

SOFTWARE

Open Access



# amica: an interactive and user-friendly web-platform for the analysis of proteomics data

Sebastian Didusch<sup>1,2,3\*</sup> , Moritz Madern<sup>1,4</sup> , Markus Hartl<sup>1,4</sup>  and Manuela Baccarini<sup>1,3</sup> 

## Abstract

**Background:** Quantitative proteomics has become an increasingly prominent tool in the study of life sciences. A substantial hurdle for many biologists are, however, the intricacies involved in the associated high throughput data analysis.

**Results:** In order to facilitate this task for users with limited background knowledge, we have developed amica, a freely available open-source web-based software that accepts proteomic input files from different sources. amica provides quality control, differential expression, biological network and over-representation analysis on the basis of minimal user input. Scientists can use amica's query interface interactively to compare multiple conditions and rapidly identify enriched or depleted proteins. They can visualize their results using customized output graphics, and ultimately export the results in a tab-separated format that can be shared with collaborators. The code for the application, input data and documentation can be accessed online at <https://github.com/tbaccata/amica> and is also incorporated in the web application.

**Conclusions:** The strong emphasis on dynamic user interactions, the integration of various databases and the option to download processed data, facilitate the analysis of complex proteomic data for both first-time users and experienced bioinformaticians. A freely available version of amica is available at <https://bioapps.maxperutzlabs.ac.at/app/amica>.

**Keywords:** Proteomics, LC-MS/MS, Web application, Data analysis, Data visualization

## Background

Mass spectrometry (MS)-based proteomics enables deep qualitative and quantitative characterization of any organism's proteome which is crucial for the understanding of the underlying cell biology, physiology and biochemistry. The constant technological advancement of instruments and data-acquisition techniques, as well as the parallel development of a broad variety of methods

such as proximity-dependent labeling for the study of weak or transient protein-protein interactions (PPIs), has expanded the scope and relevance of MS-based approaches in tackling specific biological questions. As a result, proteomic approaches have become increasingly popular, but the complex analysis of MS-based proteomics data is an obstacle for many novices in the field, complicating and delaying the interpretation of experimental outcomes.

Moreover, the many software platforms available for processing proteomics raw data and their different output formats require advanced knowledge for obtaining final interpretable results. Several attempts to solve this problem have been made in the past. The software Perseus [1],

\*Correspondence: [sebastian.didusch@univie.ac.at](mailto:sebastian.didusch@univie.ac.at)

<sup>2</sup>Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria

Full list of author information is available at the end of the article



for example, is widely used because it provides a graphical user interface and has extensive options for the analysis of both label-based and label-free methods. Other software tools, for example MSstats [2] or MSnbase [3] revolve around the R programming language. These tools have the advantage of automating many of the data processing steps that would have to be performed manually in the Perseus interface, but require knowledge of the programming language R.

Recently developed applications integrate their software into R-Shiny apps, allowing for interactive user interactions and visualizations of the output from the MaxQuant package, one of the most widely used software platforms in MS-based proteomics [4]. LFQ-Analyst [5] enables the automatic analysis of label-free data, Eatomics [6] allows for the input of enhanced experimental designs and ProVision [7] can process label-free and TMT labeled data and integrates PPI networks. These tools provide appealing solutions for first-time users but are limited to output from MaxQuant. While Protigy (available at <https://github.com/broadinstitute/protigy>) permits the upload of generic user input, none of these tools allow the systematic comparison of proteins across multiple groups or the integration and comparison of multiple proteomics experiments.

As a solution to these issues, we have developed amica, a user-friendly web-based platform for comprehensive quantitative proteomics data analysis that can automatically handle multiple database search tool outputs such as for instance data from FragPipe - a recent but increasingly popular open-source software package [8, 9] - as well as any generic tab-separated dataset such as RNA-seq data and datasets that were previously analyzed with amica. amica's built-in query interface facilitates the identification of molecular entities (proteins or RNAs) specific to biological groups. The output graphics allow immediate visualization of qualitative and quantitative differences among groups. Finally, omics data obtained by different methods (e.g. RNA-seq and mass

spectrometry) can be directly compared, allowing multi-omics integration.

**Comparison with other tools**

A comparison of amica with currently available R-Shiny apps for proteomic data analysis (LFQ-Analyst, ProVision, Eatomics and Protigy) is shown in Table 1, and underscores amica's versatility.

**Implementation**

amica is implemented as an open-source, interactive R-Shiny app that provides generic user input, quality control and differential abundance analysis for quantitative proteomics data.

amica uses established software tools and libraries and integrates various biological databases. The front page of amica and Table 1.1 in the supplementary material lists all references that should be cited, when using amica to analyze a data set and to generate output graphics.

**Upload: accepted input file formats**

amica can read in common database search tool output formats, custom formats and its own tab-separated format. It is able to achieve this by mapping file specific column names into common features present in proteomics data. These include a unique protein id, a gene name, different type of processed intensities, peptide counts, spectral counts and other common columns in proteomics output formats. A description of column names from analyzed data mapped to amica's format is shown in Table 2.

