Industry access to the literature

To the Editor:

I recently wrote about a problem that afflicts a large number of scientists at small- to mid-sized biotech companies: little or no access to a wide spectrum of the current biomedical literature[1]. The response to my article was both brisk and bleak. Industry researchers echoed my concerns, with several troubling details emerging. Shared personal subscriptions have become the primary conduit for accessing an increasingly limited number of biomedical journals. The majority of companies have no libraries to speak of and no librarians to help with literature searches. The availability of online journals is insufficient and funds for purchasing access to papers on an individual basis are limited. In one case, a company suffered a six-month setback to a drug development program because a paper was missed in an inaccessible journal. The central question that I raised in my op-ed piece was, "At a time when more and more papers are published, when information overload is a given, does a lack of access to the information become an equally large problem?" The answer from the community was a vociferous yes.

I've been fortunate to have access to world-class libraries at every stage of my career. As a result, I learned that being widely read has significant advantages. It enables the formation of new and fruitful collaborations. It facilitates your ability to make connections, to see new relationships and to partake of a bigger view. This larger vision, in turn, can lead to novel insights and spur innovative discoveries. As I noted previously, keeping up with advances in biomedicine has become increasingly difficult in recent years. The overlapping nature of disciplines within the biological sciences means that someone developing a new cancer treatment needs to stay informed about specific areas of biochemistry, genetics, toxicology, computational biology, developmental biology, cell biology, immunology and stem cell biology as well as clinical developments. This is in addition to keeping up with general trends in the biotech industry as well as technical advances in experimental reagents, devices and methodology. The number of published biological science journals has been expanding for decades, driven by both scientific societies and for-profit publishers like Nature Publishing Group (NPG). Some of these journals have grown and divided like the bacteria that they often report on. NPG, for example, publishes not just *Nature* but also *Nature Biotechnology*, *Nature Cell Biology*, *Nature Chemical Biology*, *Nature Genetics*, *Nature Immunology*, *Nature Medicine* and *Nature Neuroscience*, to name a few, and a wide spectrum of *Nature Review* journals.

There will always be costs associated with publishing articles, with the expense to be borne by the writer, the reader, the publisher or some combination thereof. The rising cost of journal subscriptions has led to the growth of open source, free online journals such as those published by the nonprofit Public Library of Science (PLoS). Although this trend is admirable, especially for those on a tight budget, the number of freely accessible journals is only a small fraction of those that you need to pay to read. PubMed Central currently provides some access to the older biomedical literature, but many of the key journals researchers would like to peruse, such as *Nature*, *Science* and *Cell*, which receive a large number of submissions but only publish a few, do not yet participate in this process. In addition, archival access to the PubMed Central journal collection is quite variable across the literature as a whole.

Years ago, a simple but inelegant solution to the journal access problem was to visit a local university or research institution's biomedical library. You could read and photocopy articles for a minimal fee, and often had access to a librarian for search queries. A variant of this workaround is still a viable option today. Although most institutions no longer purchase a substantial number of physical copies of their journals, they do provide guests computer access to their extensive online subscriptions. This process requires overcoming a degree of personal inertia to get to the library, and once there, you have to hope there is an available computer. Unfortunately, if you don't work in close physical proximity to a well-equipped medical library, then you are simply out of luck. Many scientists I know 'borrow' access codes for online university journal subscriptions from friends and family members, though this approach likely violates licensing agreements established by the journal vendors.

A concern widely voiced recently in Western countries is that they are losing their competitive edge to other countries, most notably China and India. Fingers are often pointed at perceived problems in the educational system, which many feel is chronically underfunded. It is ironic to read stories that focus on our nation's children when the biotech companies that employ many of our most highly trained and educated scientists can't afford to provide them access to the science journals that they need to do their jobs effectively. Innovative startup companies in many fields will always be financially constrained compared with their well-established brethren with whom they wish to compete. However, biotech companies are dependent on library access to a degree not seen in other technology areas, such as software.

The lack-of-access-to-information problem in biotech is worth pondering in the context of the low productivity of drug development R&D[2,3]. Of particular note is the trend within many large pharmaceutical companies (which traditionally had libraries with comprehensive literature access) to shut down their R&D departments and become increasingly reliant on small to medium-sized biotech companies (many of which don't have full access to the literature) to fill their pipelines. As a result, the burden of biological drug discovery is being pushed down from large companies with enormous research budgets to small startups with just a few coins to rub together. One cannot expect high productivity from the smaller emerging companies that collectively suffer from an inability to afford access to reports describing new discoveries in the biomedical sciences.

I can offer a few potential solutions that may provide industry scientists in small-to-medium–sized enterprises greater access

to a wider range of journals. Many biotech companies, at some point in their growth cycle, will strike a deal with a big pharma partner to help develop one or more molecules in their pipeline. These deals can be structured in a variety of ways and usually include milestone payments for the biotech partner. The smaller biotechs should ask for access to their pharmaceutical company partner's online journal subscriptions as a component of these deals. The cost to add access for 20 scientists, for example, to in-house research staffs numbering in the thousands should be minimal. An obvious added benefit of this arrangement is that the online access should facilitate the ability of the smaller biotech partner to move the joint projects as well as their own projects forward.

