

# 2019 Cornell RCR symposium

## Data acquisition and management

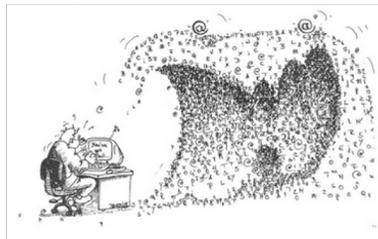
### Data acquisition and management

- Key factor in reproducibility (2017 RCR symposium)
- Poor practices can lead to research misconduct, even if no malicious intent.
- Essential for effectively getting research done.

## Outline

1. Intro – what are the issues?
2. What can go wrong? – examples from the literature
3. Journal and grant agency policies
4. Available resources at Cornell
5. Brief intro to the cases

## Big Data



The screenshot shows the homepage of the Museum of Obsolete Media. At the top left is the logo, which consists of several dark grey circles of varying sizes arranged in a grid-like pattern. To the right of the logo, the text "Museum of Obsolete Media" is displayed in a bold, sans-serif font. Below the logo and text is a search bar with the placeholder text "Search ..." and a "Search" button. A navigation menu is located below the search bar, with links for "About", "All Formats", "Audio Formats", "Video Formats", "Data Formats", "Film Formats", and "Media Preservation". The main content area features a large heading "Welcome to the Museum of Obsolete Media" followed by two paragraphs of introductory text. To the right of the text is a photograph of a white Brother MD-200 micro disc drive and its packaging. Below the photograph, the text "Brother Micro Disc and MD-200 drive" is visible. At the bottom right of the screenshot, the URL "https://obsoletemedia.org/" is provided.





<https://obsoletemedia.org/>

## Data Format Timeline

From Jacquard Loom cards to PlayStation Vita memory cards, the history of data storage on removable media. Comprising disks/discs, tape, solid-state media, ROM cartridges, and punched media.

Dates are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

### 1800s

Jacquard Loom card (1801 – 1990s)

### 1840s

Punched tape (1846 – 1980s)

### 1890s

Herman Hollerith's punched card design is used in the 1890 US census (1890)

Punched card (1890 – 1980s)

### 1940s

Aperture card (1943 – 2000s)

Open reel instrumentation and data logging tape (1949 – 2000s)

### 1950s

UNISERVO, the first magnetic tape drive for a commercial computer system is introduced (1951)

IBM introduces 7-track magnetic data tape (1952)

IBM introduces the hard disk drive with the IBM 305 RAMAC using 50 24-inch disks storing 5 MB (1956)

LEO tape (1958 – 1981)

### 2000s

The PlayStation 2, the best-selling video game console of all time, is launched. Arguably, part of its appeal is its ability to play DVD-Video discs (2000)

Sony Memory Stick / Floppy Disk Adaptor (2000 – 2001)

Sony PlayStation 2 Memory Card (2000 – 2013)

Linear Tape-Open (LTO) (2000 – )

USB flash drive (2000 – )

Sega ceases production of the Dreamcast and exits the video game console market (2001)

Microsoft enters the video game console market in the US with the introduction of the Xbox (2001)

omega Zip 250 (2001 – 2003)

omega Zip U250 (2001 – 2003)

Nintendo Game Boy Advance (2001 – 2007)

CD Cardz (2001 – 2009)

DVD+RW (2001 – )

Super DLT (2001 – )

There is concern that the discs containing the BBC Domesday Project of 1986 may already be unreadable (2002)

Pokémon mini (2002)

omega Zip 750 (2002 – 2003)

DataPlay (2002 – mid 2000s)

Nintendo GameCube memory card (2002 – 2007)

Nintendo GameCube (2002 – 2007)

xD-Picture Card (2002 – 2010)

DVD+R (2002 – )

Memory Stick Duo (2002 – )

Qualstar Corporation, the last manufacturer of 9-track open reel computer tape drives, announces the final shipment (2003)

Dell begins removing floppy disk drives from its PCs (2003)

