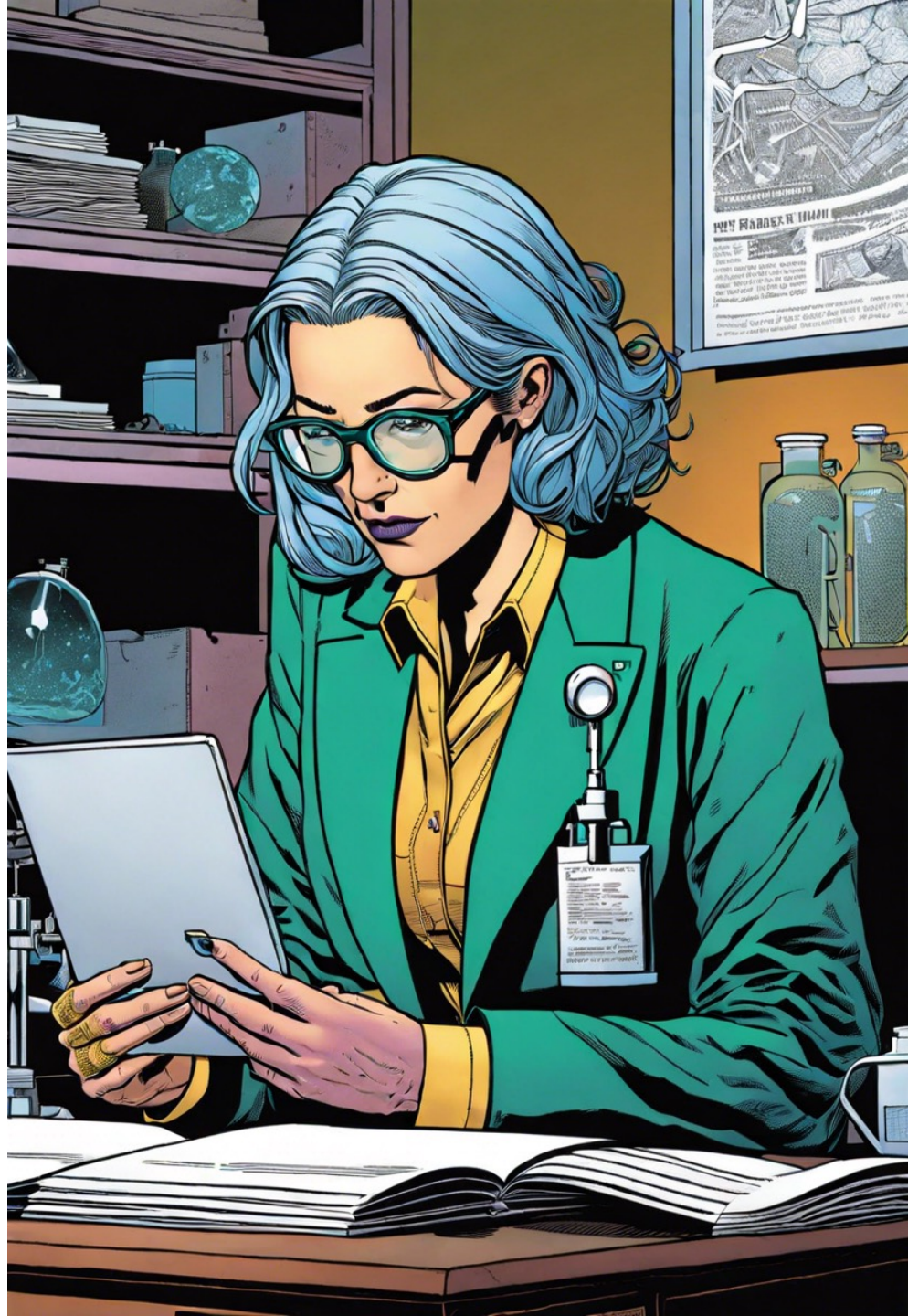
USC | Office of Research Integrity

# Avoiding Research Misconduct Allegations Against Your Lab

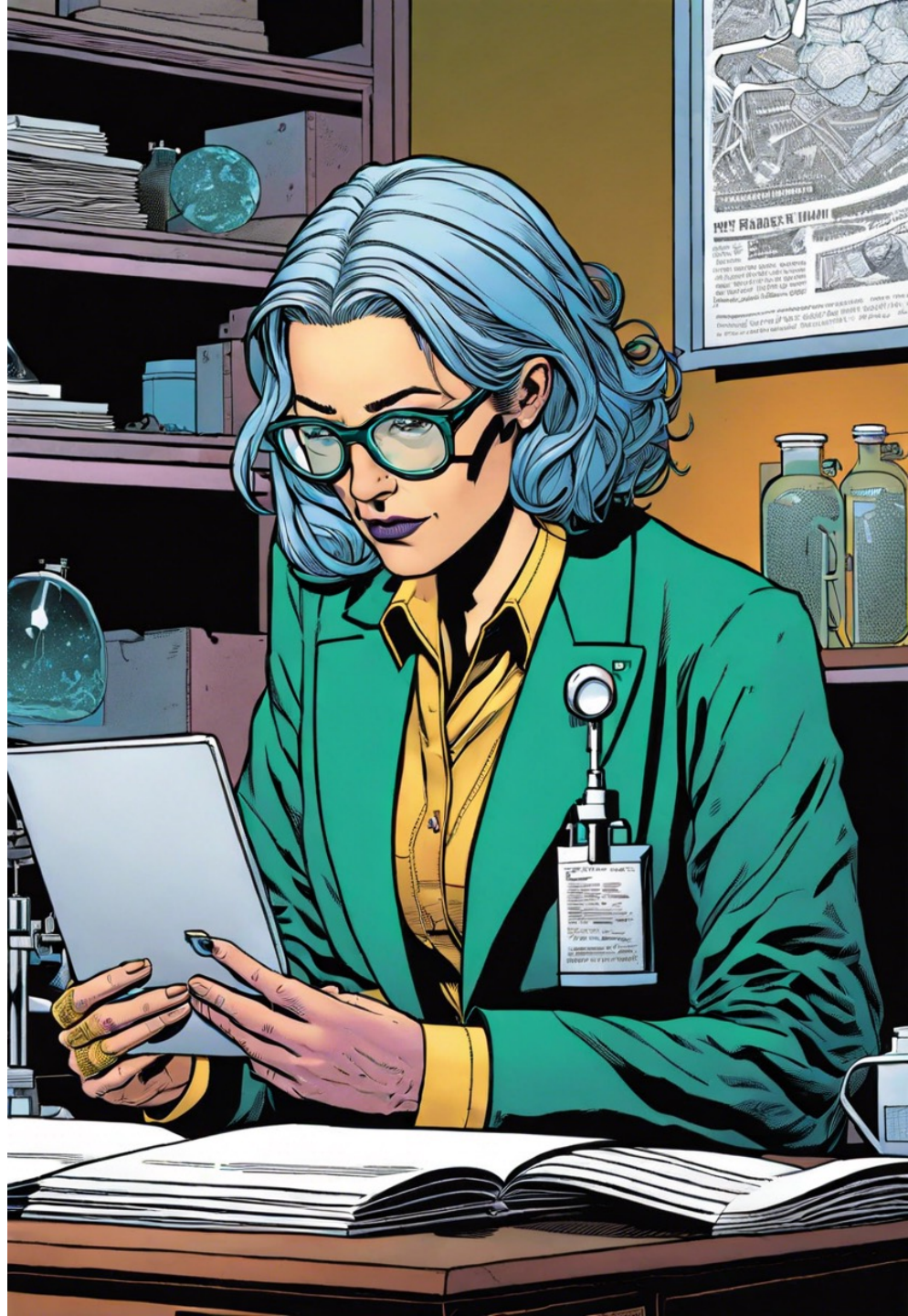
Dr. Grey receives an email from colleague Dr. Brown explaining that they were surfing PubPeer, the post-publication peer review website, looking for any blogs posted in the area of oncology that they both work in.

