How many authors does it take to publish a high profile or classic paper?

Hai Rao[a,*], Jun Che[a], Changhui Yin[a], Yunfan Zhou[a], Yifang Ma[b], and Ruilin Tian[c,*]
[a]Department of Biochemistry and [c]Department of Medical Neuroscience, School of Medicine, and [b]Department of Statistics and Data Science, College of Science, Southern University of Science and Technology, Shenzhen, People's Republic of China

ABSTRACT **Although the process of publishing a scientific paper has gotten simpler, it is increasingly difficult to publish a paper in high profile journals. We have analyzed the publishing data in the cell biology field and found several alarming trends developing over the last two decades. There is an emerging divide between scientist-run journals and professional-run high profile journals. How did this happen? What should we do? The core issue is whether the current standard for high profile journals hurts rather than helps the scientific discovery process. In this regard, we suggest that the editors and scientists should direct their focus on the potential impact and rigor of the work instead of the "perfection" or "completeness" of the study.**

Monitoring Editor
Doug Kellogg
University of California, Santa Cruz

Received: Apr 26, 2022
Revised: Jul 26, 2022
Accepted: Aug 2, 2022

The advent of the internet and technology has made the paper publication process easier and more efficient overall. Gone are the days where we stitch panels together for a figure, mail multiple copies to a journal, and wait for the fateful decision letter.

While the biggest benefit to the scientific publication process from technological advances arguably could be the speed of manuscript transfer and data dissemination, the worst by-product might have been the proliferation of scientific journals and the increasing demand of data. As a whole, is the publication enterprise now better or worse than before? Although it has vastly improved in many aspects, one important area, the acceptance by a high profile journal, is getting harder and harder. During the discussion of two classic papers, one Nobel-worthy *Science* paper on GFP expression coauthored by Chalfie and Prasher in 1994 (Chalfie *et al.*, 1994) and one highly influential *Nature* paper describing yeast two-hybrid assay by Fields in 1989 (Fields and Song, 1989), the graduate students taking the course on Scientific Thinking and Communication taught by one of us (H.R.) wondered aloud where the supplemental results are. Indeed, by current standards, these papers would not be accepted by their respective journals despite the impact.

Chalfie's paper contains three single-panel figures in two pages (Chalfie *et al.*, 1994), which would be viewed as control experiments these days. Chalfie, Prasher and three coworkers set out to determine whether GFP would produce fluorescence outside jellyfish. They expressed GFP in heterologous systems and demonstrated that GFP does not affect cell growth and can emit green fluorescence in *Escherichia coli* and *Caenorhabditis elegans* (Chalfie *et al.*, 1994). In terms of its application, Chalfie *et al.* correctly envisioned that GFP could help track gene expression, protein localization and trafficking, cell movement, etc. (Zimmer, 2009; Tsien, 2010). Fields' paper has two figures, one of which is the scheme devised and the other shows five constructs used in the experiments; the actual data is in a table (Fields and Song, 1989). In less than two pages, Fields and Song demonstrated that two separable functional domains of a transcription activator Gal4 brought back by two interacting proteins can reconstitute the transcriptional activity in yeast cells (Fields and Song, 1989). The implication was clear; given the technical limitation at that time, this simple method had revolutionized the way to look for protein-binding partners (Vidal and Fields, 2014). Classic papers don't need many figures. Both of these papers were published within 2 months of submission, despite the snail mail way of communication back then. This exemplifies how a quicker turnaround for publication results in more benefits for science.

While both of these papers stood the test of time, the editors and reviewers of these two journals today are unlikely to accept these manuscripts, visions, or predictions as they were published because they would demand the support of additional data. The

DOI:10.1091/mbc.E22-04-0140
*Address correspondence to: Hai Rao (raoh@sustech.edu.cn); Ruilin Tian (tianrl@sustech.edu.cn).
Abbreviation used: GFP, green fluorescent protein.

