Supplemental Material for
NetWeAvers: an R package for integrative biological network analysis with mass spectrometry data
McClellan EA, Moerland PD, van der Spek PJ and Stubbs AP

1. Network Analysis Tools Comparison

NetWeAvers is one of several free tools available for performing network analysis for proteomics data, each with its own (dis)advantages. For example, NetWeAvers uses a simple but general summarization method where other tools provide more sophisticated methods perhaps specific to certain data types. Table S1 compares NetWeAvers to several public network analysis tools: BioNet (Beisser *et al.*, 2010); ppiStats (Chiang *et al.*, 2013); DEGraph (Jacob *et al.*, 2010); jActiveModules (Ideker *et al.*, 2002); BiNGO (Maere *et al.*, 2005); MCODE (Bader and Hogue, 2003); R spider (Antonov *et al.*, 2010); atBioNet (Ding *et al.*, 2012).

Table S1. R packages and other tools for analyzing networks with mass spectrometry proteomics data.

	Summarization and protein-level testing?	Generic experiment type?	p-value threshold-free?	> 1 possible significant network?	Few parameter specifications?	Use of any network?
R packages						
NetWeAvers	yes	yes	yes	yes	yes	yes
BioNet	no	yes	no	no (only 1 optimal)	no	yes
ppiStats	no (testing only)	no (bait-prey only)	N/A	no (analysis is per protein)	yes	yes
DEGraph	no (testing only)	no (2 conditions only)	no	yes (if >1 network submitted)	yes	yes
Cytoscape						
jActiveModules	no (testing only)	yes	yes	yes	no	yes
BiNGO	no	yes	no	yes	yes	yes
MCODE	no	no	yes	no (no statistical scores)	no	yes
Web-based tools						
R spider *	no	yes	N/A	yes	yes	no
atBioNet *	no	yes	N/A	yes	no	no