Three different data upload options are available in amica (see Additional file 1: Fig. 2): a) a database search result such as MaxQuant's proteinGroups.txt file (see Additional file 1: Table 2.1), FragPipe's combined\_proteins.txt file (see Additional file 1: Table 2.2), b) a custom tab-separated file or c) a previously analyzed data set in amica's format. All three options require a tab-separated file denoting the experimental design in the experiment

**Table 1** Comparison with other tools

	LFQ-Analyst	ProVision	Eatomics	Protigy	amica
Generic file input				x	x
Reuploadable output				x	x
Query interface					x
Biological networks		x		x	x
Dot plots					x
Reports	x		x	x	x
Customizable output graphics	(x)	x	(x)	(x)	x
Available at public webserver	x	x	x		x

**Table 2** amica format. Column names of amica's output format. All intensities are log<sup>2</sup> transformed

Column name	Description
Majority.protein.IDs	Unique Protein ID
Gene.names	Gene names
spectraCount	Total MS MS count
razorUniqueCount	Total Razor + unique peptides
razorUniqueCount.<sample>	Razor + unique peptides per sample
iBAQ_<sample>	iBAQ intensity per sample
RawIntensity_<sample>	Raw Intensity per sample
LFQIntensity_<sample>	LFQ Intensity per sample
ImputedIntensity_<sample>	Normalized and imputed intensity per sample
Potential.contaminant	Potential contaminants have a "+" in this column
P.Value_	<i>p</i> -value for a pairwise group comparison, e.g P.Value_group1__vs__group2
adj.P.Val_	adj. <i>p</i> -value for a pairwise group comparison, e.g adj.P.Val_group1__vs__group2
logFC_	log <sub>2</sub> fold change for a pairwise group comparison, e.g logFC_group1__vs__group2
AveExpr_	Avg. intensity for a pairwise group comparison, e.g AveExpr_group1__vs__group2
quantified	Filtered proteins contain a "+" in this column

(see Additional file 1: Table 2.4), i.e. the mapping of samples to distinct biological groups. For option a) and b) a tab-separated file containing a contrast matrix that specifies the desired pairwise group comparisons to be made needs to be uploaded (see Additional file 1: Table 2.5). Finally, for custom file uploads in option b) an additional specification file, that maps common column features in proteomics output formats to amica's format needs to be uploaded (see Additional file 1: Table 2.6).

### Analysis options

Users can filter proteins on the basis of minimum count values for MS/MS counts and razor + unique peptides. Additionally, proteins can be filtered by valid values per group, e.g only proteins identified in a minimal number of replicates are kept for analysis (see Additional file 1: Fig. 3). Sample intensities (e.g raw intensities or LFQ intensities) can be selected for log<sub>2</sub> transformation and normalization. Normalization options include no normalization, quantile normalization, variance stabilization normalization (VSN) and median normalization. Two different methods for differential expression analysis can be selected, limma [10] and DEqMS [11]. Finally, three different imputation methods for missing values can be chosen: i) lowest detected value, ii) randomized sampling from a normal distribution with a user specified down-shift and width for each sample or iii) globally for all samples.

### Example data

The example data was taken from an interaction proteomics study [12]. This includes raw files for four groups from an interaction proteomics study focusing on PGRMC1,

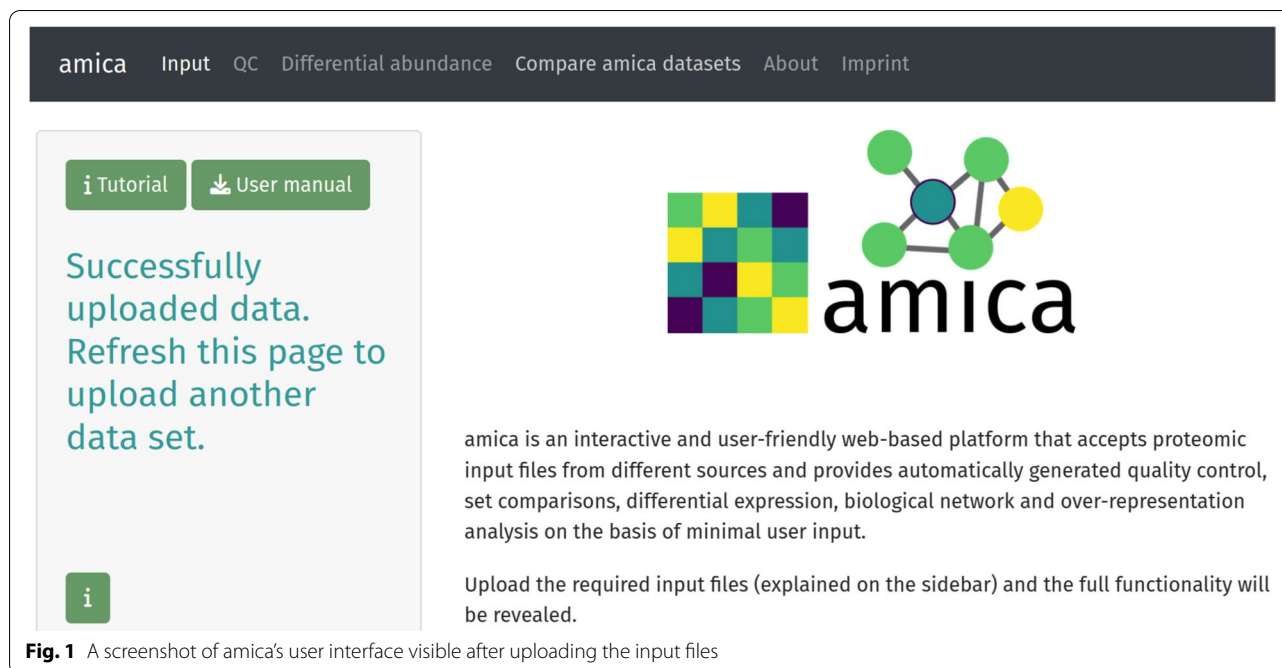
a protein from the membrane-associated progesterone receptor family with a variety of cellular functions. In this study, MIA PaCa-2 cells were stably transfected with a PGRMC1-HA plasmid and Co-IPs of PGRMC1 interacting proteins were isolated from cells expressing PGRMC1-HA, as well as from non-transfected parental MIA PaCa-2 cells as a negative control, with and without AG-205 treatment (a PGRMC1-specific inhibitor).