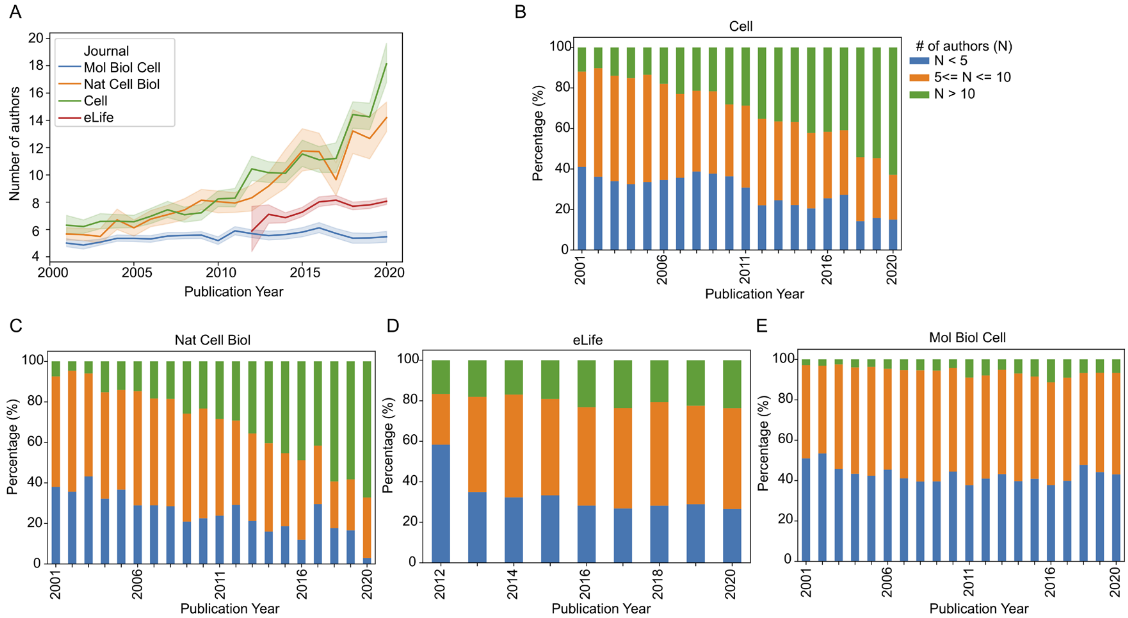

FIGURE 1: Trends in various journals in cell biology. (A) The annual average number of authors in *Cell*, *Nature Cell Biology*, *Molecular Biology of the Cell*, and *eLife* in the last 20 years. Note that *eLife* started in 2012. The shaded area represents the 95% confidence interval. (B–E) The percentage of various number of groups of authors in four journals in the years indicated. Raw publishing data from 2001 to 2020 were obtained from PubMed (https://pubmed.ncbi.nlm.nih.gov/). Nonresearch articles, including Reviews, Comment, Editorial, etc., were excluded from further analysis. A custom Python script was used for data analysis and visualization.

resulting paper may be more "solid" but the impact would be practically the same. This is the sad state we are in. The quick publication of these two papers with minimum sets of data was widely beneficial to the entire life science community, even changing the career and trajectory of some researchers. Whereas these two short papers are regarded as classics, many papers currently published in these so-called "high impact" journals are certainly more "complete" and "solid" but can hardly reach the height and broad impact of the papers on GFP and two-hybrid brought to the biological research (Tsien, 2010; Vidal and Fields, 2014).

