 ± 15.4	200.2 ± 16.5
20	259.5 ± 20.3	306.1 ± 16.8	273.8 ± 17.5
30	349.3 ± 16.1	336.6 ± 21.5	214.7 ± 23.1
50	384.0 ± 14.2	306.5 ± 11.9	207.5 ± 21.9
80	363.4 ± 25.6	197.3 ± 15.6	206.7 ± 18.2
100	221.8 ± 14.8	205.3 ± 19.2	192.2 ± 16.8

Table 3. Particle Sizes of IOCMCS/DNA Complexes Prepared in PBS

charge ratio (N/P)	IOCMCS20 (nm)	IOCMCS 100 (nm)	IOCMCS200 (nm)
1	392.4 ± 19.1	396.1 ± 12.2	383.8 ± 11.1
10	354.1 ± 15.1	393.2 ± 16.1	211.2 ± 15.2
20	289.5 ± 20.3	346.1 ± 17.5	293.8 ± 13.7
30	369.3 ± 17.1	386.6 ± 22.4	234.7 ± 21.3
50	394.0 ± 18.2	381.5 ± 12.8	227.5 ± 12.6
80	383.4 ± 15.6	297.3 ± 13.6	246.7 ± 19.4
100	241.8 ± 16.8	209.3 ± 21.2	195.2 ± 14.5

https://twitter.com/MicrobiomDigest/status/1416838153583927304

Data falsification

Table 2. Particle Sizes of IOCMCS/DNA Complexes Prepared in Water

charge ratio (N/P)	IOCMCS20 (nm)	IOCMCS 100 (nm)	IOCMCS200 (nm)
1	382.4 ± 12.1	392.1 ± 16.5	392.8 ± 15.4
10	394.1 ± 19.1	383.2 ± 15.4	200.2 ± 16.5
20	259.5 ± 20.3	306.1 ± 16.8	273.8 ± 17.5
30	349.3 ± 16.1	336.6 ± 21.5	214.7 ± 23.1
50	384.0 ± 14.2	306.5 ± 11.9	207.5 ± 21.9
80	363.4 ± 25.6	197.3 ± 15.6	206.7 ± 18.2
100	221.8 ± 14.8	205.3 ± 19.2	192.2 ± 16.8

Table 3. Particle Sizes of IOCMCS/DNA Complexes Prepared in PBS

charge ratio (N/P)	IOCMCS20 (nm)	IOCMCS 100 (nm)	IOCMCS200 (nm)
1	392.4 ± 19.1	396.1 ± 12.2	383.8 ± 11.1
10	354.1 ± 15.1	393.2 ± 16.1	211.2 ± 15.2
20	289.5 ± 20.3	346.1 ± 17.5	293.8 ± 13.7
30	369.3 ± 17.1	386.6 ± 22.4	234.7 ± 21.3
50	394.0 ± 18.2	381.5 ± 12.8	227.5 ± 12.6
80	383.4 ± 15.6	297.3 ± 13.6	246.7 ± 19.4
100	241.8 ± 16.8	209.3 ± 21.2	195.2 ± 14.5

https://twitter.com/MicrobiomDigest/status/1416838153583927304

Data falsification

Table V. Analytical results of the serum samples of Virus C patients analyzed by the standard and the developed methods [binuclear $Pt(pyc)_2$].

	[b]Found (percent of label claim ± SD)				
				Proposed method	
Serum samples no.	Reading	[a]Average found X	Standard method Average ± R.S.D. (%)	Student's t and F values	Average recovery ± R.S.D. (%)
Patient (1)	95.3, 95.5, 95.4	95.2	95 ± 1.9	$t = 3.2$, $F = 6.3$	100.2 ± 2.8
Patient (2)	88.7, 89.0, 89.6	89.1	88 ± 1.2	$t = 2.5$, $F = 4.3$	101.3 ± 3.5
Patient (3)	75.2, 75.3, 76.6	75.7	75 ± 1.3	$t = 3.4$, $F = 2.1$	100.9 ± 2.4
Patient (4)	98.0, 96.8, 96.8	97.3	99 ± 1.5	$t = 3.3$, $F = 4.6$	98.28 ± 2.5
Patient (5)	89.9, 88.9, 88.8	89.2	88 ± 1.7	$t = 3.9$, $F = 2.3$	101.4 ± 1.6
Patient (6)	92.8, 91.9, 91.9	92.2	91 ± 2.9	$t = 1.5$, $F = 6.0$	101.3 ± 0.8
Patient (7)	81.9, 81.8, 81.4	81.7	81 ± 1.3	$t = 2.1$, $F = 2.8$	100.9 ± 0.9
Patient (8)	84.5, 84.6, 83.5	84.2	85 ± 2.8	$t = 0.9$, $F = 5.8$	99.05 ± 1.8
Patient (9)	93.9, 93.8, 94.9	94.2	93 ± 1.5	$t = 1.1$, $F = 6.7$	101.3 ± 1.4
Patient (10)	97.8, 96.8, 97.0	97.2	99 ± 2.1	$t = 1.3$, $F = 7.8$	98.18 ± 0.6
Patient (11)	79.7, 78.6, 79.0	79.1	77 ± 1.7	$t = 2.5$, $F = 2.8$	102.7 ± 1.2
Patient (12)	90.8, 90.8, 91.7	91.1	92 ± 2.5	$t = 3.1$, $F = 4.8$	99.02 ± 1.6
Patient (13)	82.5, 82.6, 81.8	82.3	80 ± 0.5	$t = 1.4$, $F = 2.9$	102.9 ± 2.9
Patient (14)	78.9, 78.9, 78.6	78.8	79 ± 0.9	$t = 2.1$, $F = 2.8$	99.74 ± 2.7
Patient (15)	98.6, 98.8, 97.8	98.4	98 ± 1.4	$t = 0.6$, $F = 1.0$	100.4 ± 2.8
Patient (16)	96.7, 96.5, 96.8	96.3	98 ± 2.2	$t = 1.4$, $F = 3.7$	98.26 ± 3.2
Patient (17)	94.8, 94.9, 93.5	94.4	95 ± 1.7	$t = 1.1$, $F = 3.6$	99.36 ± 1.3
Patient (18)	98.7, 98.2, 97.9	98.2	99 ± 1.9	$t = 2.2$, $F = 4.6$	99.19 ± 2.1
Patient (19)	78.8, 78.5, 77.7	78.3	80 ± 2.9	$t = 0.8$, $F = 5.2$	97.87 ± 2.3
Patient (20)	84.8, 84.3, 83.5	84.2	87 ± 2.5	$t = 2.7$, $F = 3.6$	96.78 ± 2.1

Notes: [a]Each reading was repeated three times (X average was taking for three readings by three analysts). [b]Average of three determinations, R.S.D is relative standard deviation $(S/X) * 100$, tabulated t and F values at the 95% confidence level are 4.303 and 19, respectively. $F = s_1^2$ (standard method)/s_2^2 (proposed method) where $s_1 \geq s_2$ (s is standard deviation).

Table VI. Analytical results of the serum samples of Cirrhosis patients analyzed by the standard and the developed methods [binuclear $Pt(pyc)_2$].