Nokia N-Gage card (2003 – 2006)

miniSD (2003 – 2006)

Super Advanced Intelligent Tape (SAIT) (2003 – 2010)

<https://obsoletemedia.org/>

## Key Concepts in Data Management

- Ownership
- Collection
- Storage
- Sharing
- Protection
- Retention
- Analysis
- Reporting

*ori.hhs.gov/documents/rcrintro.pdf*

From Prof. Maurine Linder

9

## 2) What can go wrong?

The screenshot shows the homepage for the journal 'Molecular and Cellular Biology' on the ASM website. The header includes the ASM logo, the journal title, a search bar, and navigation links for Home, Articles, For Authors, About the Journal, and Subscribe. The main content area features a 'Table of Contents' for August 2017, volume 37, issue 16, and a 'Spotlight' section for 'Articles of Significant Interest Selected from This Issue by the Editors'. On the right, there is an 'In This Issue' section with a thumbnail of the journal cover and links to 'Table of Contents (PDF)' and 'Journal Cover & Description'.

From Prof. Maurine Linder

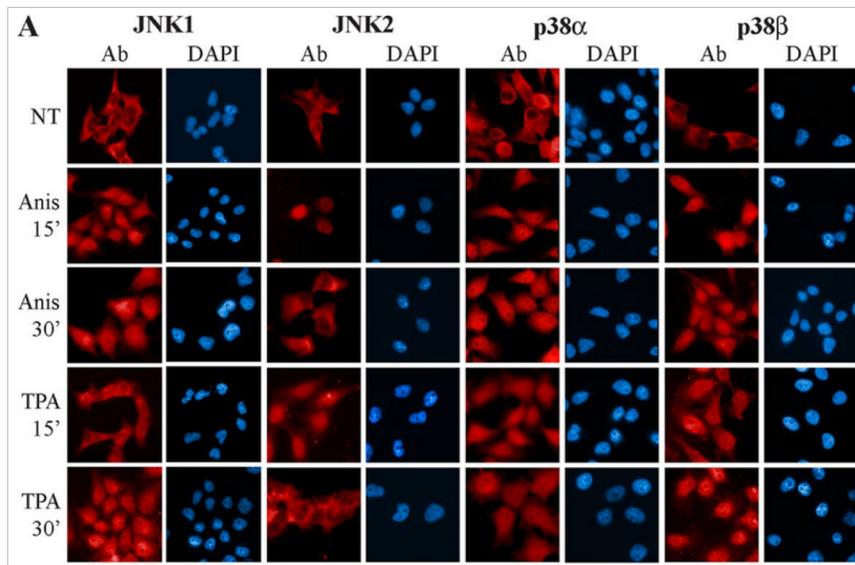
<p><b>Research Articles</b></p> <p>Research Article   Spotlight</p> <p><b>let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis</b></p> <p>Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shusheng Wang</p> <hr/> <p>Research Article</p> <p><b>Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue</b></p> <p>Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang</p> <hr/> <p>Research Article</p> <p><b>RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner</b></p> <p>Kyle Lewis, Laila Valanejad, Ashley Cast, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karma, Lubov Timchenko, Nikolai Timchenko</p> <hr/> <p>Research Article</p> <p><b>Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis</b></p> <p>Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julieta Apsa, Chi Fang Cheok, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis</p> <hr/> <p>Research Article   Spotlight</p> <p><b>DNA Polymerase Beta Participates in Mitochondrial DNA Repair</b></p> <p>P. Szykora, S. Kanno, M. Akbari, T. Kulkowicz, B. A. Bagfista, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croteau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr</p> <hr/> <p>Research Article</p> <p><b>Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment</b></p> <p>Tomoko Oyazaki-Torimaru, Tomohito Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hishiki, Hiroshi Morisaki, Makoto Suematsu, Yoji Andrew Minamishima</p> <hr/> <p>Research Article</p> <p><b>Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1</b></p> <p>Michael C. Armstrong, Sergej Sestak, Ahmed A. Ali, Hanan A. M. Sagni, Max Brown, Karen Baty, Achim Treumann, Martin Schröder</p>	<p><b>Author Corrections</b></p> <p>Author Correction</p> <p><b>Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"</b></p> <p>Eldar Zehorai, Rony Seger</p> <hr/> <p>Author Correction</p> <p><b>Correction for Lin et al., "Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway"</b></p> <p>Xi Lin, Xiangheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang</p> <hr/> <p>Author Correction</p> <p><b>Correction for Shah et al., "A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes"</b></p> <p>Zahid H. Shah, Shafig U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner</p> <hr/> <p>Author Correction</p> <p><b>Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"</b></p> <p>Kenzo Fukunaga, Youngho Kwon, Patrick Sung, Katsunori Sugimoto</p> <hr/> <p>Author Correction</p> <p><b>Correction for Copetti et al., "p65/RelA Modulates BECN1 Transcription and Autophagy"</b></p> <p>Tamara Copetti, Cosetta Bertoli, Emiliano Dalla, Francesca Demarchi, Claudio Schneider</p> <hr/> <p>Author Correction</p> <p><b>Correction for Travesa et al., "Repression of G<sub>1</sub>/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6"</b></p> <p>Anna Travesa, Tatjana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chahwan, David E. Lee, Noel F. Llewellyn, Curt Wittenberg</p> <hr/> <p><b>Retraction</b></p> <p>Retraction</p> <p><b>Retraction for Sorek et al., "Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase"</b></p> <p>Nadav Sorek, Limor Poraty, Hassana Stamborg, Elnat Bar, Efrain Lewinsohn, Shaul Yalovsky</p>	
---	--	--