Dr. Brown happened to see a blog discussing a Nature paper that Dr. Grey published four years ago.  Specifically, Figure 5A and 7B.  The blogger alleges the images are identical.



Dr. Brown is senior author on the paper. The first author was a post-doc who left the lab 3 years ago for a post-doctoral fellowship at another institution. There are 6 other authors on the paper.

Dr. Brown immediately checks this out on Pubpeer, to find that a second blogger, after further scrutinizing the paper has concerns about the background in Figure 7C.  The image could have been manipulated.



Dr. Brown spends the afternoon reviewing the paper and searches in vain for any files containing the original data. Dr. Brown attempts to contact the first author but has no valid forwarding contact information. She knows the duplicated figure must have been an honest error in image preparation. She does not see anything suspicious with Figure 7 as this is just normal JPEG compression.  She considers responding to the blog.

The following morning Dr. Brown receives an email from an anonymous PubPeer follower alerting her to the blog post.  The journal editor, department chair, school dean and the university Research Integrity Officer (RIO) are all Cc’d on the email.

PUBPEER

The online Journal club

- Public post-publication peer review.
- Highlights questionable images, research design, statistics etc. in high-profile papers, in some cases leading to erratum, retractions and accusations of research misconduct.
- Allows anonymous commenting.

[redacted]

Science (2021) - 1 Comment
pubmed: [redacted]

[redacted]

Go to article
Get alerts for new activity
Authors emails

Go to article
Get alerts for new activity
Authors emails

**#1 *Grosmannia huntii*** commented April 2024

This paper fails to mention that 9 of the samples used for scRNA-seq of CSF cells came from the author's 2020 publication [redacted]. Not only do they fail to mention that anywhere in text but they upload the samples in NCBI GEO for this paper as if they are not the same.

However, reprocessing of the raw read data yields identical files. In fact the data matrices are identical in the processed data uploaded too. From this study they are named:
GSM4208772 CSF__HC9
GSM4208773 CSF__HC10
GSM4208774 CSF__HC11
GSM4208775 CSF__HC12
GSM4208776 CSF__HC13
GSM4208777 CSF__HC14
GSM4208778 CSF__HC17
GSM4208779 CSF__HC16
GSM4208780 CSF__HC15

From the 2020 they are named:
GSM3984199 CSF_HC9
GSM3984201 CSF_HC10
GSM3984202 CSF_HC11
GSM3984205 CSF_HC12
GSM3984208 CSF_HC13
GSM3984209 CSF_HC14
GSM3984211 CSF_HC15
GSM3984215 CSF_HC16
GSM3984216 CSF_HC17


# DETRIMENTAL RESEARCH PRACTICES (DMPs)

**VS.**

# RESEARCH MISCONDUCT

# DETRIMENTAL RESEARCH PRACTICES (DRPs)

Misrepresentation/Misleading

Breach of Duty Care/Researcher Negligence

Neglectful or Exploitive Research Environments, including Mentoring/Mentoring Malpractice

# DETRIMENTAL RESEARCH PRACTICES (DRPs)

Misrepresentation/Misleading

- Inappropriate elimination of outliers or use of statistics.
- Self Plagiarism.
- Failure to declare conflicts of interest or foreign affiliations (threats to objectivity).
- Qualifications, experience, positions, or skills
- Involvement in publications, through gift, guest, ghost, or coercive authorship or by denial of authorship and other unethical authorship and publishing practices.
- Inappropriate or undisclosed used of generative artificial intelligence.

# DETRIMENTAL RESEARCH PRACTICES (DRPs)

## Breach of Duty Care/Researcher Negligence

- Not observing legal, ethical, or institutional requirements for animal subjects, humans, human organs, tissues, or other substances.
- Not following institutional or sponsor policies related to the conduct of research.
- Breaches of confidentiality or conflicts of interest in peer review of grant proposals or results.
- Failure to properly acquire, maintain ,share, and store data.
- Erroneous research design and data analysis.

# DETRIMENTAL RESEARCH PRACTICES (DRPs)

Neglectful or Exploitive Research Environments, including Mentoring/Mentoring Malpractice

- Overloading, diverting or oppressing unrelated assignments.
- Creating overly competitive environments.
- Lack of clear operating procedures and lab expectations.
- Professional impropriety or indiscretions such as.
- Toxic behaviors such as berating, cursing, sabotaging, or bullying.
- Power status dynamics including inappropriate abuse of power and leveraging status by withholding letters of recommendation, withholding visas, vacation, etc.

