The *Discussion Forum* provides a medium for airing your views on any issues related to the pharmaceutical industry and obtaining feedback and discussion on these views from others in the field. You can discuss issues that get you hot under the collar, practical problems at the bench, recently published literature, or just something bizarre or humorous that you wish to share. Publication of letters in this section is subject to editorial discretion and company-promotional letters will be rejected immediately. Furthermore, the views provided are those of the authors and are not intended to represent the views of the companies they work for. Moreover, these views do not reflect those of Elsevier, *Drug Discovery Today* or its editorial team. Please submit all letters to Joanna Owens, News & Features Editor, *Drug Discovery Today*, e-mail: Joanna.Owens@elsevier.com

To confirm or not to confirm (microarray data) – that is the question

Over the past couple of years, peer-reviewed scientific journals have begun to introduce review criteria for papers incorporating the use of microarray technology for gene expression analysis. These criteria are concerned mainly with conformance to MIAME (minimum information about a microarray experiment) guidelines [1]. However, guidelines relating to post-hybridization confirmatory studies are now being proposed. For example, editors or reviewers might typically request that the results of genomic screens are confirmed by complementary methods such as northern blot analysis.

The corroboration of all array data is clearly impractical because of the large number of genes involved. Therefore, to satisfy these guidelines, most authors normally choose to provide corroborative northern blot or quantitative reverse transcription (qRT)-PCR data for a small number of genes. The question is, at what point have sufficient corroborative experiments been carried out – two genes, five genes or perhaps 10 genes? We scientists like to think of ourselves as objective, but I doubt there are many of us who have not been guilty of cherry picking. In the case of confirming array data, scientists might be selective in their choice of genes, choosing only those that exhibit large changes or those that have been shown previously by others to be upregulated in similar models.

It is perhaps not surprising that confirmatory studies have indicated, with few exceptions, that data from DNA arrays and northern blot and/or RT-PCR analysis support each other qualitatively. In light of such reports, how can one continue to support the stance that statistically significant array data should not be treated as a genuine change until corroborated? Perhaps the strongest argument is exemplified by Affymetrix's recent woes with incorrect sequences on their mouse chip [2]. The problems originated with the use of inaccurate data from public sequence databases, a widely used source of information for making microarrays. Gene splice and tissue variants also exist, and there is the issue of cross-hybridization between closely related genes, such as members of the cytochrome P450 family. Together, these issues introduce an element of uncertainty regarding the true nature of gene sequence and hybridization events on any given array.

An additional problem is faced by those working with biopsy or laser-captured microdissection material. In these cases, the amount of extracted RNA is often insufficient to carry out both array and confirmatory experiments. Should such studies be treated more leniently by editors? It might seem so, but would it then be fair to say that the scientist working on rat liver must corroborate their array data, whereas the scientist working on human testicular biopsies is exempt?

In view of these points, should it not be the case that the requirement to corroborate data in microarray studies is judged on a case-by-case basis rather than having to conform to predetermined standards? However, if journals and referees continue to insist on the inclusion of confirmatory studies, then reasoned standards that consider at least some of the points discussed here should be adopted by the scientific community. For example, 6–10 randomly selected genes from those that are expressed could be confirmed, or two or three genes selected randomly from each of the 'upregulated', 'downregulated' and 'no change' categories could be used. Furthermore, it should be a matter of routine conduct among scientists to report how they selected their confirmatory genes, and which (if any) of their corroborative experiments failed to support the array data. Without such standards, the generation of corroborative data can be manipulated to the point where it has almost no value.

References

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

2 Knight, J. (2001) When the chips are down. *Nature* 410, 860–861

John C. Rockett
Reproductive Toxicology Division (MD-72)
National Health and Environmental Effects Research Laboratory
US EPA, Research Triangle Park
NC 27711, USA
e-mail: rockett.john@epa.gov

1359-6446/03/$ – see front matter PII: S1359-6446(03)02653-9