<p><b>Research Articles</b></p> <p>Research Article   Spotlight</p> <p><b>let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis</b></p> <p>Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shusheng Wang</p> <hr/> <p>Research Article</p> <p><b>Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue</b></p> <p>Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang</p> <hr/> <p>Research Article</p> <p><b>RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner</b></p> <p>Kyle Lewis, Laila Valanejad, Ashley Cast, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karma, Lubov Timchenko, Nikolai Timchenko</p> <hr/> <p>Research Article</p> <p><b>Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis</b></p> <p>Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julieta Apsa, Chi Fang Cheok, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis</p> <hr/> <p>Research Article   Spotlight</p> <p><b>DNA Polymerase Beta Participates in Mitochondrial DNA Repair</b></p> <p>P. Szykora, S. Kanno, M. Akbari, T. Kulkowicz, B. A. Bagfista, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croteau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr</p> <hr/> <p>Research Article</p> <p><b>Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment</b></p> <p>Tomoko Oyazaki-Torimaru, Tomohito Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hishiki, Hiroshi Morisaki, Makoto Suematsu, Yoji Andrew Minamishima</p> <hr/> <p>Research Article</p> <p><b>Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1</b></p> <p>Michael C. Armstrong, Sergej Sestak, Ahmed A. Ali, Hanan A. M. Sagni, Max Brown, Karen Baty, Achim Treumann, Martin Schröder</p>	<p><b>Author Corrections</b></p> <p>Author Correction</p> <p><b>Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"</b></p> <p>Eldar Zehorai, Rony Seger</p> <hr/> <p>Author Correction</p> <p><b>Correction for Lin et al., "Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway"</b></p> <p>Xi Lin, Xiangheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang</p> <hr/> <p>Author Correction</p> <p><b>Correction for Shah et al., "A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes"</b></p> <p>Zahid H. Shah, Shafig U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner</p> <hr/> <p>Author Correction</p> <p><b>Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"</b></p> <p>Kenzo Fukunaga, Youngho Kwon, Patrick Sung, Katsunori Sugimoto</p> <hr/> <p>Author Correction</p> <p><b>Correction for Copetti et al., "p65/RelA Modulates BECN1 Transcription and Autophagy"</b></p> <p>Tamara Copetti, Cosetta Bertoli, Emiliano Dalla, Francesca Demarchi, Claudio Schneider</p> <hr/> <p>Author Correction</p> <p><b>Correction for Travesa et al., "Repression of G<sub>1</sub>/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6"</b></p> <p>Anna Travesa, Tatjana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chahwan, David E. Lee, Noel F. Llewellyn, Curt Wittenberg</p> <hr/> <p><b>Retraction</b></p> <p>Retraction</p> <p><b>Retraction for Sorek et al., "Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase"</b></p> <p>Nadav Sorek, Limor Poraty, Hassana Stamborg, Elnat Bar, Efrain Lewinsohn, Shaul Yalovsky</p>	
---	--	---