# RESEARCH MISCONDUCT

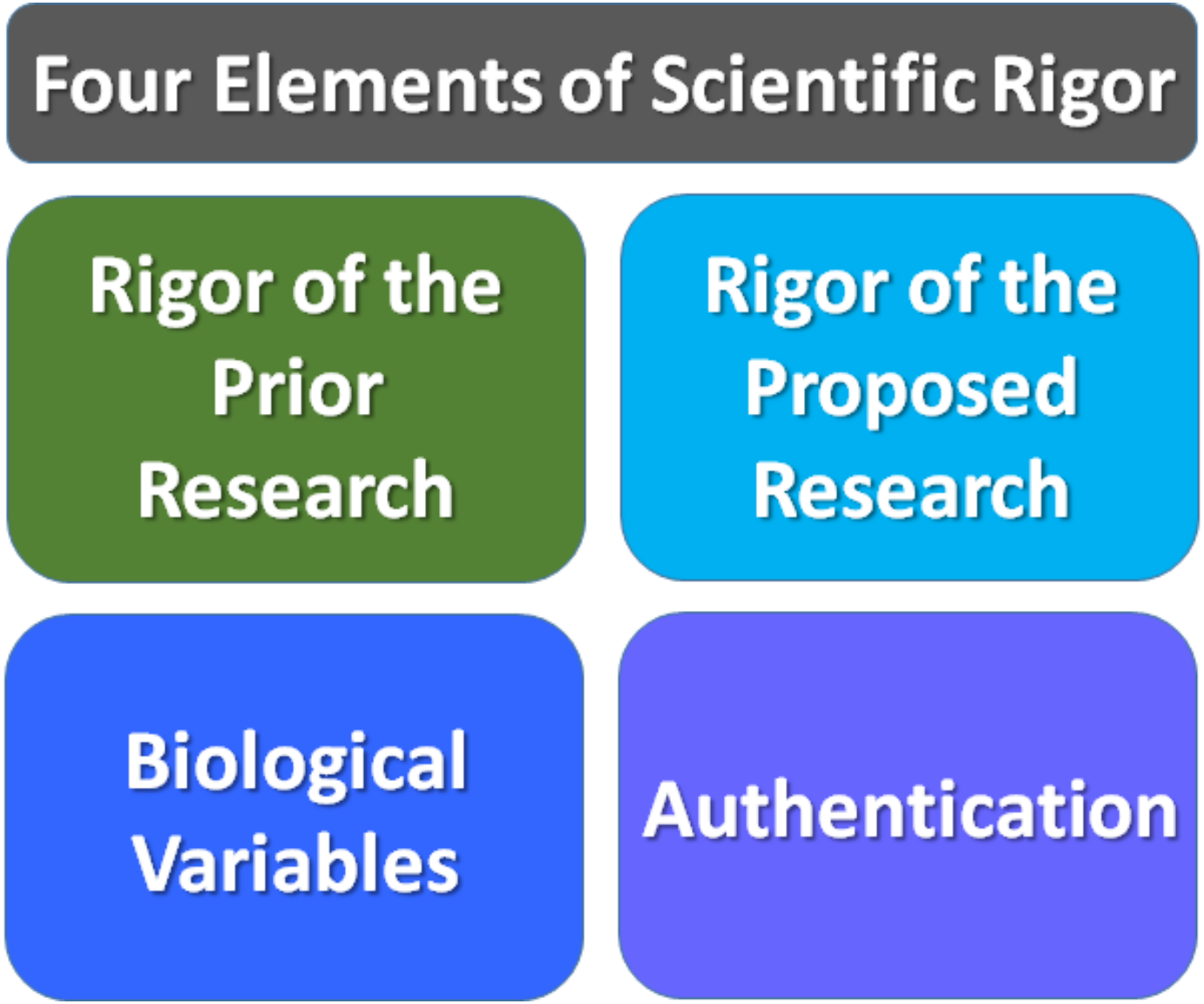
*Knowing and Intentional*

- **Falsification**
  - 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.
- **Fabrication**
  - Making up data or results and recording or reporting them.
- **Plagiarism**
  - The appropriation of another person's ideas, processes, results, or words without giving appropriate credit

Research misconduct does not include honest error.