Not coincidentally, a quick glance at major journals would show the expanding number of figures and the longer list of authors over the years (Vale, 2015; Cordero *et al.*, 2016; Fortunato *et al.*, 2018). Simply looking at the March 17 issue of *Nature*, there are 14 papers related to life science; the numbers of authors range from 2 to 61 with an average of 16.6, consistent with the trend shown below for *Cell* (Figure 1); regular and extended figures together span from 7 to 16, averaging 12.3; total panels in these figures go from 14 to 113 with 60 as the average; total pages range from 9 to 24, averaging 20.1. This snapshot is a fair representation of the current state of high profile journals, despite some deviations. For example, it is not unusual to see papers over 50 pages. Nevertheless, if these *Nature* papers were published in their full length as in the past, each issue of *Nature* would practically be a book. Perhaps for this reason, *Nature* has eliminated the research "Letter" portion as all *Nature* papers are now a longer "Article" with extended figures and supplemental data. One can't help but wonder whether many classic research papers would be published in their respective journals today with their original content.

Looking at major journals in life sciences over the last 20 years, a clear trend emerges of the number of authors for a given research paper. We chose to examine the journals in cell biology annually of *Cell*, *Nature Cell Biology*, and *Molecular Biology of the Cell* from 2001 to 2020. Each of these journals represents a different category with *Cell* being the top of its field, *Nature Cell Biology* a bit below, and *Molecular Biology of the Cell* well regarded for its solid work. Whereas *Cell* and *Nature Cell Biology* are for-profit journals run by professional editors, *Molecular Biology of the Cell* is a nonprofit journal of the American Society of Cell Biology run largely by working cell biologists. In the last two decades, the average number of authors has nearly tripled in *Cell* and *Nature Cell Biology* but remained largely flat in *Molecular Biology of the Cell* (Figure 1A). At the beginning of this millennium, the average numbers of authors per paper in these journals were close, between five and seven. It was not hard to find papers with two or three authors in any one of these journals; in fact, 40% of papers in all three journals had less than five authors. While the pattern held for the first 10 years, the split among these journals started to develop around 2012, with an accelerating pace and widening gap ever since (Figure 1A). Now, it is exceedingly rare to find a paper in *Cell* or *Nature Cell Biology* with less than five authors. Whereas the papers with one to four authors in *Molecular Biology of the Cell* still account for 40%, the numbers for *Cell* and *Nature Cell Biology* have dropped to ~15 and 5%, respectively (Figure 1, B–D). The biggest increase in these two journals is the papers with more than 10 authors, going from ~10% to ~60% in *Cell* and *Nature Cell Biology* (Figure 1, B and C). For the papers with more than 10 authors in *Cell*, it took ~6 years to go from ~10% to 20%, 3 years to jump from 20% to 30%, 5 years from 30% to 40%, 3 years from 40% to 50%, and 2 years from 50% to 60% (Figure 1B). With no change in sight, we are really not that far off from days where 90% of the papers in high profile journals will have more than 10 authors. While *Nature Cell Biology* shares the similar pattern as *Cell*, scientist-run journals, like *Molecular Biology of the Cell* and *eLife*, have remained steady (Figure 1, D and E). A rough analysis of similar journals including *Science*, *Nature*, *Genetics*, and *Journal of Biological Chemistry* reveals the same trend and divide between scientist-run journals and professional-run journals.

This trend is hardly unique only to these journals and holds true to many journals outside of life science as well (Fortunato *et al*., 2018). It begs the question: are we better off with the current standard for high profile journals? Along with the seemingly enhanced demand for the quality of papers, research misconducts and retractions are on the rise as well. Is this correlation merely coincidental or tightly coupled? We are in an era where high profile papers are directly tied to job opportunity, promotion, awards, and so on. In the quest for perfection or completion, could scientists resist the temptations? Knowing the stakes, if the referees or editors insist to have certain data despite the huge amount of solid data gathered already, is it surprising that a scientist may "wobble" a little bit to "get" it, as significant or insignificant the data itself may be? As high profile papers get longer, a number of issues may arise, including difficulty in sorting the credits of the many authors involved, the scooping of a paper during a long revision process, and the time extension for students to graduate (Vale, 2015; Cordero *et al*., 2016; Fortunato *et al*., 2018); some, but not all, of these concerns have been alleviated with the recent emergence of various preprint servers (e.g., bioRxiv; Hoy, 2020; Penfold and Polka, 2020).

