BIOINFORMATICS

Vol. 00 no. 00 2016
Pages 1–3

msVolcano: a flexible web application for visualizing quantitative proteomics data.

Sukhdeep Singh [1,*], Marco Y. Hein [2] and A.Francis Stewart[1*]

[1]Genomics, Biotechnology Center, Technische Universitaet Dresden, Tatzberg 47, 01307 Dresden, Germany
[2]Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA 94143, USA

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary: We introduce msVolcano, a web application, for the visualization of label-free mass spectrometric data. It is optimized for the output of the MaxQuant data analysis pipeline of interactomics experiments and generates volcano plots with lists of interacting proteins. The user can optimize the cutoff values dynamically to find meaningful significant interactors for the tagged protein of interest. Optionally, stoichiometries of interacting proteins can be calculated. Several customization options are provided to the user for flexibility and publication-quality outputs can also be downloaded (tabular and graphical).

Availability: msVolcano is implemented in R Statistical language using Shiny and is hosted at server in-house. It can be accessed freely from anywhere at http://projects.biotec.tu-dresden.de/msVolcano/

Contact: sukhdeep.singh@biotec.tu-dresden.de

1 INTRODUCTION

The analysis of protein-protein interactions and complex networks using affinity purification and mass spectrometry (AP/MS) is one of the most commonly used applications in proteomics. Furthermore, the data gathered using AP-MS is also high in quality Royer *et al.* (2012). Though high-resolution isotope labeling methods have been developed to detect and quantify protein-protein interactions Ong and Mann (2006), label-free approaches are gaining momentum due to their simplicity and applicability Tate *et al.* (2013). While different quantification strategies exist for label-free data, such as those based on spectral counting, methods that make use of peptide intensities (also known as extracted ion currents) are regarded as the most accurate Choi *et al.* (2012); Cox *et al.* (2014). Such methods generate quantitative profiles of peptides or proteins across samples, which can be analyzed by established statistical methods, e.g. by a modified t-test across replicate experiments Tusher *et al.* (2001).

MaxQuant is an integrated suite of algorithms for high-resolution quantitative MS data Cox and Mann (2008). Using correlation analysis and graph theory, MaxQuant detects peaks and isotope clusters as three-dimensional objects in m/z, elution time and signal intensity space. An additional module of quantification algorithms, MaxLFQ, was added to normalize the contribution of individual peptide fractions and extract the maximum available quantitative information to calculate highly reliable relative protein quantification profiles Cox *et al.* (2014); Luber *et al.* (2010). MaxQuant analyses the raw MS data and outputs a user controlled tabulated file, which contains profiles of LFQ intensities per replicate per protein identified and other meta-information about the quantity and quality of the recovered protein, with calculated statistics.

To identify interactors of a specific tagged protein of interest, termed the 'bait', in the presence of background proteins, a student's t-test or Welch's test comparing the LFQ intensities of all proteins identified in replicates of that bait with the LFQ intensities of all proteins identified in a set of negative control samples is performed, e.g. by MaxQuant's Perseus data analysis module. A volcano plot is a good way to visualize this information. When the resulting differences between the logarithmized mean protein intensities between bait and the control are plotted against the negative logarithmic p values derived from the statistical test, unspecific background binders center around zero. The enriched interactors appear on the right section of the plot, whereas ideally no proteins should appear on the left section when compared to an empty control (because these would represent proteins depleted by the bait). The higher the difference between the group means (i.e. the enrichment) and the p value (i.e. the reproducibility), the more the interactors shift towards the upper right section of the plot space, which represents the area of highest confidence for a true interaction.

Though label-free methods are nowadays as accurate as the isotope-based methods, false positives may appear enriched alongside the true positive interaction partners Eberl *et al.* (2013); Schiess *et al.* (2009). Identifying background proteins and defining a threshold that separates these from true interactors is a critical step during data analysis, and often benefits from some manual optimization.

To facilitate the analysis and presentation of AP-MS data, we present msVolcano, which is a user modulated, freely accessible web application. This takes MaxQuant output and generates a volcano plot. We implemented a recently introduced hyperbolic curve threshold Keilhauer *et al.* (2015), which can be dynamically modulated in the user interface. User also has access to the plot

*to whom correspondence should be addressed

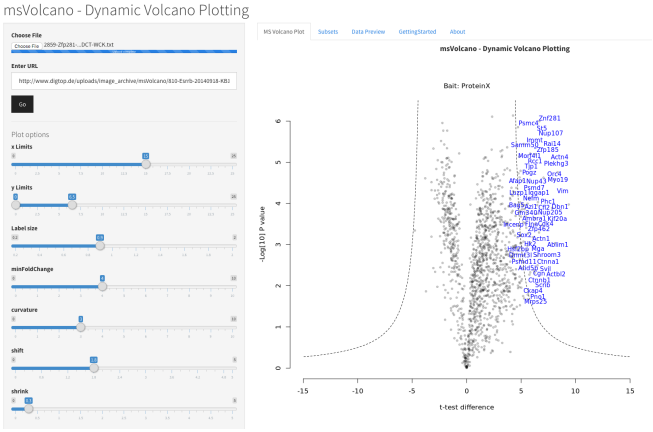

Fig. 1. The interface has several panels starting with user-controlled buttons on the left to toggle the plot aesthetics, hyperbolic curve parameters (cutoff and minFoldChange), stoichiometry options and exporting of plot and data. The user can click the "Stoichiometry" button in the interface to reflect these calculations in the plot, in terms of red circles around the labels. When user uploads a file, all LFQ columns are scanned and displayed on the left side. User now selects respective bait and control columns (minimum two) and optionally enters the name of bait in the provided text box. The plot is generated simultaneously on the main screen as the GO button is pressed, under the tab name "MS Volcano Plot". Second tab, "Stochiometry" displays the filtered input data for the significant interactors. "Data Preview" tab displays the user inputted data for scrutiny, "GettingStarted" and "About" tab displays the specific and general information about the web interface.

