Abstract

System-wide molecular profiling studies measure biological samples and organisms on a global scale, on various types of biological molecules, for example on DNA, RNA, protein and metabolite level. This results in high-dimensional omics data (genomics, transcriptomics, proteomics and metabolomics) obtained from high-throughput technologies. Dimension reduction techniques, such as principal component analysis (PCA), correspondence analysis (CA) and non-symmetric correspondence analysis (NSCA), are very important tools for the analysis of this data but they are limited to the analysis of a single level. Therefore, integrative analysis methods, such as co-inertia analysis (CIA) and multiple co-inertia analysis (MCIA), have been developed which allow simultaneous analysis of two or more data sets. MCIA is a powerful tool in the analysis of multiple high-dimensional data sets because this method enables the visualization of measurements/samples using a lower number of features/variables and therefore, facilitates the detection, representation and biological interpretation of the correlated structure within and between the different data sets. This study focuses on the comparison of sequential (PCA, CA, NSCA) and integrative analysis (MCIA) methods based on data comprising 60 human cancer cell lines at the transcript and protein level.

Surprisingly, PCA and NSCA do not show considerably higher degree of divergence among gene and protein expression data. These two methods demonstrate mainly inconsistent clustering based on protein expression. It is noteable that CA based on proteome level provided a more homogenous clustering compared to PCA and NSCA. MCIA proves that protein and gene expression profiles can be regarded as powerful molecular descriptors of different cancer tissues and integrative analysis of both provides a deeper insight on multiple layers of biological systems compared to any analysis tool alone. The most homogenous clustering and separation of the different carcinomas according to the most influential genes was achieved with the MCIA, and the different types of cancer were thus clear separated on the first two principal components. The most influential genes based on MCIA at each end of the first principal component and the second principal component provided useful information regarding the clustering of the individual cancer types. Thus, MCIA provide more information and is more powerful than corresponding sequential. Furthermore, MCIA achieves a homogenous separation of mesenchymal (glioblastoma and renal cancer cells) and epithelial (leukemia and colon cancer cells) markers which promote epithelial-mesenchymal transition (EMT) that has a significant function in cancer biology and is also involved the malignancy and metastasis of epithelial cancer cells. Mesenchymal cells have migratory and invasive characteristics which have a significant function in malignant metastasizing cancer. Thus, MCIA also yields important information regarding invasive, metastatic phenotype in cancer cells.