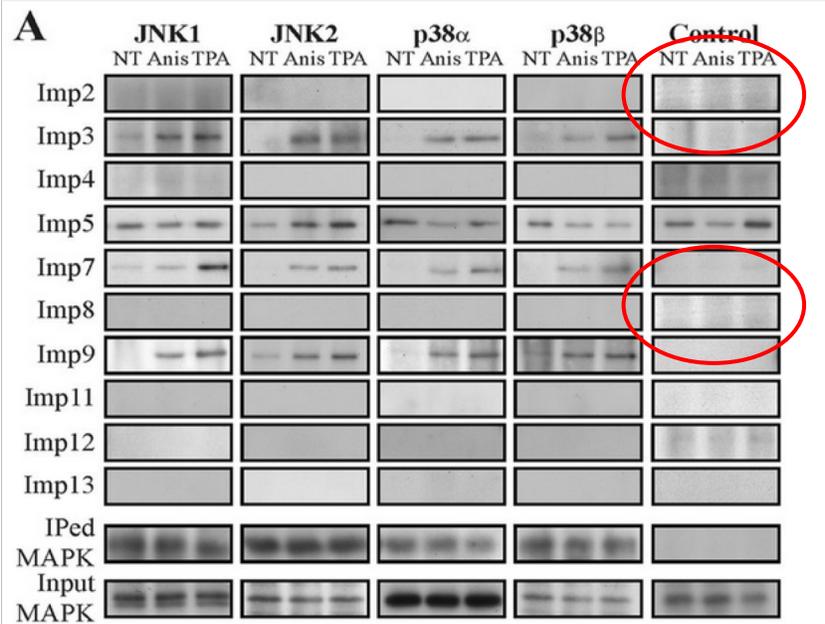
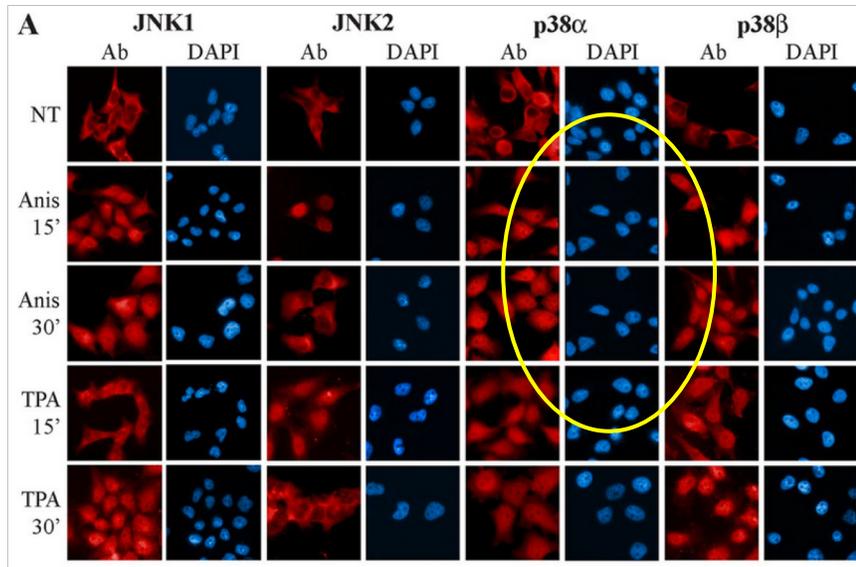
Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"

"Due to inadvertent errors in transferring images from Photoshop to Illustrator, there are **five mistakes in four figures**. We apologize for these mistakes and provide corrections below. The conclusions of the study remain unchanged by these corrections".

