Recent developments in transcriptome-oriented biotechnologies have made possible the comparative analysis of thousands of genes expression changes in parallel. These data usually consist in the measurement of gene expression under various biological conditions that can potentially provide information on the complex transcriptional activity for the biological system [1].

In such conditions, questions raised by research scientists relate to class comparison, class prediction and class discovery problems. Concerning class comparison, the aim is to select relevant genes the transcriptional changes of which are related to a clinical or biological outcome. In such a case, a major multiplicity problem arises that leads to a renewed interest for multiple comparison procedures taking into account false positives.

Until now, statistical procedures have mostly relied on the multiple testing framework in order to control false positive conclusions. In this framework, two quantities have been considered: the Family Wise Error Rate (FWER) and the False Discovery Rate (FDR). The FWER, which is the oldest criterion considered in multiple comparisons, is defined as the probability of at least one false positive conclusion over all the true null hypotheses (a null hypothesis corresponds to the lack of relationship between gene expression measurement and a response variable). The most commonly used methods are the Bonferroni and Sidák methods [2]. Extension of these two methods are based on step-down [3] or step-up principles where null hypotheses are tested sequentially [2]. In complement to the previous methods, resampling (or permutation based) procedures that make no distributional hypothesis but incorporate correlation structures and distributional characteristics (for a review see [4]) have been developed and applied to microarray gene expression study [5][6][7].
However, as argued by Benjamini and Hochberg [8], controlling the FWER in multiple testing settings may not always be appropriate. Indeed, in large-scale hypothesis generating studies such as microarray experiments, this criterion becomes so conservative that the probability of detecting any true association is, in some cases, almost nil. As an alternative and less stringent concept of error control, Benjamini and Hochberg introduced the false discovery rate (FDR). The FDR is the expected proportion of erroneously rejected null hypotheses among the rejected ones. The main reason for controlling the FDR is that it controls a quantity that is relevant and leads to more powerful procedures than those relying on the FWER. Based on this concept, they initially developed a step-up procedure under the hypothesis of independence which controls FDR at a pre-specified value. Extensions to the case of dependent tests have recently been proposed by Benjamini and Yekutieli [9].

In their seminal paper, Benjamini and Hochberg also presented other error criteria such as the conditional expectation of the proportion of false discoveries given that at least one discovery (later called “positive False Discovery Rate” (pFDR) by Storey [10]). This latter criterion was not taken into consideration by the authors since it cannot be controlled [8].

In another framework, Tusher et al. [11] have proposed to calculate a conservatively biased estimate of the pFDR using a non-parametric empirical Bayesian approach. Their procedure called “Significance Analysis of Microarray” (SAM) has been applied to microarray data analysis and implemented in the widely used SAM software [12]. Storey [13] has proposed another procedure providing a conservative point estimate for the pFDR and showed that in asymptotic setting, pFDR and FDR are equivalent. This latter author has also introduced a meaningful gene-specific quantity called the q-value [14] which measures the minimum FDR at which a gene can be called differentially expressed. This q-value corresponds to the posterior probability that a gene is not modified given that gene statistic is as extreme as the one observed for this gene in the data. These q-values, which are provided by SAM, may be used by the investigator as criteria for selecting all features with q-value less or equal to a chosen false discovery rate threshold value (for more details about SAM and q-values see, [11][14]).

Assuming independence of gene expression measurements across samples, under the null hypothesis, the distribution of the gene-based statistics (or p-values) can be considered in the framework of finite mixture model [15]. In this setting the population of genes is considered as composed of two groups of genes, those that are equally expressed between the samples, and those that are differentially expressed. Thus, the use of mixture model framework provide an efficient way for comparing the different estimators offered for the
FDR and emphasizes situations where conservative bias of non parametric FDR (SAM, Storey’s method) could be a problem.

This talk will be organised as follow. First we will present the framework of the multiple comparison procedures. Then, we will briefly describe the most used procedures. The different estimators proposed for the pFDR will be presented in a mixture model. Finally, we will present the performance of different estimators on simulated and real data.

References:


