DNA microarray experiment execution and reporting – minimum standards

Pigment Cell & Melanoma Research has been receiving an increasing number of papers that use DNA microarray technologies. For the results of such microarray studies to be acceptable for publication in this journal the standards set out below should be adhered to. These standards borrow heavily from the review by David B. Allison and co-workers (Allison et al., 2006).

Experimental Design

Biological replicates are essential; a minimum of five replicates per author-defined sample class is required. Performing technical replicates, where one sample is assessed on multiple arrays, is not an acceptable substitute for biological replicates. Pooling of samples (prior to hybridization) as an alternative to replicates is strongly discouraged as this hides inter-individual variation, abrogates estimations of variance for inference testing and risks inclusion of outlier samples which may yield misleading results.

Preprocessing

Image analysis, normalization and data transformation methods must be used and fully documented. Software and algorithms employed must be clearly referenced. Authors are encouraged to include relevant quality control data, but as there are no specific methods for quality control analysis that are widely accepted this is not compulsory.

Data filtering

The journal strongly discourages the practise of filtering data prior to inference testing unless the authors can provide sufficient rationale for doing so and demonstrate that the process does not introduce any form of selection bias to subsequent analytical steps.

Inference

The journal does not accept fold change analysis alone as a test for differential expression. It is recommended that the authors employ t-testing to select genes with specific expression properties relating to the sample classes being examined. If the authors wish to employ a fold change filter, this must be used after statistical selection has been performed. It is strongly recommended that the authors control or estimate the False Discovery Rate for class comparisons and document the softwares/algorithms used to achieve this.

Classification

Unsupervised clustering of entire datasets is generally discouraged as there is little to support the validity of the approach, reproducibility is low and it rarely contributes towards meaningful identification of differential expression between sample classes. Supervised clustering, using filtered gene sets, is acceptable but if the authors intend these filtered gene sets to be applicable to more than just their own sample set they should employ validation (i.e. by testing supervised classification on a separate sample set).

Validation

If authors wish to confirm gene expression patterns by another method (e.g. PCR) they should also employ a different sample set, as confirmation using the same sample set used to obtain relevant data is not considered sufficient validation.

Data

The journal requires that, upon acceptance of their manuscript, the authors deposit their full array data in either of two widely-used databases. These include NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and EMBL-EBI's ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/). Authors should note the relevant accession number(s) in their manuscript.

References

Allison, D. B., Cui, X., Page, G. P., and Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet 7, 55-65.