* free, not open source

2. The NetWeAvers Procedure

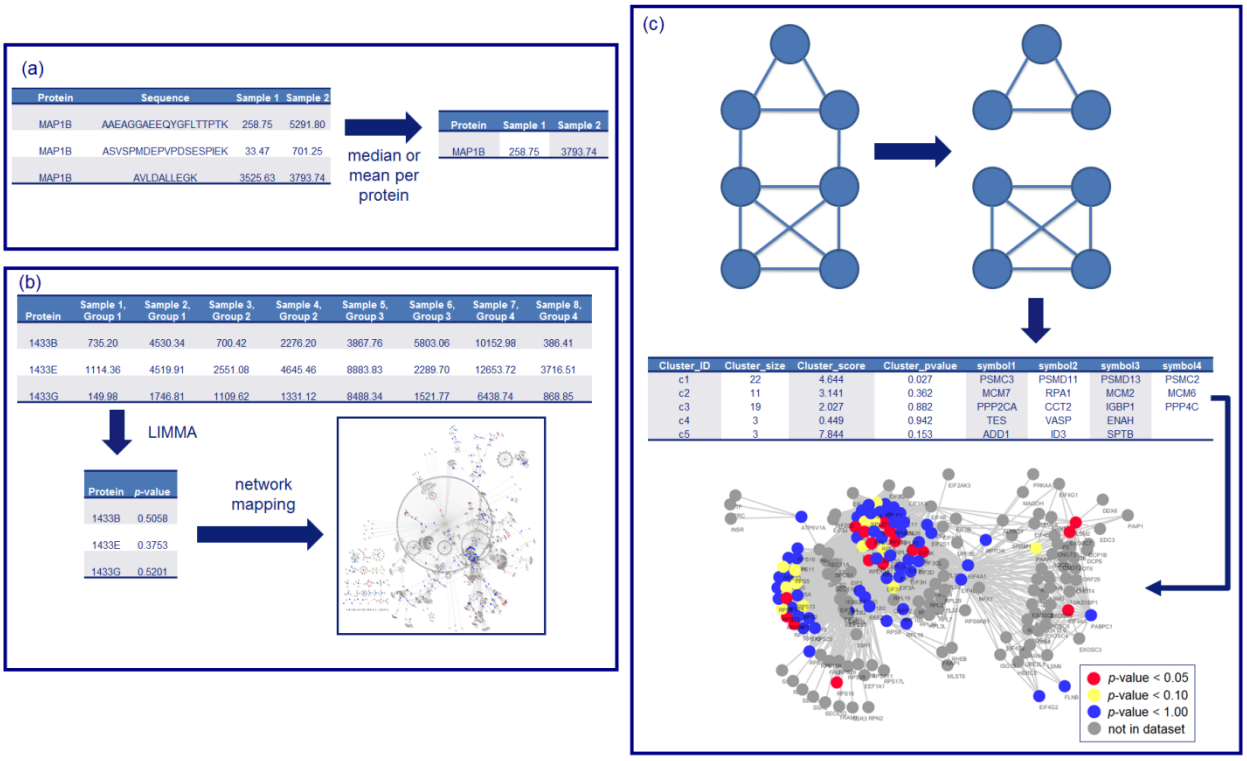


Figure S1. Main steps of the NetWeAvers procedure. (a) Unique peptide abundances or intensities are summarized per protein to obtain a single measure for each protein. (b) Linear models (R/Bioconductor package limma) are used to determine statistical significance of differential expression between groups. Protein *p*-values are mapped onto a protein-protein interaction network. (c) Network analysis first finds highly connected (dense) clusters and then calculates a weighted average of *p*-values per dense cluster. Output includes cluster size, score, *p*-value and members (proteins) of the clusters. Clusters may be viewed using e.g. Cystocape (Shannon *et al.*, 2003).

3. Details of the NetWeAvers Algorithm

Finding Dense Clusters

The function `findDenseClusters` includes options that let the user specify the minimal size of the desired network (`min_clus_size`) and the number of steps to take in random walks on the network (`steps`). See http://arxiv.org/abs/physics/0512106 for more details on the random walk algorithm and http://igraph.sourceforge.net/doc/R/walktrap.community.html for its R implementation.

Scoring Clusters

The clusters are scored using weighted *p*-values (function `scoreClusters`) where the weight, *w*, given to each protein is the inverse of the number of proteins with which the given protein interacts. A measure for each protein is given by $\ln(p\text{-value}*w^{1/c})$ where c is a user-specified value indicating how influential the weight should be. A value of $c = 1$ makes the weights as influential as the *p*-values in the algorithm, while a value less than 1 makes them highly influential such that the protein score heavily relies on the number of proteins with which the given protein interacts. A value of c that is much larger than 10 makes

the weights negligible. The default value for *c* is 10, which somewhat suppresses the weights. If, for example, the user is interested in finding clusters around a hub, then *c* should be set to a small value less than 1. Once the protein measures have been calculated, they are combined within clusters to generate a cluster score by taking either the mean or median of the individual protein measures.

Permutation Test

The null hypothesis of the permutation test (function `permTest`) is that differentially abundant proteins are present in the dense clusters at random. By default, protein names are permuted 1000 times and the cluster score is recalculated for each permutation. For a given cluster, the proportion of permuted scores greater than the cluster's observed score is the cluster *p*-value. A small cluster *p*-value may be considered statistically significant, as it indicates that the dense cluster is enriched with differentially expressed proteins.

4. Application Results

We applied the R package to mass spectrometry data from a phosphorylation study of human embryonic stem cells (hESCs, Van Hoof *et al.*, 2009). The authors performed a stable isotope labeling by amino acids in cell culture (SILAC) experiment using undifferentiated hESCs and hESCs differentiated with bone morphogenetic protein 4 (BMP4). Measurements were taken at three time points (30, 60 and 240 minutes after initiation of differentiation) with two biological replicates at each time point. The data were processed using PVIEW (Khan *et al.*, 2009, 2011); the processed dataset is available in the NetWeAvers package (`vanHoof`). See the R package vignette provided as Supplementary File 2 for the code used for summarizing the data, performing hypothesis testing on the summarized data, and running the network analysis using the protein *p*-values from the comparison of 30 and 60 minutes with the Reactome human PPI network, version 43 (http://www.reactome.org/download/current/homo_sapiens.interactions.txt.gz).

