Reviewer's report

Title: An assessment of recently published gene expression data analyses: Reporting experimental design and statistical factors

Version: 1 Date: 10 April 2006

Reviewer: Alex Sanchez

Reviewer's report:

General

This is an interesting paper that deals with an important issue: the need for using the appropriate methods in the appropriate way when analyzing gene expression data. The paper focuses on the importance of adequately reporting experimental design and statistical methods so that the validity of the results can be better judged by readers and reviewers.

The authors have examined the reporting of a series of factors: a) type of study, b) sample size calculations and power analysis, c) reporting of normalization d) description of data analysis techniques, e) discussion about missing values f) explicit statement of directionality g) explicit statement oh null and alternative hypothesis and h) reference to software tools.

---------------------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

This is a wide range of aspects and it is easy to find minor points where some suggestions can be made. However I have found that some important points have not been considered and in my opinion this omission seriously affects the quality, if not the validity of this study. I think they should be included or at least their omission should be justified.

1) It is true that clinical and biomedical studies are usually not trusted (or at least they should not) if their design does not consider aspects such as sample size or power analysis. But it is also true that in health sciences power analysis is, in general, a well established methodology. In short: there are widely accepted ways to perform it so that there should be no excuse for ommission.

However this fact is absolutely different in microarray studies. Gene expression data are complex, with many more variables that individuals. Their distributions are not clear and what is a more serious problem, these variables are not independent. They have internal dependencies which are difficult to estimate. In spite of the references cited in the paper (most of them belonging to 2005) there is not yet a
theory which is generally accepted as the "valid formulae" for computing sample size or performing power analysis. Besides the previous arguments many existing formulae require the estimation of technical and/or biological variances. The only way to have these estimates is to rely on previous or in pilot studies. Whereas this is a common practice in health studies it is often considered too expensive in microarray studies. In brief, if most studies do not include this type of calculations it is more probably due to the non-existence of valid -or accepted- methods to perform them more than to a bad practice. I think that the arguments given by the authors that "sample size and power analysis should be included in the studies" requires a re-statement because it is difficult to argue that something must be done if it is not clear how to do it.

2) The second point is related to the previous one: A common practice in microarray studies is pooling, that is to combine RNA from several individuals in one "pool" which is used to do one or more arrays. There has been considerable discussion about the convenience of pooling or not (see e.g. Witt & Mc Clure, "Statistics for Microarrays", Kerr (Biometrics 59, 822-828 December 2003) or Kendziorski et al (http://econpapers.repec.org/paper/bepjhubio/1046.htm)). It is strange that this point which is crucial in understanding the experimental design is ignored. It is yet more strange because sample size or power analysis changes according if pooling is or is not considered.

In my opinion pooling is something that should be reported. If the authors do not think so they should argue why they consider it is not important, but they should not ignore it.

3) Last but not least the authors forget, or do not consider important, to mention the need for reporting the p-values adjustment method that has been used. This is a much better established aspect that sample size computations, which are also affected by the type of error type control that the experimenter decides to apply. Non reporting the adjustment method or even worse not adjusting p-values is a serious flaw in many papers and I really do not understand why this aspect is ignored too.

In summary I think that the three points I have discussed represent important omissions. The authors should at least justify why they have decided to exclude these factors from their list what implies that they do not think that these are relevant factors to be reported.

-----------------------------------------------------------------------------------------------
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The authors include between the factors to be reported the verification of homogeneity of variances for ANOVA or t-test. Even if I agree that, when using such a test, this is something to verify, it seems to me that this point receives excessive importance. It seems to implicitly suggest that using a t-test after verifying this assumption is a good practice. However it is known that t-tests are probably the weakest analysis technique for detecting differential expression in the microarray context. It is clearly beaten by improvements such as SAM, linear models with empirical bayes corrections and many others. Perhaps it might be a good idea to comment the weakness of the approach instead of insisting so hard (it appears 7 times on the paper) on the need of verifying the homogeneity assumption.

-----------------------------------------------------------------------------------------------
Discretionary Revisions (which the author can choose to ignore)
What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:
I declare that I have no competing interests