Biomolecular Detection and Quantification 7 (2016) A1–A5

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

journal homepage: www.elsevier.com/locate/bdq

Perspective

Improving the reliability of peer-reviewed publications: We are all in it together

ARTICLE INFO

Keywords:
Research
Reproducibility
Biomedicine
qPCR
Microarrays
Next generation sequencing

ABSTRACT

The current, and welcome, focus on standardization of techniques and transparency of reporting in the biomedical, peer-reviewed literature is commendable. However, that focus has been intermittent as well as lacklustre and so failed to tackle the alarming lack of reliability and reproducibly of biomedical research. Authors have access to numerous recommendations, ranging from simple standards dealing with technical issues to those regulating clinical trials, suggesting that improved reporting guidelines are not the solution. The elemental solution is for editors to require meticulous implementation of their journals' instructions for authors and reviewers and stipulate that no paper is published without a transparent, complete and accurate materials and methods section.

The treatment of some scientific topics, particularly in biomedical research, is very much like that afforded to the catwalk fashion industry; something becomes hyped, everyone talks about it and eventually the popular press picks up the topic and generally distort its conclusions, only for the band wagon to move on to the next hot topic. Tellingly, this excitement is usually misplaced and serves more to publicize the particular authors, institutions and journals than it does to contribute to any advancement in scientific knowledge or translational benefit. In contrast, vast amounts of scientific data are published without eliciting any interest whatsoever, leaving the authors to cite their own papers in the hope that their work will, one day, become the hyped fashion. Regardless, the results and conclusions from much, if not most, of the publications of biomedical research are questionable: the majority are not reproducible [1–3] and so do not satisfy one of the fundamental requirements of scientific research. There are a number of reasons why published results cannot be reproduced:

1. The original research was carried out incorrectly, for example without sufficient regard for sample selection, template quality or inappropriate data analysis.
2. The attempts to replicate results are flawed because the information provided in the publication is not sufficiently detailed and explicit.
3. The replicating laboratories do not have sufficient understanding of the uncertainty associated with their experiments. For example, the high precision of methods like digital PCR can generate different results, but a more focused look at reproducibility may show they are all describing different parts of a data distribution, which, once understood, would allow a definition of what can actually be measured.

Any of these explanations is objectionable and results in billions of dollars being wasted every year [4]. This message is, of course, not new [5] and over the last twenty years or so there have been numerous, often high profile, publications lamenting this state of affairs and proposing solutions, most recently summarized in a review article published in this journal [6].

Why is there this apparent indifference to publication quality? Is it because detailed scrutiny of the reliability, standardization, reproducibility and transparency of methods is perceived as comparatively mundane and unexciting? Is the current peer review process inadequate to provide a reliable analysis of all techniques? In theory, there is no disagreement about the importance of the methods section of a scientific manuscript [7] or that it requires a clear, accurate [8] and, crucially, adequate description of how an experiment was carried out. In theory, it is also accepted that the aim of a methods section is to provide the information required to assess the validity of a study and hence be sufficiently detailed so that competent readers with access to the necessary experiment components and data can reproduce the results.

Certainly, despite the wealth of evidence that published methods are wholly deficient, there has never been any determined, consistent and coherent effort to address these issues and deal with their consequences. Therefore a welcome, recent effort involves the publication of a report based on the proceedings of a symposium held earlier this year, aimed at exploring the challenges and chances for improving the reliability and reproducibility and of biomedical research in the UK (http://www.acmedsci.ac.uk/policy/policy-projects/reproducibility-and-reliability-of-biomedical-research/). However, a close reading of the report suggests that it simply summarizes all of the findings and opinions that are already published and suggests the same solutions that have

http://dx.doi.org/10.1016/j.bdq.2015.11.002

been ignored until now. These include "top-down measures from journals, funders and research organisations" that aim to improve the quality of training and institute a research culture and career structure that reduces the emphasis on novelty and publication, as well as "bottom-up ones from individual researchers and laboratories" that address issues of poor study design and statistical practices, inadequate reporting of methods, and problems with quality control. What is lacking is a decisive, headline-grabbing call to action.

Some of the suggestions also imply that the authors of this report appear not to be overly familiar with existing, long standing efforts to standardize protocols and improve transparency. For example, in a section with the heading "strategies to improve research practice and the reproducibility of biomedical research" contains the suggestion that establishing standards could address some of the issues associated with reproducibility and points to the Minimum Information About a Microarray Experiment (MIAME) guidelines [9] as the exemplary standard. In fact, there are numerous "Minimum Information" standards projects following on from that paper, most of which have been registered with the Minimum Information for Biological and Biomedical Investigations initiative (http://www.dcc.ac.uk/resources/metadata-standards/mibbi-minimum-information-biological-and-biomedical-investigations), where they are collected and curated and can be accessed through a searchable portal of inter-related data standards, databases, and https://biosharing.org/standards. Complementary information is also available from the US National Library of Medicine website, which lists the organizations that provide advice and guidelines for reporting research methods and findings (https://www.nlm.nih.gov/services/research_report_guide.html). Medical research studies, in particular, are well served with reporting guidelines, for example by the EQUATOR Network, which aims to improve the reliability and value of the medical research literature by promoting transparent and accurate reporting (http://www.equator-network.org). There are reporting guidelines for many different study designs such as CONSORT (www.consort-statement.org) for randomized trials, STARD for studies of diagnostic accuracy (www.stard-statement.org/) and SPIRIT for study protocols (http://www.spirit-statement.org).