A total of 52 clusters were discovered by mapping the Van Hoof dataset *p*-values to the Reactome version 43 human protein-protein interaction (PPI) network. The clusters described in Table S1 were all significant at a level of 0.01 in the NetWeAvers network analysis. Ingenuity (Ingenuity® Systems, www.ingenuity.com) and Reactome (Croft *et al.*, 2011) were used to annotate the networks. The main function of the most significant cluster, which contains 30 proteins with *p*-values less than 0.10 (Figure S2), is transcription regulation, an essential part of embryonic stem cell (ESC) differentiation (Heng and Ng, 2010). The other significant clusters (clusters c32, c52 and c34, Figures S3-S5) also are involved in processes or pathways known to be involved in ESC differentiation, such as RNA splicing, post-translational modification, and DNA repair (Pritsker *et al.*, 2005; Cai *et al*., 2012; Maynard *et al*., 2008). Cytoscape (Shannon *et al*., 2003) was used to visualize the clusters which are shown in Figures S2-S5. The overlap between the results provided here and the results in Van Hoof *et al.* is smaller than might be expected due to the use of different databases, one a signaling pathway database and one a PPI network.

Table S2. Significant clusters from Van Hoof dataset.

Cluster_ID	Cluster_size	Cluster_score	Cluster_pvalue	Network Function
c5	190	1.412	0.000	Transcription regulation
c32	30	2.476	0.000	RNA splicing
c52	6	5.245	0.000	Post-translational modification
c34	8	2.583	0.010	DNA repair (non-homologous end-joining)

Legend for Figures S2-S5.

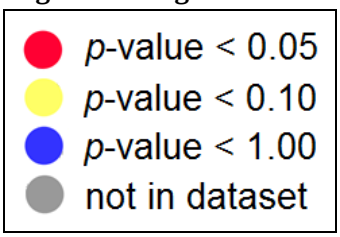


Figure S2. Cluster c5 – transcription regulation.