What's wrong with this picture?



# What's wrong with this picture?



10 months later ...

## Retraction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"

After a recent correction of some figures (E. Zehorai and R. Seger, *Mol Cell Biol* 37:e00215-17, 2017), **we were made aware of additional problems, including an extra lane in the gERK panel of Fig. 1C and inappropriate splicing in Fig. 8A.** Although these errors do not change the conclusions derived from these figures, we choose to retract this paper in the interest of maintaining accuracy in the published scientific literature. We apologize for these mistakes and any inconvenience they caused.

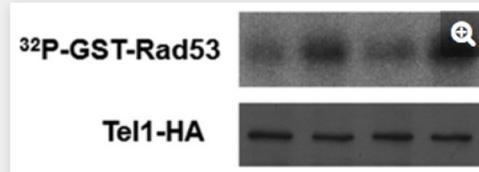
Research Articles	Author Corrections
<p><b>let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis</b> Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shuoheng Wang</p>	<p><b>Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"</b> Eldar Zehorai, Rony Seger</p>
<p><b>Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue</b> Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang</p>	<p><b>Correction for Lin et al., "Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway"</b> Xi Lin, Xiangsheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang</p>
<p><b>RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner</b> Kyoko Leno, Laila Valanegaj, Ashley Cassi, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karna, Lubov Timchenko, Nikolai Timchenko</p>	<p><b>Correction for Shah et al., "A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes"</b> Zahid H. Shah, Shaifq U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner</p>
<p><b>Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis</b> Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julietta Aprea, Chi Fang Chock, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis</p>	<p><b>Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"</b> Kenzo Fukunaga, Youshiro Kasai, Patrick Sung, Katsunori Sugimoto</p>
<p><b>DNA Polymerase Beta Participates in Mitochondrial DNA Repair</b> P. Sijona, S. Kannio, M. Albert, T. Kulkowicz, B. A. Bagfatta, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croblau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr</p>	<p><b>Correction for Copetti et al., "p65/RelA Modulates <i>BECN1</i> Transcription and Autophagy"</b> Tamara Copetti, Cosetta Bertoli, Emiliano Della Francesca Demarchi, Claudio Schneider</p>
<p><b>Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment</b> Tomoko Oyasu-Torimaru, Tomochiro Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hainki, Hiroshi Morisaki, Masako Suetaniisu, Yoji Andrew Minamishima</p>	<p><b>Correction for Travesa et al., "Repression of G<sub>1</sub>/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6"</b> Anna Travesa, Tatyana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chatwan, David E. Lee, Noel F. Loomes, Curt Wittenberg</p>
<p><b>Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1</b> Michael C. Armstrong, Sergej Seestak, Ahmed A. Ali, Hanan A. M. Sagini, Max Brown, Karen Baly, Achim Treumann, Martin Schröder</p>	<p><b>Retraction</b> <b>Retraction for Sorek et al., "Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase"</b> Nadav Sorek, Limor Poraty, Hasana Stenberg, Elnat Bar, Efrain Lewinson, Shaul Yalovsky</p>



Correction for Fukunaga et al.,  
 “Activation of Protein Kinase Tel1  
 through Recognition of Protein-  
 Bound DNA Ends”

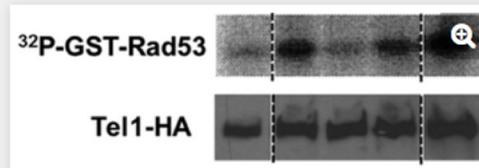


Page 1962, Fig. 2A: An incorrect Western blot image was inadvertently used. The correct version is shown below. Note that the Tel1-HA lanes are identical to the ones used in Fig. 1A and C.



