

## RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data

Silvia von der Heyde<sup>1</sup>, Johanna Sonntag<sup>2</sup>, Daniel Kaschek<sup>3</sup>, Christian Bender<sup>4</sup>, Johannes Bues<sup>2</sup>, Astrid Wachter<sup>1</sup>, Jens Timmer<sup>3,5,6</sup>, Ulrike Korf<sup>2</sup>, and Tim Beißbarth<sup>1</sup>

<sup>1</sup>Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany, <sup>2</sup>Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany, <sup>3</sup>Institute of Physics, Freiburg University, Freiburg, Germany, <sup>4</sup>TRON–Translational Oncology at the Johannes Gutenberg University of Mainz Medicine, Mainz, Germany, <sup>5</sup>BIOSS Centre for Biological Signalling Studies, Freiburg University, Freiburg, Germany, and <sup>6</sup>Freiburg Center for Systems Biology (ZBSA), Freiburg University, Freiburg, Germany

*BioTechniques* 57:125–135 (September 2014) doi 10.2144/000114205

Keywords: data analysis toolbox; open-source software; proteomics; RPPA; RPPanalyzer

Supplementary material for this article is available at [www.BioTechniques.com/article/114205](http://www.BioTechniques.com/article/114205).

Analysis of large-scale proteomic data sets requires specialized software tools, tailored toward the requirements of individual approaches. Here we introduce an extension of an open-source software solution for analyzing reverse phase protein array (RPPA) data. The R package *RPPanalyzer* was designed for data preprocessing followed by basic statistical analyses and proteomic data visualization. In this update, we merged relevant data preprocessing steps into a single user-friendly function and included a new method for background noise correction as well as new methods for noise estimation and averaging of replicates to transform data in such a way that they can be used as input for a new time course plotting function. We demonstrate the robustness of our enhanced *RPPanalyzer* platform by analyzing longitudinal RPPA data of MET receptor signaling upon stimulation with different hepatocyte growth factor concentrations.

In principle, RPPAs are miniaturized dot blot immunoassays, enabling quantitative analysis of target protein abundance, including posttranslational modifications, across large sample sets. Originally introduced by Paweletz et al. as a robust and reproducible proteomics resource (1), over the past decade the RPPA approach has become a valuable method for biomarker discovery and systems biology research (2,3). The RPPA user base is also increasing, as demonstrated by recent publications in a number of fields (1,3–11). Thus, RPPA has realized its ground-breaking potential for clinical applications as well as for the elucidation of signaling

networks. Hence, RPPA technology would benefit from a robust, freely available software tool allowing data comparison across different RPPA platforms.

To date, the available analysis tools for RPPA have been mainly limited to non standardized in-house solutions or commercial platforms such as Microvigena (Vigene Tech, Carlisle, MA). A major drawback of open-source RPPA tools such *NormaCurve* (12), *Reno* (13), and others (14) is their limitation in data preprocessing because they are restricted to data normalization steps without offering additional functions to plot the resulting data or to carry out a statistical analysis. In contrast, the

R (<http://www.R-project.org>) package *RPPanalyzer* (15) offers data preprocessing in combination with graphical and statistical analysis options. Utilization of *RPPanalyzer* is on the rise and has been documented already (4,5,16–20).

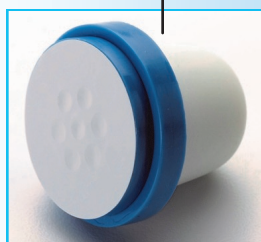
Here we describe an expanded and improved toolbox for *RPPanalyzer* providing additional useful features including a standardized workflow of upgraded data preprocessing combined with further improved graphical and statistical analyses. A novel background correction method has now been incorporated into the program. The main advantage of our background correction method is that

### METHOD SUMMARY

We describe an improved version of *RPPanalyzer* that introduces a standardized workflow for RPPA data preprocessing, a novel functionality for background noise correction, and optional graphical and statistical analysis methods for robust analysis of reverse phase protein array (RPPA) data sets.

# TAKE ACTION AGAINST CONTAMINATION WITH QMI® SAFE-SEPTUM

Sample, Inoculate Or Add Nutrients  
To Your Bioreactor Aseptically !



The QMI Safe-Septum is:

- Aseptic
- Pressure & Temperature Safe
- Pre-Sterilized
- Easy To Retrofit
- Validated



651-501-2337

Email: [info@qmisystems.com](mailto:info@qmisystems.com)

[www.qmisystems.com](http://www.qmisystems.com)

it does not require the printing of all samples as serial dilutions; instead, it estimates correction factors based on a single representative dilution series. A new time course plotting function is also incorporated that embeds enhanced variance estimations, which is useful for triplicate measurements as usually obtained from cell-based perturbation experiments.

## Materials and methods

The application of the *RPPanalyzer* (workflow depicted in Figure 1) is demonstrated using a typical systems biology RPPA data set (Figure 2A and 2B). The experimental setup was as follows: A549 cells, a human alveolar adenocarcinoma cell line, were serum-starved for 24 hours and subsequently stimulated with 6 different hepatocyte growth factor (HGF) concentrations ranging from 0 to 100 ng/mL. Samples were obtained at 6 different time points between 0 and 120 min. All experiments were done as biological triplicates, and the samples were analyzed by RPPA using antibodies directed against proteins and phosphoproteins involved in MET receptor signaling. This example RPPA data set and the corresponding R script are provided as Supplementary Material. The R package is available at the Comprehensive R Archive Network (CRAN, <http://cran.r-project.org/web/packages/RPPanalyzer>) and at R-Forge (<https://r-forge.r-project.org/projects/rppanalyzer>).

