THE JOURNAL OF CELL BIOLOGY

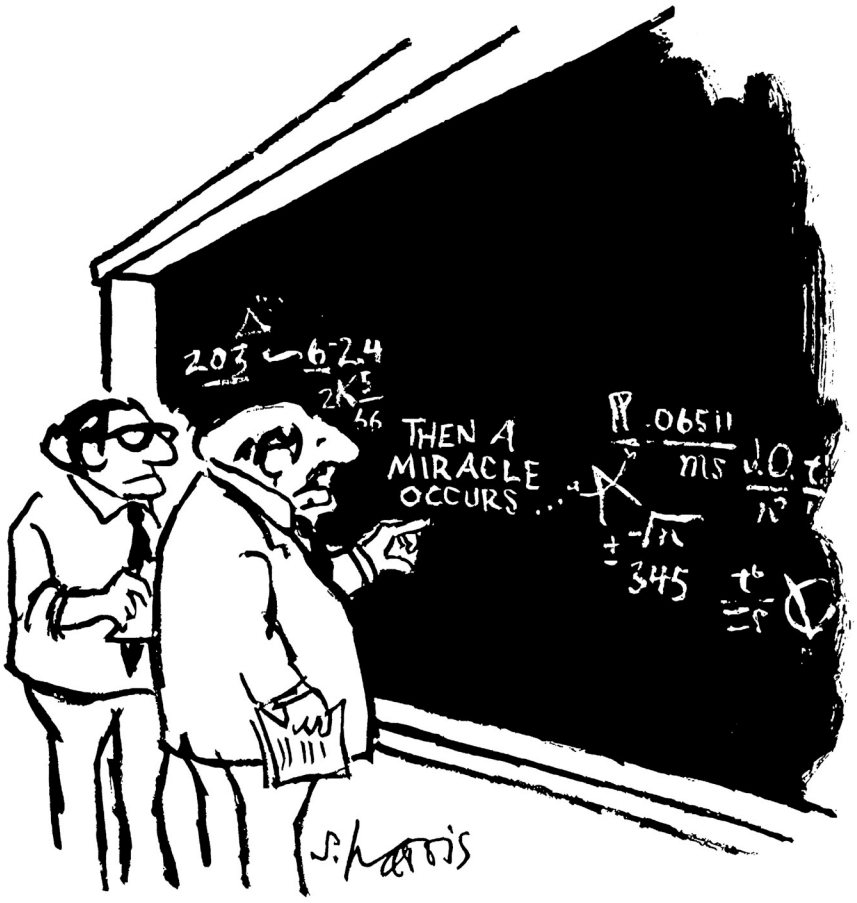
Reproducibility and cell biology

Kenneth M. Yamada[1,3] and Alan Hall[2,4]

[1]Academic Editor and [2]Editor-in-Chief, *The Journal of Cell Biology*
[3]Laboratory of Cell and Developmental Biology, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892
[4]Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065

Growing concerns about the reproducibility of published research threaten to undermine the scientific enterprise and erode public trust. Conscientious application of "best practices" for the generation and reporting of research, along with post-publication access to raw data and other research materials, will protect the integrity of the research literature.

Research reproducibility is an increasingly major concern in biomedical research. It is crucial to the scientific enterprise, not only because it underpins the accuracy and integrity of our published literature, but also because basic research increasingly contributes to the development of innovative clinical therapies. Recent accounts describe frustrating experiences of pharmaceutical companies attempting to build upon basic and translational research studies, notably in cancer biology. These companies encountered surprisingly low reproducibility (<25%) of published work (Prinz et al., 2011; Begley and Ellis, 2012). Ironically, the validity of these commentaries cannot be assessed because they do not provide primary research data. However, others have raised similar concerns about reproducibility in a significant proportion of published papers, and this has in turn raised concerns at funding agencies (Ioannidis, 2005; Landis et al., 2012; Collins and Tabak, 2014). Recent dialogue about the potential causes of low reproducibility has focused on training, journal practices, pressure to publish quickly, lack of appropriate controls, and inefficient self-correction. Are these perceived problems real, and if so how should the scientific community respond in order to enhance reproducibility?



"I think you should be more explicit here in step two."

© Sidney Harris The New Yorker Collection/The Cartoon Bank

Drawing on our many decades of service as editorial board members of *JCB*, we would like to offer a few impressions. First and foremost, our experience is that reproducibility problems most likely arise from mistakes in performing and reporting original research, or from the inability to reproduce specialized methods, rather than from scientific misconduct. One approach that provides a semiquantitative window into the issue of data integrity in individual papers was pioneered by *JCB* in 2002 (Rossner, 2002). Since then, our production team

Correspondence to Kenneth M. Yamada: kenneth.yamada@nih.gov; or Alan Hall: halla@mskcc.org

© 2015 Yamada and Hall This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/).