aesthetics and can view the original input file and its subset for significant interactors in the inbuilt browser. A publication quality PDF plot can be generated and exported along with the subset of original data limited to the significant interactors.

2 DESCRIPTION

MaxQuant is a quantitative proteomics software package designed for analyzing large mass-spectrometric data sets Cox and Mann (2008). The MaxLFQ module of MaxQuant outputs a profile of LFQ intensities, which are calculated for each protein as the best estimate satisfying all the pair-wise peptide comparisons. Importantly, this intensity profile retains the absolute scale from the original summed-up peptide intensities Cox *et al.* (2014). This should readily qualify it as a proxy for absolute protein abundance. The output file also contains information about the number of recovered peptides, protein name and annotation. The purpose of msVolcano is to implement all steps of downstream data analysis into a simple and intuitive user interface that requires no bioinformatics knowledge or specialized software. msVolcano scans for all the columns with LFQ intensities in the input and gives user the control of selecting bait and control columns. This way user can increase/decrease or select the significant replicates by analyzing the generated volcano plot. Additional information about the sequence coverage, molecular weight and the sequence length of the protein recovered is also embedded in the output file (Fig 1). msVolcano then calculates the statistical test and implements a strategy for the hyperbolic cut off, based on the given formula

$$y = \frac{c}{x - x0} \tag{1}$$

where c = curvature, x0 = minimum fold change, thus dividing enriched proteins into mildly and strongly enriched Keilhauer *et al.* (2015). The cutoff parameters can be adjusted by the user and monitored by the dynamic graphical output. The identification of protein-protein interactions, accompanied by the estimation of the stoichiometries of interacting proteins, is crucial to understand the molecular function of protein complexes Smits *et al.* (2012). Thus, optional stoichiometry calculations have been implemented in the code. We use a slightly modified version of intensity-based absolute quantification (iBAQ) estimation of protein abundance for the stoichiometry calculations with LFQ intensities, normalized by the number of theoretical peptides, as input Borner *et al.* (2014); Schwanhäusser *et al.* (2011) (Fig 1).

3 CONCLUSION

msVolcano provides a web-platform for the quick visualization of label-free mass spectrometric data and can be freely accessed globally. With the underlying hyberbolic curve parameters and other statistics, user can intuitively isolate the true protein interaction partners from the false positives, without the need of writing a code. With its ftp file input support, user can quickly analyse and re-analyse the results of the interactomics experiment present on their own personal and cloud servers and along with the calculated optional stoichiometries, all the results can be exported in publication quality tabular or graphical format.

FUNDING

This work was supported by the EU 7th Framework integrated project, SyBoSS (www.syboss.eu).

REFERENCES

Borner, G. H., Hein, M. Y., Hirst, J., Edgar, J. R., Mann, M., and Robinson, M. S. (2014). Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein–protein interactions. *Molecular biology of the cell*, **25**(20), 3178–3194.

Choi, H., Glatter, T., Gstaiger, M., and Nesvizhskii, A. I. (2012). Saint-ms1: protein–protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. *Journal of proteome research*, **11**(4), 2619–2624.

Cox, J. and Mann, M. (2008). Maxquant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, **26**(12), 1367–1372.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Molecular & Cellular Proteomics*, **13**(9), 2513–2526.

Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Molecular cell*, **49**(2), 368–378.

Keilhauer, E. C., Hein, M. Y., and Mann, M. (2015). Accurate protein complex retrieval by affinity enrichment mass spectrometry (ae-ms) rather than affinity purification mass spectrometry (ap-ms). *Molecular & Cellular Proteomics*, **14**(1), 120–135.

Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., *et al.* (2010). Quantitative proteomics

reveals subset-specific viral recognition in dendritic cells. *Immunity*, **32**(2), 279–289.

Ong, S.-E. and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (silac). *Nature protocols*, **1**(6), 2650–2660.

Royer, L., Reimann, M., Stewart, A. F., and Schroeder, M. (2012). Network compression as a quality measure for protein interaction networks. *PloS one*, **7**(6), e35729.

Schiess, R., Mueller, L. N., Schmidt, A., Mueller, M., Wollscheid, B., and Aebersold, R. (2009). Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry. *Molecular & Cellular Proteomics*, **8**(4), 624–638.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, **473**(7347), 337–342.

Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A., and Vermeulen, M. (2012). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic acids research*, page gks941.

Tate, S., Larsen, B., Bonner, R., and Gingras, A.-C. (2013). Label-free quantitative proteomics trends for protein–protein interactions. *Journal of proteomics*, **81**, 91–101.

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, **98**(9), 5116–5121.