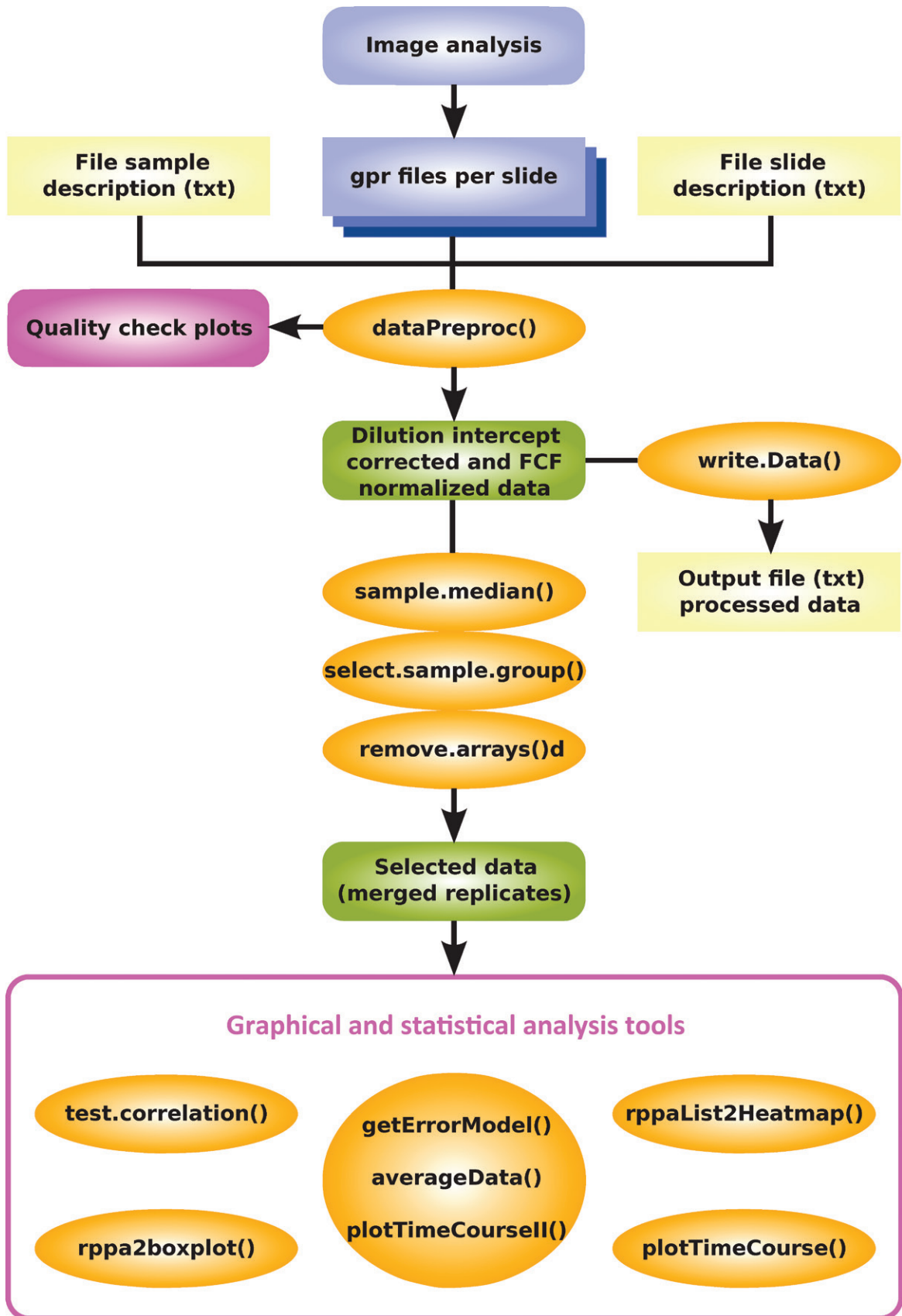
## Results and discussion

The general *RPPanalyzer* workflow is depicted in Figure 1. For this updated version, the initial data preprocessing steps were bundled in the `dataPreproc` wrapper function, leading to standardized data output. Function tasks for `dataPreproc` include:

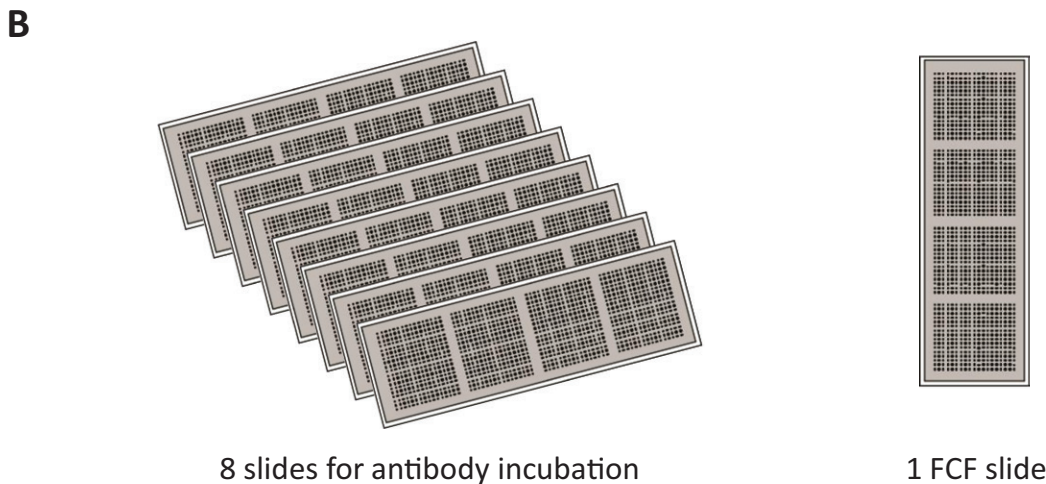
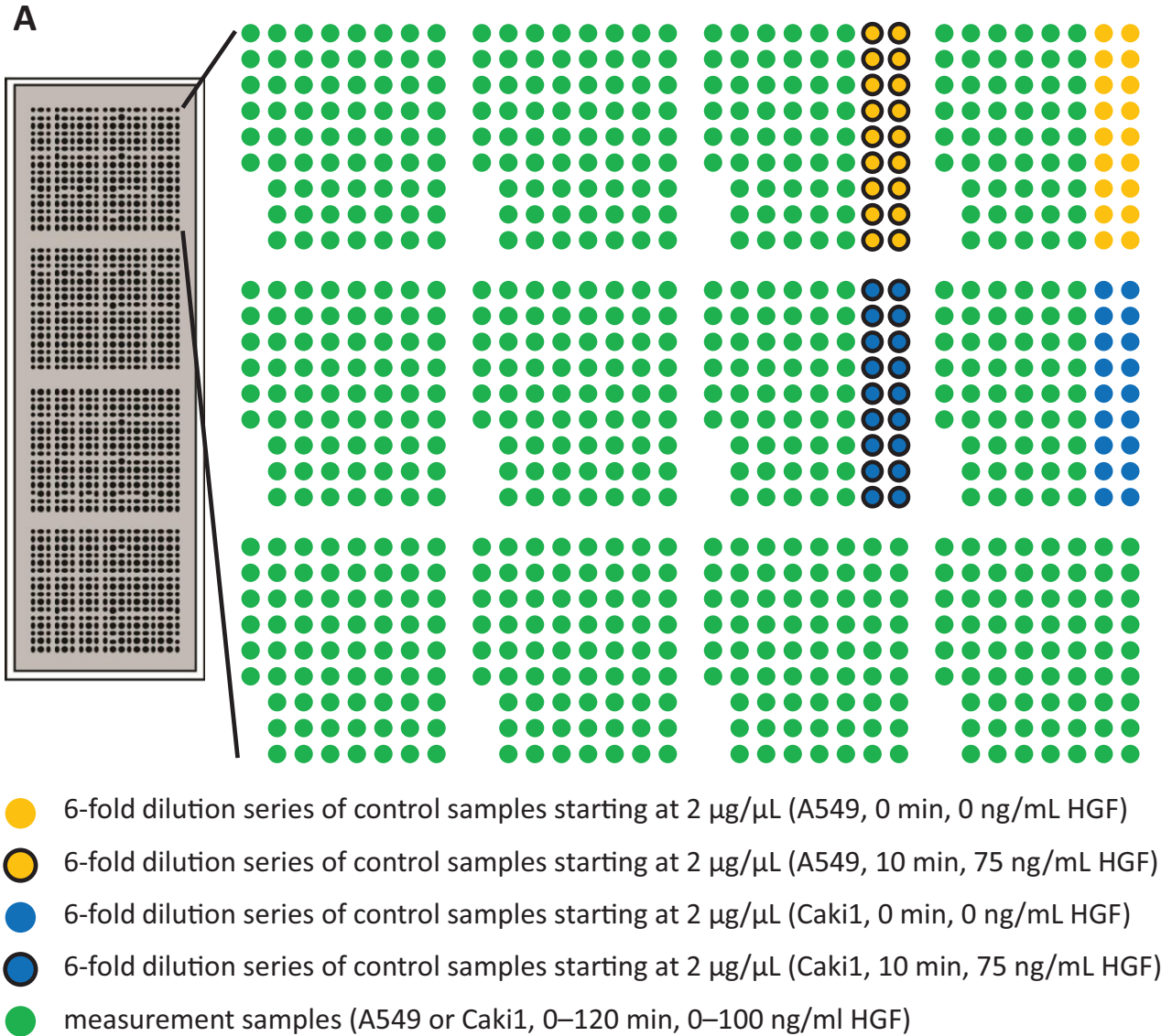
- *Import of raw data from slide, sample description text files, and corresponding gpr files.* In the text files, users provide sample-specific information and list the target proteins analyzed by RPPA. Instructions on the setup of slide and sample description files are provided in the *RPPana-*

*lyzer* vignette. The information in these files impacts all ensuing data processing steps. Hence, in case of any detected mistakes, files should be manually corrected and the workflow rerun. Spot-specific signal intensities obtained by image analysis using the commercial software GenePixPro are stored as gpr files that also contain information on array layout. *RPPanalyzer* is tailored to this format. In cases in which a user does not work with gpr files, a workaround is to save result files from other scanning systems as text files and then adjust the format to the syntax of gpr files. After adding the suffix `.gpr` to the text file generated this way, the data analysis can proceed as when processing gpr files.