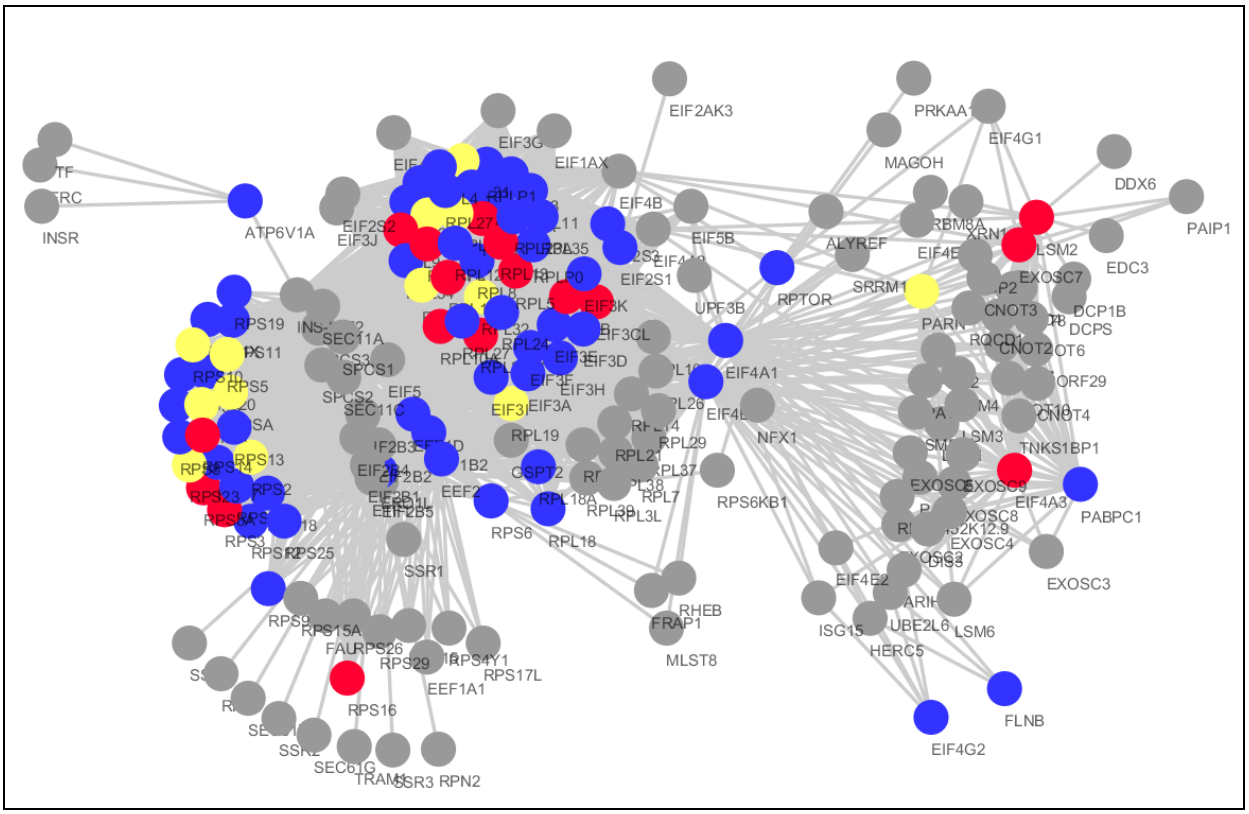


Figure S3. Cluster c32 – RNA splicing.

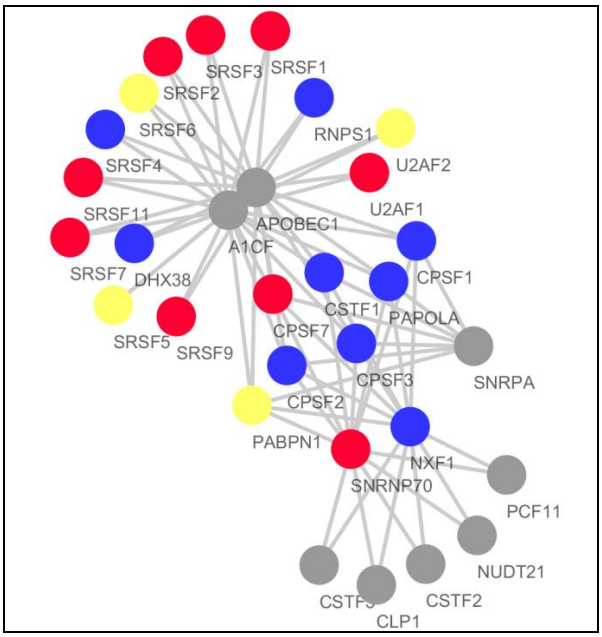


Figure S4. Cluster c52 – post-translational modification.

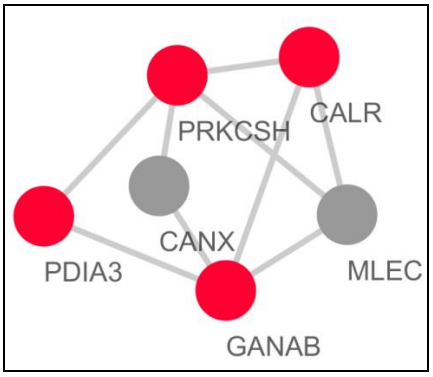
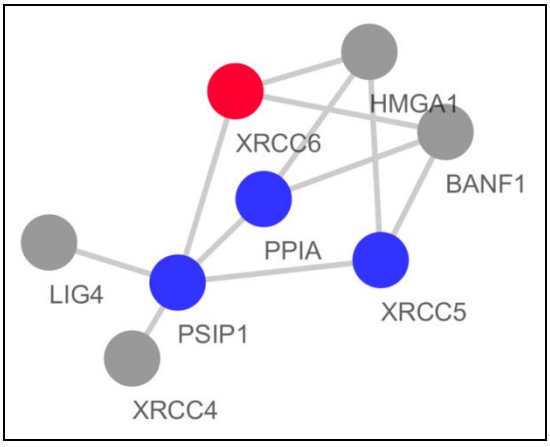


Figure S5. Cluster c34 – DNA repair.