### Data processing

MaxQuant (version 1.6.17.0) was used to analyze the raw files. As search database, UniProt UP000005640 (downloaded on 10th September 2021) was used, with Trypsin/P as proteolytic enzyme allowing for two missed cleavages. The match between runs (MBR) feature was not used. Oxidation on methionine and protein N-terminal acetylation were set as variable modifications, and Carbamidomethylation of Cysteine was set as fixed modification. Label-free quantification and normalization was performed with the MaxLFQ algorithm [13].

Additionally, FragPipe (version 16) with MSFragger [14] (version 4.0.0) and Philosopher [9] (version 4.0.0) was used to analyze the raw files. Label-free quantification and normalization was performed with the MaxLFQ algorithm by IonQuant [15] (version 1.7.5). The same search database as well as variable and fixed modifications as for MaxQuant were used for FragPipe. Peptide validation was executed by Percolator [16] (version 3.05). The MBR feature was not used.

The output from MaxQuant and FragPipe were further analyzed using the same analysis parameters in amica. Briefly, proteins with at least 2 Razor + unique peptides, at least 3 MS/MS counts, and valid values in 3 out of 5



**Fig. 1** A screenshot of amica's user interface visible after uploading the input files

replicates in at least one group were considered quantified. LFQ intensities of quantified proteins were  $\log_2$ -transformed and missing values were imputed from a normal distribution downshifted 1.8 standard deviations from the mean with a width of 0.3 standard deviations. Differential expression analysis was performed with limma. Both sets of results can be downloaded from the example datasets in amica. Only the results of the Max-Quant analysis are shown here.

## Results

amica is developed as a user-friendly web application with interactive and customizable visualizations that can be exported in a publication-ready vector graphic format.

