StatNews #74

Sample Size Determination in Microarray Studies
December 2008
Updated 2012

Microarray experiments are quite expensive in terms of material (RNA sample, chip etc.), laboratory manpower and data analysis effort. Hence, calculating appropriate sample size at the design level of the study plays a very important role. Although the importance of calculating an appropriate sample size does not differentiate a microarray study from other studies, there are a few specific considerations in microarray studies that make the sample size calculation more cumbersome. To get more information about microarray data structure, please refer to StatNews #58.

Completely randomized design and matched-pairs design are the two most commonly used experimental designs in microarray studies. In the former, we consider a control group and a treatment group and we are interested in testing if the mean expression level of a given gene in the two groups is same. In the matched-pairs design, each treatment sample is paired with one control sample, creating *n* pairs of matched observations. Within a matched pair, two observations are dependent and we are interested in testing if the mean expression levels for genes from the two groups are different.

There are four factors that play an important role in determining sample size in microarray studies. They are: the statistical model used in analyses, the effect size to be detected, the desired power, and the significance level to be used while taking into account the multiple testing problem. As hundreds or thousands of tests are conducted, simply using the significance level for a cutoff without adjusting for multiple tests will increase the chance of false positives. One approach to multiple testing is to control the family-wise error rate (FWER), which is the probability of accumulating one or more false-positive errors over a number of statistical tests.

In a list of differentially expressed genes that satisfy an FWER criterion, we can have high confidence that there will be no errors in the entire list. The simplest FWER procedure is the Bonferroni correction where the nominal significance level is divided by the number of tests. FWER criteria are very conservative, and they may substantially decrease power when the number of tests is large. An alternative approach to multiple testing considers the false discovery rate (FDR) which is defined as the expected percent of false predictions in a set of predictions. FDR criteria allow a higher rate of false positive results and thus can achieve more power than FWER procedures. The free software, 'R' has a useful package ***"sizepower"***, which can be used to calculate sample size.

For an example of a sample size calculation in microarray studies using R's "sizepower" go to: http://bioconductor.org/packages/2.0/bioc/html/sizepower.html

The following references are useful if you would like to learn more about computing sample size and power in microarray studies. Please contact the Cornell Statistical Consulting Unit for assistance during the statistical analysis phase of your study and also prior to collecting the data for assistance with the design and sample size calculations.

Author: Resmi Gupta

References:

1. Chen-An Tsai, Sue-Jane Wang, Dung-Tsa-Chen and James J. Chen, "*Sample size for gene expression microarray experiments*", Bioinformatics, Vol. 21, no. 8, 2005: 1502-1508.
2. Mei Ling Ting Lee and G.A. Whitmore, "*Power and sample size for DNA microarray studies*", Statistics in Medicine 21:3543-3570.
3. Sin-Ho Jung, Heejung Bang, Stanley Young, " *Sample size calculation for multiple testing in microarray data analysis*", Biostatistics, Vol. 6 No.1:157-169.
4. http://www.r-project.org/