5. Using NetWeAvers with Other R Packages

In order to display the flexibility of NetWeAvers, we present here an example of using alternative summarization and testing procedures, from another R package, prior to performing the NetWeAvers aglorithm. To find statistically significant dense clusters of proteins we ran NetWeAvers with default parameters on the protein ratio *p*-values from the R package isobar's vignette (Breitwieser *et al*., 2011; http://www.ms-isobar.org/isobar.pdf) and the Reactome human PPI version 43. Specifically, we performed each step in the pre-processing and testing from the vignette on the dataset `ibspiked_set1`, which is isobaric tag for relative and absolute quantitation (iTRAQ) data from human plasma with spiked-in proteins. From the object `rat.list`, which is comprised of statistics for summarized proteins, we extracted the variable `p.value.rat`, which contains the *p*-values for the ratios of isobaric tags for each protein. These protein *p*-values were input into NetWeAvers. The filtered human network did not include the spike-ins from the other species, but several clusters were significant, including those related to the complement system, platelet aggregation, cholesterol and lipoproteins, and glycolysis.

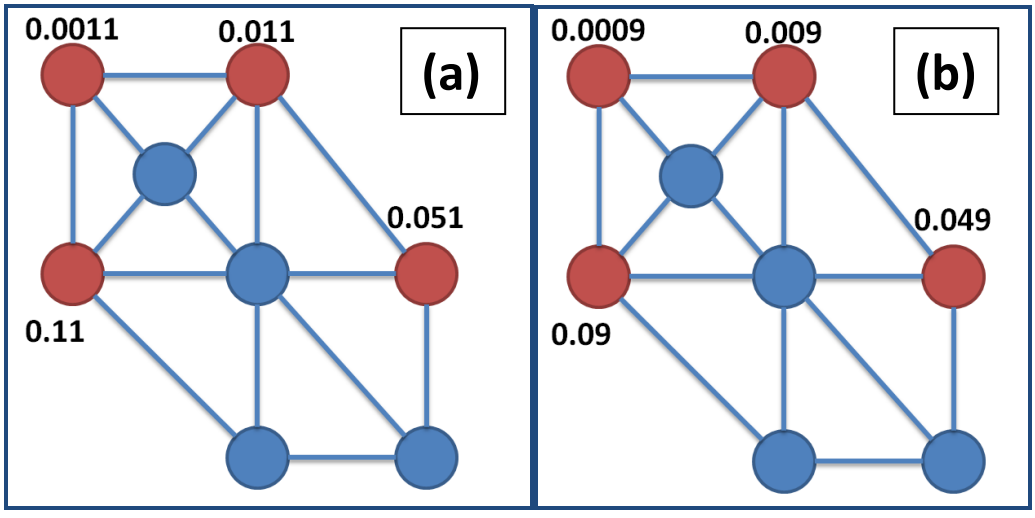
6. Null Dataset

We tested the NetWeAvers algorithm (default parameter settings) on a null data set by comparing 2 samples with themselves and shuffling the protein-protein interactions. Of the 84 dense clusters discovered, 1% had a p-value less than 0.01, 9% had a p-value less than 0.05 and 18% had a p-value less than 0.10. These values are roughly consistent with a random p-value distribution.

7. Rationale for Threshold-Free *p*-Values

Suppose we want to score the networks in Figure S6 where the *p*-values are either indicated in black (adjacent to red nodes) or are 1 (blue nodes). The networks in (a) and (b) are nearly identical, but there are slight differences in the *p*-values around typical thresholds (0.001, 0.01, 0.05, 0.10).

Figure S6. Toy example of networks containing nodes with varying *p*-values.



The networks can be scored on actual *p*-values or by using a binary classification of nodes (significant or not, based on a chosen threshold). For example, for a threshold of 0.001 the node is scored as 0 if the *p*-value is less than 0.001 and 1 if greater than 0.001. Here we use the mean of the *p*-values to demonstrate the impact of thresholding. A possibly interesting network with differentially expressed nodes will have a mean closer to 0 and closer to 1 otherwise. Table S3 shows that as the threshold increases the mean decreases, but that these values are quite different between the two networks (a) and (b). The averages without thresholding are relatively close, which seems more reasonable for such similar *p*-values in the given networks. In this latter case no subjective cutoff is required and there is less loss of information as compared to the use of dichotomized *p*-values.

