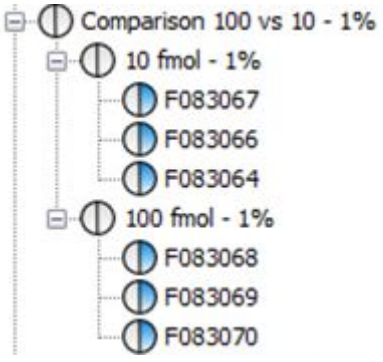
PROLINE TUTORIAL

QUANTIFY BY SPECTRAL COUNTING

I/ START PROLINE

A. EXPERIMENTAL DESIGN

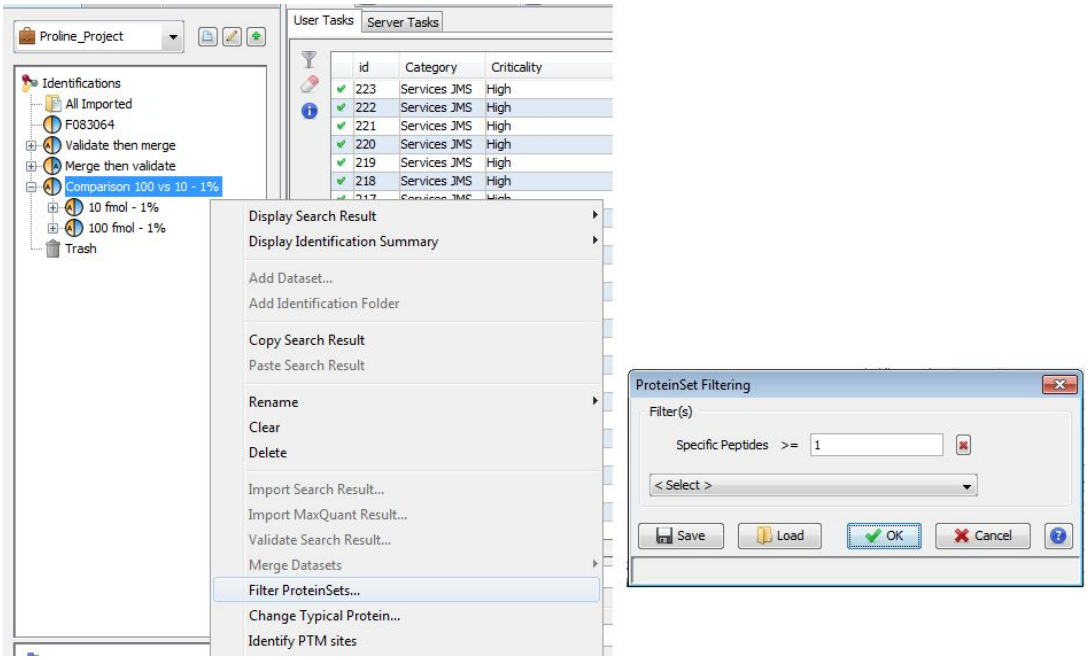
In this tutorial, all sample mascot files available on Proline website will be used. This dataset contains the MS analysis of two samples of 2 µg yeast cell lysate spiked respectively with 100fmol and 10fmol of UPS1. Samples were analyzed in triplicate by nanoLC–MS/MS on an LTQ-Orbitrap Velos mass spectrometer. For more information on samples preparation and LC-MS/MS analyses, please refer to Ramus et al., J Proteomics. 2016 Jan 30;132:51-62. doi: 10.1016/j.jprot.2015.11.011.

Action	Reproduce the following experimental design ● Create all datasets ○ *Add Dataset* Comparaison 100 vs 10 1% ○ *Add Dataset* 10 fmol - 1% ○ *Add Dataset* 100 fmol - 1% ● Import of the missing Mascot files (see Proline Basics tutorial) ● If necessary, rename the files according to the Search Result Name to reproduce the experimental design below: 

Note	*Files and datasets can be renamed manually. Files can also be renamed automatically by retrieving the Search Result name for example*

Comparaison 100 vs 10 1% is a merge by aggregation of identification summaries created at the intermediate level: **10 fmol - 1%** and **100 fmol - 1%.** These both datasets are also merged by aggregation of identification summaries (from imported result files). Validation is done using the following parameters : 1% PSM (based on score) and rank = 1. The top level dataset ("Comparison 100 vs 10 - 1%) has been filtered to retain only protein sets with at least 1 specific peptide.