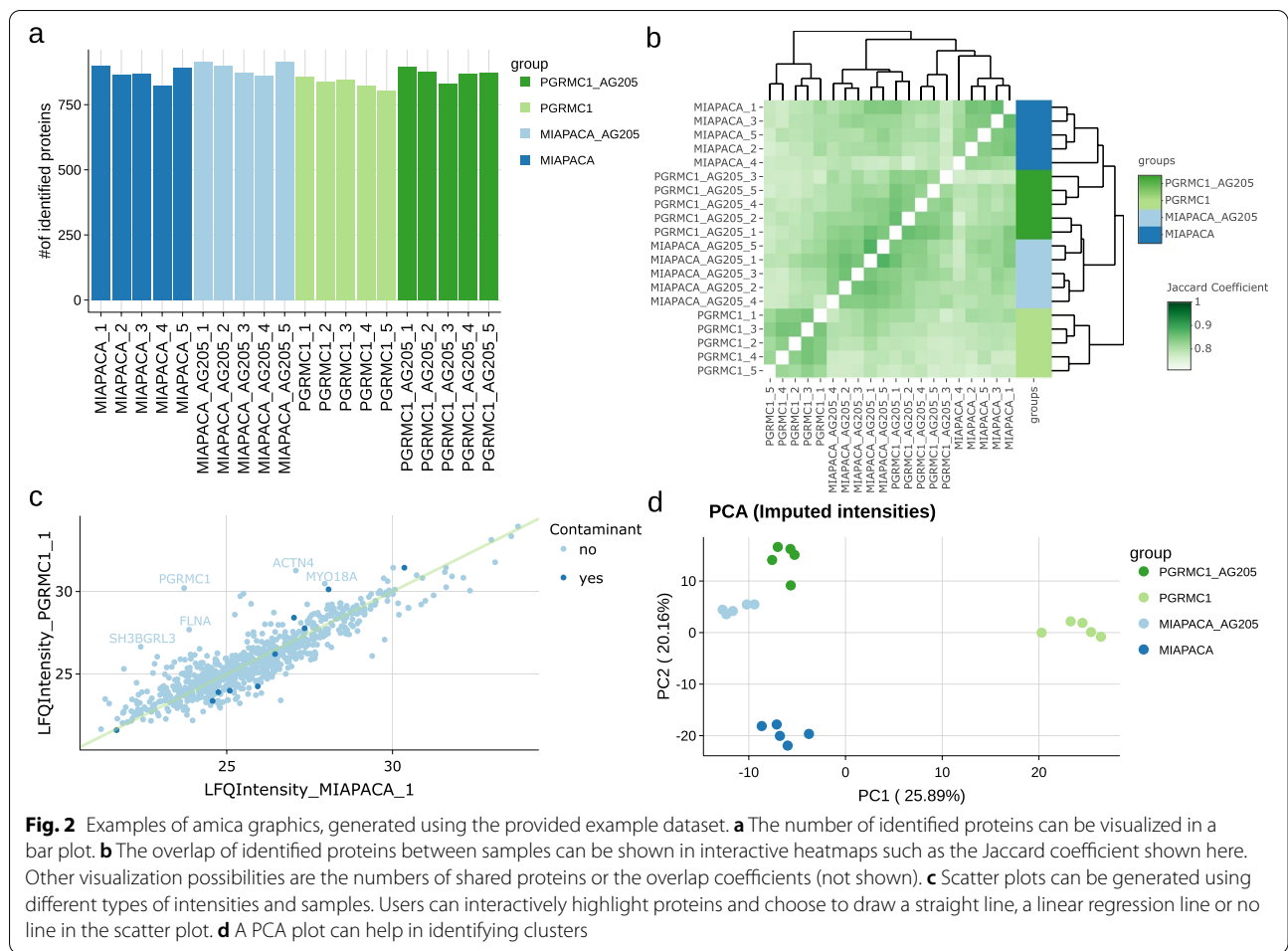
amica's landing page displays online documentation, a link to a user manual and a link to download the example data set in all allowed input formats (Fig. 1). The landing page also serves the purpose of the user input tab. After successfully uploading the required input files, amica generates a downloadable file in its custom format. This file can be used as input file for subsequent re-inspection, analysis and visualization in amica (see Additional file 1: Fig. 4).

In the input tab, users can also define the colors to be used for visualizations (see Additional file 1: Fig. 5). Color palettes can be either chosen from ColorBrewer [17] or defined by a color picker tool. This includes i) qualitative colors to label biological groups of the experimental design, ii) qualitative colors for various types of scatter plots and iii) color gradients used for heatmaps

and correlation plots. Once chosen, colors propagate through all visualizations in amica, allowing for coherent output graphics. The input tab allows the users to define a specific order in which the biological groups will be displayed on the plots axes and legends. This feature proves especially useful for visualizing time series data.

In the Quality Control (QC) tab, users can generate and inspect plots at different processing levels of the data, such as raw intensities, LFQ intensities, iBAQ intensities and normalized and imputed intensities. This makes it easy to examine the impact of common pre-processing steps (normalization and imputation). An automatically generated report containing user selected analysis parameters, and plots comparing intensities before and after normalization and imputation can be downloaded, allowing for reproducible output and analysis.

The available visualizations for the intensity distributions in density - and box plots for different samples are particularly useful in this respect (see Additional file 1: Fig. 6). A barplot for the number of identified proteins (Fig. 2a), a sample overlap heatmap of identified proteins (Fig. 2b), as well as a scatter plot (Fig. 2c) and a boxplot of coefficient of variations (see Additional file 1: Fig. 7a) make it possible, to evaluate the reproducibility of replicates. A correlation plot (see Additional file 1: Fig. 8b) and a principle component analysis (PCA) plot (Fig. 2d) allow to identify clusters in an exploratory data analysis in the QC tab. Last but not least, when iBAQ intensities are available, barplots for the percentage of missing values (see Additional file 1: Fig. 9b), the percentage of