Table S3. Average *p*-values (with and without thresholds) of nodes in networks from Figure S2.

Threshold	(a) Mean	(b) Mean
0.001	1	0.875
0.01	0.875	0.750
0.05	0.750	0.625
0.10	0.625	0.500
None	0.522	0.519

References

Beisser D, Klau GW, Dandekar T, Muller T, and Dittrich M. (2010). BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics* **26(8)**:1129-1130.

Chiang T and Scholtens D with contributions from Huber W and Wang L. (2013). ppiStats: Protein-Protein Interaction Statistical Package. R package version 1.25.0.

Jacob L, Neuvial P, and Dudoit S. (2010). Gains in Power from Structured Two-Sample Tests of Means on Graphs. arXiv techreport http://arxiv.org/abs/1009.5173.

Ideker T, Ozier O, Schwikowski B, and Siegel AF. (2002). Discovering regulatory and signaling circuits in molecular interaction networks. *Bioinformatics* **18(suppl 1)**:S223-S240.

Maere S, Heymans K, and Kuiper M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21(16)**:3448-3449.

Bader GD and Hogue CWV. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4(2)**.

Antonov AV, Schmidt EE, Dietmann S, Krestyaninova M and Hermjakob H. (2010). R spider: a network-based analysis of gene lists by combining signaling and metabolic pathways from Reactome and KEGG databases. *Nucleic Acids Res.* **38(suppl 2)**:W78-W83.

Ding Y, Chen M, Liu Z, Ding D, Ye Y, Zhang M, Kelly R, Guo L, Su Z, Harris SC, Qian F, Ge W, Fang H, Xu X, and Tong W. (2012). atBioNet- an integrated network analysis tool for genomics and biomarker discovery. *BMC Genomics* **13(325)**.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13(11)**:2498-2504.

Van Hoof D, Munoz J, Braam SR, Pinkse MW, Linding R, Heck AJ, Mummery CL, and Krijgsveld J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell 5(2):214-226.

Khan Z, Bloom JS, Garcia BA, Singh M, and Kruglyak L. (2009). Protein Quantification Across Hundreds of Experimental Conditions. *Proc. Natl. Acad. Sci.* **06(37)**:15544-15548.

Khan Z, Amini S, Bloom JS, Ruse C, Caudy AA, Kruglyak L, Singh M, Perlman DH, and Tavazoie S. (2011). Accurate proteome-wide protein quantification from high-resolution 15N mass spectra. *Genome Biology* **12**:R122.

Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, and Stein L. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39(suppl 1)**:D691-D697.

Heng JC and Ng HH. (2010). Transcriptional regulation in embryonic stem cells. *Adv Exp Med Biol.* **695**:76-91.

Pritsker M, Doniger TT, Kramer LC, Westcot SE, and Lemischka IR. (2005). Diversification of stem cell molecular repertoire by alternative splicing. *Proc. Natl. Acad. Sci.* **102(40)**:14290-14295.

Cai N, Li M, Qu J, Liu GH, and Izpisua Belmonte JC. (2012). Post-translational modulation of pluripotency. *J Mol Cell Biol.* **4(4)**:262-265.

Maynard S, Swistowska AM, Lee JW, Liu Y, Liu ST, Da Cruz AB, Rao M, de Souza-Pinto NC, Zeng X, and Bohr VA. (2008). Human embryonic stem cells have enhanced repair of multiple forms of DNA damage. *Stem Cells* **26(9)**:2266-74.

Breitwieser F, Muller A, Dayon L, Kocher T, Hainard A, Pichler P, Schmidt-Erfurth U, Superti-Furga G, Sanchez JC, Mechtler K, Bennett KL, and Colinge J. (2011). General Statistical Modeling of Data from Protein Relative Expression Isobaric Tags. *J. Proteome Res.* **10**:2758-2766.