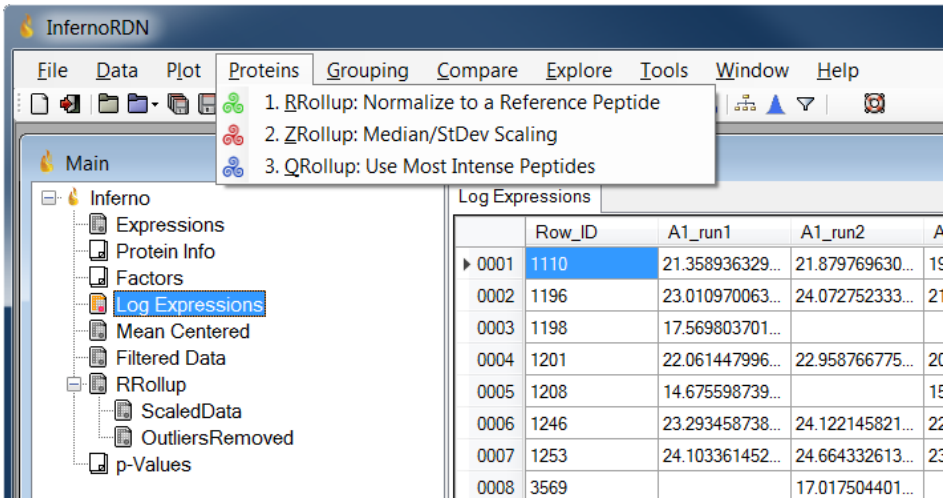
Rollup Methods



InfernoRDN has three methods for estimating protein abundances based on observed peptides, aka protein rollup.

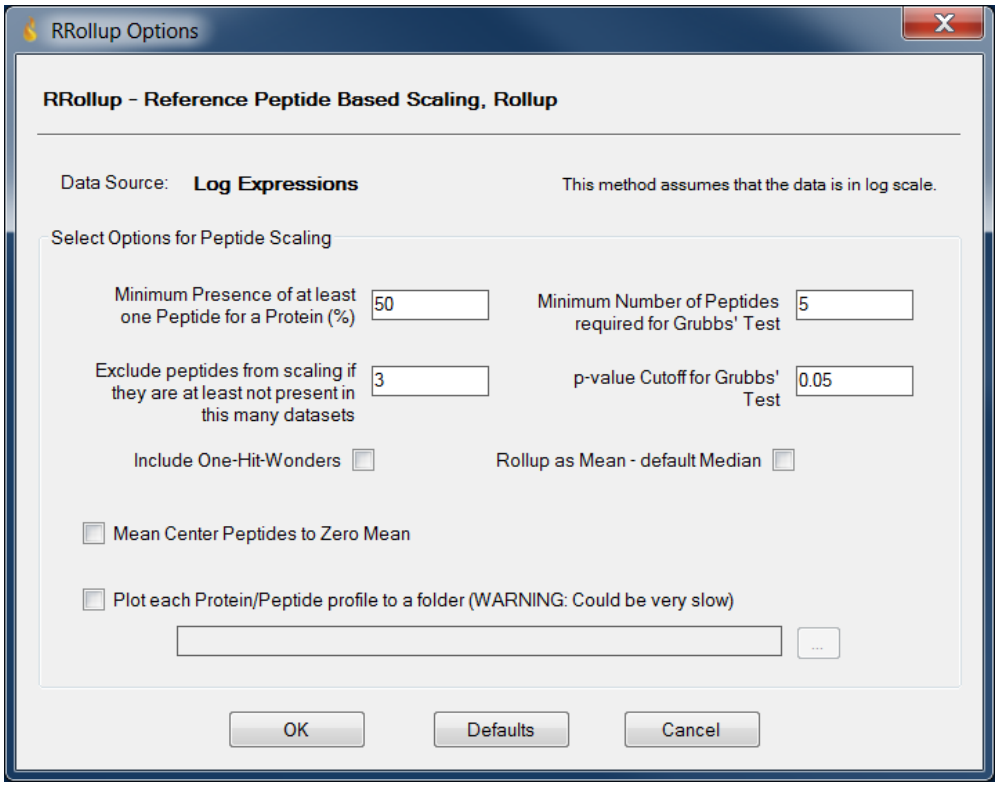
- RRollup: Reference peptide based scaling
  - Identifies a reference peptide, then computes peptide ratios relative to the reference
- ZRollup: Median / standard deviation based scaling
  - Peptides are median centered first and then scaled by the row standard deviation. Protein abundance is obtained as the median of the abundances of the peptides in the group.
- QRollup
  - Takes the top user selected percentage of peptides for each protein then computes the average (or median). In other words, compute a trimmed mean (or trimmed median).

