STATISTICAL SIGNIFICANCE OF PROTEOME ANALYSIS

Stefan Muellner; Andreas Wattenberg; Katja Aschermann; Petra Lutter

Protagen AG, Otto-Hahn-Str. 15, D-44227 Dortmund, Germany

stefan.muellner@protagen.de

Overview

Standard Operation Procedures (SOP) for each and every step whithin the work flow of proteome analysis is key for the statistical relevance of the data obtained. Since the work load of all down stream validation experiments which are build on or strongly depend on the reliability of the proteome data is huge and requires high personal capacities, a professional project management for the study design has to be employed. We have compared the data sets of four large proteome studies of different biological samples, e.g. human primary cells, bacteria, human and animal tissue, with regard to the influence of the major steps - sample generation, sample preparation, protein quantification, study design, analysis, statistics - on the coefficient of variation of the whole process.

Methods

Sample Preparation

The cell lysis of the samples was performed by sonication in presence of protease inhibitors in phosphate buffer.

Popov Assay

The protein amount of the sample prior to 2D gel electrophoresis was determined using a method according to Popov.

2D SDS-PAGE

For isoelectric focusing the method of Klose using carrier ampholytes pH 2-11 was applied. The proteins were separated according to their apparent molecular weight in a continuous buffer system. This allows the high-resolution separation of the proteins.

Protein Stain

The separated proteins were stained with silver to achieve highest sensitivity according to Heukeshoven.

Conclusion

The correct study design and standardization of the analysis workflow is crucial for the significance of the results. If these parameters are not well defined, any kind of subsequent statistical analysis will be flawed.

At Protagen AG we have developed the expertise to optimise the study design according to the needs of our customers and have standardized our proteomics workflow. The standardization was acheived by introducing a QM system with detailed SOPs and thorough training of all employees.

We have shown the success of this approach in many proteome studies with very different samples.

Results

The Challenge: Variations in Samples

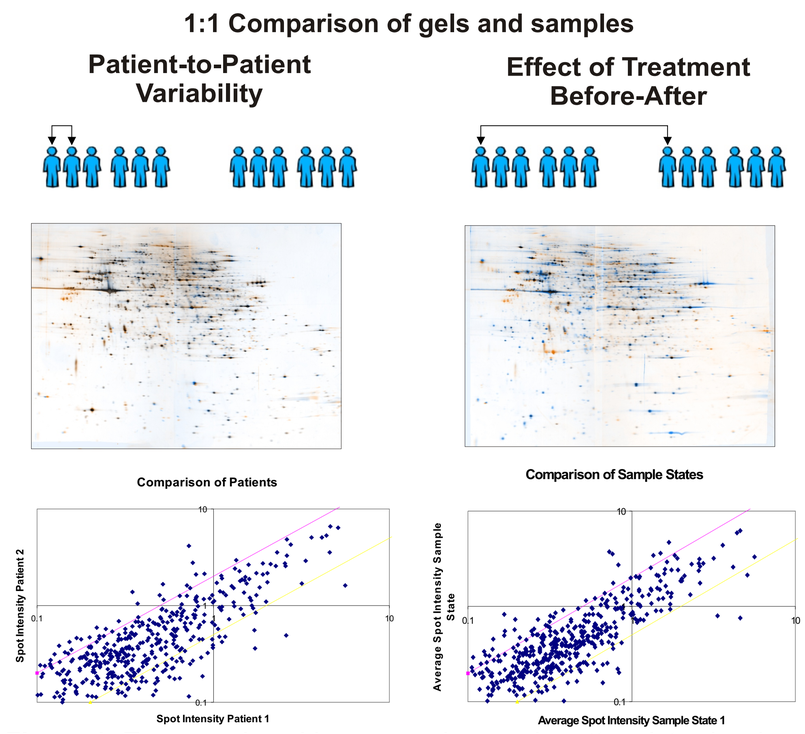

Figure 1: Two sample gel image overlays and scatter plots showing the large variation in spot intensities between two patients in one sample state (left, values averaged over 3 replicate gels) and between sample states (right, values averaged over 6 patients per sample state). Both variations are in the same order of magnitude.

One sample is not enough! Several biological replicates, a reproducible workflow and statistical analysis are needed for meaningful results. An example of such a proteome study is shown.

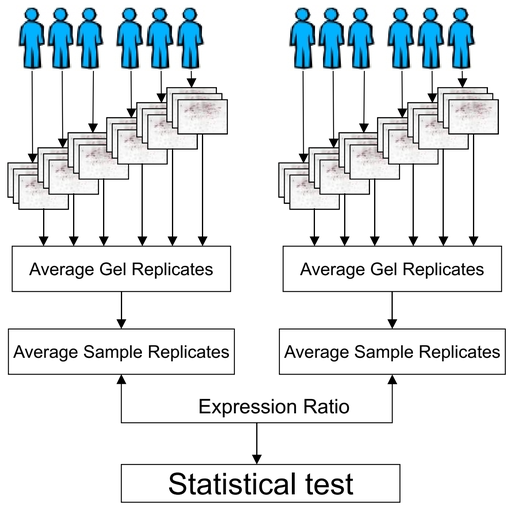

Figure 2: Experimental design of a proteome study. Several samples from different patients are used for each sample state. Replicate gels are averaged from each sample. The expression ratio for each spot is calculated from the averaged sample replicates. Subsequently, a statistical test is applied to calculate the significance.

Reproducibility of 2D PAGE Analysis

Reproducibility is key to successful proteome analysis. Here we show the reproducibility and linearity for one protein spot and the entire proteome.

