

What Defines “Priority of Discovery” in the Life Sciences?

Ronald D. Vale¹ and Anthony A. Hyman²

¹Dept. of Cellular and Molecular Pharmacology, University of California San Francisco and
The Howard Hughes Medical Institute, San Francisco, USA

²Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

Address correspondence to: vale@ucsf.edu and hyman@mpi-cbg.de

Summary

A scientist's job is to make a discovery and then broadly disseminate this new knowledge, so as to benefit the scientific mission and society overall. In exchange for releasing this knowledge, the scientist wishes to be acknowledged for their original contribution. This pact, which is embodied in the term "priority of discovery", is crucial to the process, culture and reward system of science. Yet for something so fundamental to the scientific enterprise, the rules behind "priority of discovery" are rarely discussed, written about, or even taught. Here, we break down "priority of discovery" into two steps: 1) disclosure, in which the discovery is released from an individual or small group of scientists to the worldwide community, and 2) validation, in which other scientists assess both the accuracy and importance of the work. Scientific disclosure is a discrete event, although we argue it has different rules and meaning from a public disclosure as defined by patent law. Acquiring validation is a more complex process that starts with peer review and is followed by an extended period in which the experiments are repeated and the concepts are considered. Currently, in the life sciences, both of these steps are embodied in publication through a peer-reviewed journal. However, we propose that biologists could be better served by separating the steps, with the first step of disclosure occurring through a preprint server, followed by a second step of validation occurring through a journal publication or potentially other community-recognized mechanisms that achieve the same goals. This division of these steps is embedded in the practices of the physics and mathematics communities. We also believe that "priority" is not simply a race for the first time-stamp, but rather that quality plays a critical role in how scientists ultimately judge discoveries.

Introduction

A scientist conducts experiments, analyzes the data, and then arrives at a new understanding of the natural world; we use the term “discovery” to refer to this collective process. If the work is original, the evidence accurate, and the interpretation clear and well-supported, it is expected that other scientists will acknowledge and cite the work in a reasonable manner, thus solidifying the original investigator’s “priority” for this discovery.

This straightforward and deceptively simple explanation of “priority of discovery” masks considerably complexities when reduced to practice. Indeed, disputes over “priority” have permeated every epoch of modern science. The classic article on the priority of discovery by Robert Merton (Merton, 1957) describes how many of the early scientific giants in physics and chemistry (Galileo, Newton, Hooke, Cavendish, Lavoisier, Watt, Faraday, etc) were embroiled in battles over priority. In biology, Darwin’s and Wallace’s independent conception of natural selection as a driving force for evolution provides a fascinating case study, which reveals the complexities of “priority” even when the scientists themselves act benevolently and respectfully (Merton, 1957). Debates over “priority” continue today, from the mildly aggravating “they should have cited my paper” to deliberations over Nobel Prizes. A recent article on the history of CRISPR and genome editing highlights how interpreting priority of discovery remains a complex issue and can trigger wide-spread reactions within the scientific community (Lander, 2016). Indeed, as long as human nature persists and the scientific enterprise places a premium on original work, then controversies and angst over “priority” will remain an inevitable component of science.

It is worthwhile reconsidering the issue of “priority” in light of the many changes that have occurred in the scientific enterprise over the last several decades. With many more scientists on the planet, all accessing the same information and many seeking to answer reasonably evident next questions, competitive situations arise more frequently. An original idea that so completely separates one scientist from all of his/her contemporaries, like “Relativity”, is more likely to be the exception rather than the rule. Furthermore, both the speed of discovery and communication are both much faster than they were even

twenty years ago. Darwin's decade long writing, his agonizing over what to do after discovering Wallace's similar ideas, and his eventual publication of the "Origin of the Species" seems like a glacial drama compared to how competitive situations arise and play out today. In addition, the internet provides the first major technological breakthrough in science communication since the Royal Society launched the first scientific journal, the Philosophical Transactions, in 1665. These changes lead us to re-examine such basic questions as: what is priority of discovery and what constitutes an acceptable means of communication to establish priority of discovery?

