Analyzing marginal cases in differential shotgun proteomics

Paulo C Carvalho[1,2*], Juliana S G Fischer[1], Jonas E A Perales[2], John R Yates III[3], Valmir C Barbosa[4], Elias Bareinboim[5]

[1] Center for Technological Development in Health of the Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

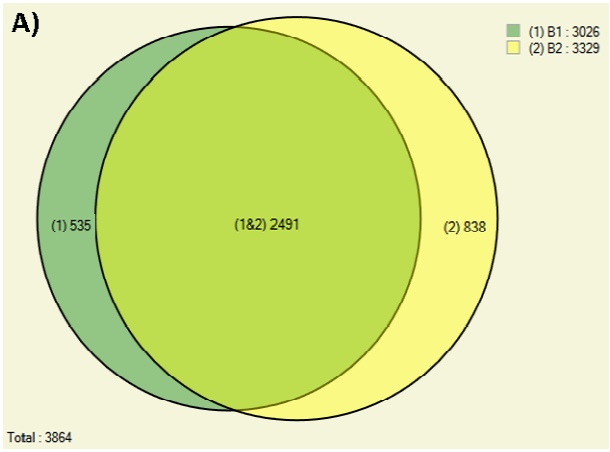
[2] Laboratory of Toxinology, Oswaldo Cruz Institute, Rio de Janeiro, Brazil

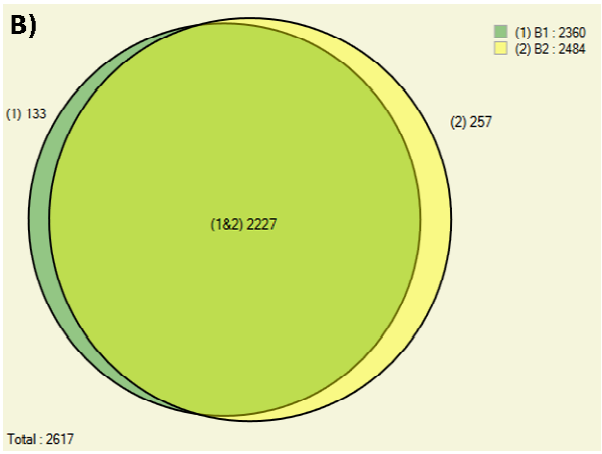
[3] Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California, USA

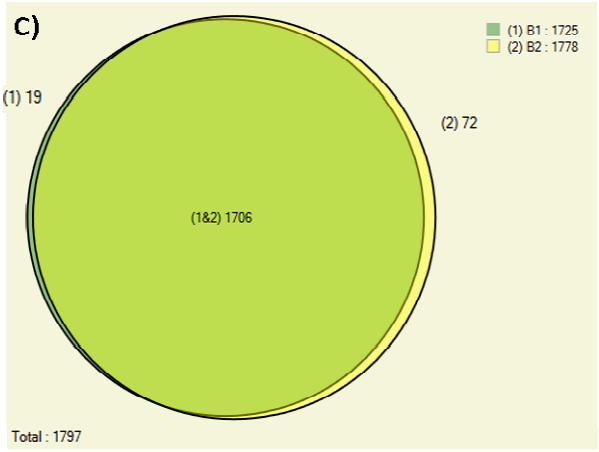
[4] Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

[5] Computer Science Department, University of California, Los Angeles, USA

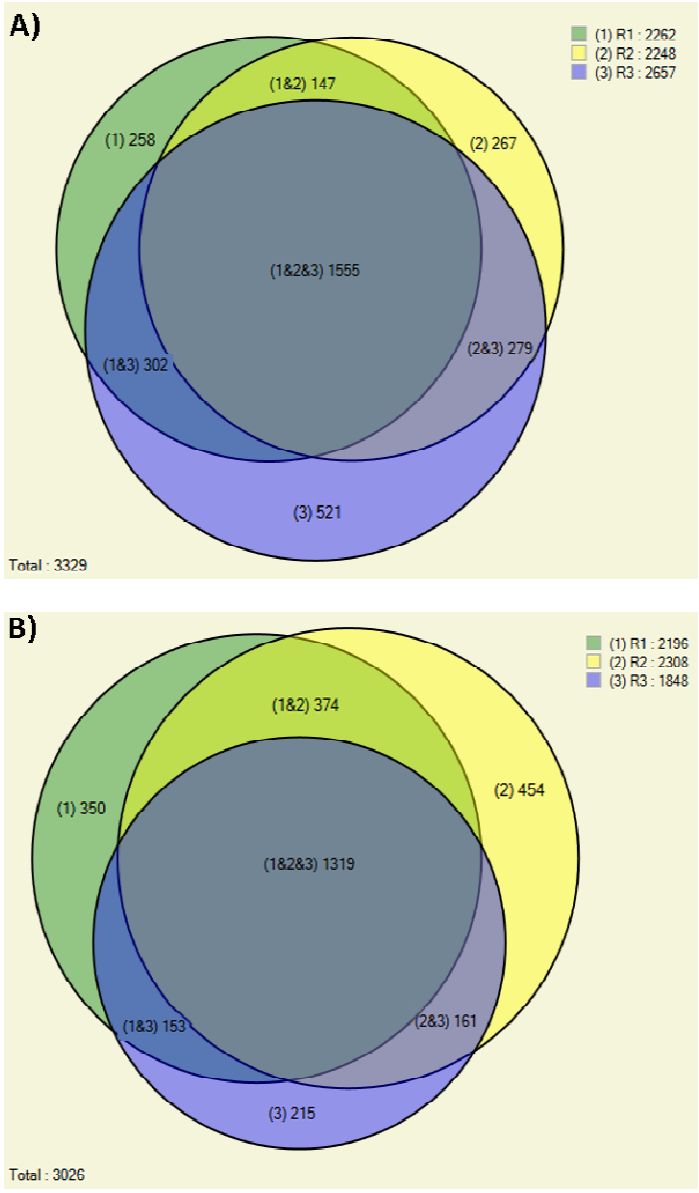
Supplementary figures and tables are given below.