	[b]Found (percent of label claim ± SD)				
				Proposed method	
Serum samples no.	Reading	[a]Average found X	Standard method Average ± R.S.D. (%)	Student's t and F values	Average recovery ± R.S.D. (%)
Patient (1)	125.3, 125.5, 125.7	125.5	125 ± 2.9	$t = 2.2$, $F = 4.3$	100.4 ± 1.8
Patient (2)	128.7, 129.0, 129.3	129.0	128 ± 3.2	$t = 3.5$, $F = 5.3$	100.7 ± 1.5
Patient (3)	135.2, 135.3, 136.0	135.5	135 ± 2.3	$t = 3.9$, $F = 5.1$	100.3 ± 2.2
Patient (4)	127.4, 126.8, 96.8	127.0	99 ± 1.7	$t = 3.1$, $F = 3.6$	97.97 ± 2.9
Patient (5)	119.5, 118.7, 118.8	119.0	88 ± 1.7	$t = 3.2$, $F = 6.3$	101.1 ± 1.3
Patient (6)	1122.7, 111.5, 111.8	112.0	91 ± 3.9	$t = 2.5$, $F = 5.0$	101.1 ± 1.8
Patient (7)	141.9, 141.2, 141.4	141.5	101 ± 1.8	$t = 4.1$, $F = 3.8$	100.5 ± 1.9
Patient (8)	144.5, 144.0, 143.5	144.0	155 ± 2.8	$t = 1.9$, $F = 4.8$	99.35 ± 1.9
Patient (9)	133.7, 133.8, 134.5	134.0	93 ± 1.9	$t = 3.1$, $F = 2.7$	101.1 ± 1.7
Patient (10)	147.5, 146.5, 147.0	147.0	109 ± 2.5	$t = 3.3$, $F = 3.8$	98.16 ± 2.6
Patient (11)	129.6, 128.4, 129.0	129.0	107 ± 1.6	$t = 3.5$, $F = 2.8$	101.9 ± 1.7
Patient (12)	140.8, 140.5, 141.7	141.0	92 ± 1.5	$t = 3.1$, $F = 1.8$	98.91 ± 2.6
Patient (13)	142.5, 142.0, 141.5	142.0	110 ± 2.5	$t = 2.4$, $F = 2.9$	101.8 ± 2.3
Patient (14)	118.9, 118.1, 118.5	118.5	119 ± 0.9	$t = 4.1$, $F = 3.8$	99.57 ± 1.7
Patient (15)	138.5, 138.6, 137.8	138.3	98 ± 1.4	$t = 3.6$, $F = 4.0$	100.3 ± 1.8
Patient (16)	116.3, 116.5, 96.3	116.0	98 ± 2.7	$t = 3.4$, $F = 3.7$	97.95 ± 3.7
Patient (17)	114.2, 114.3, 113.5	114.0	115 ± 1.5	$t = 2.1$, $F = 0.6$	99.14 ± 1.7
Patient (18)	128.7, 128.2, 127.0	127.9	129 ± 1.1	$t = 2.9$, $F = 2.6$	99.14 ± 1.1
Patient (19)	138.3, 138.5, 137.3	138.0	110 ± 2.7	$t = 3.8$, $F = 3.2$	98.18 ± 2.7
Patient (20)	134.2, 134.3, 133.5	134.0	87 ± 3.5	$t = 4.7$, $F = 5.6$	96.55 ± 2.7

Notes: [a]Each reading was repeated three times (X average was taking for three readings by three analysts). [b]Average of three determinations, R.S.D is relative standard deviation $(S/X) * 100$, tabulated t and F values at the 95% confidence level are 4.303 and 19, respectively. ($F = s_1^2$ (standard method)/s_2^2 (proposed method) where $s_1 \geq s_2$ (s is standard deviation).

https://twitter.com/MicrobiomDigest/status/1425477195141636105

Data falsification

Table V. Analytical results of the serum samples of Virus C patients analyzed by the standard and the developed methods [binuclear $Pt(pyc)_2$].

	[b]Found (percent of label claim ± SD)				
				Proposed method	
Serum samples no.	Reading	[a]Average found X	Standard method Average ± R.S.D. (%)	Student's t and F values	Average recovery ± R.S.D. (%)
Patient (1)	95.3, 95.5, 95.4	95.2	95 ± 1.9	$t = 3.2$, $F = 6.3$	100.2 ± 2.8
Patient (2)	88.7, 89.0, 89.6	89.1	88 ± 1.2	$t = 2.5$, $F = 4.3$	101.3 ± 3.5
Patient (3)	75.2, 75.3, 76.6	75.7	75 ± 1.3	$t = 3.4$, $F = 2.1$	100.9 ± 2.4
Patient (4)	98.0, 96.8, 96.8	97.3	99 ± 1.5	$t = 3.3$, $F = 4.6$	98.28 ± 2.5
Patient (5)	89.9, 88.9, 88.8	89.2	88 ± 1.7	$t = 3.9$, $F = 2.3$	101.4 ± 1.6
Patient (6)	92.8, 91.9, 91.9	92.2	91 ± 2.9	$t = 1.5$, $F = 6.0$	101.3 ± 0.8
Patient (7)	81.9, 81.8, 81.4	81.7	81 ± 1.3	$t = 2.1$, $F = 2.8$	100.9 ± 0.9
Patient (8)	84.5, 84.6, 83.5	84.2	85 ± 2.8	$t = 0.9$, $F = 5.8$	99.05 ± 1.8
Patient (9)	93.9, 93.8, 94.9	94.2	93 ± 1.5	$t = 1.1$, $F = 6.7$	101.3 ± 1.4
Patient (10)	97.8, 96.8, 97.0	97.2	99 ± 2.1	$t = 1.3$, $F = 7.8$	98.18 ± 0.6
Patient (11)	79.7, 78.6, 79.0	79.1	77 ± 1.7	$t = 2.5$, $F = 2.8$	102.7 ± 1.2
Patient (12)	90.8, 90.8, 91.7	91.1	92 ± 2.5	$t = 3.1$, $F = 4.8$	99.02 ± 1.6
Patient (13)	82.5, 82.6, 81.8	82.3	80 ± 0.5	$t = 1.4$, $F = 2.9$	102.9 ± 2.9
Patient (14)	78.9, 78.9, 78.6	78.8	79 ± 0.9	$t = 2.1$, $F = 2.8$	99.74 ± 2.7
Patient (15)	98.6, 98.8, 97.8	98.4	98 ± 1.4	$t = 0.6$, $F = 1.0$	100.4 ± 2.8
Patient (16)	96.7, 96.5, 96.8	96.3	98 ± 2.2	$t = 1.4$, $F = 3.7$	98.26 ± 3.2
Patient (17)	94.8, 94.9, 93.5	94.4	95 ± 1.7	$t = 1.1$, $F = 3.6$	99.36 ± 1.3
Patient (18)	98.7, 98.2, 97.9	98.2	99 ± 1.9	$t = 2.2$, $F = 4.6$	99.19 ± 2.1
Patient (19)	78.8, 78.5, 77.7	78.3	80 ± 2.9	$t = 0.8$, $F = 5.2$	97.87 ± 2.3
Patient (20)	84.8, 84.3, 83.5	84.2	87 ± 2.5	$t = 2.7$, $F = 3.6$	96.78 ± 2.1

Notes: [a]Each reading was repeated three times (X average was taking for three readings by three analysts). [b]Average of three determinations, R.S.D is relative standard deviation $(S/X) * 100$, tabulated t and F values at the 95% confidence level are 4.303 and 19, respectively. $F = s_1^2$ (standard method)/s_2^2 (proposed method) where $s_1 \geq s_2$ (s is standard deviation).