Comparing a Patent to Priority of Discovery

Can scientists learn from the legal profession with regard to defining priority? After all, the granting of a patent also involves assigning intellectual property based upon an original invention or discovery. A patent transfers knowledge into an asset that can be bought and sold, and many scientists and scientific institutions use patents to claim and protect their work for purposes of commercialization. However, we argue that intellectual property for a patent is fundamentally different both in process and purpose from establishing priority within the scientific community. The granting of a patent is defined by a set of written guidelines, and these rules usually differ significantly in different countries. In contrast, scientific priority is guided by the culture and practices within a scientific community, rather than by rules established by a governing authority (e.g. the NIH or the ERC). Furthermore, assigning "priority of discovery" is a practice that is applied globally rather than nationally, as it involves a body of knowledge that belongs to all of mankind. The granting of a patent is also binary; an inspector(s) decides whether or not to award a patent for the claim, and if challenged, a court will decide whether or not to uphold the patent. In contrast, a more complex, organic and democratic system is involved in establishing "priority of discovery"; there are no appointed scientific inspectors or judges, with the exception of the very tiny fraction of scientific work being considered by prize committees (and their outcomes cannot be overruled in a court). Rather than using a system of inspectors, "priority" is most commonly defined by how scientists credit one

another's work through citations in manuscripts and at meetings, and thus priority frequently takes several years, sometimes decades, to emerge.

The intent of a patent claim and priority of discovery are also fundamentally different. The primary goal of a patent is to stake territory as one's own and *exclude others* (or make them pay for using your intellectual property), thus serving a goal of commercialization. The invention only enters the public domain for free use after many years. In contrast, the disclosure of work for "priority of discovery" is intended to *encourage others* to make use of the discovery and expand upon it for non-commercial use. Not only is this aligned with mission of science, but the scientist's claim "priority" can only be firmly established when others validate (i.e. repeat) and affirm the importance (through further development) of the original discovery.

Priority of Discovery as Two-Step Process of Disclosure and Validation

If a patent is an imperfect framework for "priority of discovery", how might one better explain and define this term? Robert Merton (Merton, 1957) makes an interesting analogy between a "discovery" and "property" by considering the valuation of the "property". With generous paraphrasing and extension of his argument, a scientist's discovery is initially known and belongs to only her/himself. This is the "eureka moment", which is a thrilling reward for a scientist's hard work. However, exclusive ownership of this "property" is of little actual value, as it neither advances the scientist's reputation nor serves science as a whole. Thus, the scientist must eventually disclose the discovery. At that moment, the "property" of the single or small group of scientists is transferred to the "public trust" of science, the collective body of knowledge that benefits society overall. Once transferred, all scientists have the right to use this new knowledge and it is no longer under the control of the original scientist. Implicit in signing over his/her property, the scientist expects the scientific community to acknowledge their contribution; they would like to see a virtual "plaque" that links their name to this new piece of knowledge. In practice, this acknowledgment usually comes in the form of citations in the papers from other scientists,

which can accumulate over time. If the scientist does not receive this credit, then he/she feels that their property was lost, never to be reclaimed, and thus feels an injustice has occurred.

This relationship between the individual scientist and the scientific community highlights two steps in how “priority of discovery” becomes established. First, there is the act by which the scientist transfers knowledge to the broad scientific community, for which we use the term “disclosure”; later, we will discuss differences in how this term is used should be used in science versus patent law. In a second step, the scientific community responds to the disclosure. It establishes whether this claim of a new discovery is correct and whether it is of sufficient interest to merit attention and further development; we use the term “validation” to encompass the assessment of both the accuracy of and interest in the work. This second step, which measures the community’s response, is what establishes the scientist’s reputation and results in rewards, such career advancement and grants that enable more scientific work. However, unlike disclosure, which is an event with a defined time stamp, this second step can take variable amounts of time. Often the most novel discoveries take the longest to be acknowledged by the community. It is no coincidence that Nobel prizes are often given decades after the original discovery. Next, we discuss these two steps in establishing priority in more detail.