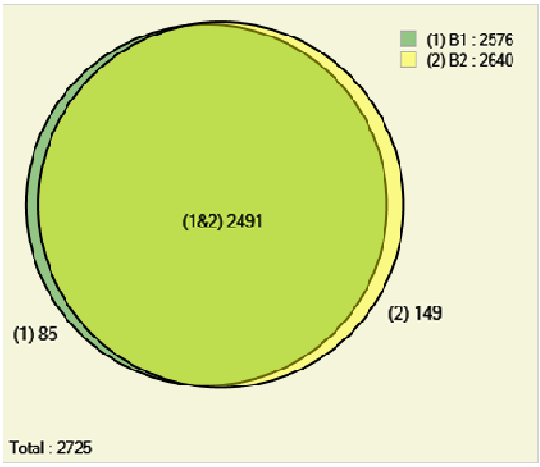




Supplementary Figure 1 – Venn Diagrams of the proteins identified by shotgun proteomics from a cell lysate in biological states B1 and B2. Panels A, B, and C consider only proteins that appeared in one or more, two or more, or in all three replicates, respectively.



Supplementary Figure 2 - Venn Diagrams of the proteins identified by shotgun proteomics from a cell lysate in biological states B1 (A) and B2 (B). R1, R2, and R3 refer to the replicates from each state.



Supplementary Figure 3 - Venn Diagram of the proteins identified by shotgun proteomics from a cell lysate in biological states B1 and B2. Proteins that could not be statistically claimed to be differentially expressed in one of the two states according to the proposed Bayesian approach (those for which *p*-value ≤ 0.05) were automatically filtered out during the generation of the Venn Diagram.

Supplementary Table I – Statistical analysis for differential expressivity in state B1 relative to state B2. Data in the table are organized according to the groups Low, Medium, High, and Very High, relative to the average signal $\bar{s}$ of each protein over the replicates in which it appears. A protein appearing in *t* replicates is counted in the corresponding row of column Num. Proteins within its group. Dividing the resulting number by the column's total for the group yields f_t. The average signal for the experiment at hand was obtained by averaging the spectral counts of a given protein. However, other quantitation strategies can be adopted as, for example, integrating extracted ion chromatograms with a label-free quantitation program such as Census (Park *et al*., 2008). PatternLab is ready to handle both label-free and label-based relative quantitation.

Low ($2 \leq \bar{s} < 2.5$)			
Num. Rep. (*t*)	Num. Proteins	Fraction (f_t)	*p*-value
1	613	0.637	0.180
2	283	0.294	0.056
3	66	0.069	0.019
Medium ($2.5 \leq \bar{s} < 4$)			
1	297	0.310	0.141
2	417	0.435	0.042
3	245	0.255	0.015
High ($4 \leq \bar{s} < 9$)			
1	168	0.176	0.112
2	185	0.193	0.033
3	604	0.631	0.011
Very High ($9 \leq \bar{s} < 896.3$)			
1	59	0.070	0.083
2	62	0.073	0.024
3	725	0.857	0.008

Supplementary Table II – Statistical analysis for differential expressivity in state B2 relative to state B1.

Low ($2 \leq \bar{s} < 2.5$)			
Num. Rep. (t)	Num. Proteins	Fraction (f_t)	p-value
1	533	0.691	0.186
2	196	0.254	0.058
3	42	0.054	0.020
Medium ($2.5 \leq \bar{s} < 4$)			
1	308	0.355	0.147
2	364	0.420	0.044
3	195	0.225	0.015
High ($4 \leq \bar{s} < 9.3$)			
1	200	0.21	0.120
2	236	0.248	0.036
3	515	0.542	0.012
Very High ($9.3 \leq \bar{s} < 611.7$)			
1	67	0.088	0.088
2	68	0.089	0.025
3	630	0.824	0.009

Reference

Park,S.K., Venable,J.D., Xu,T., and Yates,J.R., III 2008. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat. Methods* **5**:319-322.