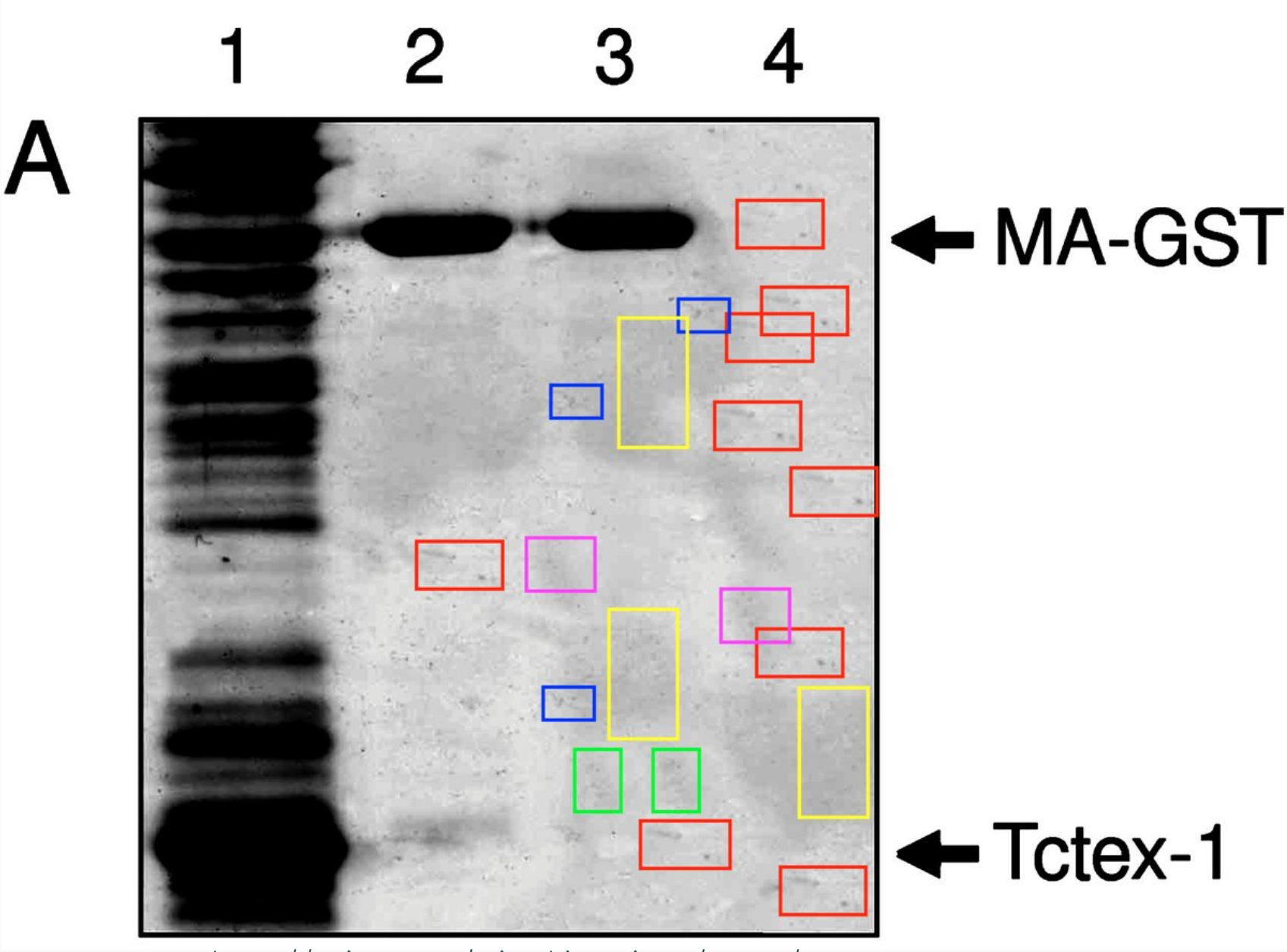
Table VI. Analytical results of the serum samples of Cirrhosis patients analyzed by the standard and the developed methods [binuclear $Pt(pyc)_2$].

	[b]Found (percent of label claim ± SD)				
				Proposed method	
Serum samples no.	Reading	[a]Average found X	Standard method Average ± R.S.D. (%)	Student's t and F values	Average recovery ± R.S.D. (%)
Patient (1)	125.3, 125.5, 125.7	125.5	125 ± 2.9	$t = 2.2$, $F = 4.3$	100.4 ± 1.8
Patient (2)	128.7, 129.0, 129.3	129.0	128 ± 3.2	$t = 3.5$, $F = 5.3$	100.7 ± 1.5
Patient (3)	135.2, 135.3, 136.0	135.5	135 ± 2.3	$t = 3.9$, $F = 5.1$	100.3 ± 2.2
Patient (4)	127.4, 126.8, 96.8	127.0	99 ± 1.7	$t = 3.1$, $F = 3.6$	97.97 ± 2.9
Patient (5)	119.5, 118.7, 118.8	119.0	88 ± 1.7	$t = 3.2$, $F = 6.3$	101.1 ± 1.3
Patient (6)	112.7, 111.5, 111.8	112.0	91 ± 3.9	$t = 2.5$, $F = 5.0$	101.1 ± 1.8
Patient (7)	141.9, 141.2, 141.4	141.5	101 ± 1.8	$t = 4.1$, $F = 3.8$	100.5 ± 1.9
Patient (8)	144.5, 144.0, 143.5	144.0	155 ± 2.8	$t = 1.9$, $F = 4.8$	99.35 ± 1.9
Patient (9)	133.7, 133.8, 134.5	134.0	93 ± 1.9	$t = 3.1$, $F = 2.7$	101.1 ± 1.7
Patient (10)	147.5, 146.5, 147.0	147.0	109 ± 2.5	$t = 3.3$, $F = 3.8$	98.16 ± 2.6
Patient (11)	129.6, 128.4, 129.0	129.0	107 ± 1.6	$t = 3.5$, $F = 2.8$	101.9 ± 1.7
Patient (12)	140.8, 140.5, 141.7	141.0	92 ± 1.5	$t = 3.1$, $F = 1.8$	98.91 ± 2.6
Patient (13)	142.5, 142.0, 141.5	142.0	110 ± 2.5	$t = 2.4$, $F = 2.9$	101.8 ± 2.3
Patient (14)	118.9, 118.1, 118.5	118.5	119 ± 0.9	$t = 4.1$, $F = 3.8$	99.57 ± 1.7
Patient (15)	138.5, 138.6, 137.8	138.3	98 ± 1.4	$t = 3.6$, $F = 4.0$	100.3 ± 1.8
Patient (16)	116.3, 116.5, 96.3	116.0	98 ± 2.7	$t = 3.4$, $F = 3.7$	97.95 ± 3.7
Patient (17)	114.2, 114.3, 115.5	114.0	115 ± 1.5	$t = 2.1$, $F = 0.6$	99.14 ± 1.7
Patient (18)	128.7, 128.2, 127.0	127.9	129 ± 1.1	$t = 2.9$, $F = 2.6$	99.14 ± 1.1
Patient (19)	138.3, 138.5, 137.3	138.0	110 ± 2.7	$t = 3.8$, $F = 3.2$	98.18 ± 2.7
Patient (20)	134.2, 134.3, 133.5	134.0	87 ± 3.5	$t = 4.7$, $F = 5.6$	96.55 ± 2.7

Notes: [a]Each reading was repeated three times (X average was taking for three readings by three analysts). [b]Average of three determinations, R.S.D is relative standard deviation $(S/X) * 100$, tabulated t and F values at the 95% confidence level are 4.303 and 19, respectively. ($F = s_1^2$ (standard method)/s_2^2 (proposed method) where $s_1 \geq s_2$ (s is standard deviation).

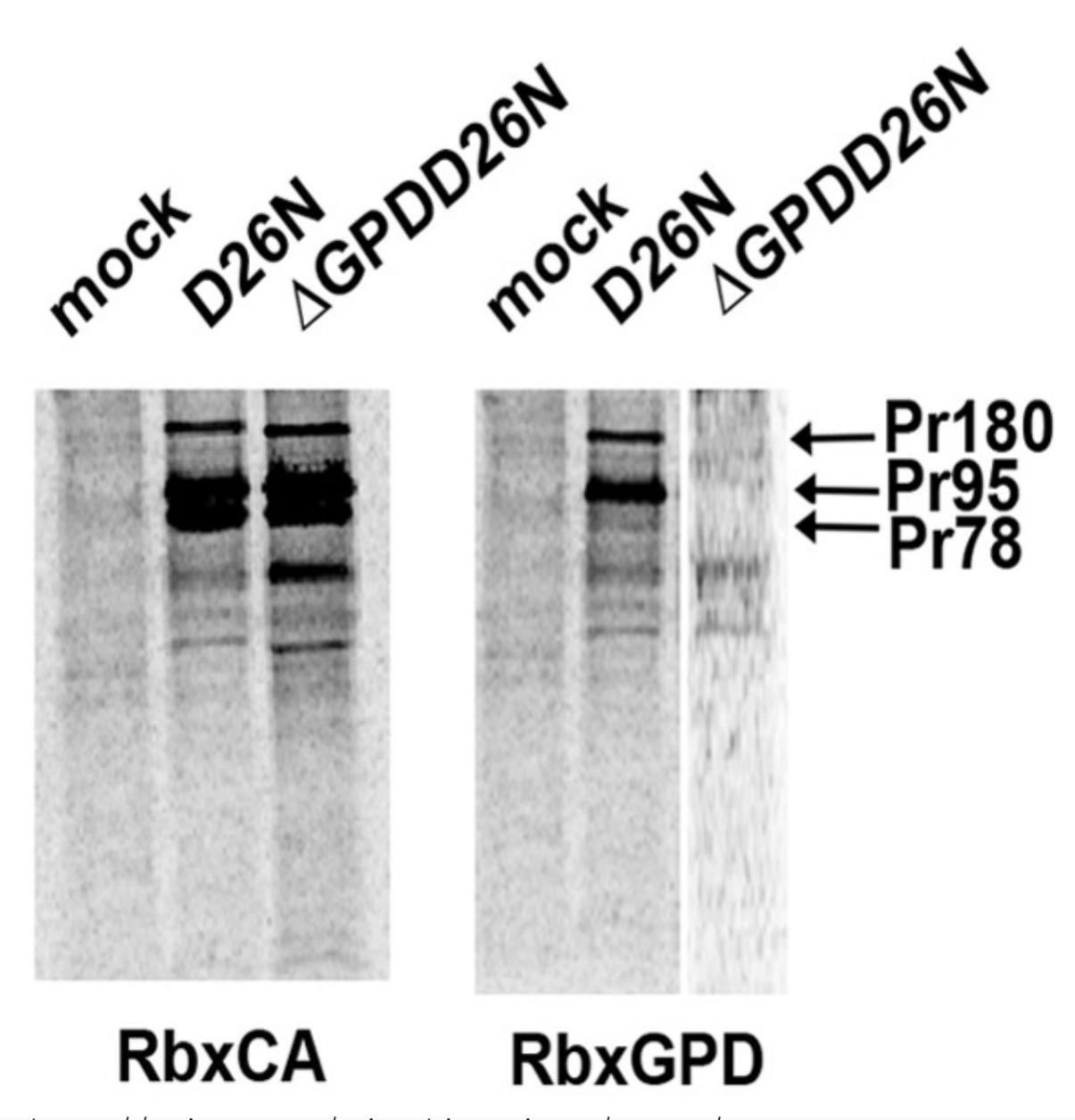
https://twitter.com/MicrobiomDigest/status/1425477195141636105