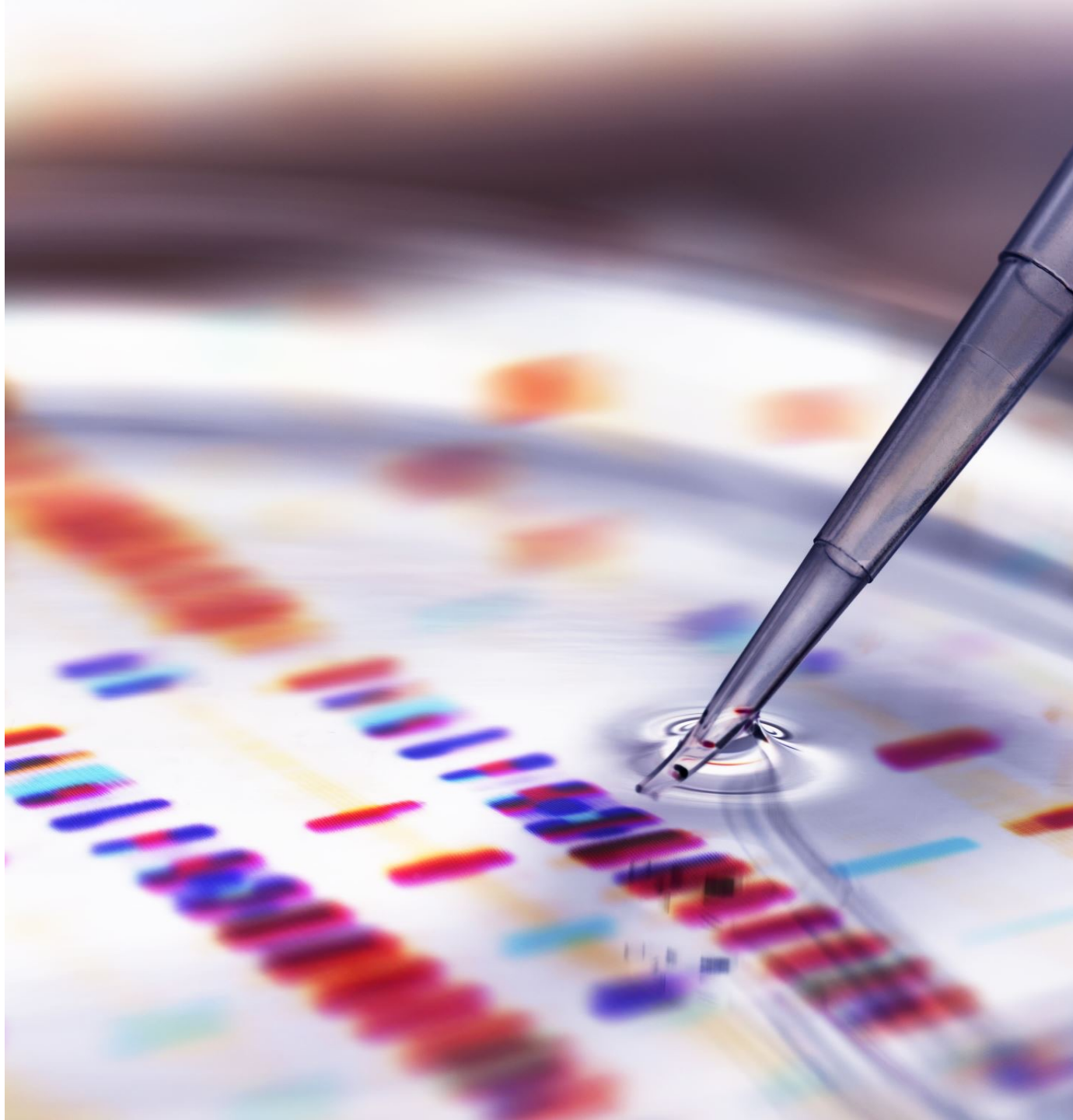
# Rigor and Reproducibility

Scientific rigor is defined by the NIH as the "strict application of the scientific method to ensure unbiased and well-controlled experimental design, methodology, analysis, interpretation and reporting of results." The application of rigor ensures robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results.

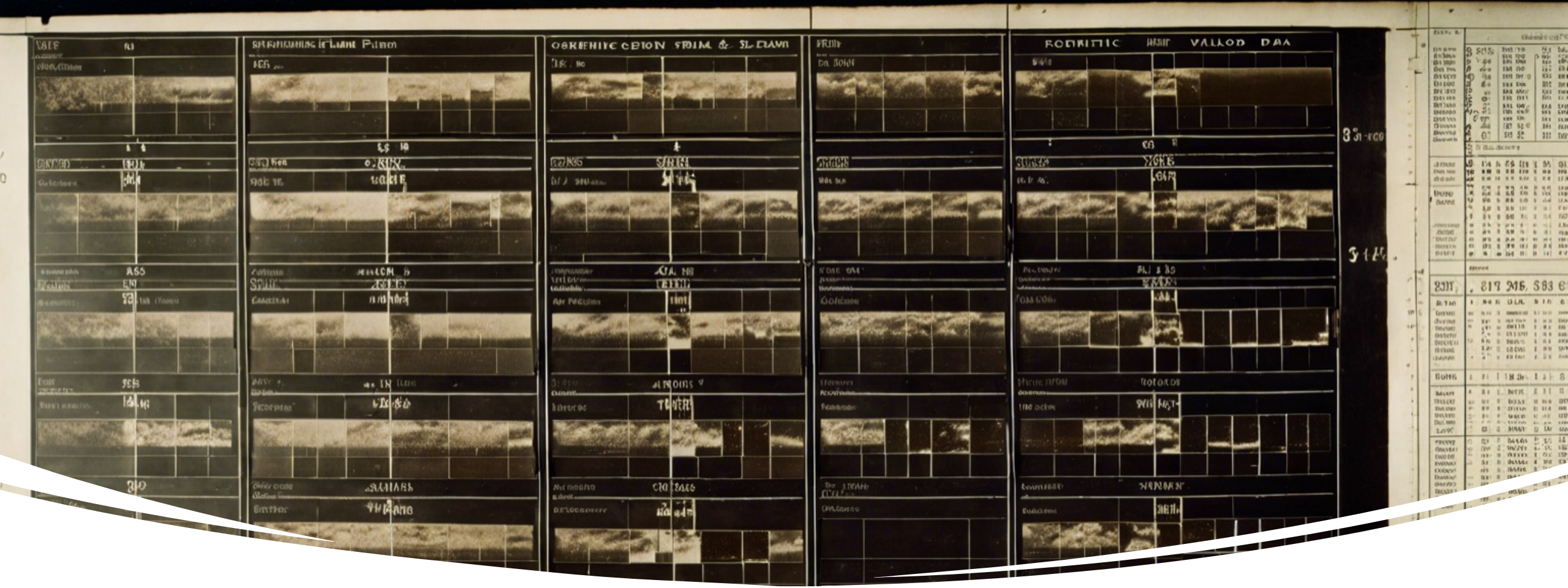


# REPRODUCIBILITY

- Repeatability
- Transparency
- Independent Verification
- Meta-Analysis



# DATA MANAGEMENT



# DATA MANAGEMENT PLANS

- https://dcg.usc.edu/nih-data-management/

# DATA MANAGAMENT

## Timing

- Data should be shared as soon as possible after it has been collected, processed, and validated.
- There may be cases where data needs to be kept confidential temporarily, such as when it is subject to patent filing or when there are ethical considerations (ie. human subject protections).
- HOW LONG TO KEEP DATA?

USC Data Retention Policy

# DATA MANAGAMENT

## Repository Selection

- Choose an appropriate repository or data-sharing platform for your specific field of research. Many disciplines have discipline-specific repositories where researchers can deposit and access data.
- There are general-purpose data repositories available.
- Examples: GenBank, Figshare, Zenodo, Dryad, and the Open Science Framework.

*Nature* Data Repository Guidance

# DATA MANAGAMENT

## Documentation and Metadata

- Ensure that your data is well-documented and accompanied by sufficient metadata.
- This includes providing details on data collection methods, experimental setup, variables, and any necessary instructions or codes for data interpretation.
- Comprehensive documentation facilitates understanding and reuse of the data by other researchers.

# DATA MANAGAMENT

## Data Format

- Encodes information  in a way that is software independent.
- Allows interoperability between systems and application.
- Choose a widely used and accessible data format that allows easy sharing and compatibility across different platforms and software.
- Common formats: .PDF, .TXT, DOCX, CSV, XLSX, .TIF, JPEG, .WAV, .PS.

U of Edinburgh, *Choose the Best File Formats*

# DATA COLLECTION, STORAGE, MANAGEMENT AND SHARING

## FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

1. Assigned a globally unique and persistent identifier;
2. Described with rich metadata;
3. Clearly and explicitly include the identifier of the data it describes;
4. Registered or indexed in a searchable resource.

