Molecular Classification of Cancer by Gene Expression Monitoring

Abstract

In this report a generic approach to cancer classification based on gene expression monitoring by DNA microarrays is described and applied to human acute leukemias as a test case. The dataset will first be analysed using data exploration techniques and clustering, for identifying and grouping similar data points. As this is a highly dimensional dataset, dimensionality reduction will be carried out to reduce the number of input variables in the training data. Then the aim of this report, to demonstrate the feasibility of cancer classification based solely on gene expression monitoring and compare the performance of miscellaneous classifiers on the dataset will be carried out. The results will demonstrate how this general strategy(proposed by Golub et al. in 1999) for discovering and predicting cancer classes for other types of cancer, independent of previous biological knowledge works.

Preprint. Under review.