Photo Manipulation



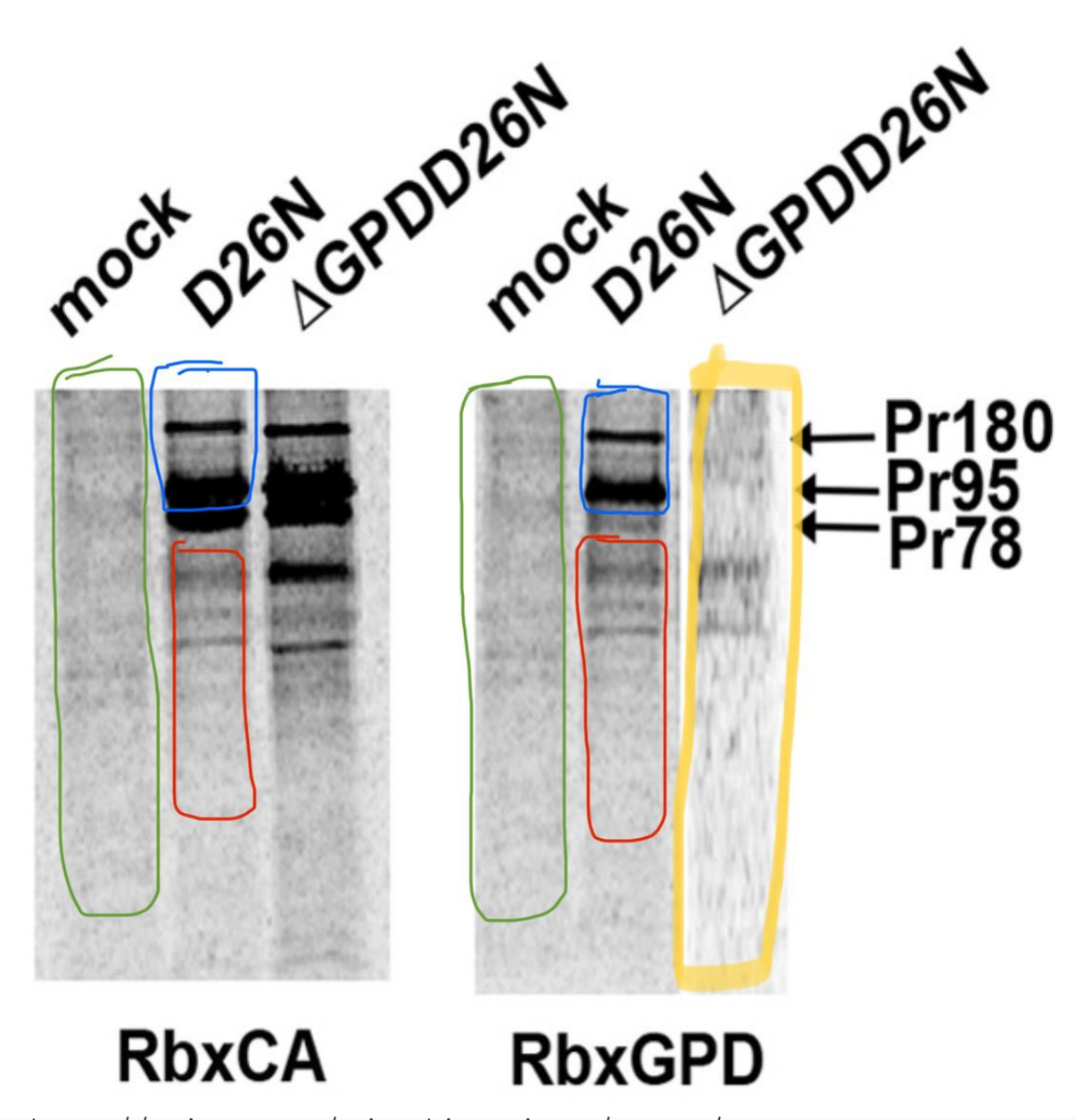
https://twitter.com/MicrobiomDigest/status/1476307790494781441

Photo Manipulation



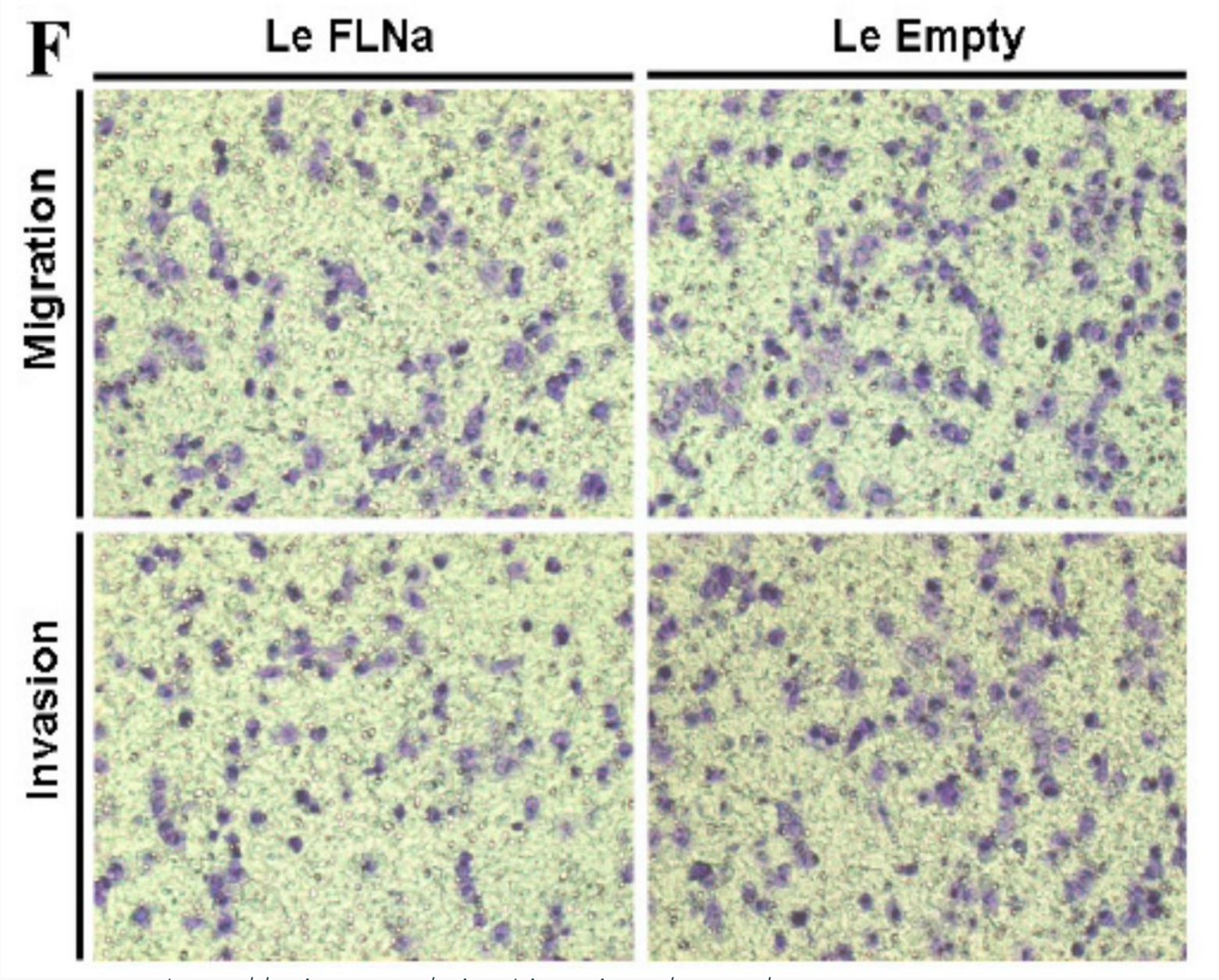
https://twitter.com/MicrobiomDigest/status/1476091603563417603