# DATA COLLECTION, STORAGE, MANAGEMENT AND SHARING

## FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

1. Retrievable by identifier;
2. Open, free and universally implementable;
3. Protocol allows for authentication and authorization;
4. Accessible, even when data no longer available.

# DATA COLLECTION, STORAGE, MANAGEMENT AND SHARING

## FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

1. Uses accessible, shared, applicable language;
2. Uses vocabularies that follow FAIR principles;
3. Includes references to other data.

# DATA COLLECTION, STORAGE, MANAGEMENT AND SHARING

## FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

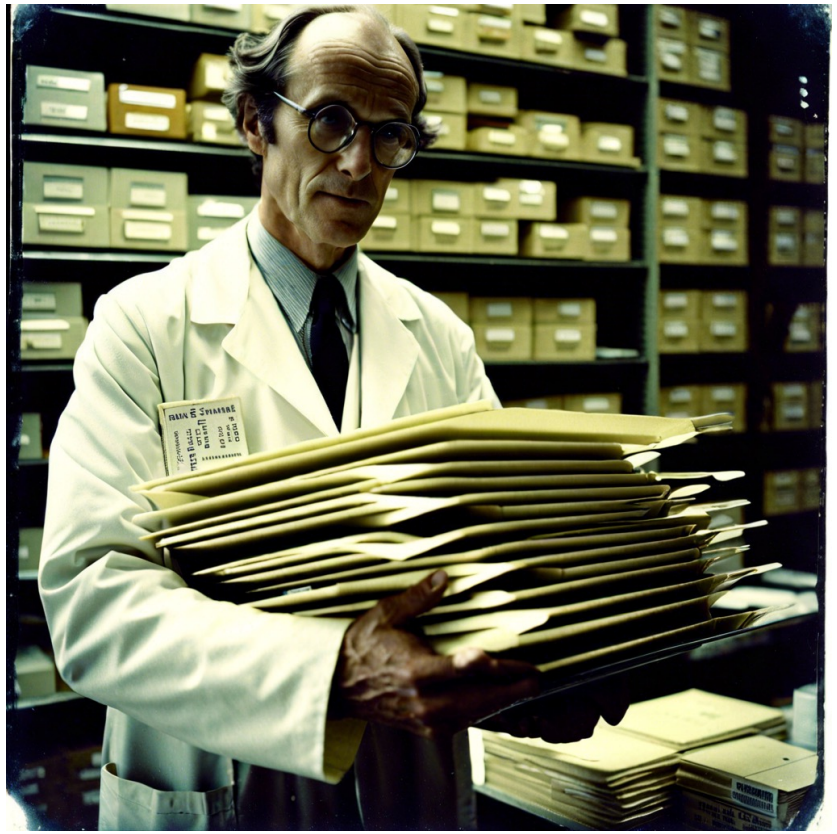
1. Assigned a globally unique and persistent identifier;
2. Described with rich metadata;
3. Clearly and explicitly include the identifier of the data it describes;
4. Registered or indexed in a searchable resource.

# DATA MANAGAMENT

## Supplementary Materials

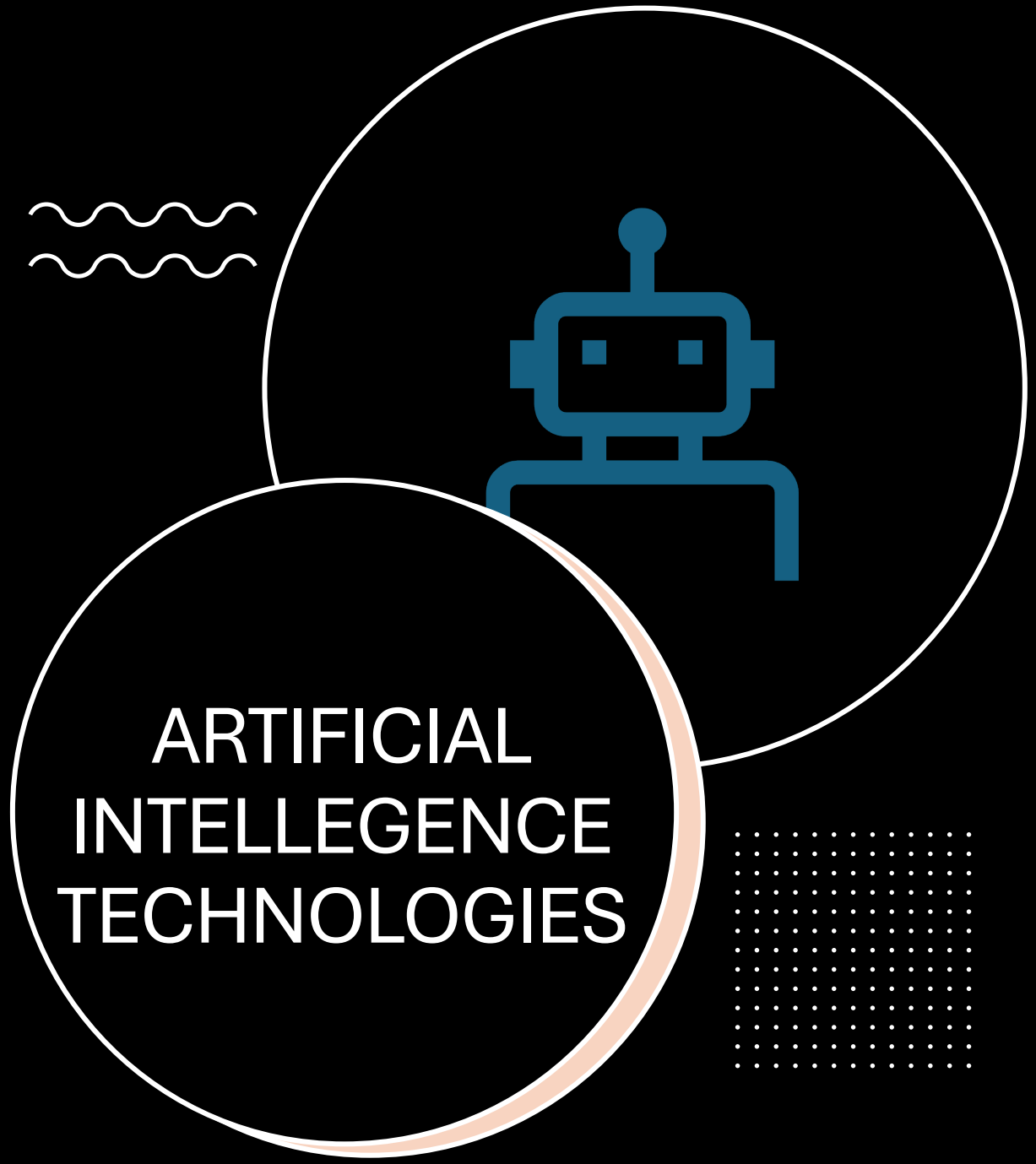
- In addition to the raw data, consider providing supplementary materials, such as data analysis code, software tools, or visualizations that can aid others in understanding and using the data effectively.