Step 1: Disclosure:

What constitutes an acceptable method of disclosure to establish priority of discovery?

A public disclosure is defined by patent law, since this event determines the time window in which the inventor can file for a patent. In US patent law, public disclosure can include oral as well as written communications, which can encompass printed and electronic publications, scientific journals, website articles, posters and conference slide presentations.

However, for “priority of discovery” within the scientific community, we argue for a different definition of a disclosure, which is more closely aligned with the principles of

science and whose objective is *to enable a fair, complete, and wide-spread transfer of knowledge from the scientist to the scientific enterprise as a whole*. As the first step towards establishing scientific priority, we propose that the disclosure of knowledge should:

1. Include all of the data and the written interpretation of the data, which together are needed to document evidence for and an awareness of a discovery.
2. Contain a complete description of the methodologies used, so that the work can be replicated and extended by other scientists and therefore is unambiguously in the public domain.
3. Be communicated by a mechanism that is widely recognized by the scientific community as a venue for broad dissemination, so that the new knowledge becomes readily apparent. Thus, an inability to find the new information cannot be used as an excuse by a competing scientist for a failure to cite the work.
4. Include a time stamp of when it became broadly disseminated. And if revised, a new time stamp applied and the earlier versions of the work retained. The venue of dissemination must be stable and long-lasting to ensure the permanence of the work.

Disclosure Through a Journal Publication

The most widely accepted practice of a disclosure for scientific priority in the life sciences is through a publication in a peer-reviewed journal, a practice that dates back to the Royal Societies in Britain (Spier, 2002). Journals provide quality control for data disclosure, have a time stamp in the form of a publication date, and are accepted as a mechanism for broad dissemination (and recently facilitated by PubMed and other online search and archive mechanisms). However, this practice comes with multiple problems, many of which are associated with using editorial and peer review as a filter to announce the discovery. By linking the disclosure to the appearance of a publication, the scientist him/herself loses control of the timing of the announcement and hands over an important component of the disclosure process to the journal. Although not the intention of this article, Eric Lander's article "The Heros of CRISPR" (Lander, 2016) illustrates this point with several examples:

CRISPR as an adaptive immune system

“Mojica went out to celebrate with colleagues over cognac and returned the next morning to draft a paper. So began an 18-month odyssey of frustration. Recognizing the importance of the discovery, Mojica sent the paper to Nature. In November 2003, the journal rejected the paper without seeking external review; inexplicably, the editor claimed the key idea was already known. In January 2004, the Proceedings of the National Academy of Sciences decided that the paper lacked sufficient ‘novelty and importance’ to justify sending it out to review. Molecular Microbiology and Nucleic Acid Research rejected the paper in turn. By now desperate and afraid of being scooped, Mojica sent the paper to Journal of Molecular Evolution. After 12 more months of review and revision, the paper reporting CRISPR’s likely function finally appeared on February 1, 2005 (Mojica et al., 2005)

“The authors proposed that the CRISPR locus serves in a defense mechanism—as they put it, poetically, ‘CRISPRs may represent a memory of ‘past genetic aggressions.’” Vergnaud’s efforts to publish their findings met the same resistance as Mojica’s. The paper was rejected from the Proceedings of the National Academy of Sciences, Journal of Bacteriology, Nucleic Acids Research, and Genome Research, before being published in Microbiology on March 1, 2005.”

“Finally, a third researcher—Alexander Bolotin, a Russian e’ migre’ who was a microbiologist at the French National Institute for Agricultural Research—also published a paper describing the extrachromosomal origin of CRISPR, in Microbiology in September 2005 (Bolotin et al., 2005). His report was actually submitted a month after Mojica’s February 2005 paper had already appeared—because his submission to another journal had been rejected.”

