Editorial

AN OPEN LETTER TO THE SCIENTIFIC JOURNALS

One of the underlying principles of scientific publication in peer-reviewed journals has been the requirement that the authors make available the data and materials necessary for a reader to reproduce the experiment or analysis and to determine whether the data support the conclusions of the authors. In many instances, such as DNA sequence or protein structure data, this has evolved into the requirement that the data underlying each published report be deposited in an appropriate international database. For microarray experiments, simply defining the appropriate data has been a challenge, because the large quantity of data generated in each experiment and the typical complexity of the ancillary information needed to interpret the results are unlike anything that has yet faced the biological research community. Databases to hold microarray data and the tools to annotate them properly are under development. As an interim solution, this guide describes the types of data that are necessary to reproduce and interpret a microarray experiment. It should go without stating that this information is only of value as long as data are available, and that every effort should be made to provide stable access to published data until such time as it is available from a public database.

The members of the Microarray Gene Expression Data (MGED — http://www.mged.org) society have been working over the past few years to solicit community input in developing standards for the publication of DNA microarray data. The authors of this guide and the MGED society as a whole represent a large cross-section of scientific community that have worked with microarrays. We are convinced of the importance of the issues described and strongly urge journals to use these recommendations when deciding whether to publish a paper using microarray data. In December 2001, we published a commentary in *Nature Genetics* in which we described MIAME — the Minimal Information About a Microarray Experiment (Brazma *et al.*, 2001). MIAME is presented as a proposed standard for representation of array data that would be sufficient to allow readers of published reports to replicate the analysis presented and to facilitate the development of novel methods of data analysis by providing access to necessary primary data.

Community response to MIAME was favorable and many instrument manufacturers, software developers, and international databases moved to adapt their systems to capture and manage MIAME-compliant data. However, by far the most common request from the community has been for a brief set of guidelines that could be used by authors, editors, and referees to try to meet the MIAME data standards. Below we present an attempt to do just that. These requirements can easily be met by adequately describing the experiment, the materials and methods used, and either (1) a relatively simple supplementary website or (2) submission of this information to one of the public repositories (ArrayExpress: http://www.ebi.ac.uk/arrayexpress or GEO: http://www.ncbi.nlm.nih.gov/geo/). Reviewers and editors should strive to help authors meet these requirements, and should ensure that there are sound reasons if a publication cannot meet them. This document in no way attempts to eliminate the need for editors or reviewers to use their judgment on both the appropriateness of the presentation and the validity of the report, but rather provides a guideline for them in their evaluation of whether or not a manuscript provides as much information as necessary for others to replicate and interpret the analysis presented.

A summary of the necessary elements, including a checklist for ease of use, is provided at http://www.mged.org.

References

Brazma,A. *et al.* (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.*, **29**, 365–371.

On the behalf of MGED, September 17, 2002

Catherine A. Ball (1), Gavin Sherlock (1), Helen Parkinson (2), Philippe Rocca-Sera (2), Catherine Brooksbank (2), Helen C. Causton (3), Duccio Cavalieri (4), Terry Gaasterland (5), Pascal Hingamp (6), Frank Holstege (7), Martin Ringwald (8), Paul Spellman (9), Christian J. Stoeckert, Jr. (10), Jason E. Stewart (11), Ronald Taylor (12), Alvis Brazma (2) and John Quackenbush (13)

(1) Stanford University
(2) EMBL — European Bioinformatics Institute
(3) Clinical Sciences Centre, Imperial College, London
(4) Bauer Center for Genomic Research, Harvard University
(5) Rockefeller University
(6) Universite D'Aix-Marseille II
(7) University Medical Center Utrecht, NL
(8) The Jackson Laboratory
(9) University of California at Berkeley
(10) University of Pennsylvania
(11) Open Informatics
(12) Center for Computational Pharmacology, University of Colorado School of Medicine
(13) The Institute of Genomic Research