

## Mostly, your results matter to others

**High-throughput datasets and analysis protocols are intrinsically difficult to referee. Community standards enforced by journals may be less effective than is widely appreciated. Greater awareness of the needs and value of secondary data users can result in higher-impact papers.**

Investigating the compliance of our publications with MIAME standards (minimum information about a microarray; Editorial, *Nat. Genet.* **38**, 1089; 2006), we found that even when authors and referees are aware of community standards and even with editors mandating both data deposition and accession linking as a condition of publication, a proportion of microarray datasets were at that time unavailable or incomplete.

Subsequently, the concept of reporting standards has been extended to proposals asking for minimum information about a proteomics experiment (MIAPE: *Nat. Biotech.* **25**, 887–893; 2007), a molecular interaction (MIMIX: *Nat. Biotech.* **25**, 894–898; 2007), a genome sequence specification (MIGS: *Nat. Biotech.* **26**, 541–547; 2008), *in situ* hybridization or immunocytochemistry (MISFISHIE: *Nat. Biotech.* **26**, 305–312; 2008), a biomedical investigation (MIBBI: *Nat. Biotech.* **26**, 889–896; 2008) and proposed facilities and standards for description and deposition of data generated by genome-wide association studies (dbGAP: *Nat. Genet.* **39**, 1181–1186; 2007 and GAIN: *Nat. Genet.* **39**, 1045–1051; 2007).

On page 149 four teams of analysts treated the findings of a number of microarray papers published in the journal in 2005–2006 as their gold standard and attempted to replicate a sample of the analyses conducted on each of them, with frankly dismal results. It can be argued that most researchers attempt replication from scratch, but if the replication of analysis of published work is impossible in most cases, in what respect can replicated experiments be fairly compared?

The findings of this Analysis should be used to improve practice rather than to criticize the authors and referees of these publications. A certain amount of both skepticism and initiative must of course be assumed on behalf of all readers and users of research publications. Equally, there must be enough goodwill and professionalism in the research community to permit critical reanalysis of research findings at any and every moment without this core scientific practice implying any personal criticism. Any scientist should be prepared to reexamine published work, one's own and one's colleagues' alike. In doing so it always helps to make clear one's needs and assumptions, and the Analysis in this

issue does indeed explain the limits of the analysts' requirements and critical aims.

Why should we consider the utility of rich datasets to researchers whose aim is reanalysis? Many experiments need to start with reanalysis, for validation or comparison. The journal needs to help our referees to spot-check the results they have been asked to examine. If we can make the papers more accessible to readers, we can make the publication and its associated dataset into a more versatile research tool for the benefit of the whole scientific community. Finally, because the spotlight is on the microarray guidelines as a model for other high-throughput methods, the recommendations of this community can be generalized to other fields.

What of the argument that a research paper is less a tool, more an advertisement published to recruit collaborators? This seems like a good idea, except that collaboration is based on mutual benefit and trust, both of which are engendered when collaborator can verify results for him or herself. In a 2004 editorial, we suggested that it is easier to make your reputation with papers that are useful to other researchers than it is to generate an equivalent number and quality of papers from the same dataset by your own efforts. In other words, good citizenship is good business (*Nat. Genet.* **36**, 1025; 2004). The Analysis finds support for this idea in its small sample—in that the more transparent papers have attracted over twice the citations gained by the papers the analysts found more difficult to replicate. It would be interesting to test this hypothesis further.

In the 2006 editorial, we encouraged authors to submit array data to the public repositories for the referees to see, a practice we now insist upon before sending papers to review. GEO and ArrayExpress have been very helpful to researchers and journals alike in providing tools and helping to implement MIAME guidelines. Other aspects of the research have evolved, too. Much of the scanner software has been upgraded to record settings and many software packages now record and output key decisions as the analysis progresses. It is now time to think how to implement the recording of key analytical decisions systematically. ■