On the in vitro reconstitution of CRISPR and the potential of gene editing:

“Siksnys submitted his paper to Cell on April 6, 2012. Six days later, the journal rejected the paper without external review. (In hindsight, Cell’s editor agrees the paper turned out to be very important.) Siksnys condensed the manuscript and sent it on May 21 to the Proceedings of the National Academy of Sciences, which published it online on September 4. Charpentier and Doudna’s paper fared better. Submitted to Science 2 months after Siksnys’s on June 8, it sailed through review and appeared online on June 28.”

These examples illustrate the unreliability of using journals for establishing the timing of disclosure. Furthermore, since the journal review process occurs “behind the curtain”, the community and history will never know exactly what a scientist discovered or how data was interpreted when he/she initially submitted the work to a journal. Perhaps an injustice occurred when the journal delayed publication, thus delaying the transmission of an important discovery. But on the other hand, perhaps there were fundamental problems in

the data or in the original interpretation which came to light during the review. We simply don't know. More generally, one of the main complaints of the current peer review systems is that the length of time it takes delays the time at which the knowledge becomes available to the wider community. We also note that disclosure of a discovery from a scientist to the public domain did not always involve "peer review", the Wallace, Darwin, Watson and Crick, and Einstein relativity papers being a just a few of numerous examples.

Disclosing by posting of a manuscript on a preprint server

In contrast to the situation described above for journals, a preprint server provides a clear, transparent, and unfiltered mechanism by which a scientist can transmit work to the community without delay. A preprint server is a tool that a scientist uses to post a complete manuscript before peer review, which is usually the same manuscript that is submitted to a journal. After a brief inspection to ensure that it is a scientific work (usually just a couple of days), the manuscript and associated data are made available to the entire scientific community through the internet without undergoing peer review (Vale, 2015). arXiv, a preprint server established in 1991 and now operated by Cornell University, is widely used in the physics, mathematics, and computational sciences communities. Recently, a similar type of server (bioRxiv) has been established by Cold Spring Harbor for the life sciences community.

A manuscript posted as a preprint could satisfy the four criteria listed above for disclosure, with certain qualifications. Since preprints are similar in content to journal manuscripts, criteria 1 and 2 can be met. However, since the initial posting is not rigorously peer-reviewed, the authors may have withheld critical data or methodologies needed for replication; such a preprint should not be considered as a fair disclosure to establish priority of a discovery. Whether criteria 3 and 4 are met depends upon the nature of preprint server. Does the preprint server have the infrastructure and ability maintain a permanent record? Does the community view the preprint server as a highly visible source of broad dissemination? Most would agree that the posting of a manuscript on a lab web site would not meet criteria 3 or 4. On the other hand, the preprint server arXiv meets those criteria for the physics community, since it is maintained by Cornell University (a

stable entity), posts ~100,000 preprints per year, and is routinely used by most physicists and mathematicians. Preprint servers that might meet criteria 3 and 4 for the life sciences are topics of active discussion, such as whether the community should support one preprint server, as well as how to make preprints more readily discoverable through PubMed or other mechanisms.

Disclosing by presenting at a scientific meeting:

Several decades ago, presenting work at a scientific meeting was often acknowledged by colleagues in the field as a reasonable mechanism for communicating a new discovery and could establish priority. In these earlier days, it was possible to assemble virtually the entire field of molecular biology in one Cold Spring Harbor Meeting. However, this is no longer possible. Furthermore, the amount of data and methodologies presented in a meeting talk or poster is usually insufficient to meet criteria 1 and 2 and is generally not retained in the form of a permanent record (criteria 4).

While meetings fall short as a mechanism for establishing scientific priority, we also recognize their substantial benefit for the scientist and the scientific process. Scientists get feedback on their work, which helps them to improve their study, and meeting presentations are outstanding training experiences for students and postdocs. Unfortunately, meetings are becoming increasingly filled with published work, as scientists are concerned about presenting unpublished work, in part due to the uncertainty of how long it will take to get their work published in a journal. However, if preprints become accepted as a basis for priority, then more scientists might be more open to sharing work prior to journal publication, recognizing that they have greater control of disclosure or in some cases might have already disclosed the work through a preprint server.