If it were simply a matter of developing standards, then the state of the peer-reviewed literature would not be as scandalous as it is. The real problem stems from the lack of application of those standards. This is most easily demonstrated by looking at, arguably, the most widely used molecular techniques, real-time PCR (qPCR) and reverse transcription (RT)-qPCR. These methods have found supporting roles as part of a huge number of publications in every area of the life sciences, clinical diagnostics, biotechnology, forensics and agriculture. qPCR-based assays are usually described as simple, accurate and reliable. This is true, but only if certain technical and analytical criteria are met. It is especially important to emphasize that the accuracy of results is critically dependent on the choice of calibration, whether this be a control sample or a calibration curve. This method is easily abused and one particularly egregious example is provided by its use to detect measles virus in the intestine of autistic children. Numerous, independent replication attempts, including those carried out by the original authors, failed to reproduce the original data and an analysis of the raw data, carried out as part of the US autism omnibus trial in Washington DC, revealed that the conclusions were based on fallacious results obtained by a combination of sample contamination with DNA, incorrect analysis procedures and poor experimental methods [10,11]. A paper publishing these data remains to be retracted 13 years after publication. While this delay is typical, it is totally unacceptable and results in an underestimation of the role of fraud in the ongoing retraction epidemic [12,13].

A typical problem associated with qPCR assay variability is illustrated in Fig. 1, which demonstrates that qPCR assays can behave significantly different under different experimental conditions. As the data demonstrate, at the higher target DNA concentration both assays generate reliable data. However, at the lower concentration, results are reliable from only one of the assays (A), with the ΔCq of 5.98 ± 0.21 between two different target DNA concentrations being in line with the expected value for the dilution factor. In contrast, the results of the other assay (B) are much more variable, (ΔCq of 8.29 ± 1.65) and also do not accurately reflect the dilution factor. The report by Dr. Andreas Nitsche in this issue shows that some assays are particularly sensitive to variability in different buffers and even different batches of same buffer. If assay behavior is not thoroughly assessed such that experimental conditions are simulated, prior to carrying out real-life tests, this can lead to false results and confound any potential conclusions.

The MIQE guidelines, published in 2009 [14], are among the most cited molecular recommendations (nearly 4000 citations vs around 3500 for the MIAME guidelines published in 2001). They describe the minimum information necessary for evaluating qPCR experiments and include a checklist comprising nine sections to help guide the author to the full disclosure of all reagents, assay sequences and analysis methods and so help to minimise this kind of variability or potential inaccuracy. The guidelines suggest appropriate parameters for qPCR assay design and reporting, and have become widely accepted by both the research community and, especially, the companies producing and selling qPCR reagents and instrumentation. Implementation of these guidelines has been demonstrated to result in the publication of more complete and transparent papers, although the majority of qPCR-based papers continue to provide inadequate information on experimental detail [15].

There can be no doubt that there are a vast number of unreliable and incorrect results published that have been generated by qPCR, a relatively simple technique. This begs the obvious question of how reliable the results are that have been obtained using significantly more demanding methods. An example is digital PCR (dPCR), which involves the dilution and partitioning of target molecules into large numbers of separate reaction chambers so that each contains either one or no copies of the sequence of interest [16]. A comparison of the number of partitions in which the target is detected vs those in which it is not, allows quantitative analysis without the need for a calibration curve. Hence data analysis can be not just more precise, but also more straightforward than with qPCR. However, there are additional parameters that any reader of a publication using this technology needs to be aware of, most obviously the mean number of target copies per partition, the number of partitions, individual partition volumes and the total volume of the partitions measured. Hence the necessary requirement for the publication of the digital PCR MIQE guidelines, which address known requirements for dPCR that were identified during the early stage of its development and commercial implementation [17]. Expression microarrays and next generation sequencing incorporate an additional layer of complexity. Whilst the parameters required to ensure reliable qPCR and dPCR results are reasonably few, those required to assess the validity of expression microarrays or RNA sequencing are significantly more complex. There have been several papers investigating the effects of technical and bioinformatics variability of RNA-seq results [18–21] and standards for RNA sequencing [22,23] (http://www.modencode.org/publications/docs/index.shtml) as well as Chromatin immunoprecipitation and high-throughput DNA sequencing (ChIP-seq) [24] are being developed, but again there is no decisive push for their universal acceptance.

There is a correlation between the number of retractions and the impact factor of a journal [12]. While this could be due to the