Introduction

The recent development of mass spectrometry-based proteomics techniques is becoming possible to perform in-depth analysis of the entire proteome with protein identification of few thousand proteins. It is several magnitudes lower than to detect the proteins even at the single cell level.

Methods

To test ProteinCenter software, we used data from a recent publication “Deep proteome and transcriptome mapping of a human cancer cell line” [1]. The current study was performed using the Thermo Fisher Scientific ProteinCentre XQ system. The LC-MS/MS study of HeLa cells and comparison to transcriptome data was carried out on the ProteinCenter for a set of low-abundance proteins. These proteins were expressed on the level up to 50 copies per cell in the range of 0.2 copies per cell for Centriolin up to 3.3E+7 for Cofilin-1. Highly abundant proteins (top 100) are mainly of cytoplasmic origin and involved in cell growth, differentiation and development. On the level of molecular function they are involved in ATP binding, structural molecule activity, catalytic activity.

Results

Statistical Analysis of HeLa Proteome

Very often, proteomics and transcriptomics experiments are performed in different laboratories at different times and using different experimental setups. As a result, the models of correlation between the protein and mRNA levels differ. Therefore, the most reliable pathway analysis should be performed using large-scale data derived during three experiments could introduce a level of bias. In the best case scenario, the same experimental setup is used to perform the experiments or to perform the same experimental setup. That is why we selected the very well controlled study of HeLa cell analysis and the reference protein dataset (data from [2]). The advantage of this approach was obvious: first, the bias was lower than 10% and second, the accuracy of this statistical analysis can be higher than 90%. The main advantage of this approach is that the data from study is not sensitive to the level of proteomic analysis of datasets.

Overview

Purpose: Demonstration of fast and effective way to perform statistical and bioinformatic analysis of large scale proteome and transcriptome studies.

Methods: Set of bioinformatic data-mining tools

Conclusion

While the knowledge of over-represented terms or pathways is helpful, it does not directly link protein or transcriptome networks and regulatory network analysis. Our approach to improve this is the combination of network analysis, pathway analysis and expression pattern analysis. To improve this is the combination of network analysis, pathway analysis and expression pattern analysis. To improve this is the combination of network analysis, pathway analysis and expression pattern analysis. The final results are not discussed here and the focus is on the correlation between proteins and transcriptome that are not only related to the selected study, but also to other studies and experimental setups.

References