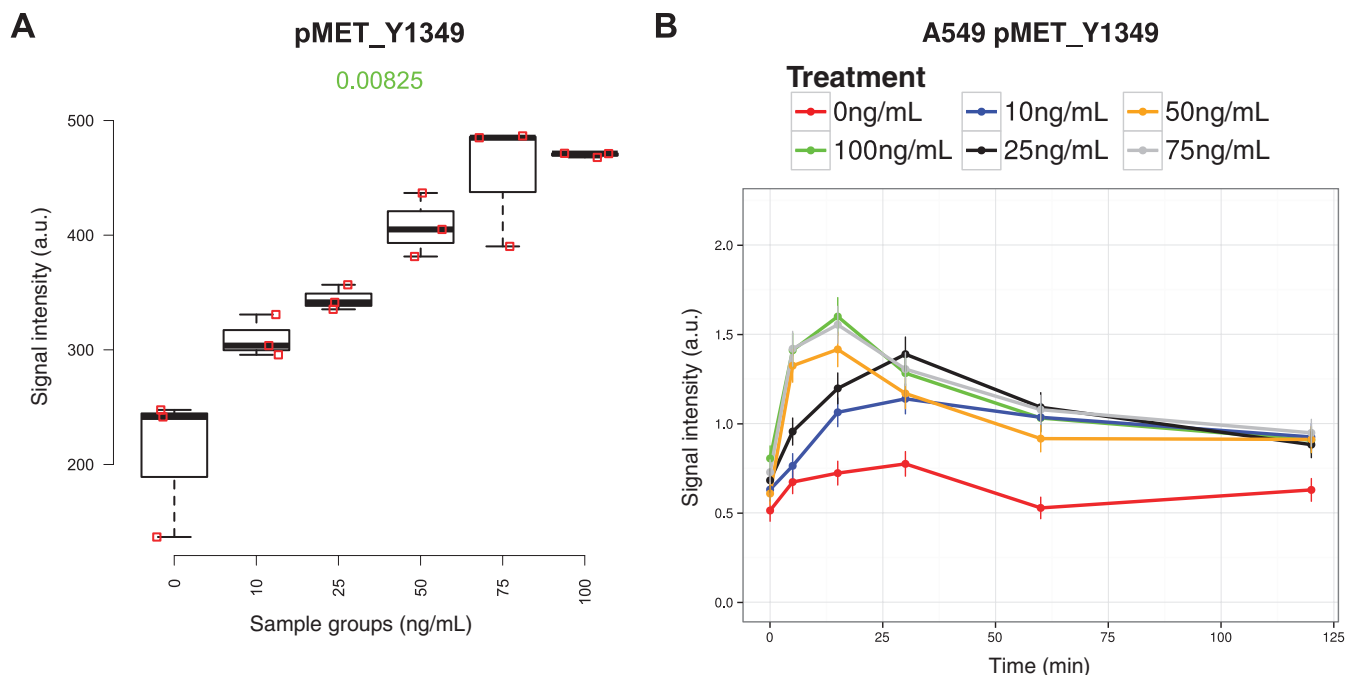
- *Background noise correction using the new `correctDilInterc` function, as described next.*
- *Generation of plots to assess data quality.* These include: (i) *QC\_dilutioncurve\_raw.pdf* (raw signal intensities of serially diluted control samples to check linearity of target protein detection), (ii) *QC\_target\_VSblank\_normed.pdf* (scatter plot of background noise corrected and normalized target signals versus “secondary antibody only” signal to assess bias caused by the secondary antibody), and (iii) *QC\_qqPlot\_normed.pdf* (quantile-quantile plot to control whether data are normally distributed).
- *Spot-specific normalization of signal intensities based on the total protein concentration determined by the FCF method (21).* Replicate slides, usually 1 slide out of 8 to 10, are stained with the dye Fast Green FCF to determine the total protein concentration of each individual lysate spot (Figure 2B). The signal intensity of FCF readout ( $f_i$ ) is determined for each individual spot. The target protein specific signal intensities ( $x_i$ ) obtained from slides probed with different detection antibodies are corrected for technical variance by dividing antibody signals by the factors  $f_i$ . Afterward, the corrected spot intensities are multiplied by the median  $f_m$  of the corresponding



**Figure 1. RPPanalyzer Toolbox workflow.** The recommended workflow for reverse phase protein array (RPPA) data analysis starts with the import of the image analysis output files as well as the slide and sample description files. The data preprocessing returns dilution intercept corrected data, which were normalized by total protein concentration using the Fast Green FCF approach (21). Data preprocessed this way can then be analyzed using diverse graphical and statistical methods implemented in the RPPanalyzer toolbox.



**Figure 2. Exemplary reverse phase protein array (RPPA) layout.** (A) Samples (control and measurement) are spotted as four identical subarrays on a series of nitrocellulose coated glass slides. All samples are spotted as technical triplicates. Only control samples are spotted as 6-fold dilution series. (B) Slide numbers 1 to 8 are used for incubation with target-specific antibodies; 4 subarrays allow for using 4 different antibodies per slide. Slide number 9 is used for total protein staining using the Fast Green FCF method (21).