[Open in new tab](#) | [Download powerpoint](#)

Page 1962, Fig. 2C: A Western blot image was inadvertently duplicated and presented. Since we could not locate the original image, a set of data directly relevant to the experiment is shown below.



### 3) Journal/funding agency policies

 AMERICAN SOCIETY FOR MICROBIOLOGY Journals

**ASM Ethical Publication Checklist for Authors**

**AVOID Authorship Disputes**

- Authors included on the byline meet the [ICMJE criteria for authorship](#).
- Everyone who meets the ICMJE criteria for authorship has been included in the byline.
- All authors agree to the order of authorship.
- All authors have reviewed the manuscript, whether primary submission or resubmission, and have approved submission to this journal.
- The corresponding author and/or the primary investigator(s), if different, have examined the raw data represented in the manuscript, affirm that such representations accurately reflect the raw data, and will ensure that the original data are preserved and retrievable for at least six years following publication.

**DISCLOSE Conflicts of Interest**

- All funding sources and nonfinancial support (including equipment, reagents, and services) have been acknowledged in the manuscript.
- All financial and personal interests that may be affected by the publication of this study have been declared in both the manuscript submittal letter and in the acknowledgment section of the manuscript.

**AVOID Duplicate Submissions and Publications**

- This manuscript is not currently under consideration for publication by another journal.

**If this manuscript is a resubmission of a previous submission to this or any other ASM journal:**

- The former manuscript control number has been included in the cover letter and on the submission form.
- Point-by-point responses to reviews of the previously submitted manuscript have been uploaded as a separate file.
- A complete copy of the revised manuscript, showing all changes, has been included as a marked up manuscript.

**AVOID inappropriate image manipulation**

- Submitted figures and tables show original data.
- Linear adjustments to contrast, brightness, and/or color, if used, are applied to the entire image.
- Unless fully disclosed, explained and demarcated (see the following bullet points), no portions of any figures have been:
  - Removed or deleted
  - Concealed
  - Duplicated (cut and pasted)
  - Added
  - Selectively enhanced
  - Repositioned
- Nonlinear adjustments made to images (e.g., changes to gamma settings) are fully disclosed and explained in the figure legend.
- Appropriate “[tooling](#)” has been used to demarcate individual images that have been compiled to make a figure. Descriptions and explanations for any such compilation are described in the figure legend.

## Journal policies – ASM / Molecular and Cellular Biology

- **Image manipulation.** Digital images submitted for publication may be inspected by ASM production specialists for any manipulations or electronic enhancements that may be considered to be the result of scientific misconduct based on the guidelines provided.

## PLoS Journals “Data availability” guidelines

- Acceptable Data-Sharing Methods
- Unacceptable Data Access Restrictions
- Explanatory Notes and Guidance
- Recommended Repositories
- Repository Inclusion Criteria
- FAQs for Data Policy
- PLOS Data Advisory Board

<https://journals.plos.org/plosbiology/s/data-availability>

## PLoS Journals “Data availability” guidelines

- Acceptable Data-Sharing Methods
- Unacceptable Data Access Restrictions
- Explanatory Notes and Guidance
- **Recommended Repositories**
- Repository Inclusion Criteria
- FAQs for Data Policy
- PLOS Data Advisory Board

### Cross-disciplinary repositories

- › Dryad Digital Repository
- › figshare
- › Harvard Dataverse Network
- › Open Science Framework
- › Zenodo

### Repositories by type

Biochemistry	Neuroscience	Social Sciences
Biomedical Sciences	Omics	Structural Databases
Marine Sciences	Physical Sciences	Taxonomic & Species Diversity
Model Organisms	Sequencing	Unstructured and/or Large Data

<https://journals.plos.org/plosbiology/js/data-availability>

#### Biomedical Sciences

- › [The Cancer Imaging Archive \(TCIA\)](#)
- › [ImmPort](#)
- › [Influenza Research Database](#)
- › [National Addiction & HIV Data Archive Program \(NAHDAP\)](#)
- › [National Database for Autism Research \(NDA\)](#)
- › [PhysioNet](#)
- › [SICAS Medical Image Repository](#)

