Abstract

As high-throughput data measurements in biomedical studies have become routine, the challenges shifted from data generation to data analysis. In particular, the integration of multiple omics data sets is an intriguing but ambitious task.

This thesis comprised three specific aims. First, integrative analysis methods shall be applied to multiple omics data sets. Second, existing integrative analysis methods shall be investigated. Third, the development of a novel integrative pathway enrichment approach (IPEA).

For the first aim, (multiple) co-inertia analysis (MCIA) was applied to *A. gambiae* and to the cross-species comparison of the expression systems *P. pastoris* and Chinese hamster ovary (CHO) cells. In the former case, a high structural concordance between the hemocyte transcriptome and the granulocytic proteome could be shown. In the latter case, a number of secretion and ribosome relevant target genes and proteins were identified by a detailed characterization of four production strains.

For the second aim, three integrative analysis methods (MCIA, generalized singular value decomposition (GSVD) and integrative biclustering (IBC)) were applied to the transcriptome and proteome of the parasite *P. falciparum*. From the intersection of these results, a network of biological processes was derived which characterizes the parasite's life cycle stages and unifies numerous findings from the past 25 years of research in a single analysis. Additionally, a traditional gene set enrichment analysis (GSEA) was applied to validated target genes of two sets of human microRNAs. The 36, respectively 35, enriched neuron related biological processes were almost identical between the two sets, although the overlap in the corresponding miRNA lists was below 50%.

For the third aim, in order to overcome flat gene list limitations of the traditional GSEA, we developed a novel integrative pathway enrichment analysis (IPEA). Our IPEA approach combines scores from a multivariate analysis with pathway specific scores based on network topology. Enriched pathways computed by IPEA are characterized by biologically relevant concordance between the measured data and the intrinsic structure of the pathways. IPEA visualizes the results as a double bipartite graph of activated features and enriched pathways. Applied to 38 matched tumor and stroma samples from ovarian cancer patients, IPEA reveals an unprecedented view of the cross-talk between tumor and stroma, suggesting new targets for the treatment of ovarian cancer, e.g. CTNNB1, ERBB4 and SMAD4 which have already shown their potential in the therapy of other cancers.