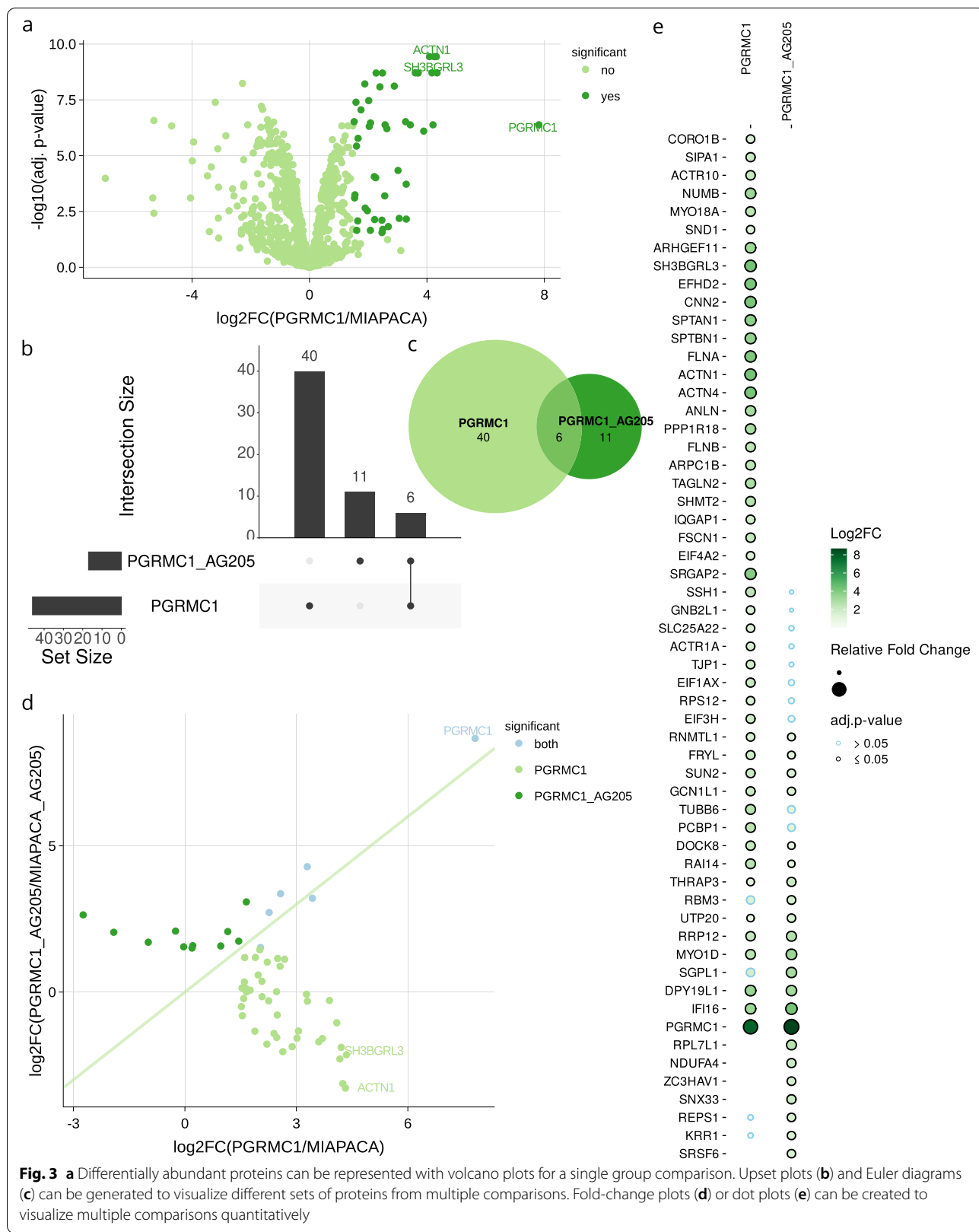
intensity of the most abundant proteins, and the percentage of intensity of potential contaminants per sample allow for the detection of outliers (see Additional file 1: Fig. 11).

The Differential Abundance tab is the heart of amica, enabling the analysis and visualization of quantitative proteomics data. As a first step, users can specify common thresholds to subset protein groups of interest, such as  $\log_2$  fold change thresholds, thresholds on  $p$ -value or adjusted  $p$ -value, and select only enriched, only reduced, or all differentially abundant proteins in pairwise group comparisons. In the next step, users can select single or multiple pairwise group comparisons to which these thresholds should be applied. An automatically generated differential abundance report summarizing the results can be downloaded.

Differentially abundant proteins can be visualized as volcano (Fig. 3a) - and MA - plots (see Additional file 1: Fig. 12b) for single group comparisons. Unlike other R-Shiny apps for proteomics data analysis, amica can generate UpSet plots [18] (Fig. 3b) and Euler diagrams

(Fig. 3c) for visualizing the overlap of significant proteins from multiple selected group comparisons. Differentially abundant proteins are displayed in a data table that can be exported as a csv file for further analysis (see Additional file 1: Fig. 14). The columns of the data table show the gene names of proteins, statistical information such as  $\log_2$  fold changes,  $p$ -values, adjusted  $p$ -values, and a binary sign (yes or no) of significance for all user-selected comparisons. In addition to the filters applicable to each column, users can use the gene name column to further subset the data table by specifying search patterns using regular expressions. amica's unique query interface allows to immediately inspect differentially abundant proteins that are part of a specific process or pathway (see Additional file 1: Section 7.2).