Action	• Validate the six search results using the described parameters. Note: Validation can be done on multiple search results • Merge the resulting identification summary at intermediate and top level of the dataset hierarchy. • Filter the top level identification summary to retain only protein sets with at least 1 specific peptide.

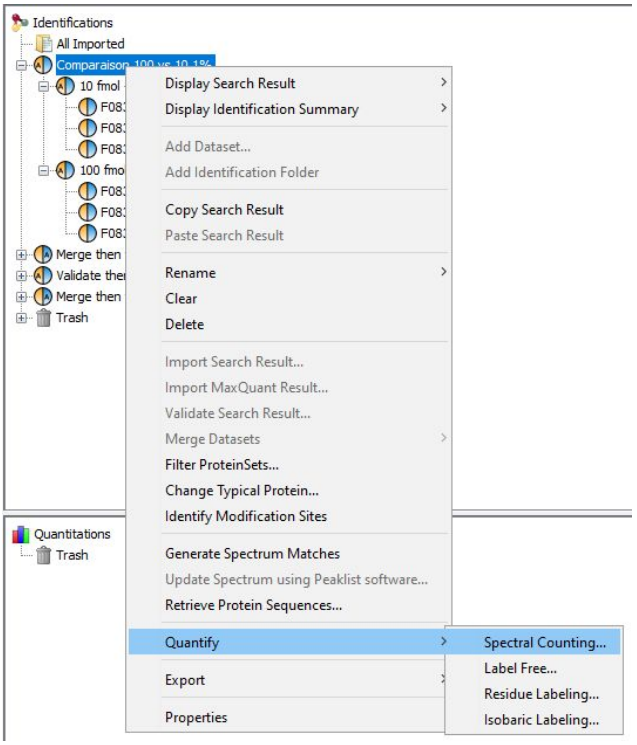


II/ Compare samples by Spectral Counting

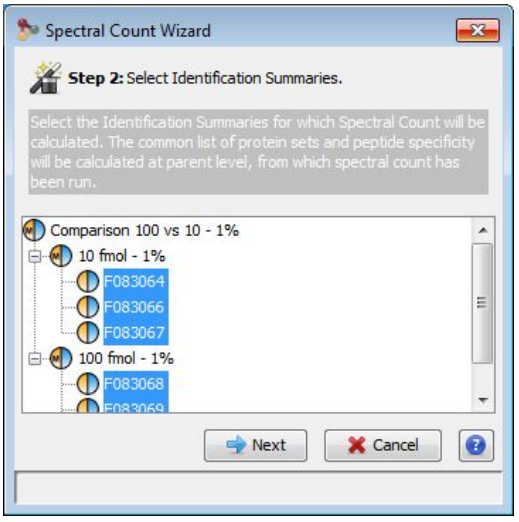
A. Run spectral counting

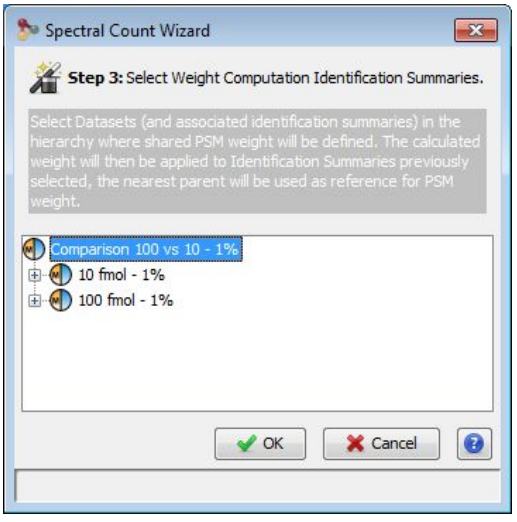
To run the SC comparison, you should have a merged dataset containing child datasets (which may be also merged datasets or identification datasets). Actually, only identification summaries merge could be used to execute SC. (SC is the abbreviation of Spectral Counting)

To execute SC comparison, right-click on merged dataset and select *Quantify* > *Spectral Counting...*



The opened dialog allows you to specify a name and a description for the comparison. On the second dialog box, Step 2, select the dataset on which you would like to perform the Spectral Count and finally choose the dataset where shared peptides spectral count weights will be calculated.



