International Biometric Society

Statistical methodology for the analysis of dye-switch microarray experiments

Tristan Mary-Huard [1], Julie Aubert [1], Nadera Mansouri-Attia [2], Olivier Sandra [2] and Jean-Jacques Daudin [1]

[1] UMR AgroParisTech/INRA 518, Paris, France
[2] UMR INRA/ENVA/CNRS 1198, Jouy en Josas, France

In individually dye-balanced microarray designs, each biological sample is hybridized on two different slides, once with $Cy3$ and once with $Cy5$. While this strategy ensures an automatic correction of the gene-specific labelling bias, it also induces dependencies between log-ratio measurements that must be taken into account in the statistical analysis.
We present two original statistical procedures for the statistical analysis of individually balanced designs. These procedures are compared with the usual ML and REML mixed model procedures proposed in most statistical toolboxes, on both simulated and real data.
Random terms taking into account the array and the sample effects must be included in the statistical model at the gene level for dye-switch experiments. We showed on simulations that the naive paired T-test, which does not take into account the biological sample effect, leads to more false positives than expected, especially when the biological sample effect is high. It may be safely used only when the biological variance is lower than the technical variance. The REML estimate for mixed model provides an approximatively correct false positive rate, at the price of high computational complexity, lack of convergence for low or medium sample sizes and sometimes spurious results. To the contrary, the method we propose is easy to implement and not computationally intensive. The method is protected against spurious results, leading to a more robust and powerful analysis than REML when the biological variability is high and the number of samples low, an usual situation in microarray experiments.