# RISKS AND BENEFITS OF POOLING BIOLOGICAL SAMPLES IN GENE EXPRESSION STUDIES

L. Edler, A.Benner, C. Ittrich

*German Cancer Research Center, Heidelberg, Germany*

Email: edler@*dkfz.de*

The determination of high-dimensional gene expression information is frequently performed as companion protocol both in experimental bioassays as well as in clinical studies. Because of the high gene chip costs, standard statistical "one-subject-one array" designs with replicates are prohibited even in small studies. Therefore, and in particular for bioassays, various pooling techniques have been used, e.g. mixing samples of subjects and tissues. This raises questions on the optimal experimental design when the number of subjects $n_s$, the number of arrays $n_a$, and the number of subjects $r_j \leq n_s$, pooled into array j, j=1,...,$n_a$, and the total number $n_p$ of pools can be chosen deliberately. Assuming the two components of biological and technical variance to be known, optimal $n_s$-$n_a$-$n_p$ designs could be constructed by some research groups for situations where the data are balanced. This presentation will elaborate on the limitations of the designs proposed by juxtaposing them to current practice. It is argued that the simplification of only two variance components may be too strict not accounting for further subcomponents occurring both in the biological and the technical part of the experiment, e.g. pure biological (genetic) and tissue sampling and preparation. Further limitations to be considered exist for randomization at various levels of replication as well as due to the presence of a large number of endpoints and possible correlations between those.