BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 19 2010, pages 2480–2481
doi:10.1093/bioinformatics/btq443

Systems biology

Advance Access publication August 12, 2010

QuantProReloaded: quantitative analysis of microspot immunoassays

Anika Jöcker[1,*], Johanna Sonntag[1], Frauke Henjes[1], Frank Götschel[1], Achim Tresch[2], Tim Beißbarth[3], Stefan Wiemann[1] and Ulrike Korf[1]

[1]Division of Molecular Genome Analysis, German Cancer Research Center, Heidelberg, [2]Gene Center and Department of Biochemistry, Ludwig-Maximilians-University Munich, Munich and [3]Department of Medical Statistics, University of Göttingen, Göttingen, Germany

Associate Editor: Joaquin Dopazo

ABSTRACT

Summary: Protein microarrays are well-established as sensitive tools for proteomics. Particularly, the microspot immunoassay (MIA) platform enables a quantitative analysis of (phospho-) proteins in complex solutions (e.g. cell lysates or blood plasma) and with low consumption of samples and reagents. Despite numerous biological and clinical applications of MIAs there is currently no user-friendly open source data analysis software available with versatile options for data analysis and data visualization. Here, we introduce the open source software QuantProReloaded that is specifically designed for the analysis of data from MIA experiments.

Availability and implementation: QuantProReloaded is written in R and Java and is open for download under the BSB license at http://code.google.com/p/quantproreloaded/

Contact: a.joecker@dkfz.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 17, 2010; revised on July 22, 2010; accepted on August 2, 2010

1 INTRODUCTION

A robust experimental platform to assess the abundance and activity of many proteins in parallel, in biological and clinical samples is a prerequisite to obtain a better understanding of signaling networks. Thus, a careful and systematic analysis of samples from healthy and diseased patients promises to facilitate also the detection of new biomarkers for diagnosis and prognosis.

Microspot immunoassays (MIAs) (Korf *et al.*, 2008) can determine the abundance as well as the activity of different proteins and other analytes in parallel. The MIA technology is comparable to the enzyme-linked immunosorbent assay (ELISA) approach, but is superior in terms of its lower sample and reagent consumption and a vastly improved dynamic range (Korf *et al.*, 2008). Capture antibodies that are specific for a particular analyte or post-translational modification (e.g. phosphorylation) are immobilized on a solid surface subdivided into several sub-arrays (blocks). Each block is incubated with a different biological sample so that the immobilized antibodies capture their specific analytes. In the detection step, suitable antibodies directed against a different epitope of the respective target are utilized contributing to the high specificity of this approach. Finally, antibodies labeled with near-infrared fluorescent dyes are used to visualize the abundance of a particular detection antibody in each of the spots (Korf *et al.*, 2008). In addition to the samples themselves, a predefined number of blocks is incubated with a serial dilution of standard analytes of known concentration. These so-called 'calibrators' are used to fit a calibrator curve to estimate the target concentration in the measurement samples.

Despite the great potential of the MIA technology, the analysis and visualization of data is still a bottleneck because no comprehensive software tool is available allowing statistical data analysis as well as data visualization. ProMat (White *et al.*, 2006), a tool designed for the analysis of ELISA microarrays, can in principal also be employed for certain MIA applications; however, ProMat cannot handle time-course data. We have previously published basic algorithms for the analysis of MIA data (Korf *et al.*, 2008) and released a first version of QuantPro. However, this version was not implemented with a user-friendly interface and the functions integrated in QuantPro were not robust against outliers.

The new open source tool, QuantProReloaded, was written in R and Java and tailored towards the statistical analysis of time-resolved and non-time-resolved MIA data. QuantProReloaded is based on the ideas of QuantPro, but was completely redesigned to streamline the analysis workflow and to make it applicable to a wide range of protein array designs. A new method was implemented into QuantProReloaded to calculate how accurately the analyte concentration of the calibrators can be determined using a re-estimation approach. This helps to identify the valid concentration range for a certain antibody pair. Besides that, this method allows also to identify a suitable function for appropriate fitting of the calibrator curve and to detect experimental outliers.

2 METHODS

QuantProReloaded is an open source software consisting of functions written in R and embedded into a intuitive and platform-independent user interface written in Java. rJava/JRI (http://www.rforge.net/rJava/) is used in the middle layer to enable the interaction between R and Java.

As data input a tab-delimited 'slide-description file', listing analytes, detection antibodies, as well as summarizing block-specific information (calibrator or measurement) are required. In case time-course experiments were performed this file should also list the individual time points as part of the sample description. QuantProReloaded can read any kind of tab-delimited image analysis output file. If the format is not known, the user can edit

*To whom correspondence should be addressed.