Step 2: Validation

The role of journals and peer review

If preprint servers can be used to disclose scientific work and the medium of print has virtually disappeared, what is purpose of a journal? And why do virtually all physicists, who post on and search for material on arXiv also submit their work routinely to a journal?

The answer is that the disclosure of a discovery is of little value, unless the work is seen, discussed, repeated, and cited by other scientists. The scientist can accomplish the disclosure on her/his own, but needs assistance in achieving these next steps in establishing priority, which is the validation by other scientists. Currently, the peer review system through a journal serves this role. One reads a published journal article knowing that two or three fellow scientists have spent some amount of time carefully examining the work and looking for obvious errors in the experiments and/or interpretation. Thus, peer review provides a service to the scientist by conveying to the scientific community and public that the work has undergone a first step of validation. However, it is important to recognize that journal-initiated peer review is a first step and not the final word on validation and establishing priority of discovery. Published papers in high profile, peer-reviewed journals have been proven wrong or in rare cases even falsified; conversely rejected papers have later won Nobel Prizes. Thus, priority of discovery ultimately involves a broader scrutiny from the scientific community after publication, a process that historically can evolve over years. This scrutiny, validation, and recognition often become evident in the form of citations. A provocative discovery is initially marked by a flurry of citations. Important discoveries that stand the test of time continue to be highly cited within a field, while those that were wrong or flawed generally fall by the wayside. The other reason that scientists use journals is to achieve visibility for their work, which is not achieved by the democracy of preprint servers. Some journals are widely read while others are not. Indeed, the primary function of an editor is to assemble an interesting set of papers in a particular field. Thus, publishing work in a widely read journal gets it on the fast track to been seen and discussed, although there is no guarantee that such attention will be long lasting.

Using peer review and high profile journals as a “first pass” for validation and visibility respectively is far from perfect, as has been discussed by many (Alberts et al., 2008:

Krumholz, 2015). The small numbers (2-4) of scientists contributing to peer review leads to highly variable outcomes of validation and decision-making regarding which work should be visible and recognized. Importantly, the heightened competition of scientists for rewards (e.g. grants and jobs)(Alberts et al., 2014) has strained a system that was established in the 18th century and seemed to work reasonably well until just a couple of decades ago. Although some have proposed that journals and peer review as currently practiced should be abandoned as soon as possible, we would argue against this. The need for a system of validation in the establishment of priority and discovery, which has existed for the past three hundred years, has only become more pronounced as scientific work has increased. Currently, journals are making interesting adjustments to make peer review more transparent and consensus-driven (e.g. the review process at eLife) and thus are not standing still in the internet era.

However, we recognize the opportunity for change in the future. As we have discussed, post-publication validation of accuracy and assessment of quality has always been part of the process of priority of discovery, but has been poorly organized. Currently, many bright people are attempting to organize this effort through post-publication, internet-based commentary and social media, which should be applauded and accelerated. Post-publication review can be considered to occur after a preprint, after journal publication, or both. These mechanisms also might represent an opportunity to move the assessment of discoveries away from what some may regard as an old boys network of elite journals and plenary talks at scientific meetings. However, importantly, as scientists, we need to evaluate outcomes from these new validation efforts, which are as much experiments in human nature than they are in technology. Any new system should be assessed with real data, rather than perceptions, to establish whether it improves the validation of a discovery or if it introduces new opportunities for gaming the system with unintended consequences.

Priority is more than just a time stamp

Our breakdown of priority into two steps, disclosure and validation, reveals an inherent tension between the two steps- one emphasizes speed and the other emphasizes quality.