In order to compare two conditions quantitatively, amica can generate a scatterplot of  $\log_2$ -fold changes on the x-axis and the y-axis (fold-change plots; Fig. 3d). Significant changes are highlighted in color. amica integrates dot plots for the visualization of quantitative and statistical information of more than two conditions (Fig. 3e). In

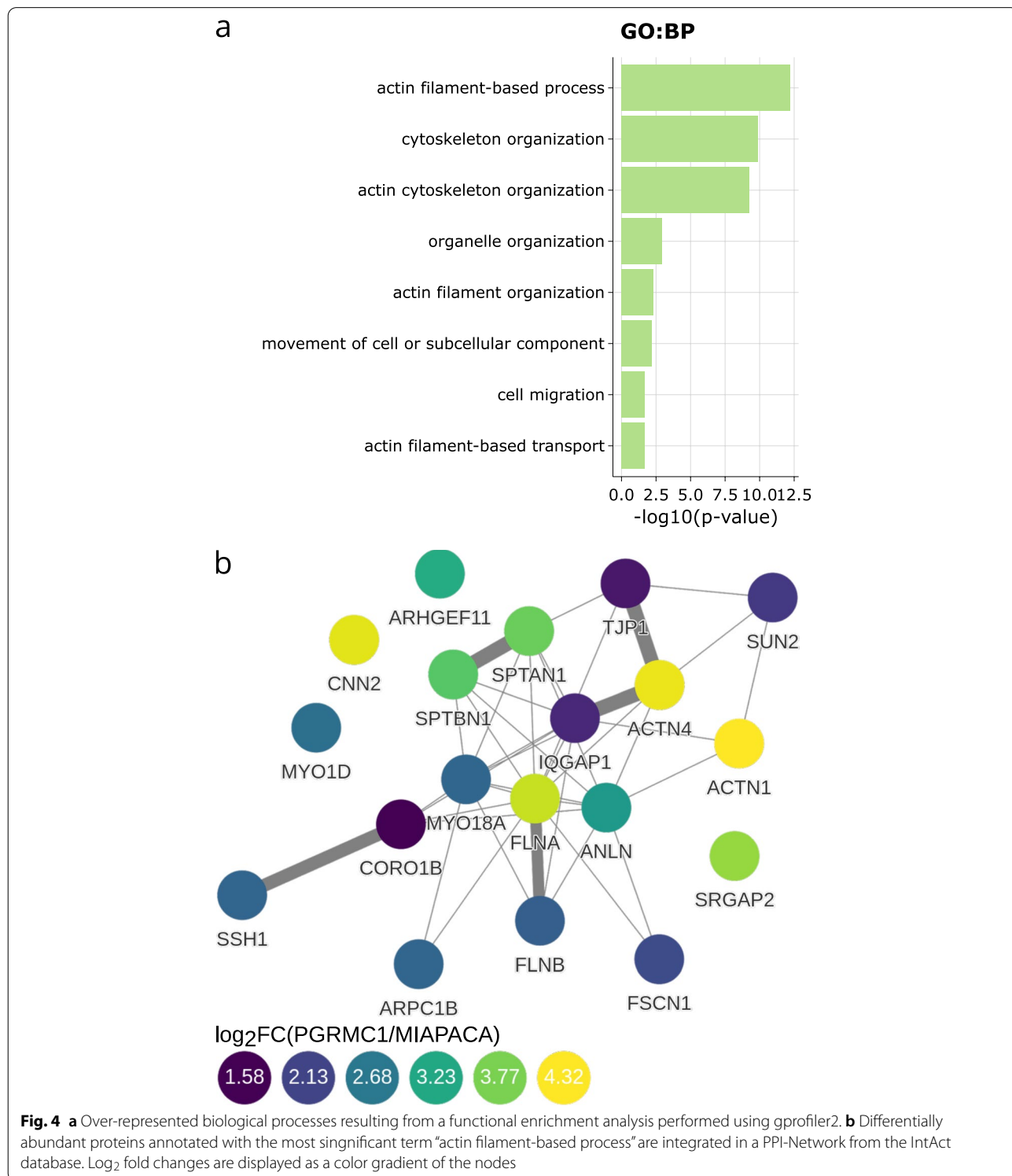


**Fig. 3** a Differentially abundant proteins can be represented with volcano plots for a single group comparison. Upset plots (b) and Euler diagrams (c) can be generated to visualize different sets of proteins from multiple comparisons. Fold-change plots (d) or dot plots (e) can be created to visualize multiple comparisons quantitatively

this type of plot, every row corresponds to a protein, and every column to a pairwise group comparison. Proteins are displayed as circles, whose sizes and colors can be either selected as intensity or log<sub>2</sub> fold change. The line

color of the circle shows the statistical significance of a protein in a group comparison.

In addition to visualizing proteins across multiple groups, interactive heatmaps [19] (see Additional file 1:



**Fig. 4** **a** Over-represented biological processes resulting from a functional enrichment analysis performed using gprofiler2. **b** Differentially abundant proteins annotated with the most significant term “actin filament-based process” are integrated in a PPI-Network from the IntAct database. Log<sub>2</sub> fold changes are displayed as a color gradient of the nodes

Fig. 15) can be created to compare and cluster many proteins, or profile plots can be created for a single protein (see Additional file 1: Fig. 20).

