Comparative Study of Global Protein Turnover in Tissues and Cell Lines

Martin Damsbo, Jacob Poder, Erik Nielsen, Christian Ravnsborg, Alexandre Podtelejnikov
Thermo Fisher Scientific, Odense, Denmark
Overview

Purpose: Demonstration of fast and effective way to perform statistical, comparative, and bioinformatical analysis of large-scale proteome and transcriptome studies

Methods: Set of bioinformatic data-mining tools

Results: Comprehensive bioinformatic analysis of data derived from large-scale LC-MS/MS proteomics studies of protein turnover and their comparison, including the data of non-mass-spectrometric origin

Introduction

While the success of the mass-spectrometry-based proteomics approaches in analysis of protein expression is undisputable, more fundamental questions like global protein turnover remain a challenging task. The proteome is extremely dynamic and exists in constant flux. Most current proteomics studies present one particular proteome state at a given time and under given conditions, and do not describe the level of protein synthesis and degradation. Recent attempts to study protein turnover based on the stable isotope labeling use different approaches and different algorithms for data processing. Here we performed the comparative analysis of rates of protein turnover from these different approaches from different tissues and cell lines using Thermo Scientific™ ProteinCenter™ software.

Methods

ProteinCenter software enables scientists to perform complex comparison studies of protein turnover derived from experimental data, available databases, and literature. Protein turnover data sets were obtained from literature sources: A study of M. pneumoniae proteome based on LCMS based on an approach involving stable isotope-labelled reference peptides (1), Analysis of proteome dynamics in the mice based on 15N enriched algae diet (2), SILAC-based quantitative mass spectrometry data from fractionated HeLa cells (3), Pulse SILAC study with protein turnover rates based on relative isotope abundance (RIA) in living mice (4). We also selected data from a non-mass-spectrometric method based on reporter-dependent approach (5). Genomic data was logged into ProteinCenter software by using Entrez gene accession codes and corresponding taxonomy. Comparative statistical analysis of data sets was performed based on Benjamini-Hochberg correction of p-values with false discovery rate (FDR) values at 5%. Statistical analysis was performed with XLSTAT®2013 software package (http://www.xlstat.com).

ProteinCenter software is a protein-centric tool that integrates the contents of a large number of public protein sequence databases (Figure 1), experimental protein identifications, and a number of bioinformatics tools. It contains a non-redundant database with more than 25 million proteins from the major public protein databases – distilled from 210 million accession codes from past and present versions of public databases.

FIGURE 1. Consolidated database. Genomic databases supported by ProteinCenter software are highlighted by the circle.
Results

Proteome and Transcriptome Analysis in Bacteria Cell Lines

Despite enormous success of mass-spectrometry-based proteomics methods in characterization and quantification of proteins and their modifications in cell lines and tissues at different conditions, our fundamental understanding of a proteome and its dynamics is currently rather poor. There are no clear explanations of limited or even a lack of correlation between mRNA expression and protein abundance values. In order to explain the relationship between a transcriptome and a proteome, the dynamic processes involved in protein turnover have to be better understood.

Simplistically speaking, the protein abundance depends on two major processes – protein synthesis and protein degradation. While protein synthesis can be assumed as a function of the mRNA concentration, the protein concentration, over time, will be equal to the rate of translation minus the rate of degradation, and can be characterized by first order kinetics.

Proteins exhibit tremendous variability in their lifetime and there is significant heterogeneity even within proteins that have similar functions. Approximately 30 attempts (6) have been made recently to study global protein turnover using metabolic labeling followed by advanced mass spectrometric analysis.

Here we are focused on facilitating the analysis of protein and mRNA expression as well as protein turnover by using features of ProteinCenter software package. First we selected the study of integrated large-scale average abundance data for mRNA and proteins with turnover rates in the bacterium M. pneumoniae (1).

A modest correlation between $k_{\text{translation}}$ and $k_{\text{degradation}}$ with Pearson's correlation coefficient of 0.428 was observed. We applied protein profiling based on a soft clustering algorithm (Figure 2) to uncover proteins and protein groups involved in related processes and pathways based on protein expression along a 4-day (AQR11–AQR14) time course (AQR10–control). Such an approach allows us to isolate significant protein variables, and then perform separate statistical or bioinformatic analysis of the selected set.

**FIGURE 2.** Molecular function (GO) profiling based along a 4-day time course

- **GO 1:** [Molecular function activity]
  - AQR10
  - AQR11
  - AQR12
  - AQR13
  - AQR14

- **GO 2:** [Molecular function activity]
  - AQR10
  - AQR11
  - AQR12
  - AQR13
  - AQR14

- **GO 3:** [Molecular function activity]
  - AQR10
  - AQR11
  - AQR12
  - AQR13
  - AQR14

- **GO 4:** [Molecular function activity]
  - AQR10
  - AQR11
  - AQR12
  - AQR13
  - AQR14
Heat map can also assist the visualization of the comparison of transcriptome and proteome data. Figure 3 shows protein (AQR1–3) and mRNA (AQR4–6) fold change in response to heat shock, osmotic stress, and DNA damage correspondingly. Heat maps can be generated either on protein or gene level as well as according to Gene Ontology classification or pathways.