Photo Manipulation



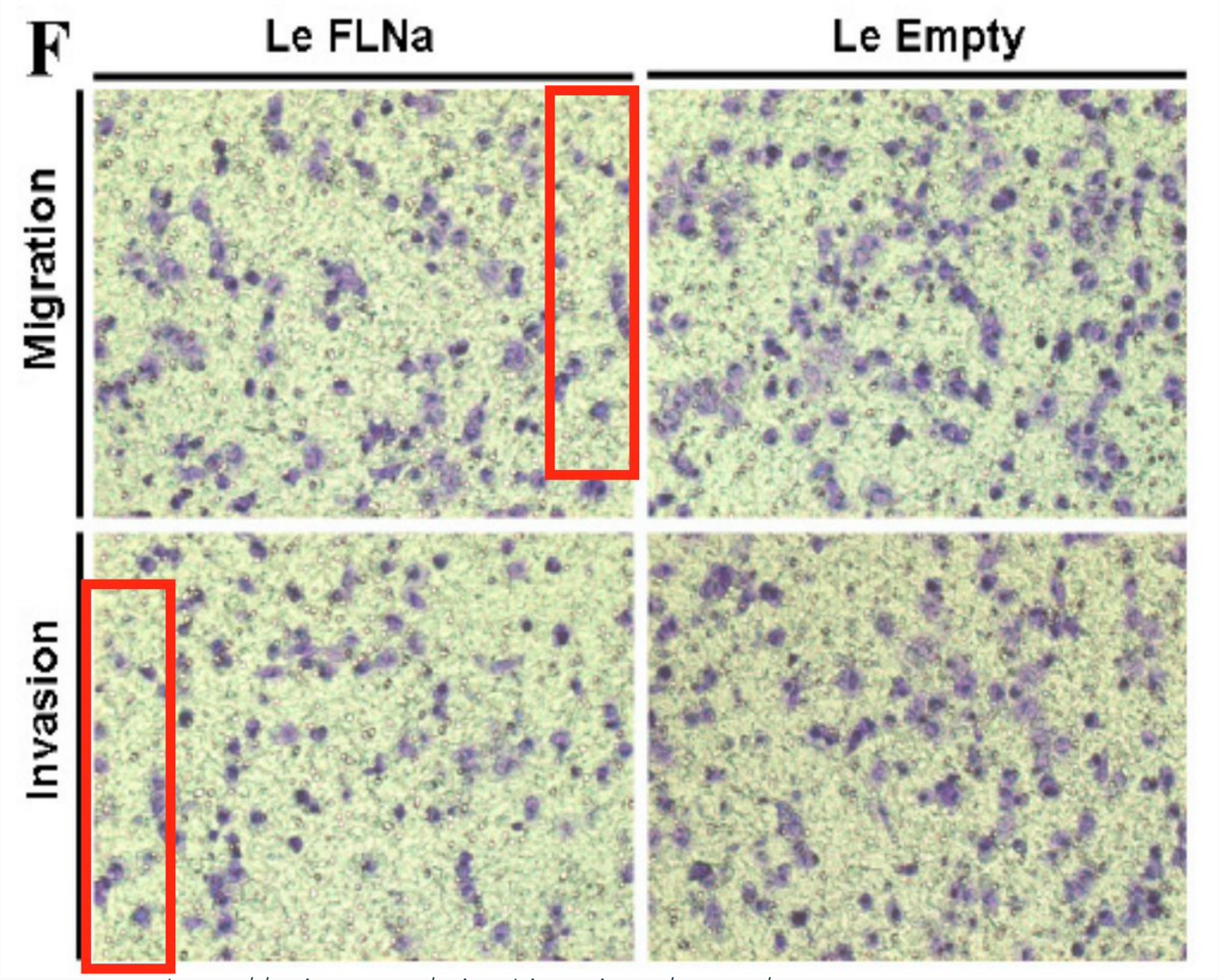
https://twitter.com/MicrobiomDigest/status/1476091603563417603

Photo Manipulation



https://twitter.com/MicrobiomDigest/status/1475898917250207746

Photo Manipulation



https://twitter.com/MicrobiomDigest/status/1475898917250207746

Photo Manipulation

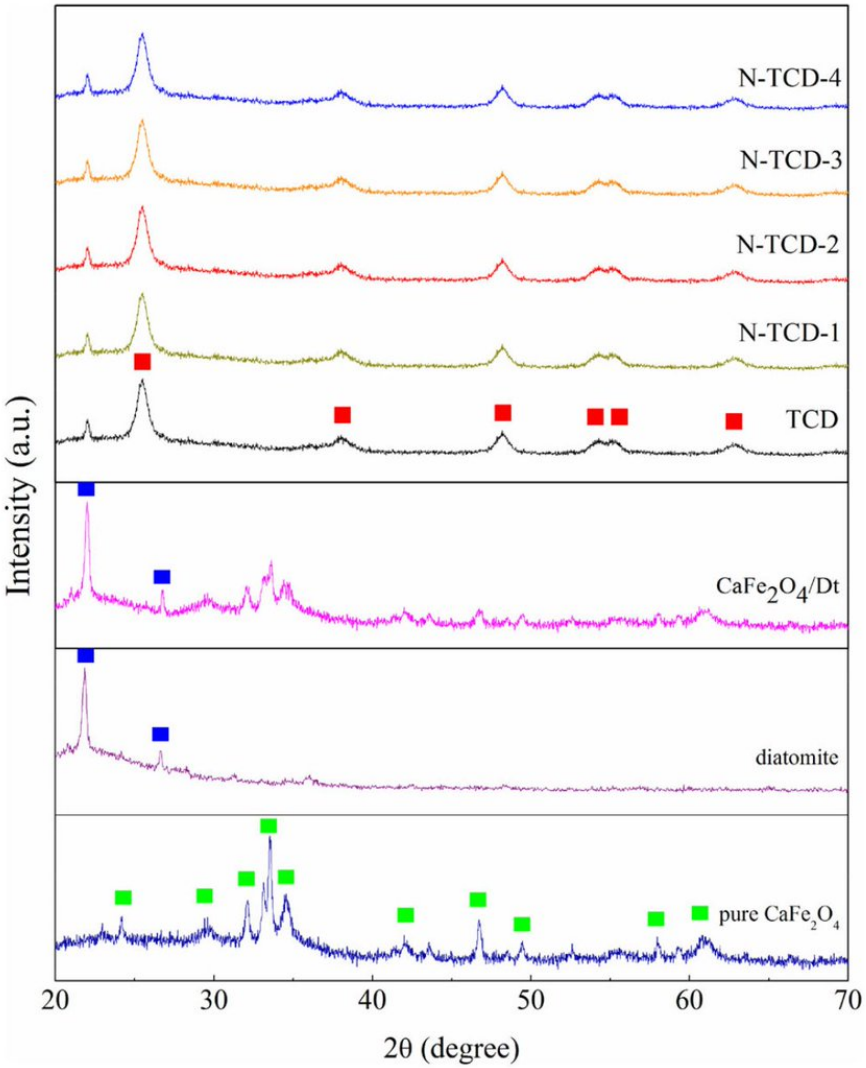
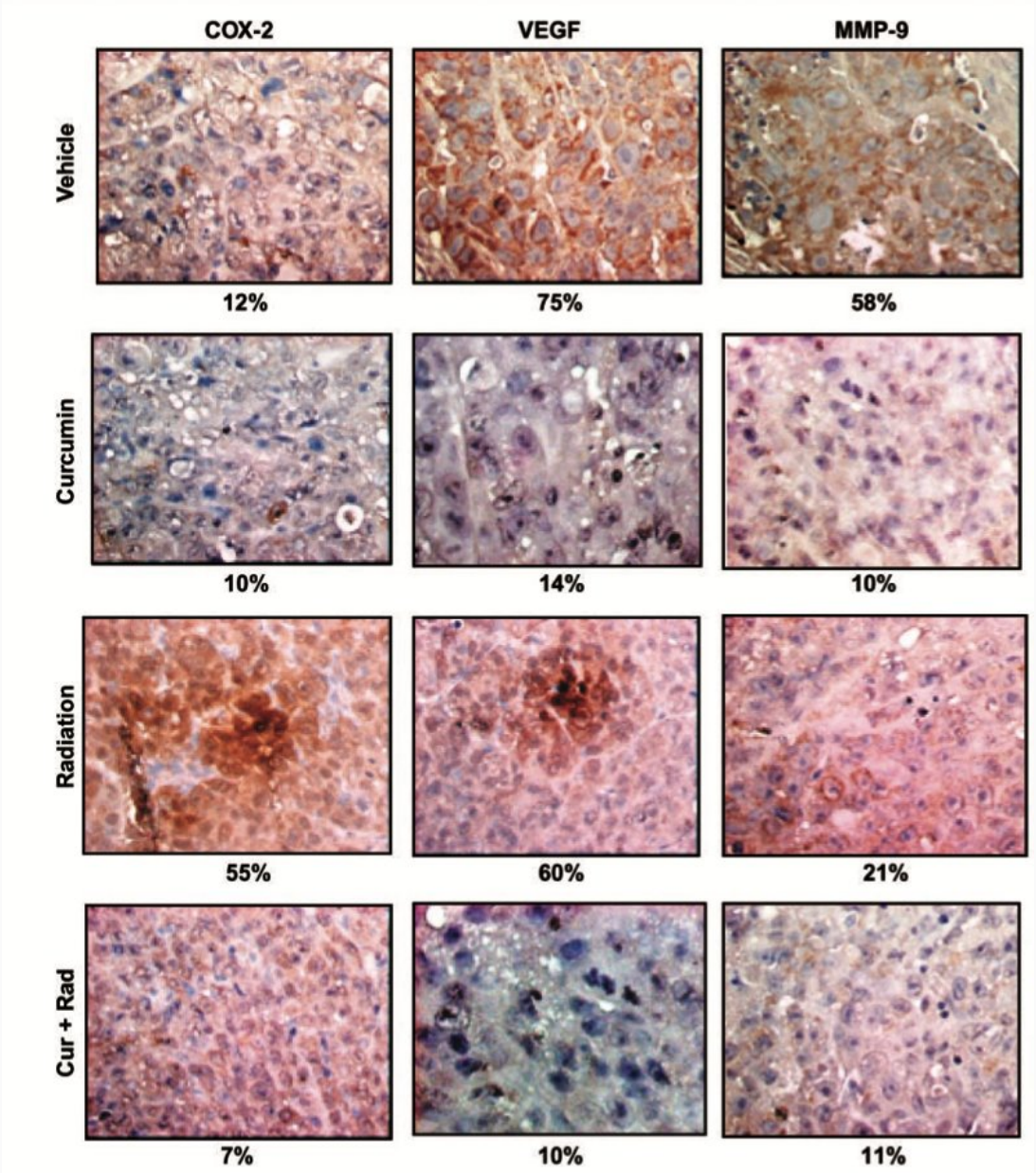