Downloaded from http://rupress.org/jcb/article-pdf/209/2/191/1587811/jcb_201503036.pdf by guest on 08 February 2026

The Rockefeller University Press
J. Cell Biol. Vol. 209 No. 2 191–193
www.jcb.org/cgi/doi/10.1083/jcb.201503036

and editorial staff have screened image and gel data in >4,000 papers that had been approved for publication by both external reviewers and our own scientific editors. Over these past dozen years, ~15% of papers (~600) sent to production for publication contained inappropriate presentation of data. Most were subsequently corrected by the authors after reformatting their primary data to ensure confidence in its authenticity. Only 1% of papers (~40) could not be published due to serious discrepancies, which might reflect the level of scientific misconduct in manuscripts submitted to *JCB*. There is still a significant gap that needs to be addressed between the 15% of problems involving author errors in presentation and the 50% or more of irreproducibility suggested in some commentaries.

The analysis from Madsen and Bugge in this issue provides a new type of hard data identifying concerns about reproducibility. The authors looked for agreement among 200 publications addressing a single, heavily studied and important research question. They directly compared findings from numerous peer-reviewed, published studies to determine the identity of the cell or tissue type that produces the proteases mediating matrix degradation in four common tumors. There were major discrepancies between the conclusions of these studies, even among those that appeared to use appropriate controls. Madsen and Bugge (2015) identify a series of possible sources of this vexing lack of reproducibility and recommend approaches for future studies.

In light of this analysis and various other publications and commentaries, what can the cell biology community—and particularly we at *JCB*—do to help us solve this troubling problem of reproducibility?

Best practices

We suggest that authors, reviewers, and editors seriously consider adopting the following best practices for reproducibility. Many of these used to be integral to any serious study, but now seem to be given less consideration.

Provide complete methodology. Each study should provide detailed descriptions of methods to permit replication by other laboratories, and authors should be prepared to share all unique research materials after publication. Both are currently mandated by *JCB* and by many funding agencies. Experiments should include a sufficient number of independent replications when practical, sufficiently large sample sizes with convincing magnitudes of effects (or no effect), and, when appropriate, other best-practice approaches including randomization, observer blinding, validation of cell lines, and appropriate statistical analysis as described in recent guidelines from journal editors (http://www.nih.gov/about/reporting-preclinical-research.htm).

Apply independent approaches. Key conclusions should be evaluated and supported if possible by independent means of analysis; for example, beyond showing imaging data, providing quantification by immunoblotting, testing conclusions by genetic manipulation, and ideally providing at least some insight into mechanisms.

Deposit primary data. The raw data underlying each published conclusion should ideally be readily available to both reviewers and readers after publication. There are two major benefits from such public deposit of raw data: (1) The research community can be assured that the study rests on sufficiently strong data, and it will reduce the temptation to show only the best results ("cherry picking") or inappropriately manipulate data. (2) Other researchers may be able to use that data for further analysis, of course under appropriate guidelines analogous to those in place for primary genomic data. A major question is where should the large amounts of primary data be archived? Depositing of complex data in public databases such as GenBank, the Gene Expression Omnibus (GEO), Peptide Atlas, and the Protein Data Bank is well established. However, there are very few repositories for primary imaging data, as well as the numerical data used to generate tables and graphs. Ideally, the raw or minimally processed images or other forms of primary data underlying each of the repeats of key experiments (not merely the figures shown) should be deposited along with associated metadata descriptive information. *JCB* has led the way in this regard by hosting the JCB DataViewer: a cross-platform repository for large amounts of raw imaging and gel data (Williams et al., 2012), and potentially other forms of data for its published manuscripts. At present, data deposition is recommended but not mandated. The JCB DataViewer currently contains 4 terabytes of data, and it can hold considerably more. More generally, many philosophical and practical issues concerning publication of raw data are under wide discussion, including the types of raw imaging data appropriate for deposition, the need for standardizing the data presented, and centralized databanks (e.g., see Kratz and Strasser, 2014). Ultimately, funding agencies or academic institutions should consider supporting large primary image data repositories for the full range of biomedical journals.

Resolve failures to reproduce. Researchers who encounter discrepancies between their conclusions and published work, or those whose work cannot initially be replicated, should make good-faith efforts to resolve the differences by working together with other laboratories to try to determine the sources of nonreproducibility. Even though this cooperative approach may sometimes prove difficult for interpersonal or political reasons, it should be attempted. Such comparisons can lead to novel collaborative findings of broad interest (Wolf et al., 2013; Hines et al., 2014).

Common sources of nonreproducibility

Three specific technical approaches are likely to be the source of much experimental irreproducibility:

Antibodies. Studies that depend on antibodies require validation to establish the specificity and sensitivity of the reagent. Western immunoblotting generally only validates the ability of an antibody to recognize an SDS-unfolded protein, and not necessarily the native 3D protein structure often needed for immunofluorescence (IF) and immunoprecipitation (IP) studies. Consequently, appropriate specificity controls with careful validation are essential. This is clearly more important for IF, as confirmation of molecular weight is not available as in Western blots and IPs. It is not good practice to assume antibody specificity just because the company selling it calls

Downloaded from http://rupress.org/jcb/article-pdf/209/2/191/1587811/jcb_201503036.pdf by guest on 08 February 2026

it "specific." Experiments using several independent antibodies to the same molecule are more convincing than using a single antibody.