**Figure 3. Examples of graphical output.** (A) Boxplot of MET receptor phosphorylation (Y1349) signal intensity at the 15 min time point, revealing dependence on hepatocyte growth factor (HGF) concentration. The green number represents the *P* value result from a Kruskal-Wallis rank sum test, indicating significant difference between expression values of the six sample groups. (B) Time course plot of MET receptor phosphorylation (Y1349) over 120 min after HGF stimulation at different concentration levels.

normalizer subarray to scale data back to the original signal range. This can be summarized as normalized spot intensity  $x_{FCF} = x_i / f_i * f_m$ .

- *Returning a list of four items to users.* The first item lists raw data, that is, matrices of foreground and background signal intensities as well as information describing sample localization and providing keywords to identify individual samples as part of the downstream analysis. The second item is analogously built up but lists corrected foreground signal intensities generated using

the `correctDilinterc` function. If negative values are obtained after background subtraction applying the `correctDilinterc` function, the absolute minimum plus one is added, as for most applications (e.g., comparison of multiple conditions), a small positive value is required even for probes which are within the background levels. The third item is also structured as the first two but holds dilution intercept corrected and FCF normalized foreground data. The final item defines the directory for output storage.

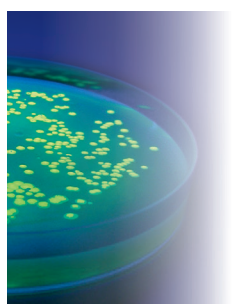
- *Raw data export to a text file in table format.* Any other processed data (e.g., normalized data) can be exported manually using the `write.Data` function.

All output files are stored in a folder labeled with the date of analysis at the input files location.

After preprocessing, the next step is to merge technical replicates (`sample.median` function), select data subsets of interest (`select.sample.group` function), and then remove arrays or target proteins that are not required during downstream analysis, or data that did not pass the quality check (`remove.arrays` function).

Finally, different R functions can be applied for statistical analysis and graphical representation of the preprocessed RPPA data. According to the toolbox character, this output layer is not limited to the five implemented methods but can be easily expanded and tailored to a user's specific needs by new or modified R functions. The methods currently available in the *RPPAnalyzer* package are:

- `test.correlation`: A correlation plot to compare protein expression



## Plasmid & Minicircle DNA Service

- Customized plasmids and *minicircles*
- *InStock* reporter genes
- AAV helper & packaging plasmids

[PlasmidFactory.com](http://PlasmidFactory.com)



PlasmidFactory GmbH & Co. KG | Meisenstraße 96 | D-33607 Bielefeld | Germany | Fon ++49 521 2997350

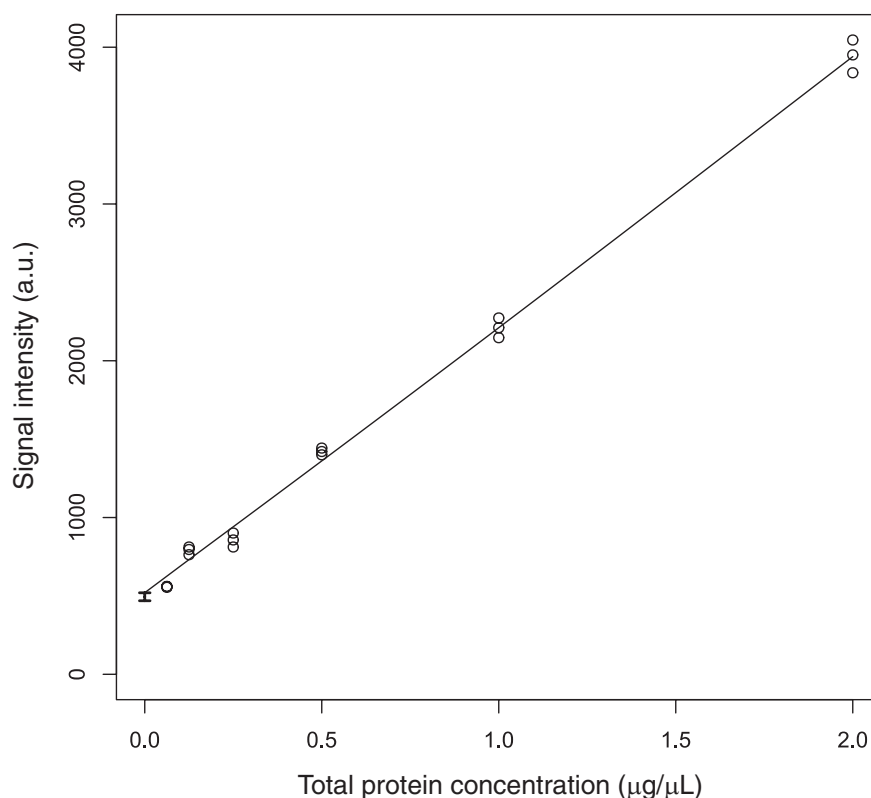
and a numeric sample attribute, including statistical testing according to Pearson, Spearman or Kendall.