Time is easy to understand and quantify and has been the most emphasized component of scientific priority. As an extreme interpretation, the secretary of the French Academy of Science Francois Arago in the 19th century stated that “questions of priority may depend on weeks, on days, on hours, on minutes”, which some have referred to as the Arago Effect for establishing priority (Merton, 1957). Naturally, the timing of disclosure is important, since it convinces the community that one’s work is original and not just derivative of another scientist’s discovery. In fact in European Patent law, priority is given to first to file. However, for the academic scientist, racing to be first at the expense of quality is a recipe for disaster, especially if it becomes a repeating pattern, as this behavior tarnishes a scientist’s reputation overall. Furthermore, we believe that a simple time-based “winner take all” philosophy is not in the best interest of science, in contrast to what others have claimed (Strevens, 2003). Nor does the Arago viewpoint rule the day in practice, as there are plenty of examples in which the scientific community recognized similar discoveries that appeared close to one another in time. Furthermore, quality, and not just being first or tied for first, plays an important role in the ultimate assessment of a work by the scientific community. Darwin and Wallace provide a case in point. Both scientists are recognized for their independent idea of natural selection and its role in evolving new species. But even in their lifetimes, it became broadly recognized, including by Wallace himself, that it was the outstanding corpus of evidence in his masterpiece “Origin of Species” that associated Darwin’s name most widely with the theory of evolution.

Conclusions and Recommendations

In this article, we argue that priority of discovery in biomedicine is established in two different phases: 1) A disclosure phase, in which the scientist relinquishes his/her new knowledge in a complete form to the broad scientific community, and 2) a validation phase in which the scientist seeks confirmation and recognition for her/his work from the community. Currently, both of these steps are tied together in the peer review publication process. However, we suggest that both the scientist as well as the scientific community would be better served by unwrapping this process in two independent steps. In addition to expediting disclosure, this may help the process of journal-based peer review, since it

allows the authors to have experiments checked and commented on by other scientists, and where necessary to correct the findings without the embarrassment of retraction after journal publication.

We feel that the first step of disclosure is best served by the simplicity of a preprint, which allows the rapid transmission of work from one scientist to the community. We have outlined four criteria that should be considered in this transfer step. These criteria could be fulfilled by a preprint server or by a similar service offered by a publishing company. However, a number of important questions in implementation remain to be resolved. How should preprints be made discoverable to enhance their visibility and allow them to serve most effectively as a mechanism of disclosure? Is the function of disclosure best served by one or many preprint servers? Regarding the latter question, we feel that a single preprint repository in the life science, supported by funding agencies and scientific institutions, has many advantages, with the 25 year success of arXiv providing a strong proof-of-principle of this non-profit model. A single effort, supported by all parties of the biology community, could achieve the trust and critical mass of content that is needed to encourage biologists to upload to a preprint server and lead to the cultural shift from present day behavior.

Validation, the second step of priority, is much more complex than disclosure. Perfect solutions are less obvious, since scientific quality is difficult to judge and thus not easily quantitated or accomplished by crowd sourcing. Journals, which have established a considerable infrastructure for peer review, provide the present-day gold standard for validation, albeit imperfect. But new mechanisms for validation could arise in the future that better serve the scientist who is seeking to make a claim of priority of their discovery.

References

Alberts, B., Hanson, B., and Kelner, K.L. (2008). Reviewing Peer Review. *Science* 321, 15.

Alberts, B., Kirschner, M.W., Tilghman, S., and Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *PNAS* 111, 5773–5777.

Lander, E.S. (2016). The Heroes of CRISPR. *Cell* 164, 18–28.

Krumholz, H.M. (2015). The End of Journals. *Circulation: Cardiovascular Quality and Outcomes* 8, 533–534.

Merton, R.K. (1957). Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review* 22, 635-659.

Spier, R. (2002). The history of the peer review process. *Trends Biotech.* 20, 357-358.

Strevens, M. (2003). The role of the priority rule in science. *The Journal of Philosophy* 55–79.

Vale, R.D. (2015). Accelerating scientific publication in biology. *Proceedings of the National Academy of Sciences* 112, 13439–13446.