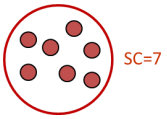


Action	Run Spectral Counting using ● "SC Compare 1%" as name ● compute the SC value for each of the six datasets ● top level dataset as dataset where sharing peptides is considered

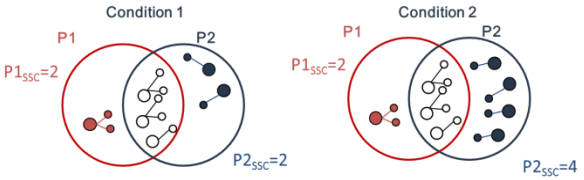
B. Computed values

The algorithm implemented in Proline compute three different spectral count values :

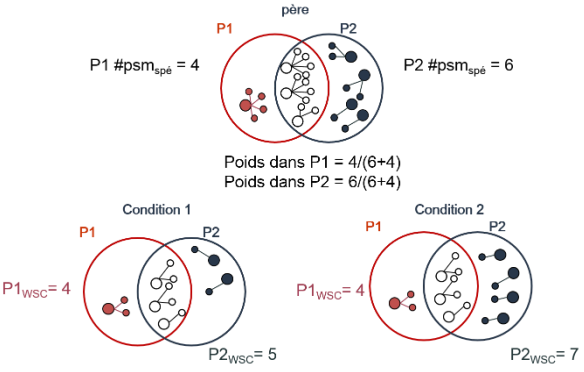
1/ **Basic SC** : is the total number of MS/MS validated peptide spectrum match (PSM) of all peptides matching the considered protein set.



2/ **Specific SC** : Is the total number of validated PSM of peptides that are **specific** to the considered protein set. Shared peptides (peptides shared by different protein sets) are excluded from the count. The peptide specificity is calculated from the top level identification summary to ensure that the uniqueness of the protein set the peptide belongs to is not modified by an additional peptide identification from another result summary.

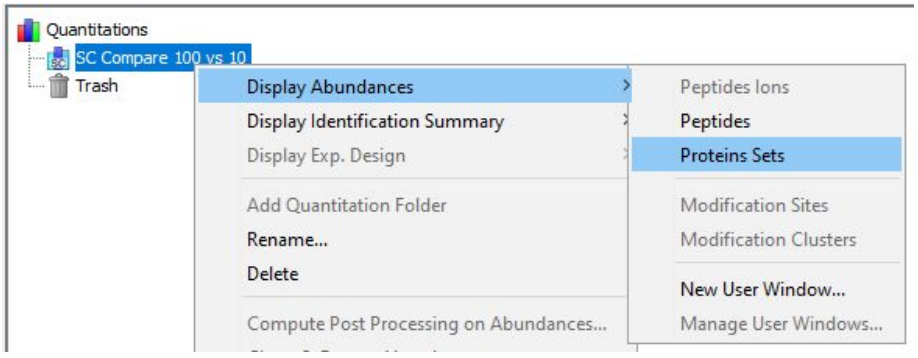


3/ **Weighted SC** : is based on all identified peptides, but taking into account that spectral count of shared peptides must be apportioned to reflect the contribution of each protein set. The weight of these respective contributions is based on the number of specific peptides of each considered protein set at the top level identification summary.

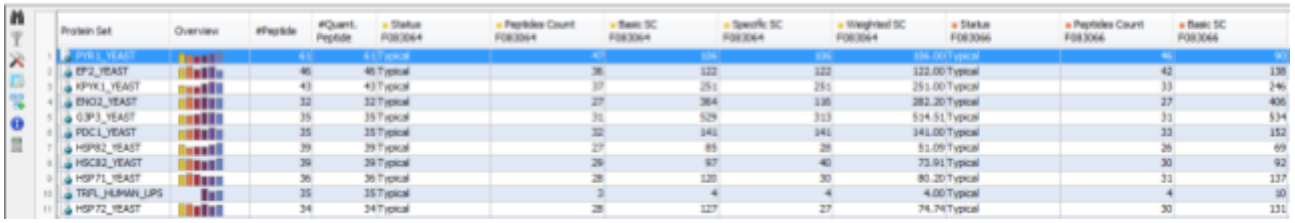


C. Result visualization

Once finished, a new dataset appears in the “Quantitations” panel (lower part of the left window). To visualize SC results, right-click on this dataset and select *Display Abundances* then *Proteins Sets*.