What happens, though, to scientists working for unpartnered companies, or with partners that can't or won't provide such access? Another solution would be for small biotechs to form some type of cooperative or consortium that would provide biomedical journal access. The consortium could be modeled on the cable TV business, which sells a range of tiered services to different customers. The consortium would purchase online subscriptions to a core group of journals and members of the group would pay a set fee per number of employees in their organization to gain access. As with the cable business, those wanting to have access to additional tiers of journals (e.g., those with a focus on clinical studies, oncology or veterinary science) would pay additional fees. The charge for the basic access tier needs to be widely affordable, and the arrangement would benefit both the journal publishers and the researchers. The journals that participated in providing access to consortium members would gain additional (albeit reduced) revenues from organizations that otherwise could not afford their subscriptions. On the other side of the equation, scientists working at the consortium companies would gain access to journals that they otherwise could not afford. This should help facilitate innovation and stimulate both basic as well as translational research. It would be preferable if access to the older scientific literature (say, greater than two years old) were provided for free, and for tiered subscriptions to newly published articles to be the focus of what consortium members are paying for.

The biotech industry's advocacy group, the Biotechnology Industry Organization (BIO; Washington, DC), would be a logical candidate to create and manage such a consortium within the auspices of their own organization. Providing this service would have benefits in three specific areas. It would help motivate small companies to join BIO, thereby strengthening the organization and its ability to assist its members. Companies who joined the consortium would bolster the ability of their scientists to do cutting edge R&D work in creating new drugs. Journal publishers would increase their revenues by drawing in a much larger pool of subscribers, which is why they could afford to give them a price break. As an alternative to BIO, a pioneering information services company might garner a strong financial reward for creating and operating an online journal access consortium.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

Stewart Lyman

Lyman BioPharma Consulting LLC, Seattle, Washington, USA.
e-mail: lymanbiopharma@comcast.net

1. Lyman, S. *Xconomy/Seattle* (18 January 2011) <http://www.xconomy.com/seattle/2011/01/18/what-does-biotech-really-suffer-from-information-overload-or-underload/?single_page=true>
2. Kling, J. *Nat. Biotechnol.* **29**, 197–200 (2011).
3. Samaha, L. *Investing by Suite 101* (9 July 2010) <http://www.suite101.com/content/estimating-drug-probabilities-of-success-in-fda-clinical-trials-a259518>

Sequencing technology does not eliminate biological variability

To the Editor:

RNA sequencing technology provides various advantages over DNA microarrays. For example, it is possible to measure alternative transcription[1] or measure transcription for noncoding regions[2] *de novo*. Another potential advantage is low technical variation[2–4]. This has led to rapid adoption of the technology and a recent surge of publications[5]. We would like to caution, however, that the euphoria surrounding the technology has led many of these publications to discount the influence of biological variability, forgetting perhaps that unwanted variability in gene expression measurements is not due only to measurement error. Gene expression is a stochastic process[6] and is known to vary between units considered to be of the same population, for example, in samples from a specific healthy tissue across individuals[7]. In a typical experiment, variation in gene expression measurements [Var(Expr)] can be decomposed[8] as the following:

Var(Expr) = Across Group Variability + Measurement Error + Biological Variability

'Across Group Variability' is the variation in gene expression due to the groups under consideration in an experiment. For example, it is well known that gene expression profiles for tumor samples differ from expression profiles for matched healthy controls[9]. This type of variability can be measured by comparing samples from different biological groups and is typically the outcome of interest. The second component of gene expression variation, 'Measurement Error', can be estimated with technical replicates—different aliquots of the same sample measured with a technology multiple times. This is the type of variation that may be reduced with technology improvements[4]. Well-known sources of technical variability in both sequencing and microarray studies are laboratory[10,11] and batch[12] effects. The third component of expression variation is true 'Biological Variability', which can be measured only by considering expression measurements taken from multiple biological samples within the same group. Regardless of the technology used to measure expression levels, the true gene expression levels will vary among individuals because expression is inherently a stochastic process[6]. In an experiment where the group comparison is of primary interest, both measurement error and biological variation may be confused with the outcome of interest: the estimated difference in expression between groups.

To illustrate how biological variability among individuals within the same group is not eliminated by sequencing technology, we collected public data from two of the only RNA-sequencing experiments with a large number of biological replicates, $n = 60$ and $n = 69$, respectively[13,14]. We compared a subset of these sequencing data ($n = 43$ and 51 samples, respectively) with microarray data from two different platforms[15,16]. In each comparison, the exact same cell lines were analyzed on both technologies. In study one, $m = 14{,}797$ genes had expression measurements from both sequencing and microarrays on all samples. In study two, $m = 7{,}157$ genes had expression measurements from both technologies on all samples (**Supplementary Methods**).

For each expressed gene in each of the two studies, we calculated an estimate of the variability in expression levels across