Figure 2. XRD patterns of pure $CaFe_2O_4$, Dt, $CaFe_2O_4$/Dt, undoped TCD and various N-TCD (green squares, peaks of spinel $CaFe_2O_4$; red squares, peaks of anatase TiO_2; blue squares, peaks of Dt).

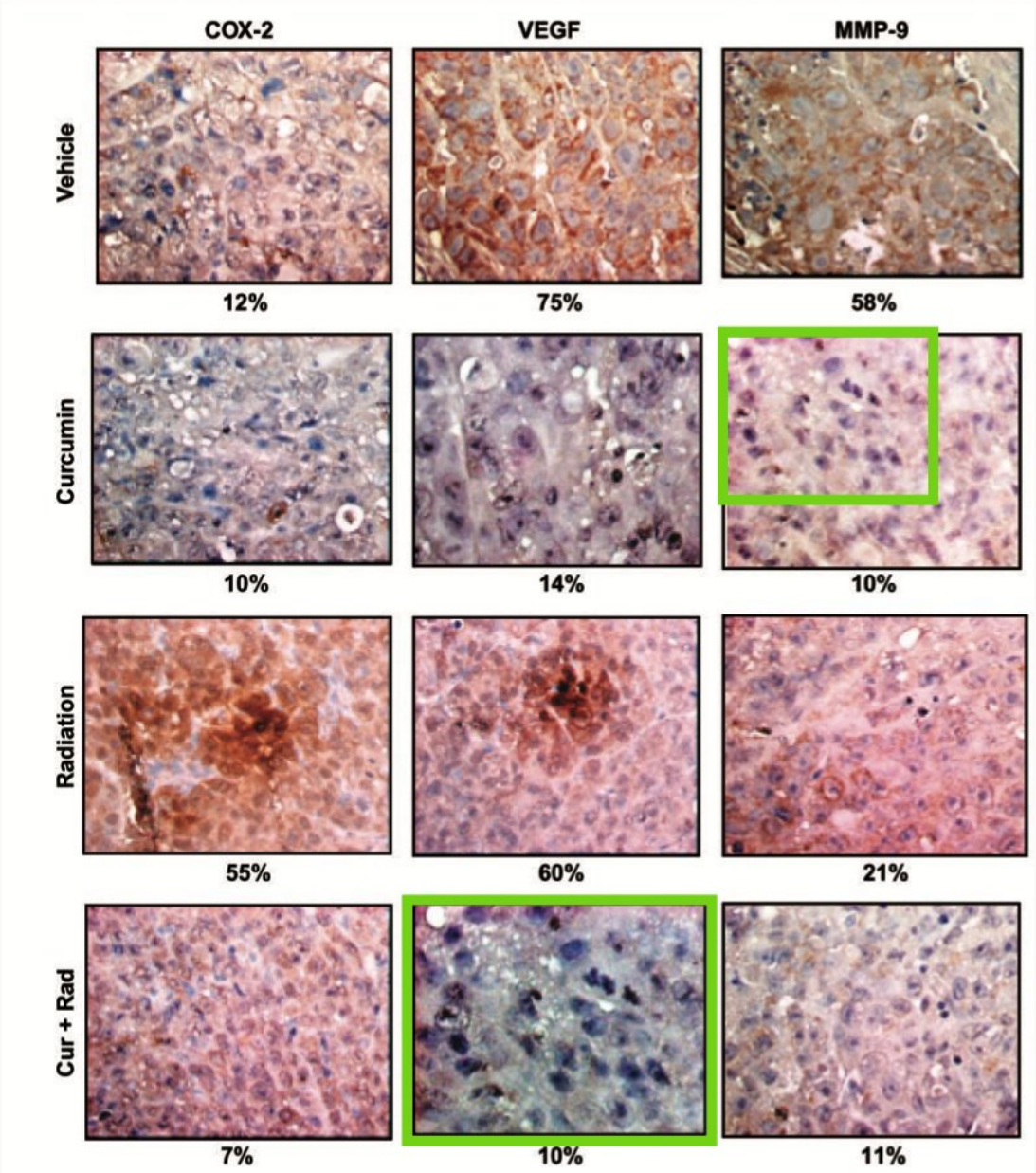
https://twitter.com/MicrobiomDigest/status/1475649414190362625

Photo Manipulation



https://twitter.com/MicrobiomDigest/status/1480660402874638337

Photo Manipulation



https://twitter.com/MicrobiomDigest/status/1480660402874638337

Photo Manipulation

Photoshoping used to be a common/requested practice

Going further

HARKING Hypothesizing After the Results are Known (Kerr, 98)
presenting a post hoc hypothesis in the introduction of a research report as if it were an a priori hypothesis

P-hacking A manifesto for reproducible science

- spurious correlations/multiple comparisons
- systematic bias toward shiny results
  - overhacking (hack data until < 0.05)
  - selection bias (only report the smallest one)
  - selective debugging (change test, look for bugs until < 0.05)
  - publication bias
  - (Bonferroni or other) corrections is super hard to apply

Cherry picking and not reporting failed attempts

Possible Answers

- $p < 0.0001$? No more p values ? Freakonometrics
- Academic journals increasingly shift to pre-registration mechanism on platforms like OSF.io
- Meta-analysis may help reaching a consensus ?