FIGURE 3. Heat maps of transcriptome and proteome data in response to heat shock, osmotic stress, and mitomycin-induced DNA damage

Proteome Analysis of HeLa Cell Lines

A general limitation of liquid chromatography-mass spectrometry (LC/MS) strategy is that it does not discriminate between peptides arising from different protein isoforms encoded by the same gene. A recent attempt to overcome that obstacle and determine turnover rates in different subcellular compartments in HeLa cell lines has been made (3) by combining cell fractionation and the separation of intact proteins by chromatography, prior to enzyme digestion and peptide identification by mass spectrometry. Turnover rates were calculated for more than 6,000 identified proteins. We performed GO analysis on the groups with the lowest and highest protein turnover rates in order to determine which biological functions were over-represented in selected datasets. Cellular localization analysis showed that plasma membrane part, transport vesicle, Golgi apparatus part, and endosome were all over-represented in the group of proteins with slow turnover. Among under-representative cellular components were mitochondrial matrix, nucleosome, and ribosomal subunit. The results of GO molecular function are presented in Figure 4.

We used individual peptide information to distinguish different isoforms of the proteins by means of “indistinguishable proteins” clustering. Figure 5 shows an alignment of one of these clusters. These two isoforms have slightly different protein turnover rates of 12.0/day and 13.1/day.

FIGURE 4. Over-representation analysis of GO molecular function for proteins with slower and fast protein turnover ratios

FIGURE 5. Fragment of cluster alignment of 60S ribosomal protein L32
Mass-Spectrometry-Based Methods Versus GFP Fusion Approach

Although the most advanced protein turnover studies are based on stable isotope labeling and mass spectrometric RIA measurements, a number of non-mass-spectrometric methods are also emerging. For example, protein turnover can be monitored by coupling flow cytometry with microarray technology to track the stability of individual proteins within a complex mixture (4). It is a reporter-dependent approach named GPSP (global protein stability profiling) by the authors. After transducing cells with a bicisitronic retroviral vector encoding RFP (red fluorescent protein) immediately downstream of an internal ribosome entry site and an upstream cDNA of the gene of interest fused to GFP (green fluorescent protein), they determined the RFP/GFP ratio for each library member by flow cytometry. The RFP/GFP ratio is then converted to a protein half-life value algorithmically, using measured values from a panel of GFP variants with biochemically defined half-lives. Here, we compare this GPSP to a pulse-chase SILAC™ metabolic labeling followed by mass spectrometric determination of protein turnovers. Both datasets were imported into ProteinCenter software (protein and gene identifiers), and compared and clustered. The total of 2,824 shared proteins group underwent Pearson correlation analysis. Figure 6 shows relatively poor correlation. The correlation matrix was r=0.148 with p value < 0.0001. First of all, such poor correlation can be explained by the origin of the samples. SILAC-based analysis was conducted using HeLa cell lines, while GPSP data was obtained from 293T embryonic kidney cells. Another explanation can be that GPSP underestimates protein stability by interfering with protein assembly due to a tagging approach that causes spuriously high degradation rates.

FIGURE 6. Protein half-life correlation: GFP fusion vs. SILAC metabolic labeling

Protein Turnover Analysis of Tissues

While calculation of protein turnover rates in cell lines is relatively easy to manage by the measurement of relative isotope abundances, it is becoming a significantly more complex task for tissue or the whole organism analysis. A number of different methods and algorithms are currently being used to deconvolute data to achieve greater accuracy. We performed comparative analysis of rates of protein turnover determined by 15N enriched algae diet and 13C6-lysine and in different tissues and fluid: brain, liver, and blood (2) followed by bioinformatics statistical analysis by means of GO and participation in KEGG™ pathways. Figure 7 presents the comparison of liver, blood, and brain proteins. Overall rates of protein turnover varied significantly. Proteins observed in brain had longer turnover rates in comparison with blood or liver proteins. At the same time, we noticed that proteins with similar turnover rates share similar GO distribution. Over-representative analysis was able to detect high specificity for GO categories. While cytoplasmic proteins that represent a wide variety of biological functions showed a broad distribution, mitochondrial proteins and some ribosomal proteins showed similar values, probably due to their participation in stable protein complexes.

FIGURE 7. Venn diagram of protein comparison
Figure 8 demonstrates an example of ribosome pathway, with an implementation of quantitative coloring for KEGG pathway proteins, highlighted in accordance with their turnover rates.

FIGURE 8. Visualization of ribosome pathway

Protein half-life values obtained for the 15N- and SILAC-labeled mice studies of brain (6) showed rather significant correlation with Pearson coefficient values close to 0.8. It should be taken into account that different algorithms (peak area integration vs. RIA, correspondingly) were applied in these studies.

Conclusion

- We demonstrated the software tools (Gene Ontology, over-represented analysis, pathways analysis, and overall statistics) that can assist data analysis of protein turnover studies based on transcriptome and proteome data.
- Metabolic labeling and, in particular, pulse-labeling SILAC followed by mass spectrometric analysis provide the most accurate measurement of protein turnover rates of thousands of proteins and currently it is the method of choice for measuring protein degradation rates.
- Despite enormous complexity, variability, and diversity, as well as labor intensity, we believe that protein turnover analysis will become an essential tool for proteome characterization.

References

4. Ruhs, Aaron; Konzer, Anne; Cemić, Franz; Hemberger, Jürgen; Braun, Thomas; Krueger, Marcus. Pulse SILAC for absolute protein turnover rates in living animals. Personal communication.