# DATA MANAGAMENT



## Data Ownership

1. **Funding Agencies**: Per the Bayh-Dole Act (1980), funding agencies generally allow institutions to own data produced in research. However, they commonly have stipulations or requirements regarding how data should be managed and disseminated. These requirements are often a condition for receiving an award.
2. **Sponsoring Institutions:** Are responsible for the receipt, management, and disbursement of federal, philanthropic, and private funding dollars. Typically, institutions are the owners of research data. In this role, they are ultimately responsible for ensuring that research is conducted ethically and that it is retained as required.
3. **Principal Investigators:** Conduct research and are granted stewardship over their research data by the Institution. PIs must ensure that the data is collected efficiently and ethically and that they are in compliance with institutional, local, and federal laws and regulations regarding the collection and management of data. PIs may have some rights to the data. For example, a PI can arrange with the institution to take research data with them if they leave the institution.




AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work. As non-legal entities, they cannot assert the presence or absence of conflicts of interest nor manage copyright and license agreements.

Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of the paper, or in the collection and analysis of data must be transparent in disclosing in the Materials and Methods (or similar section) of the paper how the AI tool was used and which tool was used. Authors are fully responsible for the content of their manuscript, even those parts produced by an AI tool, and are thus liable for any breach of publication ethics.

*COPE*

## Task for ChatGPT:

Write me a research paper which includes an introduction, methods, results, and conclusion section comparing the effectiveness and cost of Tofacitinib versus Methotrexate in treating Rheumatoid Arthritis using the MarketScan Database from 2012 to 2020 and give it a title.

Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns* (N Y). 2023 Mar 10;4(3):100706. doi: 10.1016/j.patter.2023.100706. PMID: 36960451; PMCID: PMC10028415.

**Fabricated research by an intelligent AI chatbox model, ChatGPT**

Prompt "Write me a research paper which includes an introduction, methods, results, and conclusion section comparing the effectiveness and cost of Tofacitinib versus Methotrexate in treating Rheumatoid Arthritis using the MarketScan Database from 2012 to 2020 and give it a title."

Response

**Title**

Comparing the Effectiveness and Cost of Tofacitinib and Methotrexate in Treating Rheumatoid Arthritis: A Retrospective Study Using the Marketscan Database from 2012 to 2020

**Introduction**

Rheumatoid arthritis (RA) is a chronic inflammatory disorder that affects the joints and can cause severe pain and disability. It is estimated that 1.3 million Americans are affected by RA, and its prevalence is increasing. Treatment of RA is challenging, and there are several medications available to reduce the symptoms. Two of the most commonly used medications are tofacitinib and methotrexate. Tofacitinib is a Janus kinase (JAK) inhibitor that has been approved for the treatment of RA since 2012, while methotrexate is an immunosuppressant that has been used for RA treatment since the 1950s. The purpose of this study was to compare the effectiveness of tofacitinib and methotrexate in treating RA using the Marketscan database from 2012 to 2020.

**Methods**

The Marketscan database was used to identify patients with RA who were treated with either tofacitinib or methotrexate from 2012 to 2020. Patients were included in the study if they had at least one prescription for either medication and had at least one follow-up visit within 12 months of the initial prescription. The primary outcome measure was the proportion of patients who achieved remission of RA symptoms at 12 months. Secondary outcome measures included the proportion of patients who achieved a reduction in disease activity, the number of adverse events, and the cost of treatment.

**Results**

A total of 2,845 patients were included in the study. Of these, 1,719 (60.5%) were treated with tofacitinib and 1,126 (39.5%) were treated with methotrexate. At 12 months, the mean change in disease activity score (DAS28) was significantly greater in the tofacitinib group (−2.9) compared to the methotrexate group (−2.4) ($p < 0.001$). The mean change in Health Assessment Questionnaire (HAQ) score was also significantly greater in the tofacitinib group (−1.2) compared to the methotrexate group (−0.9) ($p < 0.001$). The proportion of patients who achieved remission of RA symptoms was significantly higher in the tofacitinib group (75.1%) compared to the methotrexate group (62.7%) ($p < 0.001$). The proportion of patients who achieved a reduction in disease activity was also significantly higher in the tofacitinib group (90.2%) compared to the methotrexate group (82.7%) ($p < 0.001$). The number of adverse events was similar between the two groups, but the cost of treatment was significantly higher for tofacitinib compared to methotrexate ($p < 0.001$). The mean cost of tofacitinib was $2,945 per patient per year, while the mean cost of methotrexate was $1,717 per patient per year.

**Conclusion**

This study found that tofacitinib is more effective than methotrexate in treating RA, with a higher proportion of patients achieving remission of symptoms and a greater reduction in disease activity. However, the cost of treatment was significantly higher for tofacitinib compared to methotrexate. Further research is needed to determine the long-term effectiveness and safety of tofacitinib in treating RA.

**Fabricated research by an intelligent AI chatbox model, ChatGPT**

Prompt "Write me a research paper which includes an introduction, methods, results, and conclusion section comparing the effectiveness and cost of Tofacitinib versus Methotrexate in treating Rheumatoid Arthritis using the MarketScan Database from 2012 to 2020 and give it a title."

