ABSTRACT

By using the gene shaving method, it is possible to examine gene expression data and identify relevant gene clusters that require further investigation.

INTRODUCTION

The aim of this study was to describe and characterize the expression pattern of the human LTR1 gene and its associated LTR2 gene in relation to the expression of a novel gene in the human brain and to compare the expression profile of the LTR1 gene in the brain with that of the LTR2 gene in the brain of a mouse model of epilepsy.

Materials and Methods:

In this study we used the LTR1 gene as a marker for the presence and absence of a novel gene in the human brain. The expression pattern of the LTR1 gene was examined in vivo in the rat brain with or without the use of an LTR2-specific polyclonal antibody (polyclonal) and the expression pattern of the LTR2 gene in the mouse brain was examined in vitro A new statistical approach, gene shaving, aims to identify groups of elements in a biological sample that exhibit coherent expression patterns and are optimal for different properties of the variation in their expression. Figure 1 demonstrates the dataset we used for our study, which included 4673 gene expression measurements on 48 patients with diffuse large B-cell lymphoma (DLCL). These data were previously described in detail. Column labels refer to different patients, and the rows correspond to genes. The order of these rows and columns is arbitrary. The use of clustering methods to arrange genes in a systematic manner, with similar-genes-separate orders being placed close to each other (see later section for details) has been proposed by some authors. As illustrated in Figure 2, we have applied hierarchical clusterering to both genes and samples, which produces non-unique orderings that ensure branch crossing in the corresponding dendrogram. Figure 2 shows the original data, including its order of rows and columns. A certain level of organization is demonstrated in Figure 2, and this approach can be employed to identify connections between the genes and samples. However, fine details may be lost with any data that is reduced in size. For instance, suppose the expression patterns of one subset of genes, such as those associated with the speed of tumor growth or the differentiation of immune responses, are shown in a second way; yet another method, known as two-way hierarchical clustering, attempts to find unified reordering of the samples for all genes. Gene shaving is an approach to obtaining small and coherent gene clusters that vary widely across samples. Figure 3 illustrates three gene groups for DLCL data, which are found using shaving. Some genes within each cluster are located in close proximity to each other in the hierarchical clustering of Figure 2, while others are quite distant. The samples in Figure 3 have been ordered based on average gene expression values, and the variance measures at the top of each cluster are discussed later. We use an automatic method to determine the size of the clusters, which is not like a hard way through the large number of genes. The three cluster-average genes in each group are reasonably uncorrelated, while the correlation of one gene to another is measured by the mean of their cluster sizes. The article is structured as follows: Gene shaving (with the method described above), the gap estimate of cluster size, predicting patient survival using gene cluster averages (included below), and finally some generalizations to the conclusion of 'Conclusions'. Gene Shaping (details below): This article

delves into the treatment of gene shaving, with the section discussing the methods used for deciding on the appropriate size of the cluster to be used and how best to use them both, followed by the discussion of statistical analysis.

CONCLUSION

Remarkable conclusions Our proposal involves the use of shaving methods to isolate intriguing gene clusters from DNA microarray experiments. These methods can be either unsupervised or supervised by information about available samples, such as class labels or survival time. The shaving techniques aim to identify genes with high variation across the samples and coherence across them, which cannot be achieved by simple clustering or thresholding of individual genes based on sample variation. We have developed our model-based shaving method, which allows for the inclusion of other prognostic factors to assist in finding intriguing gene clusters. If a specific outcome is available for each sample, the method searches for matched genes in the group with corresponding column average genes, who may influence the outcome and potentially other contributing factors. These real-valued expression levels are the focus of our discussion on microarray data xij. However, other arrays can also generate different types of data. For instance, some array systems identify single-nucleotide polymorphisms (SNPs) and use one of the two independent, unmarked digits in the array that produces the resulting data: a random selection of SNP values based on an algorithm called "pragmata". To perform this, we simply apply principal component shaving