For each replicate, the table columns indicate:
protein status/ Peptides Count / Basic SC / Specific SC / Weighted SC



	Protein Set	Overview	#Peptide	#Quant. Peptide	Status F083064	Peptides Count F083064	Basic SC F083064	Specific SC F083064	Weighted SC F083064	Status F083066	Peptides Count F083066	Basic SC F083066
1	PYR1_YEAST		61	61	Typical	42	106	106	106.00	Typical	46	90
2	EF2_YEAST		46	46	Typical	36	122	122	122.00	Typical	42	138
3	KPYK1_YEAST		43	43	Typical	37	251	251	251.00	Typical	33	246
4	ENO2_YEAST		32	32	Typical	27	384	186	282.20	Typical	27	406
5	G3P1_YEAST		35	35	Typical	31	529	313	514.51	Typical	31	534
6	PDC1_YEAST		35	35	Typical	32	141	141	141.00	Typical	33	152
7	HSP82_YEAST		39	39	Typical	27	65	28	51.09	Typical	26	69
8	HSC82_YEAST		39	39	Typical	28	97	40	73.91	Typical	30	92
9	HSP71_YEAST		36	36	Typical	28	120	30	80.20	Typical	31	137
10	TRFL_HUMAN_UPS		35	35	Typical	3	4	4	4.00	Typical	4	10
11	HSP72_YEAST		34	34	Typical	28	127	27	74.74	Typical	30	131

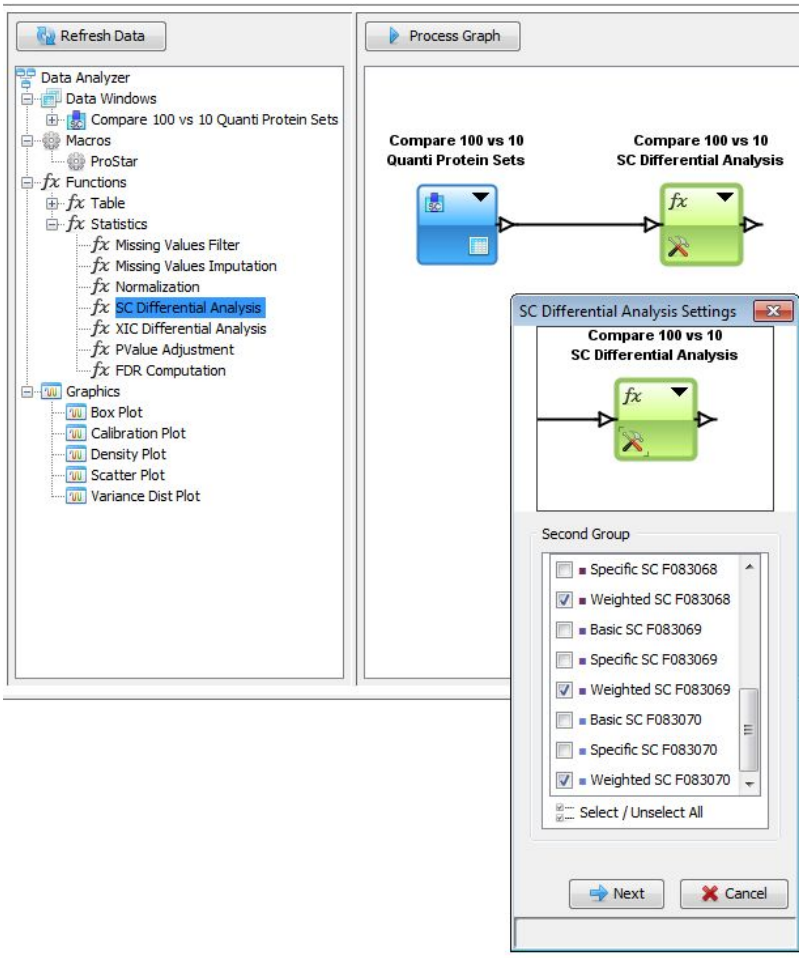
Note	*Every table in Proline can be customized by clicking on the*  icon to select visible and invisible columns.

