Construction of Metagenes by Conditional Factor Analysis

Andreas Mayr, Djork-Arne Clevert, Andreas Mitterecker, An De Bondt, Willem Talloen, Hinrich Göhlmann, Sepp Hochreiter

Gene selection for classification microarray data is important to improve a classifier, to find new drug targets, or to design predictive gene signatures.

However, gene selection has a very bad reputation. One reason is the "selection bias" of early publications, where genes have been selected or normalized using the validation set. Now, normalization and selection is done separately on different training sets to evaluate a method. Another reason for the bad reputation is the small intersection of gene sets selected by different methods. Besides these problems, new array generations make selected gene sets obsolete as they do not work on new platforms (e.g. Affymetrix' U95, U133+, and array plates).

In order to combine the selection results on different training sets, to combine gene sets selected by different methods, and to transfer gene signatures across platforms, we propose the construction of "metagenes" by conditional factor analysis. A "metagene" of a chosen gene is a hidden factor of the expression values which varies around the expression value of this gene. Increasing the variation around the expression of the chosen gene allows estimating the factor by including more genes that are related to the chosen gene. The given variance of a "metagene" determines its size (number of included genes) and its robustness.

We apply our method to Affymetrix gene chips, where we compare and unify different genes signatures and even transfer genes signatures across platforms.