How did we get here? Who is to blame? Although the trend holds true for nearly all journals, its rise is often steeper with high profile journals (Figure 1; Vale, 2015; Fortunato *et al*., 2018), which tend to be for-profit and run by professional editors. A submitted manuscript is first evaluated for its suitability for the respective journal. Whereas the decision is largely made by working scientists in nonprofit journals, it is mostly rendered by professional editors in high profile for-profit journals. Interestingly, in these highly selective journals, this process endows quite a bit of power to the editors, many of whom are relatively young and turned to the profession often after a few years in a reputable laboratory as a postdoctoral fellow. On the one hand, these editors can grow to have a broader view and deeper understanding of science. On the other hand, the editors can play an outsized role in shaping the direction and trend of science and its fields. As high profile papers often lead to better jobs and more professional recognition, these editors are treated as potential "king-makers" in some corners of the scientific community.

The next line of the decision process is mainly based on the reviews from working scientists in relevant areas though the editors may push through or block the paper at times. There is little doubt that the list of additional experiments and the time from submission to acceptance has gotten longer and longer over the years. One of the authors (H.R.) remembers how a senior faculty commented 20 years ago, that as a Nationa Institutes of Health study section member, he regarded a *Cell* paper as two *Journal of Biological Chemistry* papers in his evaluation process. This was not a unique view then. When H.R. was a graduate student around 25 years ago, he and his thesis advisor at one time considered combining two "regular" manuscripts in preparation to submit to *Cell*. A *Cell* paper now can easily be split into five or more quality papers. Indeed, a manuscript from H.R. was reviewed but rejected by *Molecular Cell* around 10 years ago, but was later developed into four quality papers, including one in *Proceedings of the National Academy of Sciences of the United States of America*.

The most important part of the publication process (i.e., data preparation, manuscript submission, review) is run by scientists. However, we have managed to make it increasingly difficult to publish in high profile flagship journals (Vale, 2015; Fortunato *et al*., 2018). Are the papers in these prestigious journals better quality and more solid than they were 20 years ago? Looking at the GFP and two-hybrid papers above, we know the impact of the paper is not based on the number of authors or figures in a paper. Some journals like *eLife* and *PLos Biology* have tried to turn the tide with some success (Barbour and Patterson, 2006; Schekman *et al*., 2012). It is worth noting that the trends we have seen are continuing upwards with no sign of slowing down or plateauing (Figure 1). We are really going to see the research articles in high profile journals with more than 50 figures and 100 pages consistently in the not too distant future, aren't we? After all, 20 years ago, few would think a *Nature* paper would have this many authors and figures in the 2020s.

How are we going to get out of this trap we set up for ourselves and future generations? Scientists are the main player in this system and should make it right, instead of being dictated by auxiliary players. An obvious way is to devalue the high profile for-profit journals. Keep in mind that the journals were created to disseminate information, not for obtaining prestige and awards. A sign of a reputable institution is a fair evaluation system that emphasizes the novelty and impact of their scientists and their work, not the meaningless metrics (e.g., impact factor or CiteScore) of the journals published. In 2013, hoping to rid of damaging effects by the scientific publishing industry, R. Schekman called for boycotting *Cell, Nature*, and *Science*. This is the right path to take for any scientist, but difficult and especially risky for young scientists without institutional support. The challenge here lies in an effective system to evaluate scientists and their research, which requires time and patience, and can be done right by their peers instead of journals or bureaucrats.

Another way to get out of this trap is to bring the anonymous editors of all journals to the front. In many scientist-run journals (e.g., *Journal of Biological Chemistry*, *Molecular Biology of the Cell*), the editors handling the manuscripts are indicated along with the papers, which can serve as a seal of approval from a senior scientist. On the other hand, the professional editors dealing with the manuscript at the highly selective journals are not listed in the publication. The editors often play a significant role behind the scene in terms of working with the authors and reviewers to usher the manuscript through the process and sometimes make key decisions along the way. However, they are often not appropriately recognized for their work. Listing their names in the papers gives them the recognition they deserve and also heightens the responsibility/accountability associated.