- `rppa2boxplot`: Allows for visualization of the data as boxplots for defined sample groups. In addition, statistical rank sum tests are applied. In case of a reference group to be tested against, two-sample Wilcoxon tests are performed, and the *P* values are adjusted for multiple testing according to Benjamini and Hochberg. To test for a general difference among all samples, a Kruskal-Wallis rank sum test is applied. The *P* values are displayed in the boxplot graph, as shown in the example data in Figure 3A.
- `rppaList2Heatmap`: Allows for the visualization of hierarchical clustering as a heatmap, adding specific column side colors to mark groups of selected phenodata. Before applying this function, the data should be logarithmized using the `logList` function. Scaling and viewing the data on a logarithmic scale have the effect of variance stabilization and are necessary for statistical tests and for the computation of distance measures that have the requirement that the data should look normally distributed.
- `plotTimeCourse`: Allows for the visualization of time course data. Different plotting options can be specified, such as smoothed spline fits through the data.

In addition to the `dataPreproc` function, four other functions have been added to this latest version of *RPPAnalyzer*.

- `correctDilinterc`: *Correction of background noise*. This function corrects signal intensities obtained for a protein of interest by subtracting an intercept estimated for a total protein concentration of 0  $\mu\text{g}/\mu\text{L}$ . This function does require that the array contains a dilution series that passed the quality check mentioned before. The function is integrated in the data preprocessing function `dataPreproc`. The y-intercepts of the dilution

## A549 PRAS40 2691



**Figure 4. Dilution series plot visualizing the signal intercept estimation.** In this example, a y-intercept of 453 a.u. was estimated as background noise for an antibody directed against PRAS40. Sample signal intensities will be background corrected by subtraction of this value.

series for corresponding samples are derived using `dilSeriesID`, a parameter defined in the sample description file, subarray-specific information, detection antibody and spotting run as summarized in the slide description file. Actual samples are defined as *measurement* in the *sampledescription.txt*, and serially diluted samples are specified as *control*.

A major advantage of this method is that it is independent of dilution series prepared from all samples as described by the original RPPA approach (1). Correction factors are obtained from representative samples chosen ahead of array printing during design. In our example data set, actual samples represent 6 different time points ranging from 0 to 120 min, whereas the corresponding dilution represents a 10 min time point only, as highlighted in Figure 2A. To link a sample

of interest with a certain control dilution series from a panel that mostly includes several dilution series (e.g., representing different cell lines or conditions), the corresponding identifier has to be entered in the column `dilSeriesID` of the *sampledescription.txt*. The selection of the sample to be used as dilution series strongly depends on the particular experimental setup.

For near-infrared (NIR) detection, signal and protein concentration are detectable over a linear range for the majority of detection antibodies. Hence, intercept subtraction will produce a parallel shift. For non-NIR fluorescence-based detection approaches, signals frequently saturate at higher protein concentrations. In this case, smoothing splines as nonparametric fits allow us to deal with several different RPPA detection methods and are

applied for extrapolation to zero concentration. The uncertainty of the extrapolated intercept is estimated by nonparametric bootstrapping.

The estimated intercepts  $I$  are conditional on the factors antibody, slide, and sample, of which the latter is defined by *dilSeriesID*. Linear models are established for the following hypotheses:

$$I \sim \text{const.} \quad [\text{Eq. 1.1}]$$

$$I \sim \text{const.} + \text{antibody} \quad [\text{Eq. 1.2}]$$

$$I \sim \text{const.} + \text{antibody} + \text{slide} \quad [\text{Eq. 1.3}]$$

$$I \sim \text{const.} + \text{antibody} + \text{slide} + \text{sample} \quad [\text{Eq. 1.4}]$$

The estimated uncertainties of the intercepts are used as weights. The provided bar plot of the residual sum of squares (RSS) (*anovaIntercepts\_Output.pdf*) should be used to choose the model with the smallest RSS favoring less complexity. For example, if the bars of model 1.3 and 1.4 are the smallest and are equally high, model 1.3 should be preferred because the sample in 1.4 does not provide additional information. Based on the chosen model, the *measurement* intercepts are predicted and are then subtracted from the raw intensities. The function additionally generates plots of the dilution series and related intercept estimations (*getIntercepts\_Output.pdf*), as shown in Figure 4.

- `getErrorModel`: *Estimation of signal variance*. Signal variance can be estimated from technical replicates. The variance estimator for triplicates is generally very poor because it is  $\chi^2$  distributed with  $n-1$  degrees of freedom. In the case of  $n = 3$ , this is an exponential distribution. Therefore, information from several different triplicates is combined to estimate parameters of an error model.

A variance versus signal plot characteristic for RPPA data reveals the following:

- The signal dependency of the variance depends on the factors subarray, detection antibody, and median FCF normalizer value. Slides probed with different detection antibodies are scanned with different scanner settings to yield an optimal image for data analysis. Signals are scaled by the median FCF value.
- Signal variances can increase with the signal strength, are constant, or have (almost) zero variance at zero signal.

The error model used to account for these observations is  $\sigma^2 = \sigma_0^2 + S^2\sigma_{rel}^2$ , where  $S$  is the signal strength,  $\sigma_0^2$  denotes the variance at zero signal, and  $\sigma_{rel}^2$  reflects the relative error. To estimate  $\sigma_0^2$  and  $\sigma_{rel}^2$  from the variance versus signal plot, a maximum likelihood approach, exploiting the  $\chi^2$  distribution of the variances, is applied. This is due to the strong nonnormality of the triplicate variance estimator.

## Safe & Secure Weighing

### Secura®

Reduce Your risk with integrated and intelligent protection systems.