Small molecule inhibitors. Most small molecule inhibitors used today seem to be designated "specific," yet it is unlikely that any small molecule really fits that description. Our view is that people use this word carelessly or, even worse, to obfuscate and thereby avoid using alternative approaches to exploring a pathway. Some insight into this problem can be seen in various comparative studies, notably with kinase inhibitors (Bain et al., 2007).

RNAi. RNAi approaches have revolutionized our ability to interrogate pathways in diverse organisms, but artifactual results due to "off-target" effects have been widely reported in signaling pathways (Schultz et al., 2011) and in cell cycle control (Tschaharganeh et al., 2007). Despite this, papers still appear describing a phenotype induced by a single RNAi. An initial rule of thumb has been to use at least two siRNAs for each gene to demonstrate similar effects, but while this is expedient and often satisfies reviewers, it is likely to account for much irreproducibility in the literature. Three or four independent siRNAs are safer, and of course, the best approach is to show "rescue" from the RNAi effects after experimental expression of a cDNA containing an siRNA-resistant mutation.

A final comment relates to the importance of transmitting general awareness of the problem of reproducibility, and of having this discussion with trainees as they enter the research community. We need to better educate our students, postdoctoral fellows, and staff about these issues. The National Institutes of Health is holding workshops on research reproducibility (e.g., http://videocast.nih.gov/summary.asp?Live=15277&bhcp=1) and is supporting research toward this aim through the R25 funding mechanism. We at *JCB* will continue to draw upon the expertise and judgment of our scientific editorial board, our reviewers, and our production and editorial staff at the Rockefeller University Press to ensure that only the highest-quality work is published.

References

Bain, J., L. Plater, M. Elliott, N. Shpiro, C.J. Hastie, H. McLauchlan, I. Klevernic, J.S. Arthur, D.R. Alessi, and P. Cohen. 2007. The selectivity of protein kinase inhibitors: a further update. *Biochem. J.* 408:297–315. http://dx.doi.org/10.1042/BJ20070797

Begley, C.G., and L.M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature*. 483:531–533. http://dx.doi.org/10.1038/483531a

Collins, F.S., and L.A. Tabak. 2014. Policy: NIH plans to enhance reproducibility. *Nature*. 505:612–613. http://dx.doi.org/10.1038/505612a

Hines, W.C., Y. Su, I. Kuhn, K. Polyak, and M.J. Bissell. 2014. Sorting out the FACS: a devil in the details. *Cell Reports*. 6:779–781. http://dx.doi.org/10.1016/j.celrep.2014.02.021

Ioannidis, J.P. 2005. Why most published research findings are false. *PLoS Med.* 2:e124. http://dx.doi.org/10.1371/journal.pmed.0020124

Kratz, J., and C. Strasser. 2014. Data publication consensus and controversies. *F1000 Res.* 3:94.

Landis, S.C., S.G. Amara, K. Asadullah, C.P. Austin, R. Blumenstein, E.W. Bradley, R.G. Crystal, R.B. Darnell, R.J. Ferrante, H. Fillit, et al. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 490:187–191. http://dx.doi.org/10.1038/nature11556

Madsen, D.H., and T.H. Bugge. 2015. The source of matrix-degrading enzymes in human cancer: Problems of research reproducibility and possible solutions. *J. Cell Biol.* 10.1083/jcb.200101025

Prinz, F., T. Schlange, and K. Asadullah. 2011. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10:712. http://dx.doi.org/10.1038/nrd3439-c1

Rossner, M. 2002. Figure manipulation: assessing what is acceptable. *J. Cell Biol.* 158:1151. http://dx.doi.org/10.1083/jcb.200209084

Schultz, N., D.R. Marenstein, D.A. De Angelis, W.Q. Wang, S. Nelander, A. Jacobsen, D.S. Marks, J. Massagué, and C. Sander. 2011. Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-β pathway and reveal microRNA regulation of TGFBR2. *Silence*. 2:3. http://dx.doi.org/10.1186/1758-907X-2-3

Tschaharganeh, D., V. Ehemann, T. Nussbaum, P. Schirmacher, and K. Breuhahn. 2007. Non-specific effects of siRNAs on tumor cells with implications on therapeutic applicability using RNA interference. *Pathol. Oncol. Res.* 13:84–90. http://dx.doi.org/10.1007/BF02893482

Williams, E.H., P. Carpentier, and T. Misteli. 2012. The JCB DataViewer scales up. *J. Cell Biol.* 198:271–272. http://dx.doi.org/10.1083/jcb.201207117

Wolf, K., M. Te Lindert, M. Krause, S. Alexander, J. Te Riet, A.L. Willis, R.M. Hoffman, C.G. Figdor, S.J. Weiss, and P. Friedl. 2013. Physical limits of cell migration: control by ECM space and nuclear deformation and tuning by proteolysis and traction force. *J. Cell Biol.* 201:1069–1084. http://dx.doi.org/10.1083/jcb.201210152

Downloaded from http://rupress.org/jcb/article-pdf/209/2/191/1587811/jcb_201503036.pdf by guest on 08 February 2026