

# Pathway analysis and transcriptomics improve protein identification by shotgun proteomics from samples comprising small number of cells - a benchmarking study

Jing Sun, Guang Lan Zhang, Siyang Li, Alexander R. Ivanov, David Fenyo, Frederique Lisacek, Shashi K. Murthy, Barry L. Karger, Vladimir Brusic

School of Medicine

## Abstract

**Background:** Proteomics research is enabled with the high-throughput technologies, but our ability to identify expressed proteome is limited in small samples. The coverage and consistency of proteome expression are critical problems in proteomics. Here, we propose pathway analysis and combination of microproteomics and transcriptomics analyses to improve mass-spectrometry protein identification from small size samples. **Results:** Multiple proteomics runs using MCF-7 cell line detected 4,957 expressed proteins. About 80% of expressed proteins were present in MCF-7 transcripts data; highly expressed transcripts are more likely to have expressed proteins. Approximately 1,000 proteins were detected in each run of the small sample proteomics. These proteins were mapped to gene symbols and compared with gene sets representing canonical pathways, more than 4,000 genes were extracted from the enriched gene sets. The identified canonical pathways were largely overlapping between individual runs. Of identified pathways 182 were shared between three individual small sample runs. **Conclusions:** Current technologies enable us to directly 10% of expressed proteomes from small sample comprising as few as 50 cells. We used knowledge-based approaches to elucidate the missing proteome that can be verified by targeted proteomics. This knowledge-based approach includes pathway analysis and combination of gene expression and protein expression data for target prioritization. Genes present in both the enriched gene sets (canonical pathways collection) and in small sample proteomics data correspond to approximately 50% of expressed proteomes in larger sample proteomics data. In addition, 90% of targets from canonical pathways were estimated to be expressed. The comparison of proteomics and transcriptomics data, suggests that highly expressed transcripts have high probability of protein expression. However, approximately 10% of expressed proteins could not be matched with the expressed transcripts.

**Original language** English

**Article number** S1

**Journal** [BMC Genomics](#)

**Volume** 15

**State** Published - Dec 8 2014

Sun, J., Zhang, G. L., Li, S., Ivanov, A. R., Fenyo, D., Lisacek, F., ... Brusic, V. (2014). Pathway analysis and transcriptomics improve protein identification by shotgun proteomics from samples comprising small number of cells - a benchmarking study. *BMC Genomics*, 15, [S1]. DOI: [10.1186/1471-2164-15-S9-S1](https://doi.org/10.1186/1471-2164-15-S9-S1)