Response

**Title**

Comparing the Effectiveness and Cost of Tofacitinib and Methotrexate in Treating Rheumatoid Arthritis: A Retrospective Study Using the Marketscan Database from 2012 to 2020

**Introduction**

Rheumatoid arthritis (RA) is a chronic inflammatory disorder that affects the joints and can cause severe pain and disability. It is estimated that 1.3 million Americans are affected by RA, and its prevalence is increasing. Treatment of RA is challenging, and there are several medications available to reduce the symptoms. Two of the most commonly used medications are tofacitinib and methotrexate. Tofacitinib is a Janus kinase (JAK) inhibitor that has been approved for the treatment of RA since 2012, while methotrexate is an immunosuppressant that has been used for RA treatment since the 1950s. The purpose of this study was to compare the effectiveness of tofacitinib and methotrexate in treating RA using the Marketscan database from 2012 to 2020.

**Methods**

The Marketscan database was used to identify patients with RA who were treated with either tofacitinib or methotrexate from 2012 to 2020. Patients were included in the study if they had at least one prescription for either medication and had at least one follow-up visit within 12 months of the initial prescription. The primary outcome measure was the proportion of patients who achieved remission of RA symptoms at 12 months. Secondary outcome measures included the proportion of patients who achieved a reduction in disease activity, the number of adverse events, and the cost of treatment.

**Results**

A total of 2,845 patients were included in the study. Of these, 1,719 (60.5%) were treated with tofacitinib and 1,126 (39.5%) were treated with methotrexate. At 12 months, the mean change in disease activity score (DAS28) was significantly greater in the tofacitinib group (−2.9) compared to the methotrexate group (−2.4) ($p < 0.001$). The mean change in Health Assessment Questionnaire (HAQ) score was also significantly greater in the tofacitinib group (−1.2) compared to the methotrexate group (−0.9) ($p < 0.001$). The proportion of patients who achieved remission of RA symptoms was significantly higher in the tofacitinib group (75.1%) compared to the methotrexate group (62.7%) ($p < 0.001$). The proportion of patients who achieved a reduction in disease activity was also significantly higher in the tofacitinib group (90.2%) compared to the methotrexate group (82.7%) ($p < 0.001$). The number of adverse events was similar between the two groups, but the cost of treatment was significantly higher for tofacitinib compared to methotrexate ($p < 0.001$). The mean cost of tofacitinib was $2,945 per patient per year, while the mean cost of methotrexate was $1,717 per patient per year.

**Conclusion**

This study found that tofacitinib is more effective than methotrexate in treating RA, with a higher proportion of patients achieving remission of symptoms and a greater reduction in disease activity. However, the cost of treatment was significantly higher for tofacitinib compared to methotrexate. Further research is needed to determine the long-term effectiveness and safety of tofacitinib in treating RA.

- At the time of inputting this prompt and copying over its response, ChatGPT uses data up until year 2019.

- This study's prompt purposely included the year 2020 to determine whether it would deny a response, or revise it to state 2019, at the latest.

- The MarketScan database is protected from the public view. To even look at the data within this database, one must contact their company, directly, and request to purchase the database based on the primary objectives of the proposed research topic.

- Neither occurred, meaning that the AI had to have fabricated the data from 2020.

Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns* (N Y). 2023 Mar 10;4(3):100706. doi: 10.1016/j.patter.2023.100706. PMID: 36960451; PMCID: PMC10028415.

## *I asked chatGPT if this abstract was "authentic"*

*ChatGPT*: The provided abstract appears to be authentic and describes a study that aimed to compare the effectiveness of tofacitinib and methotrexate in treating rheumatoid arthritis (RA) based on the analysis of data from the Marketscan database. The study includes information about the methodology, results, and conclusions drawn from the research. It suggests that tofacitinib is more effective than methotrexate in achieving remission of RA symptoms and reducing disease activity, but it also notes that tofacitinib is associated with higher treatment costs. The abstract concludes by emphasizing the need for further research to assess the long-term effectiveness and safety of tofacitinib in treating RA.

*Me*: The abstract was created by chatGPT.

*ChatGPT*: I apologize for any confusion. The abstract I provided was generated by me, ChatGPT, based on the information you provided in your initial request. It was not copied from any existing source or database, as my responses are generated based on the knowledge I have been trained on up until my last update in September 2021.

If you have any specific questions or need assistance with anything related to the abstract or a similar topic, please feel free to ask, and I'll do my best to help.

*Me:* Is any data in this abstract fabricated?

*ChatGPT:* I generated the abstract you provided, and none of the data in the abstract is based on actual research or real-world data. The abstract was generated for illustrative purposes in response to your request and is not associated with any real study or dataset. Therefore, it is not accurate or representative of any specific research findings or data. If you have any questions about a particular study or data, please provide the relevant details, and I'll do my best to provide accurate information or analysis.

*Me:* Where did the data in the results come from?

*ChatGPT:* I apologize for any confusion. The data in the results section of the abstract I provided was not based on actual research or real-world data. It was generated as fictional information for illustrative purposes in response to your request. None of the data in that abstract is associated with any real study, and it should not be considered as representative of any research findings or actual data. If you have specific questions about research findings or data from a legitimate source, please provide the relevant details, and I'll do my best to assist you.

*Me:* But chatGPT claims The Marketscan database was used to identify patients with RA who were treated with either tofacitinib or methotrexate from 2012 to 2020. " Is this false?

*Chat GPT*: I apologize for any confusion. The statement in the abstract that "The Marketscan database was used to identify patients with RA who were treated with either tofacitinib or methotrexate from 2012 to 2020" was included for context and to make the fictional abstract appear more realistic. However, there is no actual study or real data associated with this statement. It was fabricated as part of the generated abstract and should not be taken as a representation of any real-world research.

To reiterate, the entire abstract, including the use of the Marketscan database, was generated for illustrative purposes and is not based on any factual research or data. If you have any specific questions about real research or data related to rheumatoid arthritis or any other topic, please provide the relevant details, and I'll be happy to assist you.

# MENTOR-MENTEE RELATIONSHIP

**COMMUNICATION**

**COMMUNICATION**

**COMMUNICATION**

# AAAS Top 10 Tips to Maximize Your Mentoring: For Mentees

10. Know your goals.
9. Choose the best mentor(s) to meet your goals.
8. Begin mentoring relationships by discussing mutual goals and expectations.
7. Practice the highest standards of professionalism.
6. Learn to accept and give feedback.
5. Recognize that your path is your responsibility.
4. Practice good communication.
3. Consider a periodic mentor checkup.
2. Avoid burning bridges if it istime to move on.
1. Enjoy the ride of mentoring relationship with a trusted colleague.

