



Published in final edited form as:

Nat Methods. 2008 December ; 5(12): 991. doi:10.1038/nmeth1208-991.

Much room for improvement in deposition rates of expression microarray datasets

Scott A Ochsner¹, David L Steffen², Christian J Stoeckert Jr³, and Neil J McKenna¹

¹Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

²Bioinformatics Research Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

³Microarray and Gene Expression Data Society and Department of Genetics, University of Pennsylvania, 418 Guardian Drive, Philadelphia, Pennsylvania 19104, USA

To the editor: *Nature Methods'* editorial¹ 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

nmckenxna@bcm.edu.

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

(iii) processed, normalized expression values. 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 Oncomine3, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Thou shalt share your data. *Nat Methods*. 2008; 5:209.
2. Brazma A, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001; 29:365–371. [PubMed: 11726920]
3. Rhodes DR, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9:166–180. [PubMed: 17356713]

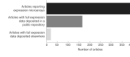


Figure 1.
Rate of deposition of published microarray datasets in online repositories in 2007.