Microarray Gene Expression Data Analysis in Cancer Research

Prof. Eric Y. Chuang, Taiwan

Following the development of advanced technologies in biomedical science, large quantity of basic and clinical data can be generated and obtained effectively through new techniques. For instance, DNA microarray technology affords the opportunity to interrogate the expression of thousands of genes in a single experiment. Not surprisingly, the simultaneous study of expression profiles of a huge population of genes present on DNA microarrays is challenging. It is important to utilize appropriate approaches to analyze data generated from microarray experiments in order to produce biologically meaningful interpretation. Therefore, two microarray studies with different approaches will be discussed in this presentation.

First, traditionally, gene expressions, monitored at different time points or under different conditions, were analyzed by clustering genes having similar expression profiles and by a comparison of distributions, line graphs of a clustered gene expression. However, many of previous studies did not critically address the statistical significance of the observed differences with respect to measurement accuracy. A single time course study involves multiple arrays measured under the same conditions at different time points. In principal, systematic study of expressions of successive time points can provide an estimate of reliability. However, a true replicate would additionally cover the variations that can occur within biological samples. Thus, we investigated time course experiments with multiple replicates to obtain the statistical significance for every measured expression ratio and to verify reproducibility. The global gene expression of breast cancer cells exposed to three different oxidants: MEN, HP, and TBH was followed by microarray analysis in three replicate experiments. RNA collected after treatment (at 1, 3, 7, and 24 h) rather than after a single time point, enabled an analysis of gene expression patterns. Using a 17K microarray, template-based clustering and multidimensional scaling analysis (MDS) of the gene expression over the entire time course identified 421 genes as being either up- or down-regulated by the three oxidants. In contrast, only 127 genes were identified for any single time point and a 2-fold change criterion. Surprisingly, the patterns of gene induction were highly similar among the three oxidants; however, differences were observed, particularly with respect to *p53*, *IL-6*, and heat-shock related genes. Replicate experiments increased the statistical confidence of the study, whereas changes in gene expression patterns over a time course demonstrated significant additional information *versus* a single time point. Analyzing the three oxidants simultaneously by template cluster analysis identified genes that heretofore have not been associated with oxidative stress.

Secondly, the hereditary form of retinoblastoma (Rb) is associated with a germ line mutation in one RB allele and is characterized by the occurrence of multiple, bilateral Rb tumors and a predisposition to the development of second cancers. In an earlier study, we observed an unexpected hypersensitivity to ionizing radiation in skin fibroblasts derived from unaffected parents of children with hereditary Rb. In at least four of these five families, there was no family history of Rb, indicating a new germ line mutation. We hypothesized that the increased parental cell sensitivity to radiation may reflect the presence of an as yet unrecognized genetic abnormality occurring in one or both parents of children with Rb. Therefore, we applied 20K microarrays to determine whether differences in gene expression profiles occurred in the unaffected parents of patients with hereditary Rb relative to normal individuals. A distinct difference was observed in the patterns of gene expression between unaffected Rb parents and normal controls. By use of the prediction analysis for microarrays (PAM) and principal component analysis (PCA), significant differences between the two groups were identified when as few as nine genes were analyzed. Further study of this phenomenon may offer a new insight into the genetic mechanisms of Rb and perhaps more broadly in cancer biology.

Finally, knowing how to use appropriate software tools on powerful computers for microarray data analysis has become a necessity for biologists to obtain reliable biological information as well as to facilitate identifying new genes or targets for further investigation.