The three methods are described in detail on the following pages.

RRollup - Reference Peptide Based Scaling

Note that this method correctly works only with log transformed data.

A reference peptide which has the most presence across all the datasets, is chosen from the group of peptides that belong to a protein. If there are multiple candidates, the most abundant one is chosen. Then the ratios of peptide abundances with respect to the reference are computed (since the data is assumed to be in log scale, the differences are used) and their median is used as a scaling factor. Protein abundance is obtained as the median of the resulting peptide abundances.

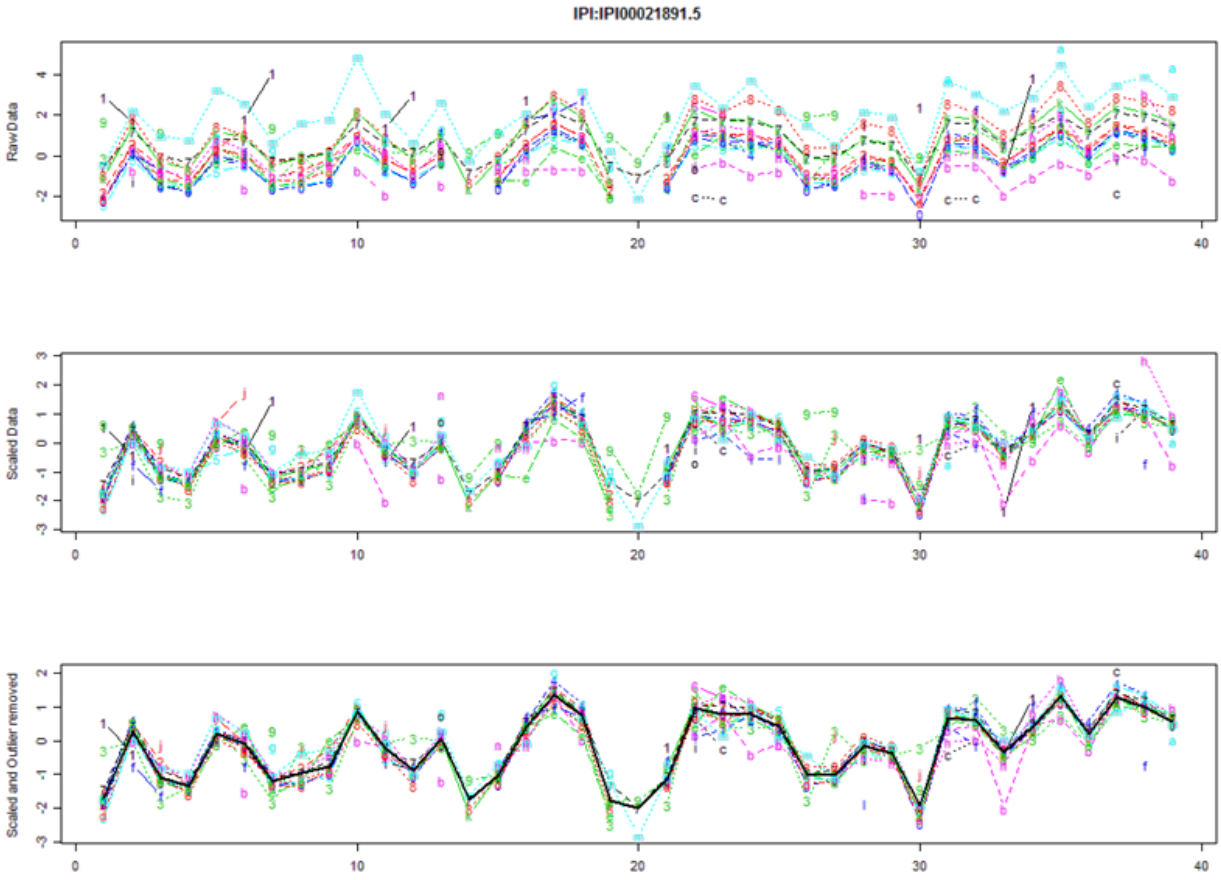


Minimum Presence of at least one Peptide for a Protein: Peptides that have too many missing values below this percentage are dropped.