Any estimate of the damage ?

E. Bik: Hard to tell but some papers have made a lot of damage.

Measles Mumps Rubella vaccine causes autism in young children

- Wakefield et al. *Ileal-lymphoid-nodular hyperplasia, nonspecific colitis, and pervasive developmental disorder in children.* Lancet. 1998. [RETRACTED]
- "*Despite the small sample size (n=12), the uncontrolled design, and the speculative nature of the conclusions, the paper received wide publicity, and MMR vaccination rates began to drop because parents were concerned about the risk of autism after vaccination.*" The MMR vaccine and autism: Sensation refutation, retraction, and fraud (2011)
- 12 years of investigation: Falsified data

People don't believe in vaccine and Wakefield contributed to that.

- HCQ paper from Didier Raoult has done a lots of damage as well. Still not retracted.

Collective Response

E. Bik: Often many fraudulent papers from the same person. Sad because often little action/response from the journals.

PubPeer post-publication peer review and whistleblowing platform

- Brandon Stell and Boris Barbour (CNRS) in 2012
- Allows anonymous post-publication; moderated
- Position of the CNRS CEO, Antoine Petit:

Utiliser un site anonyme pour faire part de ses doutes est une démarche dont j'ai du mal à comprendre le sens scientifique. Chacune et chacun est libre de consulter de tels sites, ou même d'y contribuer. Cela relève de la vie privée. Mais à ce titre, cela ne peut se faire sur le temps de travail ou en mobilisant des moyens appartenant au CNRS. Et une telle démarche ne sera évidemment jamais cautionnée par le CNRS

Retractration watch reports on retractions of scientific papers

- As of January 2020, 21 792 items
- Separate list of retracted articles that add to misinformation about the pandemic

Scientific Integrity (Plagiarism)

Editors requirements

Last time I submitted an article to ACM TOMACS, I had to testify my paper to:

- be the authors' own original work, which has not been previously published elsewhere
- not be submitted to more than one journal for consideration (ensuring it is not under redundant simultaneous peer review)
- properly credit the meaningful contributions of co-authors and co-researchers,
  - be appropriately placed in the context of prior and existing research,
  - reflect the authors' own research and analysis and do so in a truthful and complete manner.

Is this really plagiarism ?

The new means of communications have broken down the barriers between people and knowledge.

We can access a quasi unlimited amount of information. An utter most important skill is the ability to filter and aggregate information.

Is this really plagiarism ?

The new means of communications have broken down the barriers between people and knowledge.

We can access a quasi unlimited amount of information. An utter most important skill is the ability to filter and aggregate information.

- *I've read wikipedia and the description of ... is so good that there is no way I can write anything better*
  Why not use it ? After all, wikipedia is public domain...
- *I've read an internal report of a PhD student of the team and his introduction is just what I need*
- *The figure of this other report is just great so I've used it*
- *Deadlines came around more quickly than expected, I had to produce something*
- *I used ChatGPT for my introduction so it's an original work*

People often use terms like "copying" and "borrowing", which disguises the seriousness of the offense.

What is plagiarism?

According to the Merriam-Webster Online Dictionary, to ”plagiarize” means:

- to steal and pass off (the ideas or words of another) as one’s own
- to use (another’s production) without crediting the source
- to commit literary theft
- to present as new and original an idea or product derived from an existing source.

What is plagiarism?

According to the Merriam-Webster Online Dictionary, to ”plagiarize” means:

- to steal and pass off (the ideas or words of another) as one’s own
- to use (another’s production) without crediting the source
- to commit literary theft
- to present as new and original an idea or product derived from an existing source.

In other words, plagiarism is an act of fraud. It involves both stealing someone else’s work and lying about it afterward.

But can words and ideas really be stolen?

According to many governmental laws, the answer is yes. All of the following are considered plagiarism:

- turning in someone else's work as your own
- copying words or ideas from someone else without giving credit
- failing to put a quotation in quotation marks
- giving incorrect information about the source of a quotation
- changing words but copying the sentence structure of a source without giving credit
- copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not

But can words and ideas really be stolen?

According to many governmental laws, the answer is yes. All of the following are considered plagiarism:

- turning in someone else's work as your own
- copying words or ideas from someone else without giving credit
- failing to put a quotation in quotation marks
- giving incorrect information about the source of a quotation
- changing words but copying the sentence structure of a source without giving credit
- copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not

The fraud is generally more about lying about the content of your work than about dispossessing someone from something.

How to avoid plagiarism?

Most cases of plagiarism can be avoided, however, by citing sources.

Simply acknowledging that certain material has been borrowed, and providing your audience with the information necessary to find that source, is usually enough to prevent plagiarism.

How to avoid plagiarism?

Most cases of plagiarism can be avoided, however, by citing sources.

Simply acknowledging that certain material has been borrowed, and providing your audience with the information necessary to find that source, is usually enough to prevent plagiarism.

That is why I need to confess now that most of previous slides are plain copies from http://www.plagiarism.org/, just like the next slides. 😉

Citations

A ”citation” is the way you tell your readers that certain material in your work came from another source.

It also gives your readers the information necessary to find that source again, including:

- information about the author
- the title of the work
- the name and location of the company that published your copy of the source
- the date your copy was published
- the page numbers of the material you are borrowing

URLs are convenient but not perenial

Why should I cite sources?

Giving credit to the original author by citing sources is the only way to use other people's work without plagiarizing. But there are a number of other reasons to cite sources:

- Citations are extremely helpful to anyone who wants to find out more about your ideas and where they came from
- Not all sources are good or right – your own ideas may often be more accurate or interesting than those of your sources. Proper citation will keep you from taking the rap for someone else's bad ideas
- Citing sources shows the amount of research you've done
- Citing sources strengthens your work by lending outside support to your ideas

Doesn't citing sources make my work seem less original?

Not at all. On the contrary, citing sources actually helps your reader distinguish your ideas from those of your sources. This will actually emphasize the originality of your own work.

But do not cite too much things. Only cite the work you have actually read!

Plagiarism variants

Self-plagiarism

- Duplicate publication
- Text recycling and Copyright infringement
- Salami-slicing publication (”least publishable units”)

Citation plagiarism *willful or negligent failure to appropriately credit other or prior discoverers, so as to give an improper impression of priority* aka citation amnesia, disregard syndrome, and bibliographic negligence

– From https://en.wikipedia.org/wiki/Scientific_misconduct

Authorship, guest authorship, and ghostwriting

Authorship responsibility

> *All authors of a scientific publication are expected to have made reasonable attempts to check findings submitted to academic journals for publication.*

Ghostwriting e.g., to hide a conflict of interest

- Invisibilization and Silencing (engineers/technicians, young researchers, ...). Yet, quite important for carreer.
- What is a substential contribution ?

Guest/gift authorship to increase publication list

- You may actually buy authorship for $\approx$ 10K€

Possible answer CRediT – Contributor Roles Taxonomy

How far can this go ?

Article generation SCIgen

- Prevalence of nonsensical algorithmically generated papers in the scientific literature
  - *We estimate the prevalence of SCIgen-papers to be 75 per million papers in Information and Computing Sciences*
  - *Only 19% of the 243 problematic papers were dealt with: formal retraction (12) or silent removal (34)*

Detecting

- Errors (semi-automated fact-checking of nucleotide sequence reagents)
- Fraud

  (the possibility of systematic research fraud targeting under-studied human genes)

Papermills The full-service paper mill and its Chinese customers

"*Predatory*" journals Predatory journals: no definition, no defense

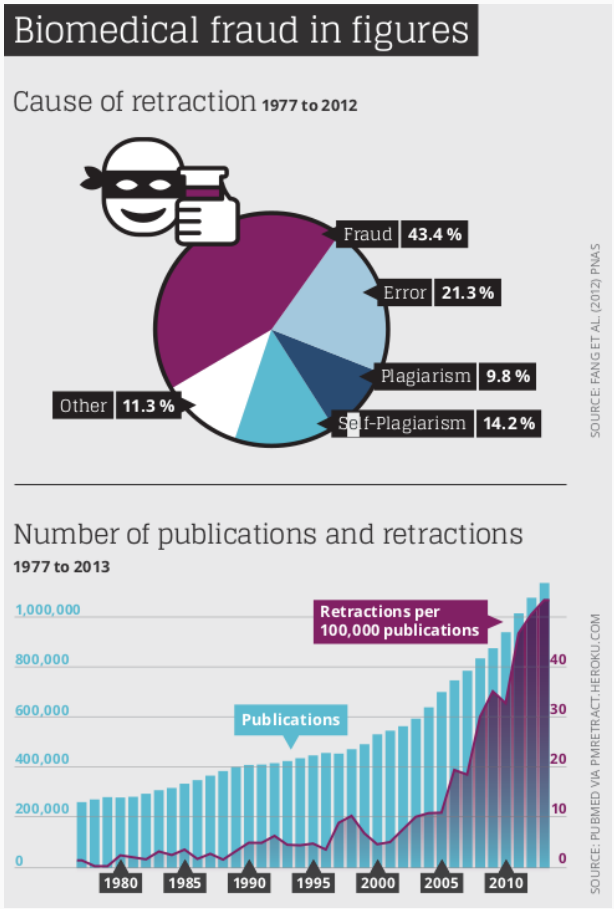
- Just pay for fake reviews and editorial work

Discussion on Research Organization

Is this new ?