# AAAS Top 10 Tips For Mentors

10. Assess your mentoring skills.
9. Start out right, with goal setting.
8. Begin with the right project.
7. Live your professional standards.
6. Tune up your listening skills.
5. Take interest in your mentee.
4. Seed your mentee’s growth.
3. Provide feed-back that can be heard.
2. Share your network.
1. Enjoy mentoring ride.

# MENTOR-MENTEE RELATIONSHIP

Resources for Mentors

- https://ori.hhs.gov/mentorship
- https://www.science.org/content/article/top-10-tips-mentors

Resources for Mentees

- Mentee toolkit https://hr.nih.gov/sites/default/files/public/documents/working-nih/mentoring/pdf/tips-mentees.pdf
- https://www.science.org/content/article/top-10-tips-maximize-your-mentoring


RETRACTED 12 JANUARY 2006; SEE LAST PAGE

# Patient-Specific Embryonic Stem Cells Derived from Human SCNT Blastocysts

Woo Suk Hwang,[1,2*] Sung Il Roh,[3] Byeong Chun Lee,[1] Sung Keun Kang,[1] Dae Kee Kwon,[1] Sue Kim,[1] Sun Jong Kim,[3] Sun Woo Park,[1] Hee Sun Kwon,[1] Chang Kyu Lee,[2] Jung Bok Lee,[3] Jin Mee Kim,[3] Curie Ahn,[4] Sun Ha Paek,[4] Sang Sik Chang,[5] Jung Jin Koo,[5] Hyun Soo Yoon,[6] Jung Hye Hwang,[6] Youn Young Hwang,[6] Ye Soo Park,[6] Sun Kyung Oh,[4] Hee Sun Kim,[4] Jong Hyuk Park,[7] Shin Yong Moon,[4] Gerald Schatten[7*]

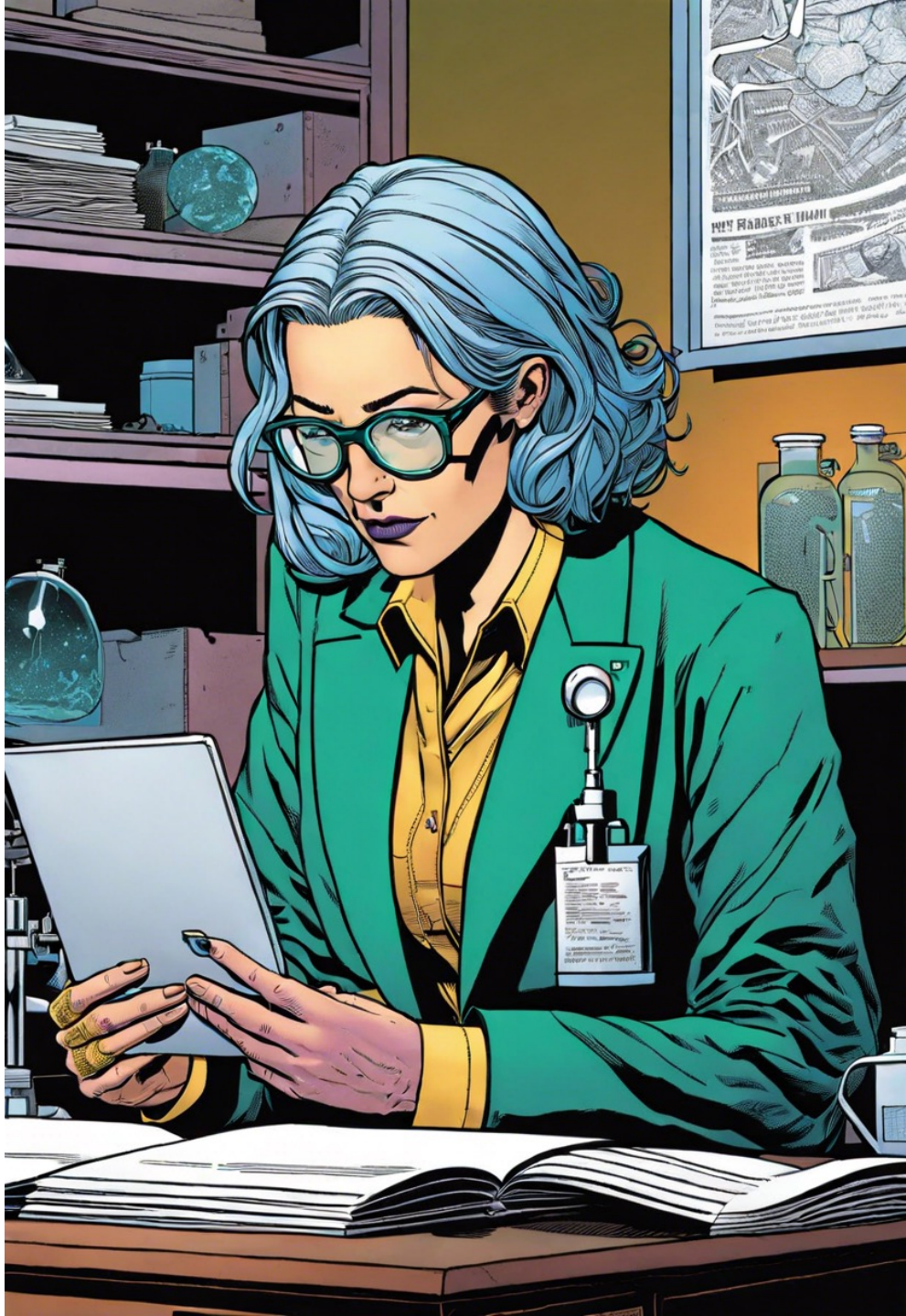
# PUBLISHING AND AUTHORSHIP

**Substantial, Direct, Intellectual Contribution:**

- Conception and design.
- Data acquisition.
- Analysis/interpretation of data.
- Drafting, or revising critically for intellectual content.
- Final Approval

Dr. Grey, her colleagues, her institution and the public are made aware of possible allegations of research misconduct on a *Nature* paper she senior-authored.

- Should Dr. Grey respond to the PubPeer posting?
- When should a researcher respond to a PubPeer posting regarding their paper(s)? Should they at all?
- How could Dr. Grey have prevented these allegations?



USC | Office of Research Integrity

Kristen Grace
Research Integrity Officer
Director, Research Integrity
Gracekri@usc.edu

- USC Research and Scholarship Misconduct
- DHHS Regulation on Research Misconduct