# Unlimited BioImaging

Colorimetric  
Bioluminescence  
Fluorescence  
Chemiluminescence  
Multiplex



The **BioSpectrum Imaging System** with its advanced capabilities enables a wide range of imaging applications.

**Unlimited** in its ability to grow with your lab's developing life science research needs *today and into the future.*



An Analytik Jena Company

UVP, LLC: USA (909) 946-3197  
UK +44 (0)1223 420022

Learn more at  
[uvp.com/biospectrum.html](http://uvp.com/biospectrum.html)

- `averageData`: *Averaging biological replicates.*

Frequently, experiments provide two kinds of replicates: technical and biological. In this context, *biological replicate* means that spotted lysates result from equal treatment, and *technical replicate* means that the identical lysate is spotted several times on the same slide.

Averaging is done for each detection antibody separately. It is based on the assumption that a true dynamic behavior  $y_s(t)$  for a given stimulation  $s$  exists. The observed dynamics  $S_{js}(t)$  for each biological replicate  $j$ , that is, factor (biological replicate + slide), and stimulation  $s$  differ systematically from  $y_s(t)$  by a scaling factor  $s_j$ . In addition,  $s_j * S_{js}(t)$  differs statistically from  $y_s(t)$  because of biological variability. To estimate the true dynamics and the scaling factors, the objective function (see Equation 2 below) for discrete measurement time points  $t_i$  is minimized. The response is weighted by the variance from the error model evaluated at the “true” value  $y_{is}/s_j$ . Strictly speaking, this is an approximation because each biological replicate has its own true dynamic behavior. A direct estimate of these signaling dynamics would be the mean of the technical replicates for each biological replicate, that is,  $S_{ijs}$  itself. Depending on the value of  $\sigma_{i,j,s}^2$ , erroneous estimates can have a huge impact because smaller values are automatically favored. In contrast, the approximation by  $y_{is}/s_j$  is more robust.

- `plotTimeCourseII`: *Visualization of time course data.*

This function is applied after transforming the preprocessed data by the `getErrorModel` and `averageData` functions, and an example of this function is shown in Figure 3B.

In conclusion, we have simplified, enhanced and standardized RPPA data analysis by extending the functionality of the existing *RPPAnalyzer* package through the introduction of new approaches for variance estimation, background noise correction, and time course data visualization. In addition to this new functionality, the package was also streamlined by easing and improving several already existing functions. Although we recommend the use of the standard preprocessing steps, the newly introduced toolbox approach described here does allow integration of add-on functions that can be tailored to the specific needs of users.

## Author Contributions

J.B. developed the new time course plotting function under the supervision of U.K.; D.K. developed the new functions for background correction as well as variance estimation under the supervision of J.T.; J.S. provided the RPPA example data and was involved in the conception of the software update under the supervision of U.K.; S.v.d.H. implemented the corresponding upgrades into the R package under the supervision of T.B.; C.B. is the current maintainer of the R package at CRAN; A.W. was involved in discussions and is currently one of the package maintainers at R-Forge. All authors contributed to conception, writing, and editing of the manuscript.

## Acknowledgments

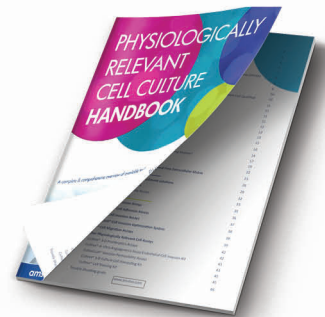
We thank Heiko Mannsperger for his initiative to develop the first version of the R package *RPPAnalyzer*. In addition, we thank all users of the *RPPAnalyzer* for valuable input to continuously improve its functions. We are grateful to Corinna Becki and Sabrina Schumacher for excellent technical assis-

$$x^2(\{s_j\}, \{y_{is}\}) = \sum_{i,j,s} \frac{(s_j S_{ijs} - y_{is})^2}{s_j^2 \left[ \sigma_{ijs,0}^2 + \left( \frac{y_{is}}{s_j} \right)^2 \sigma_{ijs,r}^2 \right]} = \sum_{i,j,s} \frac{\left( S_{ijs} - \frac{y_{is}}{s_j} \right)^2}{\sigma_{ijs,0}^2 + \left( \frac{y_{is}}{s_j} \right)^2 \sigma_{ijs,r}^2} \quad [\text{Eq. 2}]$$



# PHYSIOLOGICALLY RELEVANT CELL CULTURE HANDBOOK

- Popular 3D techniques
- Solutions to common problems
- How to obtain optimal 3D cell culture matrix for different applications



FREE ONLINE  
DOWNLOAD

[www.amsbio.com/PRCC.aspx](http://www.amsbio.com/PRCC.aspx)

tance. This work was supported by the Medical Systems Biology program (grant 0315396 of the BreastSys project, and grant 0316173 of the eBio Metasys project) of the German Federal Ministry of Education and Research.

## Competing interests

The authors declare no competing interests.