#### Marine Sciences

- › [SEA scieNtific Open data Edition \(SEANOE\)](#)

#### Model Organisms

- › [The Arabidopsis Information Resource \(TAIR\)](#)
- › [Eukaryotic Pathogen Database Resources \(EuPathDB\)](#)
- › [FlyBase](#)
- › [Mouse Genome Informatics \(MGI\)](#)
- › [Rat Genome Database \(RGD\)](#)
- › [SmedGD](#)
- › [VectorBase](#)
- › [WormBase](#)
- › [Xenbase](#)
- › [Zebrafish Model Organism Database \(ZFIN\)](#)

#### Neuroscience

- › [Functional Connectomes Project International Neuroimaging Data-Sharing Initiative \(FCP/INDI\)](#)
- › [German Neuroinformatics Node/G-Node \(GIN\)](#)
- › [NeuroMorpho.org](#)
- › [OpenNeuro](#)

#### Omics

- › [ArrayExpress](#)
- › [Biological General Repository for Interaction Datasets \(BioGRID\)](#)
- › [Database of Interacting Proteins \(DIP\)](#)
- › [dbGAP](#)
- › [The European Genome-phenome Archive \(EGA\)](#)
- › [Gene Expression Omnibus \(GEO\)](#)
- › [GenomeRNAi](#)
- › [GPM DB](#)
- › [IntAct Molecular Interaction Database](#)
- › [MetabolLights](#)
- › [Metabolomics Workbench](#)
- › [NURSA](#)
- › [PeptideAtlas](#)
- › [ProteomeXchange](#)
- › [Proteomics Identifications \(PRIDE\)](#)

# Funding agency requirements

‘In discussing retention requirements, one often hears that institutions must retain records on federal awards “for three years”. There is just enough truth in that statement to get one into serious trouble. The key question is “three years from when?”’

NIH: “the retention period is three years from the date the final financial report is submitted”.

NSF: “records must be retained for three years after the submission of all required reports (research and other special reports)”.

[https://ori.hhs.gov/education/products/rcradmin/topics/data/tutorial\\_11.shtml](https://ori.hhs.gov/education/products/rcradmin/topics/data/tutorial_11.shtml)

## 4) Cornell resources: data.research.cornell.edu

The screenshot shows the homepage of the Cornell University Research Data Management Service Group. The page is organized into several sections:

- RESEARCH DATA MANAGEMENT SERVICE GROUP:** A header section with a navigation menu (Home, About, Services, Data Management Planning, Best Practices) and a brief description of the group's collaborative role in data management.
- How do I ...:** A dropdown menu with a selected option for "New Data Storage Finder", which includes a "Browse data storage" button.
- RECENT NEWS:** A section titled "Now available: OSF for Institutions at Cornell" dated 14 Dec 14, 2018, describing the Open Science Framework (OSF) as a free scholarly web tool.
- SERVICE SPOTLIGHT:** A section titled "Programming and coding" that offers help with programming from a variety of service providers on campus.
- CORNELL SERVICES:** A list of services including Collaboration tools, Data collection and analysis, Data sharing, High performance computing, Intellectual property and copyright, Metadata, Security, privacy, and confidentiality, Storage, backup, and recovery, and Support for researchers in the Life Sciences.
- BEST PRACTICES:** A list of best practices including Data citation, File formats, File management, Glossary of data management terms, Guide to writing a Data Management Plan (DMP), Guide to writing "readme" style metadata, Introduction to intellectual property rights in data management, Metadata and describing data, Preparing tabular data for description and archiving, Sharing data, and Support for researchers in the Life Sciences.
- CONSULTATIONS:** A section stating that ROMSG consultants are available year-round to answer questions, with contact options for in-person, Zoom, email, or phone.
- TWEETS:** A section titled "Tweets by @Cindmag".

### Describe your data

Answer these questions to help identify data storage services that are suitable for your needs. Checking these boxes will change the list of available services. If you are uncertain how to answer, leave the question blank to maximize your resulting options.