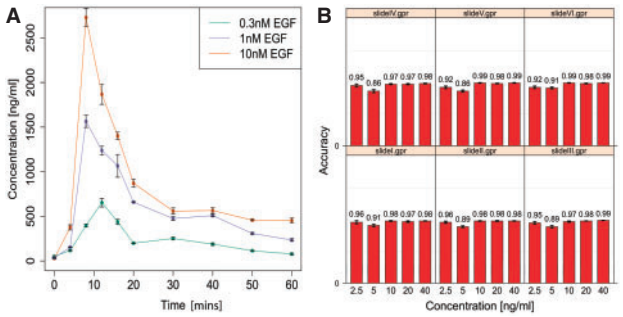

Fig. 1. (**A**) Time-series plot and (**B**) Accuracy plot.

and rename the 'default.properties' template file of the QuantProReloaded software package to describe the new format. Two applications are provided in QuantProReloaded: the *accuracy analysis* and the *measurement analysis*.

The *accuracy analysis* determines the accuracy of the calibration by cross-validation of the calibrator data points. Thus, 60% of the calibrator data points of each concentration are used to fit a calibrator curve and the remaining data points are used to re-estimate the protein concentration. The overall accuracy is one minus the normalized distance between the real protein concentration and the estimated protein concentration. Thus, a result of '1' indicates that there is no difference between the real protein concentration and the estimated protein concentration. Several robust linear and non-linear functions are included as option to fit a calibrator curve and other functions can easily be integrated. The resulting *accuracy plot* visualizes how accurately the analyte concentration can be re-estimated using a certain algorithm to fit the calibrator curve. In addition, the calibrator curve displaying the single calibrator data points is shown.

The *measurement analysis* is designed to estimate the target protein concentration of the samples and can be used independently of the *accuracy analysis*. A calibrator curve is calculated using all calibrator data points and serves as basis for the calculation of the analyte concentration in each sample. Spots can be excluded from the analysis by marking them in the image analysis software (e.g. Spotfinder and GenePix) and setting the 'Remove marked spots' option to 'Yes'. The *measurement analysis* produces two tables as output, the first one contains the estimated analyte concentrations for all samples and the second one lists the median as well as the median absolute deviation for all analyte concentrations of all technical and biological replicates. Furthermore, a calibrator plot displaying the calibrator curve, the calibrator data points, a confidence interval as well as the data points of measurement samples is generated as well. For further information about the displayed confidence intervals can be found in Supplementary Material. In case of dynamic measurements, a time-course plot is returned to facilitate a fast comparison between the median concentration levels of a particular protein at different time points.

Data from the time-course example in Figure 1 result from dose-dependent measurements of receptor tyrosine kinase signaling. In particular, the human breast cancer cell line (MCF7) was exposed to different concentrations of epidermal growth factor (EGF) and the formation of phosphorylated ERK was monitored. Six slides with 16 blocks were used to analyze the measurements and scanned with an Odyssey Infrared Imaging System at a resolution of 21 μm at 700 nm. Spot signal intensities were quantified using the GenePix Pro 5 software and the results were saved as .gpr files (see Supplementary Material for further details on the protocol).

3 RESULTS

3.1 Accuracy analysis

After completing the slide-description file (see Supplementary Material), QuantProReloaded was started and the *accuracy analysis* was chosen. In principal, this analysis can be done only for a certain antibody pair at a time, but additional result windows can be opened to compare different antibody pairs. In the example, data are shown for the phospho-ERK-specific antibody pair named ERK_CST9106. We choose a 'linear fit' to calculate the calibration curves. As we noticed slide effects in this particular experiment, the software option to analyze arrays separately was selected. To obtain accurate results, the cross-validation repetitions were set to a high number (in this case: 1000). Plots and tables generated by QuantProReloaded are stored in a selected output directory and are displayed in a new window. Calibrator plots for each of the slides show an estimated calibrator. This particular option for data visualization enables the user to identify unreliable spots or slides that have to be removed from the final analysis. The *accuracy plot* is available in a second tab and shows how accurate the analyte concentration could be re-estimated. As can be seen in Figure 1B, in this instance almost all values are greater than 0.9 and, therefore, a linear correlation between spot intensity and protein concentration was assumed.

3.2 Measurement analysis

To estimate the dynamics as well as dose-dependent effects of phospho-ERK formation, a new '*measurement analysis*' was started. The same antibody combination was selected as used for the *accuracy analysis* as well as for the linear fit. The output window consists of two tabs (*calibrator plot* and *time-series plot*). The calibrator plot was skipped because of the linear correlation identified in 3.1 and the 'time-series' tab was selected (Fig. 1A). To sum up, Figure 1 confirms a positive correlation between EGF concentration and phospho-ERK formation for the entire time period. Furthermore, the phospho-ERK concentration reaches its maximum after 8 min for the two highest EGF concentrations (1, 10 nM) compared to a peak at 12 min for the lowest, EGF concentration (0.3 nM).

Funding: BMBF project IG Cellular Systems Genomics (grant number 01GS0864) in the NGFN-Plus program; BreastSys (grant number 0315396B) of the Medical Systems Biology program; MOGLI (grant number 0315394B).

Conflict of Interest: none declared.

REFERENCES

Korf,U. *et al.* (2008) Quantitative protein microarrays for time-resolved measurements of protein phosphorylation. *Proteomics*, **8**, 4603–4612.

White,A.M. *et al.* (2006) ProMAT: protein microarray analysis tool. *Bioinformatics*, **22**, 1278–1279.