# Much room for improvement in deposition rates of expression microarray datasets

**To the Editor:** *Nature Methods*' editorial[1] of March 2008 asserts that the deposition of supporting raw microarray datasets is "routine." However, our retrospective study shows this not to be the case.

We surveyed papers from the 2007 issues of 20 journals (alphabetically: *American Journal of Pathology, Blood, Cancer Research, Cell, EMBO Journal, Endocrinology, FASEB Journal, Journal of Biological Chemistry, Journal of Endocrinology, Journal of Immunology, Molecular and Cellular Biology, Molecular Endocrinology, Molecular Cell, Nature, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Methods, Proceedings of the National Academy of Science of the United States of America* and *Science*), retrieved with a Medline search for the terms "microarray/s OR genome-wide OR expression profile/s OR transcription profile/profiling." After removing false positives, we searched the full text of the papers for reference to deposition of a microarray dataset.

The rate of deposition of datasets was less than 50% (**Fig. 1** and **Supplementary Data** online), indicating that many researchers do not deposit datasets and/or many journals are not positioned to give effect to their own policies on deposition. Regrettably, federal funding institutes are not empowered to facilitate this process.

A notable obstacle to deposition in public microarray repositories is the effort required to deposit these data, which, owing to their highly contextual nature, have a more complex metadata structure than sequence data. This impediment persists even as repositories strive to simplify submissions while encouraging compliance with minimum information about a microarray experiment (MIAME)[2] standards. Although microarray datasets are most useful to bioinformaticians in their raw, unnormalized forms, which facilitate cross-comparison with other datasets, processed datasets are more useful to the bench scientist. Moreover, unless a description of the experimental details is available, neither form of the data are biologically interpretable.

We accordingly urge repositories to require deposition by authors of (i) at least MIAME-compliant metadata and, where possible, as detailed a set of experimental parameters as is required to make the data fully interpretable, (ii) the raw unnormalized intensity values, and (iii) processed, normalized expression values.

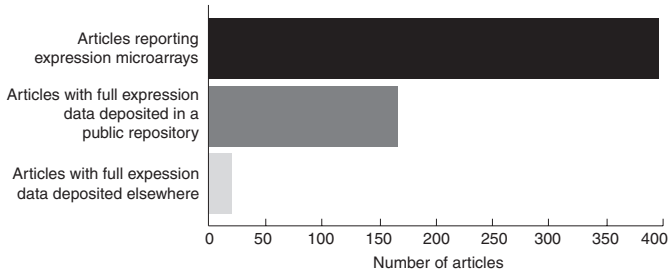


**Figure 1** | Rate of deposition of published microarray datasets in online repositories in 2007.

We propose adoption by journals of the GenBank sequence deposition model, requiring a statement in the manuscript identifying a repository and accession number at the time of submission, with the record embargoed until acceptance of the paper. To facilitate the tasks of journal staff, reviewers and repository curators, this statement could be positioned on the manuscript title page where other essential information is typically found. Lastly, improved communication between repositories and journals would ensure that dataset embargoes are lifted in a timely manner after acceptance of the paper.

Seven years after the elaboration of the MIAME principles, the emerging discipline of microarray meta-analysis, exemplified by the cancer gene expression resource Oncomine[3], continues to be hobbled by the mundane, time-consuming and often fruitless exercise of tracking down annotated full datasets. We call for a renewed collective effort from researchers, publishers and funding organizations to redress this situation and secure these data-rich research resources for posterity.

*Note: Supplementary information is available on the Nature Methods website.*

**Scott A Ochsner[1], David L Steffen[2], Christian J Stoeckert Jr[3] & Neil J McKenna[1]**

[1]Department of Molecular and Cellular Biology, and [2]Bioinformatics Research Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [3]Microarray and Gene Expression Data Society and Department of Genetics, University of Pennsylvania, 418 Guardian Drive, Philadelphia, Pennsylvania 19104, USA.
e-mail: nmckenna@bcm.edu

1. Anonymous. Thou shalt share your data. *Nat. Methods* **5**, 209 (2008).
2. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
3. Rhodes, D.R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–180 (2007).

**CORRECTED AFTER PRINT 26 NOVEMBER 2008**

# Erratum: Much room for improvement in deposition rates of expression microarray datasets

Scott A Ochsner, David L Steffen, Christian J Stoeckert Jr & Neil J McKenna
*Nat. Methods* 5, 991 (2008); published online 25 November 2008; corrected after print 26 November 2009.

In the version of this article initially published, the e-mail address of the corresponding author Neil J. McKenna was incorrect. The correct e-mail address should be nmckenna@bcm.edu. The error has been corrected in the HTML and PDF versions of the article.