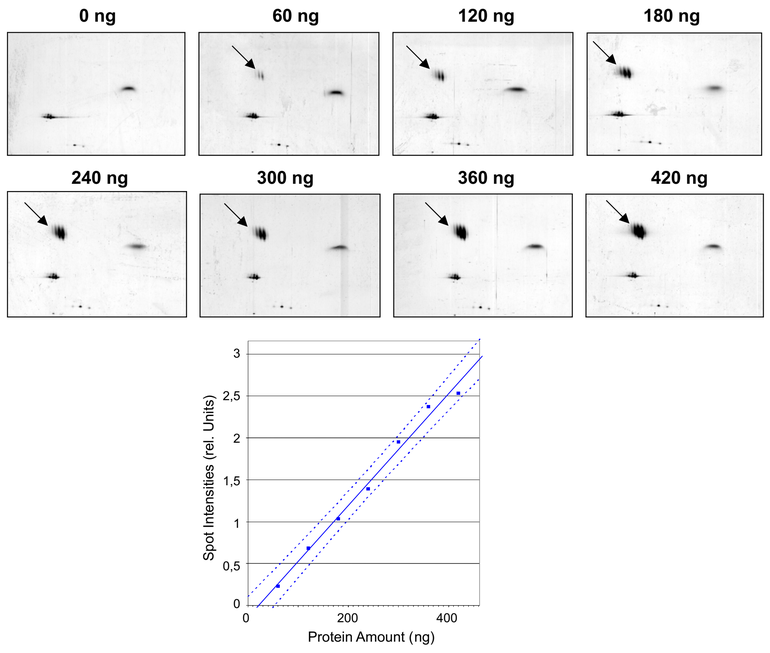

Figure 3: Systematic evaluation of the 2D gel quantification. Glucose Oxidase was spiked into a proteome and quantified. The reproducibility and the linear range could be demonstrated. All analyses were carried out using SOPs in a QM system.

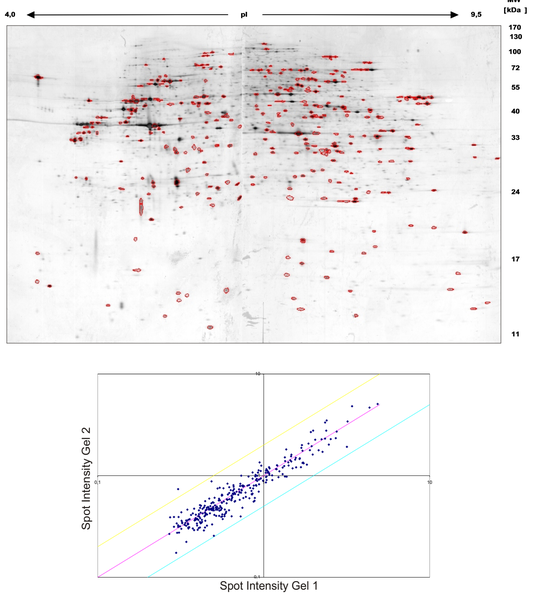

Figure 4: Reproducibility of spot quantification. The gel electrophoresis, staining and image analysis are carried out according to SOPs. The proteome-wide reproducibility can clearly be shown in the scatter plot The average %CV of the averaged spot intensities is 19%.

Statistical Analysis of Results

Having understood the variability of the samples, drafted a proper study design and ensured a reproducible analysis, the statistical analysis of results is mandatory in order to obtain biologically relevant results. With full control of these parameters is now possible to calculate the effect, e.g. of the sample size on the number of differentially detected spots. In the table below, the main characteristics of four different proteome studies are compared.

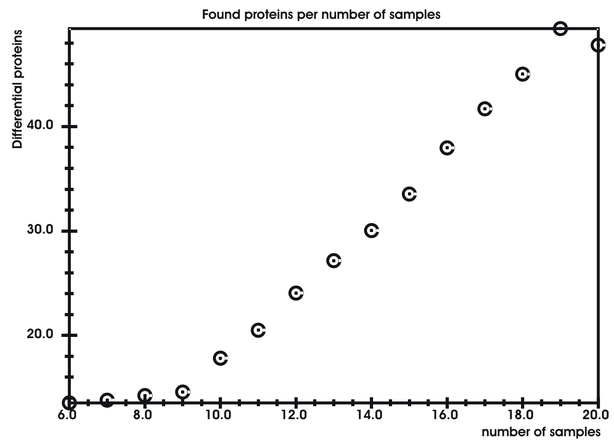

Figure 5: Influence of sample number on proteome analysis. A proteome study was conducted with 10 control and 10 treated samples. The graph was derived from experimental data by randomly reducing number of gels. The number of differential spots is determined by statistical analysis. The number of detected differences increases with number of biological replicates.

	T-Cell	Customer1	Customer2	Customer3
# states	**4**	**2**	**2**	**4**
Biological replicates	**4**	**10**	**3**	**6**
Technical replicates	**3**	**2**	**3**	**3**
Total # gels	**48**	**40**	**18**	**72**
Sample origin	**Cell culture T-cells**	**Tissue**	**Bacterium**	**Cell culture Tissue**
Statistical evaluation	**Yes**	**Yes**	**Yes**	**Yes**
# Differential spots	**157**	**71**	**23**	**454**
Significant ratio	**± 2.0**	**± 1.7**	**± 1.7**	**± 2.0**
%CV	**21**	**14**	**17**	**25**

Figure 6: Comparison of four different proteome studies. The samples analyzed in these studies are very different with regard to concentration, buffer, origin, amount. The %CV of the spot quantification ranges from 14-25%. This range directly reflects the heterogeneity of the samples used in the study.