## References

1. Paweletz, C.P., L. Charboneau, V.E. Bichsel, N.L. Simone, T. Chen, J.W. Gillespie, M.R. Emmert-Buck, M.J. Roth, et al. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20:1981-1989.
2. Kornblau, S.M., N. Singh, Y. Qiu, W. Chen, N. Zhang, and K.R. Coombes. 2010. Highly phosphorylated FOXO3A is an adverse prognostic factor in acute myeloid leukemia. *Clin Cancer Res.* 16:1865-1874.
3. Uhlmann, S., H. Mannsperger, J.D. Zhang, E.-Á. Horvat, C. Schmidt, M. Küblbeck, F. Henjes, A. Ward, et al. 2012. Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Mol. Syst. Biol.* 8:570.
4. Sonntag, J., C. Bender, Z. Soons, S. von der Heyde, R. König, S. Wiemann, H.-P. Sinn, A. Schneeweiss, et al. 2014. Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer. *Transl. Proteomics* 2:52-59.
5. Ummanni, R., H.A. Mannsperger, J. Sonntag, M. Oswald, A.K. Sharma, R. König, and U. Korf. 2013. Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery. *Biochim Biophys Acta.* 1844:950-959.
6. Li, J., Y. Lu, R. Akbani, Z. Ju, P.L. Roebuck, W. Liu, J.-Y. Yang, B.M. Broom, et al. 2013. TCPA: a resource for cancer functional proteomics data. *Nat. Methods* 10:1046-1047.
7. Chavez-Macgregor, M., S. Liu, D. De Melo-Gagliato, H. Chen, K.-A. Do, L. Pusztai, W. Fraser Symmans, L. Nair, et al. 2014. Differences in gene and protein expression and the effects of race/ethnicity on breast cancer subtypes. *Cancer Epidemiol. Biomarkers Prev.* 23:316-323.
8. Hennessy, B.T., Y. Lu, A.M. Gonzalez-Angulo, M.S. Carey, S. Myhre, Z. Ju, M.A. Davies, W. Liu, et al. 2010. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clin. Proteomics* 6:129-151.
9. Federici, G., X. Gao, J. Slawek, T. Arodz, A. Shitaye, J.D. Wulfschlegel, R. De Maria, L.A. Liotta, and E.F. Petricoin 3rd. 2013. Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with "-OMIC" data fields and therapeutic response signatures. *Mol. Cancer Res.* 11:676-685.
10. Gonzalez-Angulo, A.M., B.T. Hennessy, F. Meric-Bernstam, A. Sahin, W. Liu, Z. Ju, M.S. Carey, S. Myhre, et al. 2011. Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. *Clin. Proteomics* 8:11.
11. Shankavaram, U.T., W.C. Reinhold, S. Nishizuka, S. Major, D. Morita, K.K. Chary, M.A. Reimers, U. Scherf, et al. 2007. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integrative microarray study. *Mol. Cancer Ther.* 6:820-832.
12. Troncale, S., A. Barbet, L. Coulibaly, E. Henry, B. He, E. Barillot, T. Dubois, P. Hupé, and L. de Koning. 2012. NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. *PLoS ONE* 7:e38686.
13. Li, B., F. Liang, J. Hu, and X. He. 2012. Reno: regularized non-parametric analysis of protein lysate array data. *Bioinformatics* 28:1223-1229.
14. Chiechi, A., C. Mueller, K.M. Boehm, A. Romano, M.S. Benassi, P. Picci, L.A. Liotta, and V. Espina. 2012. Improved data normalization methods for reverse phase protein microarray analysis of complex biological samples. *Biotechniques* 0:1-7.
15. Mannsperger, H.A., S. Gade, F. Henjes, T. Beissbarth, and U. Korf. 2010. RPPAnalyzer: analysis of reverse-phase protein array data. *Bioinformatics* 26:2202-2203.
16. Aleskandarany, M.A., O.H. Negm, E.A. Rakha, M.A.H. Ahmed, C.C. Nolan, G.R. Ball, C. Caldas, A.R. Green, et al. 2012. TOMM34 expression in early invasive breast cancer: a biomarker associated with poor outcome. *Breast Cancer Res. Treat.* 136:419-427.
17. Alshareeda, A.T., O.H. Negm, N. Albarakati, A.R. Green, C. Nolan, R. Sultana, S. Madhusudan, A. Benhasouna, et al. 2013. Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer. *Breast Cancer Res. Treat.* 139:301-310.
18. Jozefczuk, J., K. Kashofer, R. Ummanni, F. Henjes, S. Rehman, S. Geenen, W. Wruck, C. Regenbrecht, et al. 2012. A systems biology approach to deciphering the etiology of steatosis employing patient-derived dermal fibroblasts and iPS cells. *Front Physiol.* 3:339.
19. Mannsperger, H.A., S. Uhlmann, C. Schmidt, S. Wiemann, O. Sahin, and U. Korf. 2010. RNAi-based validation of antibodies for reverse phase protein arrays. *Proteome Sci.* 8:69.
20. Nelson, L.D., C. Bender, H. Mannsperger, D. Buergy, P. Kambakamba, G. Mudduluru, U. Korf, D. Hughes, et al. 2012. Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Mol. Cancer* 11:38.
21. Loebke, C., H. Sueltmann, C. Schmidt, F. Henjes, S. Wiemann, A. Poustka, and U. Korf. 2007. Infrared-based protein detection arrays for quantitative proteomics. *Proteomics* 7:558-564.

Received 21 February 2014; accepted 20 July 2014.

Address correspondence to Tim Beißbarth, Statistical Bioinformatics, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany. [Tim.Beissbarth@med.uni-goettingen.de](mailto:Tim.Beissbarth@med.uni-goettingen.de)

To purchase reprints of this article, contact: [biotechniques@fosterprinting.com](mailto:biotechniques@fosterprinting.com)