Detecting Broad-Band and Selective Correlation Patters among Gene Expression and Drug Activity Data

W. Dubitzky[1], D. Berrar[1], M. Granzow[2], R. Eils[1]

[1]Intelligent Bioinformatics Systems, German Cancer Research Center; [2]phase-it, Intelligent Solutions AG

It is of fundamental scientific interest to understand the mechanisms of the effects of anticancer drugs on tumor tissue. A major effect of these drugs on tumor cells is growth inhibition. This can be measured by a sulphorhodamine B assay that detects the changes in total cellular protein after 48 hours of drug application [Weinstein et al. 1997]. Cancer cells can be characterized by their gene expression profile using cDNA microarrays. This technique produces a view upon the activity status of up to several thousand genes in one single experiment. A new emerging scientific field called pharmacogenomics aims at linking pharmacological and gene expression research. A key element of pharmacogenomic research is to identify and characterize new targets for developments in drug therapy through the understanding of genes involved in drug action and metabolism.

A study of the impact of 1,400 chemical compounds and the gene expression profiles of 1,376 genes from 60 cancer cell lines was described by [Scherf et al. 2000]. The gene expression data comprise mainly ESTs of known and unknown function given by the negative logarithm of the ratio between the red and green fluorescence of the signals. The drug data set consists of the measurement of drug activity expressed as the negative logarithm of the concentration required to inhibit cell growth by 50% in comparison with untreated control cells (GI_{50}). The cancer cell line (NCI60) comprise 9 categories or classes: colorectal, renal, ovarian, breast, prostate, lung, central nervous system, leukemia and melanoma cancers. The task is now to analyze the data sets and to detect correlations between gene expression patterns and drug activity. One problem in the constellation of the data sets described above is the fact that they were collected independently from each other. An advanced approach would include gene expression data without treatment and after application of drugs. However, data of this kind are not publicly available to date.