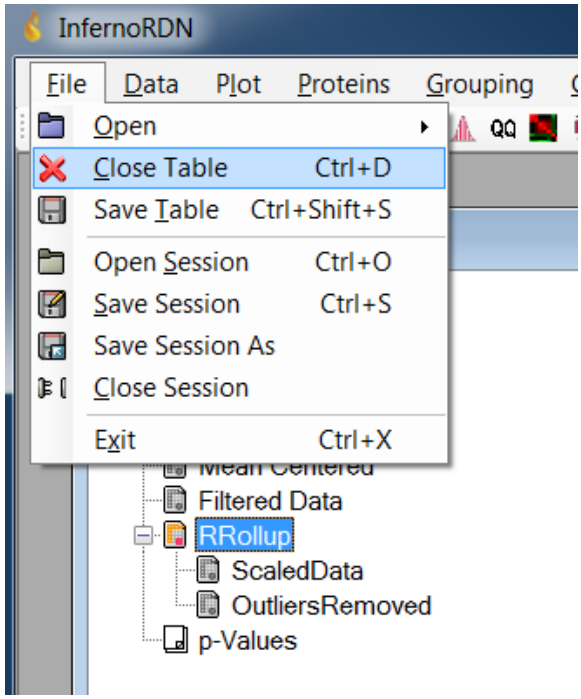
Exclude peptides from scaling if they are at least not present in this many datasets: Within a group of peptides for a specific protein, the ones that do not overlap well (controlled by this value) are not scaled but they are kept to calculate the final protein abundance.

Include Single peptide/protein matches (i.e. 'One-Hit-Wonders'): Protein with only one observed peptide will be included in the final list of proteins. The rationale behind this is that if a particular protein may have only one peptide but it may be quite abundant and present throughout giving some strong confidence on the presence of the protein.

If the plotting is enabled using the checkbox, the scaling results will be plotted for each peptide group as follows:

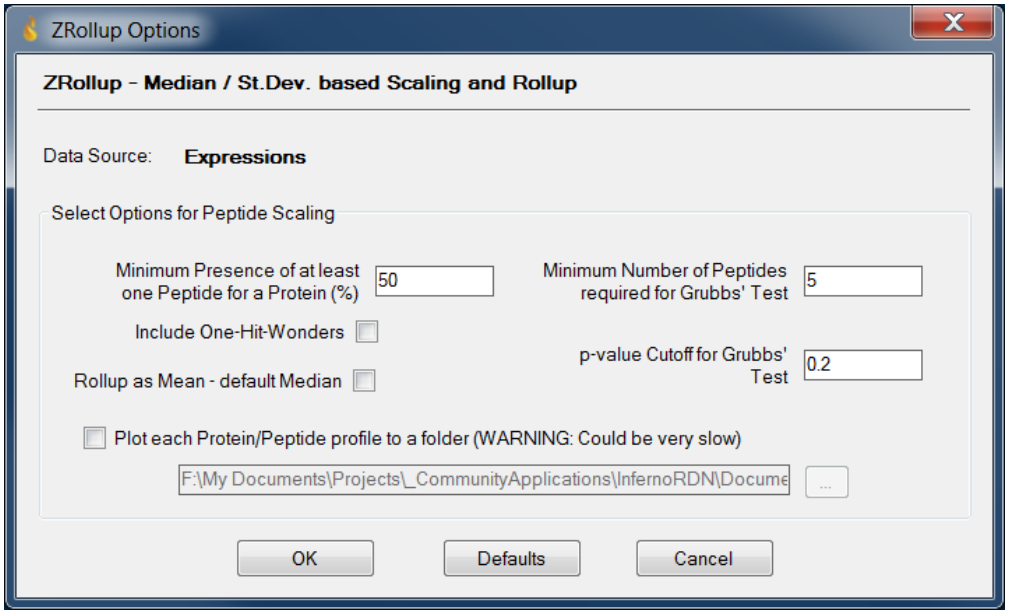


To redo RRollup, you must first delete the RRollup table. Select RRollup in the data table list, then choose File, Close Table. Alternatively, use Ctrl+D



ZRollup - Median / Standard Deviation Based Scaling

Peptides are median centered first and then scaled by the row standard deviation. Protein abundance is obtained as the median of the abundances of the peptides in the group.