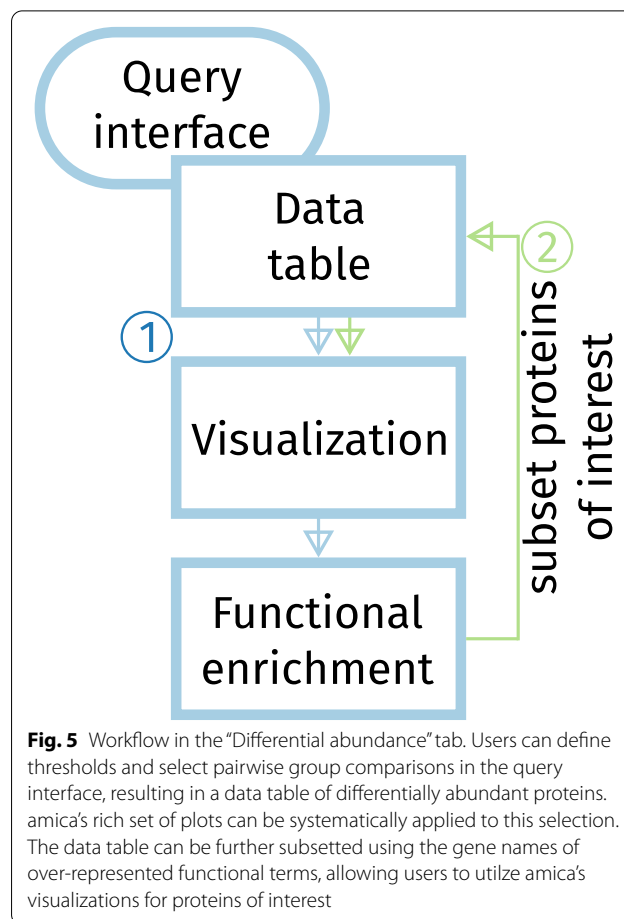
For human proteomics studies, amica allows users to map proteins to a PPI network from the IntAct database [20] (Fig. 4b), and enables the retrieval of subcellular localization predictions from the humancellmap database [21]. As an example, selecting a particular cellular compartment in the interactive web interface will highlight all proteins that map to this localization. This information can be downloaded in gml format for visualization in a network visualization tool such as Cytoscape [22] (a tutorial on visualizing amica's output in Cytoscape is described in Additional file 1: Section 7.3).

A functional enrichment analysis of differentially abundant proteins using gprofiler2 [23] assists in building hypotheses on the underlying biology. Users can select the organism of the experiment and its corresponding data sources (see Additional file 1: Fig. 21). The output of gprofiler2 is shown as a manhattan plot (see Additional file 1: Fig. 22a) and as a barplot (Fig. 4a) for a user selected data source. Additionally, a downloadable output table with all data sources, term names,  $p$ -values and corresponding gene names is displayed (see Additional file 1: Fig. 23). The gene names from this table can be used to construct a regular expression to subset the query data table, allowing for the visualization of proteins belonging to a specific gene set.

Finally, the "Compare amica datasets" tab gives the possibility to upload previously analyzed datasets to be compared with the current data input (see Additional file 1: Fig. 24). As a key column, the two datasets can be combined by protein ID or gene name. After successfully uploading a second amica file, a scatter plot and a correlation plot becomes available for correlation analysis of the combined dataset (see Additional file 1: Fig. 25). In addition, a selection box will appear at the top of the Differential Abundance tab, allowing users to use amica's query interface on the integrated datasets (see Additional file 1: Fig. 26).

## Conclusions

amica is a versatile software tool for the analysis, visualization, and interpretation of MS-based proteomics data. amica's user-friendly interface provides a customizable data analysis workflow, and the results of the analysis can be conveniently exported, shared and reloaded into the amica environment for re-inspection at a later time. The data analysis workflow in amica includes quality control and standard differential expression testing, as well as the integration of PPI networks and pathway and gene ontology enrichment analysis for differentially abundant proteins. This latter feature allows the identification of



**Fig. 5** Workflow in the "Differential abundance" tab. Users can define thresholds and select pairwise group comparisons in the query interface, resulting in a data table of differentially abundant proteins. amica's rich set of plots can be systematically applied to this selection. The data table can be further subsetted using the gene names of over-represented functional terms, allowing users to utilize amica's visualizations for proteins of interest

process-relevant entities that can be used to subset the data table in an iterative manner (Fig. 5). The combination of all these features in one single application will help researchers to focus on, and interpret the biology underlying the results of their proteomics experiment. The code for the application and online documentation can be found at <https://github.com/tbaccata/amica> and the software is available at <https://bioapps.maxperutzlabs.ac.at/app/amica>.