[Clear Answers](#)

**1. What is the classification of your data?**

Public

Sensitive / Moderate Risk

Confidential or Restricted / High Risk

HIPAA-Regulated

**2. Do you need backups, snapshots or replication of your data?**

I need one or more backup/snapshot copies of the data, and need to be able to restore data from previous points in time (high durability).

I need to have replicate copies of the data to minimize downtime (high availability).

**3. How much data do you have and how fast will it grow?**

Unlikely to exceed 1TB in 2 years

Greater than 1TB or likely to exceed in 2 years

**4. Do you have special performance needs?**

I am likely to have more than 1,000 files in a single directory within two years.

My data interactions demand high transaction or transfer rates.

**5. How are you expecting to access the data?**

Select data storage services you would like to compare. [Select All](#) [Clear Selections](#)

<p><b>Amazon Web Services Elastic Block Store</b></p> <p>Storage for use with Amazon EC2</p>	<p><b>Amazon Web Services Elastic File System</b></p> <p>Storage for use with multiple Amazon EC2 instances</p>	<p><b>Amazon Web Services Glacier</b></p> <p>Cloud based archival storage</p>	<p><b>Amazon Web Services S3</b></p> <p>Flexible, scalable object storage</p>	<p><b>BioHPC Cloud Storage for BioHPC lab computing services</b></p>
<p><b>CAC Archival Storage</b></p> <p>Single copy, non-mountable storage</p>	<p><b>CAC Red Cloud Storage</b></p> <p>Storage for Red Cloud compute instances</p>	<p><b>CISER Research Servers</b></p> <p>Storage for CISER computing environment</p>	<p><b>CUGIR</b></p> <p>Publicly shared geospatial data storage</p>	<p><b>CUL eCommons</b></p> <p>Publicly shared data repository</p>
<p><b>Cornell Box</b></p> <p>Online file sharing and collaboration</p>	<p><b>Cornell Restricted Access Data Center</b></p> <p>Storage for GRADC (confidential) computing environment</p>	<p><b>EZ-Backup Static Storage</b></p> <p>Archival storage and backup storage</p>	<p><b>Google Drive</b></p> <p>Cornell G Suite file storage and sharing</p>	<p><b>Kaltura Video Platform Service</b></p> <p>Flexible, scalable video and multi-media storage (customizable)</p>
<p><b>Kaltura Video on Demand</b></p> <p>Video and multi-media storage</p>	<p><b>LabArchives</b></p> <p>Online electronic lab notebook</p>	<p><b>Open Science Framework</b></p> <p>Online project management repository</p>	<p><b>Shared File Services</b></p> <p>File sharing between users and computers</p>	<p><b>Shared File Services - Confidential</b></p> <p>File sharing between users and computers for confidential data (non-WCM)</p>
<p><b>WCM Block Storage</b></p> <p>High performance storage attached to centrally hosted servers (WCM only)</p>	<p><b>WCM File Sharing</b></p> <p>Secure network storage (NFS/CIFS) for research computing (WCM only)</p>	<p><b>WCM Red Cloud Secure Storage</b></p> <p>Secure storage for Red Cloud compute instances (WCM only)</p>	<p><b>WCM Secure Remote Archive</b></p> <p>Secure, single copy, non-mountable storage (WCM only)</p>	

## 5) Intro to case studies

### Case 1: Managing research data.

- Prof Carr and his unavailable image analysis computer program.
- “How does this general scenario apply to your own type of research”?
- Unspecified: how old is the paper?
- Unspoken: Involves student/advisor relationship.

**Case 2: Data sharing.**

- Henryk and his large database of infectious agent outcomes.
- “How does this general scenario apply to your own type of research”?
- Unspecified: How many others contributed to the database?
- Unspoken: involves postdoc/advisor relationship.

Thank you! CU Office of Research Integrity and Assurance

Amita Verma  
Myles Gideon  
Bridget MacRae  
Guilaine Senecal