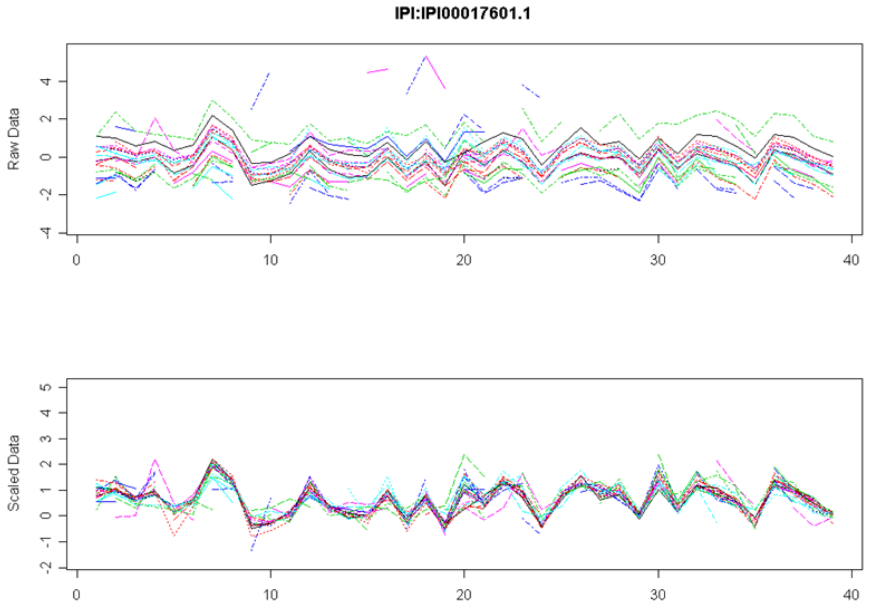


Minimum Number of Peptides required for Grubb's Test: Grubb's test for outliers is performed for peptide groups in each dataset. This value controls the minimum number required to perform the test.

p-value Cutoff for Grubb's Test: This controls the cutoff value for removing outliers determined using Grubb's test.

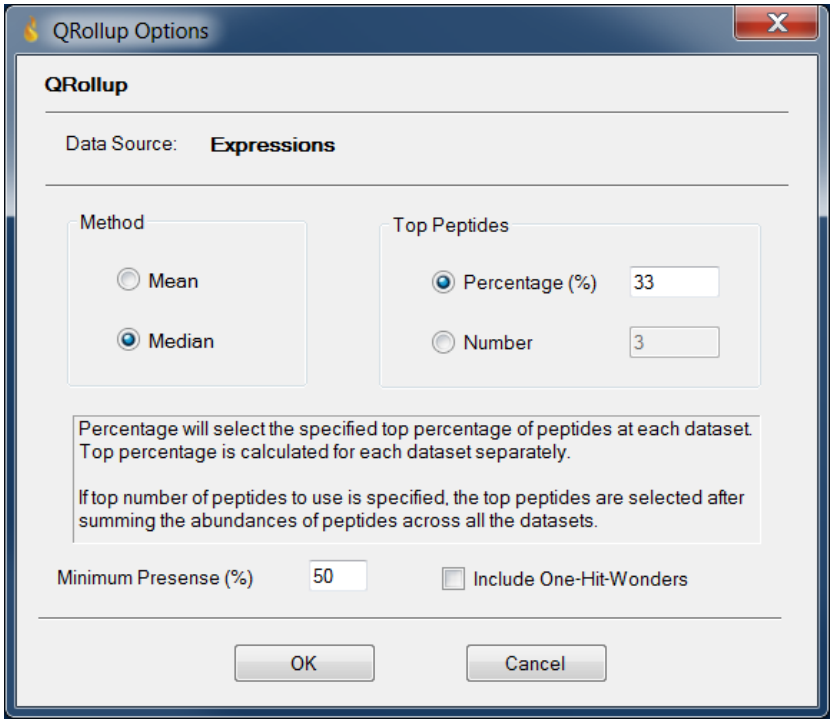
Include Single peptide/protein matches (i.e. 'One-Hit-Wonders'): Same as in the first method.

If the plotting is enabled using the checkbox, the scaling results will be plotted for each peptide group:



QRollup method

QRollup method simply takes the top user selected percentage of peptides and averages (either mean or median) to obtain the protein abundance. The parameters to control are:



Threshold: All peptides that have abundance values above this percentage at each dataset are used for protein abundance calculation.

Minimum Presence: Minimum presence of the most present peptide in a peptide group.

Method: Method to calculate Protein abundance from peptides. Either mean or median of the peptides in the group.

Include Single peptide/protein matches (i.e. 'One-Hit-Wonders'): Same as in first method.