D. Statistical analysis

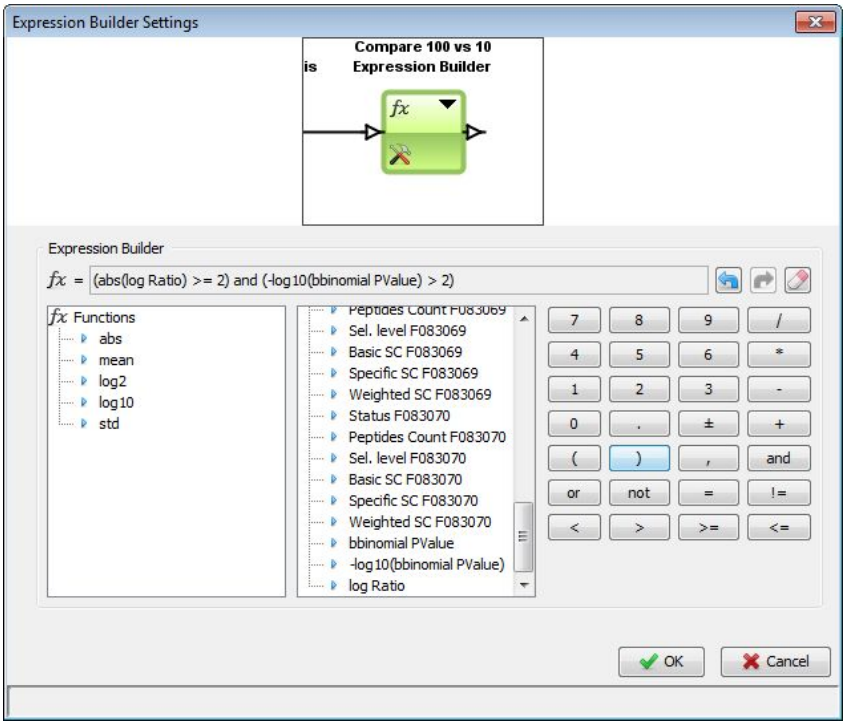
In this section, we will perform the statistical beta binomial test [1] on the **weighted spectral count** data and display the test results.

(1) Pham, T. V., Piersma, S. R., Warmoes, M., and Jimenez, C. R. (2010) On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* 26, 363–369

Action	• Open *Display Abundances* then *Proteins Sets* and click on  • The DataAnalyser window opens and a box indicating « SC Compare 1% Quanti Protein Sets » appears on the right side of the window. • Add the « SC Differential Analysis » function to the workflow from Functions > Statistics in DataAnalyser tree, (drag & drop or double click the function) and connect the two boxes. • Run the statistical function, select the columns belonging to each group to be compared



Action	• To easily identify proteins of interest, add a column to the table with the « Expression Builder » (from Functions > Table in DataAnalyser tree). Connect the output of SC Differential Analysis box to the entry of the Expression Bulder box. : Expression of the new column: proteins with log ratio <= -2 or >= 2 and pvalue <0.01 (use abs() function as shown in the screenshot below)



Action	• Visualize results as a Volcano plot: add a graphical view(using icon) to the newly computed table, select scatter plot and choose log Ratio as x axis and -log10(bbinomial pvalue) as y axis. • In the table, select rows (protein sets) with a non null value in the column that have been added with the « Expression Builder »(use filter button) and visualize those protein sets in the scatter plot.

Note	*In Proline, selection can be "transferred" from a view to another view by using the icon. In the plot, right click on the selected points and create a group containing these points.*

Bravo ! 49 proteins sets have been identified as differentially expressed by the spectral counting approach and the beta binomial statistical test. Among these 49 proteins, 43 proteins out of the 48 UPS1.


Graphic : Scatter Plot
X Axis: log Ratio
Y Axis: -log10(bbinomial PValue)
Differential proteins
-log10(bbinomial PValue)
log Ratio