How so? Why now? Why is this important? What can we do about it?

The Battle against Scientific Fraud in the CNRS International Magazine






Galileo (data fabrication), **Ptolemy** (plagiarism), **Mendel** (data enhancement), **Pasteur** (rigorous but hid failures), ...

Scientific misconduct is obviously wrong but it's not new!

- Every domain has its black sheep
- The publish or perish pressure is a pain

Motivation for cheating

Career pressure Funding depends on good reputation, which depends on the publication of high-profile scientific papers $\rightsquigarrow$ Publish or perish

Ease of fabrication Results are often difficult to reproduce accurately, being obscured by noise, artifacts, ... Checking is hard

- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)

Monetary Gain The most lucrative options for professionals are often selling opinions

- Academics are badly paid so a small infringement to the deontology code could help...

Power structure *solve this or I'll find another postdoc to do the job*

- High pressure to deliver and incentive to cheat
- Experiments do not always work, cells do not always grow, sometimes it works, sometimes it does not, sometimes it does not work as expected

Quantity matters more than quality, especially when you have to write papers to get a position.

People are pushed to deliver the impossible

Bibliometry

Science builds on the ability to measure, right ?

- If we could measure the quality of science (scientist, group, lab, universities), we would have a rational way of managing budget

h-index captures both productivity and impact (Google Scholar)

- Do you know, Ike Antkare one of the great stars in the scientific firmament ?
- Game the system: encourages self-citation, reviewer-coerced citation, which impoverishes scientific communities

Conference/journal ranking http://www.conferenceranks.com/ (CORE, ERA, QUALIS)

Shanghai Ranking Academic Ranking of World Universities

This is prescriptive measurement: e.g., "*one B2/A2 paper every year or a A1 paper every two year at least for an academic*"

Hiring committees and research institutes massively sign the San Francisco Declaration on Research Assessment (DORA, 2012)

Too much noise is also deleterious

Release early, release often

- Goodhart's Law: Are Academic Metrics Being Gamed? (Fire 2019)
  - AI: over 1,000 ranked journals ($\times 10$ in 15 years)
  - Shorter papers with increasing self references
  - More and more papers without any citation
  - Sharp increase of the number of new young authors publishing at a much faster rate
- Slowed canonical progress in large fields of science

  > *When the number of papers published per year in a scientific field grows large, citations flow disproportionately to already well-cited papers; the list of most-cited papers ossifies; new papers are unlikely to ever become highly cited, and when they do, it is not through a gradual, cumulative process of attention gathering; and newly published papers become unlikely to disrupt existing work*

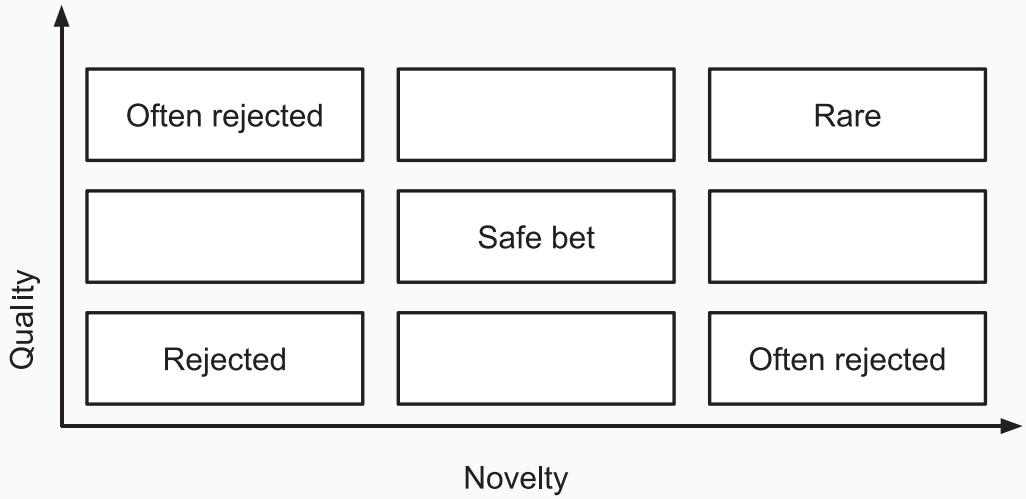
- Moshe Vardi recently recalled Shannon's 1956 one-page article trying to curb the information theory hype

  > *Authors should submit only their best efforts, and these only after careful criticism by themselves and their colleagues. A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are no credit to their writers and a waste of time to their readers.*

How serious is reviewing taken really depends on the community

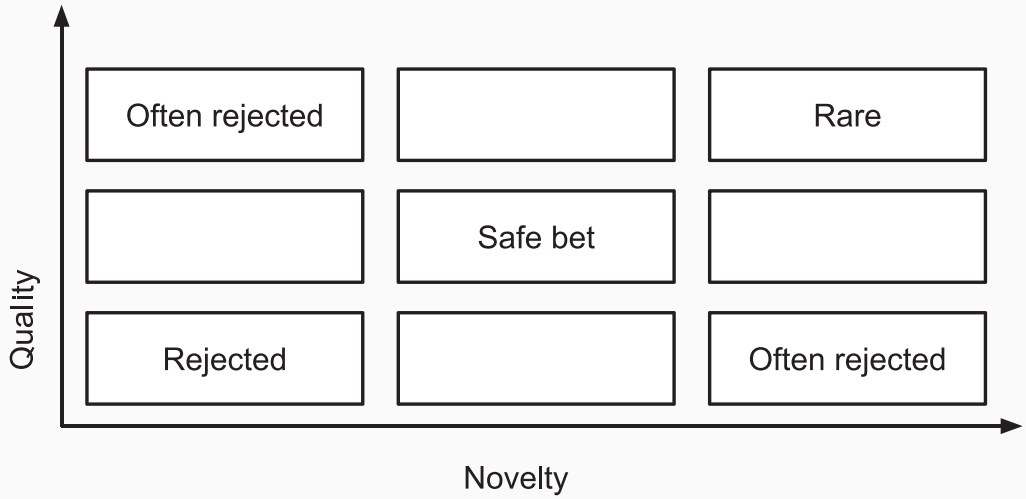
New publication/reviewing modes ?

The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic Guide to Assessing Empirical Evaluations, *TOPLAS* 2016



New publication/reviewing modes ?

The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic Guide to Assessing Empirical Evaluations, *TOPLAS* 2016



Open science data papers, negative results, small pieces with https://science-octopus.org/, reproductions, open reviews with f1000research, software papers with JOSS, ...

New sharing modes but beware of performative reproducibility

- No one has the time or patience to read a 100-page pre-analysis plan and compare it with the later publication
- Need for robust, sustainable change in underlying cultural values (knowledge reliability, reward, hiring process, etc.)

Deleterious impact of liberal and bureaucratic research management

In the past: steering of research done by Research institutes

2003 ANR (call for proposal, less recurrent budget for labs and institutes)

2007 ERC grants (peer-reviewed excellence)

2013 HCERES (evaluation of ~~"academics"~~, teams, labs, institutes in France)

2021 LPR, ~~CNU~~

2022 PEPR (massive funding on strategic topics)

Short-sighted madness that does not leave time to the academia

On the time spent preparing grant proposals: an observational study of Australian researchers