## Availability and requirements

**Project name:** amica.

**Project home page:** <https://bioapps.maxperutzlabs.ac.at/app/amica>

**Operating system(s):** Platform independent.

**Programming language:** R.

**Other requirements:** none for the webserver, R version 4 or higher for local installation.

**License:** GNU GPL version 3.0.

**Any restrictions to use by non-academics:** none.

amica's source code, online documentation and a formatted example data set are available on <https://github.com/tbaccata/amica>.



## Abbreviations

MS: Mass spectrometry; PPI: Protein-protein interaction; MS/MS: Tandem-mass spectrum; LFQ: Label-free quantification; VSN: Variance Stabilization Normalization; QC: Quality Control; iBAQ: Intensity-based absolute quantification; MBR: Match between runs.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-09058-7>.

**Additional file 1: Supplementary Material — File 1.** Link to a user manual demonstrating every available function and output graphics, as well as extensive tutorials.

## Acknowledgements

The authors thank the Baccharini lab, the Max Perutz Labs MS facility, Jörg Menche and his group, Matthias Bärnthaler and the IT team for feedback and helpful comments.

## Authors' contributions

Conceptualization: S.D, M.B, M.H and M.M. Software development: S.D and M.M. Implementation of webserver: S.D. Testing and feature requests: S.D, M.B, M.H and M.M. Writing: S.D, M.B, M.H and M.M. The authors read and approved the final manuscript.

## Funding

This work was supported by Austrian Research Fund (W1261 to MB, SFB F70, F7007 to MH). The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

The example dataset was downloaded from the ProteomeXchange Consortium (<http://www.proteomexchange.org/>) via the PRIDE partner repository with the dataset accession number: PXD016455.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Max Perutz Labs, Vienna Biocenter Campus (VBC), Dr.-Bohr-Gasse 9, 1030 Vienna, Austria. <sup>2</sup>Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria. <sup>3</sup>University of Vienna, Center for Molecular Biology, Department of Microbiology, Immunobiology and Genetics, Dr.-Bohr-Gasse 9, 1030 Vienna, Austria. <sup>4</sup>University of Vienna, Center for Molecular Biology, Department of Biochemistry and Cell Biology, Dr.-Bohr-Gasse 9, 1030 Vienna, Austria.

Received: 3 May 2022 Accepted: 28 November 2022

Published online: 09 December 2022

## References

- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat Methods*. 2016;13(9):731–40.
- Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*. 2014;30(17):2524–6.
- Gatto L, Lilley KS. MSnbase—an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*. 2012;28(2):288–9.
- Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*. 2016;11(12):2301–19.
- Shah AD, Goode RJ, Huang C, Powell DR, Schittenhelm RB. LFQ-analyst: an easy-to-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with MaxQuant. *J Proteome Res*. 2019;19(1):204–11.
- Kraus M, Mathew Stephen M, Schapranow MP. Eatomics: Shiny exploration of quantitative proteomics data. *J Proteome Res*. 2020;20(1):1070–8.
- Gallant JL, Heunis T, Sampson SL, Bitter W. ProVision: a web-based platform for rapid analysis of proteomics data processed by MaxQuant. *Bioinformatics*. 2020;36(19):4965–7.
- Teo GC, Polasky DA, Yu F, Nesvizhskii AI. Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. *J Proteome Res*. 2020;20(1):498–505.
- da Veiga Leprevost F, Haynes SE, Avtonomov DM, Chang HY, Shanmugam AK, Mellacheruvu D, et al. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat Methods*. 2020;17(9):869–70.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47.
- Zhu Y, Orre LM, Tran YZ, Mermelekas G, Johansson HJ, Malyutina A, et al. DEqMS: a method for accurate variance estimation in differential protein expression analysis. *Mol Cell Proteom*. 2020;19(6):1047–57.
- Teakel SL, Ludescher M, Thejer BM, Poschmann G, Forwood JK, Neubauer H, et al. Protein complexes including PGRMC1 and actin-associated proteins are disrupted by AG-205. *Biochem Biophys Res Commun*. 2020;524(1):64–9.
- Cox J, Hein MY, Lubner CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteom*. 2014;13(9):2513–26.
- Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods*. 2017;14(5):513–20.
- Yu F, Haynes SE, Nesvizhskii AI. IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. *Mol Cell Proteom*. 2021;20:100077.
- MacCoss MJ, Noble WS, Käll L, et al. Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. *J Am Soc Mass Spectrom*. 2016;27(11):1719–27.
- Neuwirth E. RColorBrewer: ColorBrewer palettes; 2014. Available: <https://cran.r-project.org/package=RColorBrewer>. Accessed 05 Dec 2022.
- Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938–40.
- Gallili T, O'Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*. 2018;34(9):1600–2.
- Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 2014;42(D1):D358–63.
- Go CD, Knight JD, Rajasekharan A, Rathod B, Hesketh GG, Abe KT, et al. A proximity-dependent biotinylation map of a human cell. *Nature*. 2021;595(7865):120–4.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and comparisons of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.