Last but not the least, the impact of a paper mostly comes from its significance, not the amount of the data. Historically, scientific breakthroughs usually come from small laboratories; yet the publishing trend with high profile journals favors big laboratories. Maybe we should go back to our roots; it used to be the other way around where the papers in high profile journals like *Science* and *Nature* actually often had less figures and pages than those in journals like *Molecular Biology of the Cell* or *Journal of Biological Chemistry*. In assessing a manuscript for a high profile journal, scientists should go back in time to put more value in potential impact and novel insights over "perfection," which is at odds with the nature of research which is to establish, revise, and destroy the paradigms or "truth." Studies suggest that too many papers in a short time may hurt rather than promote the field (Chu and Evans, 2021). Similarly, it is possible that large amounts of data in a high profile paper may restrict rather than stimulate our imagination and potential creative extension of the work. Interestingly, in the graduate student training today, we ask our students to pick out the most critical data in the paper they read. Even in the *Nature* papers that average 12 figures with 60 panels total, it is often enough to pick out 3 figures with 6 panels to comprehend the essence of the paper. Do we really need 12 figures with 60 panels? The authors and reviewers already deemed some of the data as not as critical, which were published in the supplemental section; few scientists would spend much time downloading them, not to mention reading them. It is perhaps better to include critical, but not ancillary, data that conveys the novelty, impact, and rigor of the work. Speedy publication of a high impact paper was key to the success of high profile journals, and we should keep it that way.

ACKNOWLEDGMENTS

We thank R. Schekman, J. Polka, P. Rodgers, D. Kellogg, and Y. Rao for insightful inputs and S. Rao for editorial assistance. H.R. is supported by the National Natural Science Foundation of China (Grant no. 82170159). R.T. is supported by grants from the National Natural Science Foundation of China (Grants no. 32100766 and no. 82171416).

REFERENCES

Barbour V, Patterson M (2006). Open access: the view of the Public Library of Science. J Thromb Haemost 4, 1450–1453.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994). Green fluorescent protein as a marker for gene expression. Science 263, 802–805.

Chu JSG, Evans JA (2021). Slowed canonical progress in large fields of science. Proc Natl Acad Sci USA 118, e2021636118.

Cordero RJ, de Leon-Rodriguez CM, Alvarado-Torres JK, Rodriguez AR, Casadevall A (2016). Life science's average publishable unit (APU) has increased over the past two decades. PLoS One 11, e0156983.

Fields S, Song O (1989). A novel genetic system to detect protein-protein interactions. Nature 340, 245–246.

Fortunato S, Bergstrom CT, Borner K, Evans JA, Helbing D, Milojevic S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, *et al.* (2018). Science of science. Science 359, doi:10.1126/science.aao0185.

Hoy MB (2020). Rise of the Rxivs: how preprint servers are changing the publishing process. Med Ref Serv Q 39, 84–89.

Penfold NC, Polka JK (2020). Technical and social issues influencing the adoption of preprints in the life sciences. PLoS Genet 16, e1008565.

Schekman R, Patterson M, Watt F, Weigel D (2012). Launching eLife, Part 1. Elife 1, e00270.

Tsien RY (2010). The 2009 Lindau Nobel Laureate Meeting: Roger Y. Tsien, Chemistry 2008. J Vis Exp 35, 1575.

Vale RD (2015). Accelerating scientific publication in biology. Proc Natl Acad Sci USA 112, 13439–13446.

Vidal M, Fields S (2014). The yeast two-hybrid assay: still finding connections after 25 years. Nat Methods 11, 1203–1206.

Zimmer M (2009). GFP: from jellyfish to the Nobel prize and beyond. Chem Soc Rev 38, 2823–2832.