Data-tab (uploading data)

As input, proteiNorm expects tab-separated peptide (optional) and protein data (not on logarithmic scale) as produced by software such as MaxQuant, where each row represents a peptide or protein and the column names of the measured intensities (samples) beginning with “Reporter intensity corrected” followed by an integer and an optional label (e.g. “Reporter intensity corrected 5 TMT2”) for TMT experiments. The column names for samples in label free experiments should begin with “Intensity” followed by an integer (“Intensity 01”). In addition, for peptide file, proteiNorm expects the following columns: “id”, “Protein group IDs”, “Leading razor protein”, “Gene names”, “Reverse”, “Potential contaminant” (see Table 1 for an example). And for protein file, proteiNorm expects the following column: “id”, (optional: “Reverse”, “Potential contaminant”, “Only identified by site”) (see Table 2 for an example)

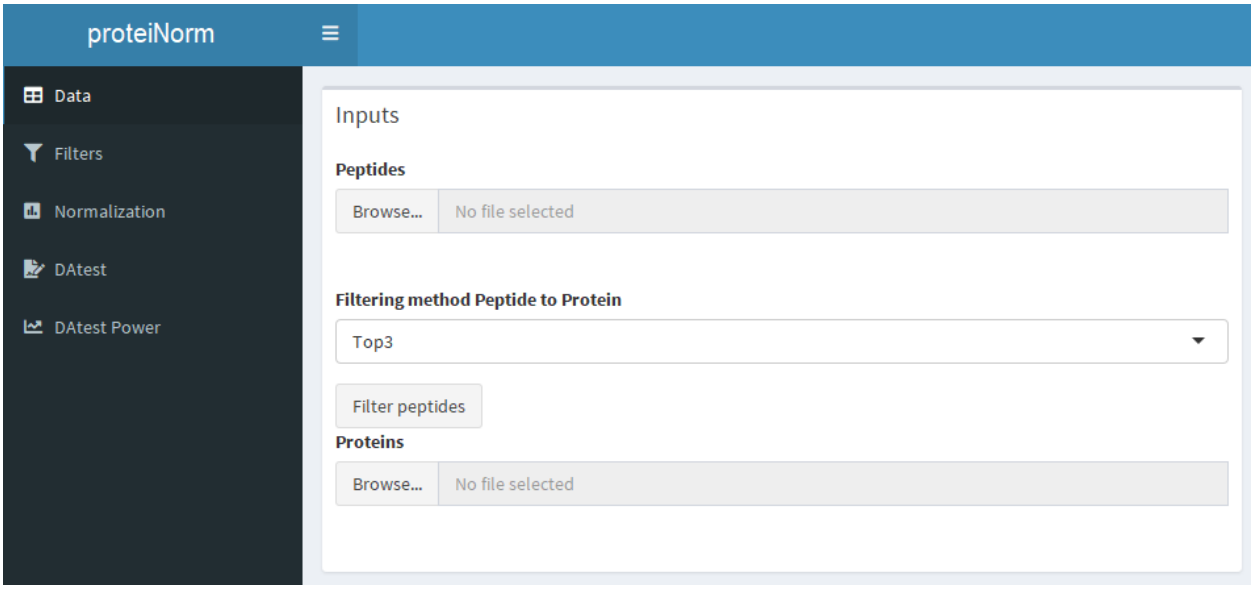

Figure 1 Data-tab. Peptide and/or protein files (protein required) can be uploaded. Peptide file can be filtered using “Top3” method to create a filtered protein file.

Uploading a peptide-file will provide option to filter peptides using the “Top3” methods and export a filtered protein-file (using the “Filter peptide” button) (see Figure 1).

Table 1 Example of "peptide.txt" file

Leading razor protein	Gene names	Reporter intensity corrected 1 TMT1	Reporter intensity corrected 2 TMT1	Reporter intensity corrected 1 TMT2	Reporter intensity corrected 2 TMT2	Reverse	Potential contaminant	id	Protein group IDs
G3UY42	Pabpn1	10441	0	10354	16336			0	1302
A2ARS0	Ankrd63	221690	93505	0	0			1	665
P63085	Mapk1;Erk2	0	0	139070	69007			2	1872
D3Z3G6	Mapk3	0	0	0	0			3	1020
Q3TLR3	Trim32	0	0	15950	0		+	4	2128
Q8BKC5	Ipo5	0	0	0	0			5	2689
Q9CX34	Sugt1	0	0	0	7502.7			6	3017
D3Z4J5	Get4	725980	487850	0	0			7	1025
E9PU87	Sik3	0	0	0	0			8	1084
B1AX98	Lrrc47	5738.5	0	0	0	+		9	776

Table 2 Example of "proteinGroup.txt" file

id	Reporter intensity corrected 1 TMT1	Reporter intensity corrected 2 TMT1	Reporter intensity corrected 1 TMT2	Reporter intensity corrected 2 TMT2	Reporter intensity corrected 3 TMT2
G3UY42	141855.5	274230	70327	74433	69362
A2ARS0	1534300	688728.3	1018670	848550	934076.7
P63085	7850967	6062200	14272167	15162567	13325467
D3Z3G6	1325385	904430	3375350	4139050	3376250
Q3TLR3	17013	6679.7	235733.3	399395	391580
Q8BKC5	770910	729753.3	631380	488136.7	647203.3
Q9CX34	NA	NA	325425	187800.9	210552.7
D3Z4J5	725980	487850	NA	NA	NA
E9PU87	346169	217685.7	212900	212770	214720
B1AX98	583256.7	607100	289972	348293.3	301390

After uploading a protein-file, meta information for the recognized intensity columns can be specified. This include groups, batches and optional custom sample names (must be unique). See Figure 2 for an example.

	Protein.Sample.Names	Custom.Sample.Names	Group	Batch
1	Reporter.intensity.corrected.1.TMT1	NA_T_1	NA_T	1
2	Reporter.intensity.corrected.2.TMT1	NA_T_2	NA_T	1
3	Reporter.intensity.corrected.3.TMT1	NA_T_3	NA_T	1
4	Reporter.intensity.corrected.4.TMT1	NA_1	NA	1
5	Reporter.intensity.corrected.5.TMT1	NA_2	NA	1
6	Reporter.intensity.corrected.6.TMT1	NA_3	NA	1
7	Reporter.intensity.corrected.1.TMT2	S_T_1	S_T	2
8	Reporter.intensity.corrected.2.TMT2	S_T_2	S_T	2
9	Reporter.intensity.corrected.3.TMT2	S_T_3	S_T	2
10	Reporter.intensity.corrected.4.TMT2	S_1	S	2
11	Reporter.intensity.corrected.5.TMT2	S_2	S	2
12	Reporter.intensity.corrected.6.TMT2	S_3	S	2

Figure 2 Meta information example

Filter-tab

The filter tab is designed to identify outlier samples and exclude unwanted samples or samples with poor quality. Therefore, the user can evaluate the distribution of intensities of each sample from the peptide and protein data (Figure 3) and the PCA plots, which can be color-coded by group or batch (Figure 4).

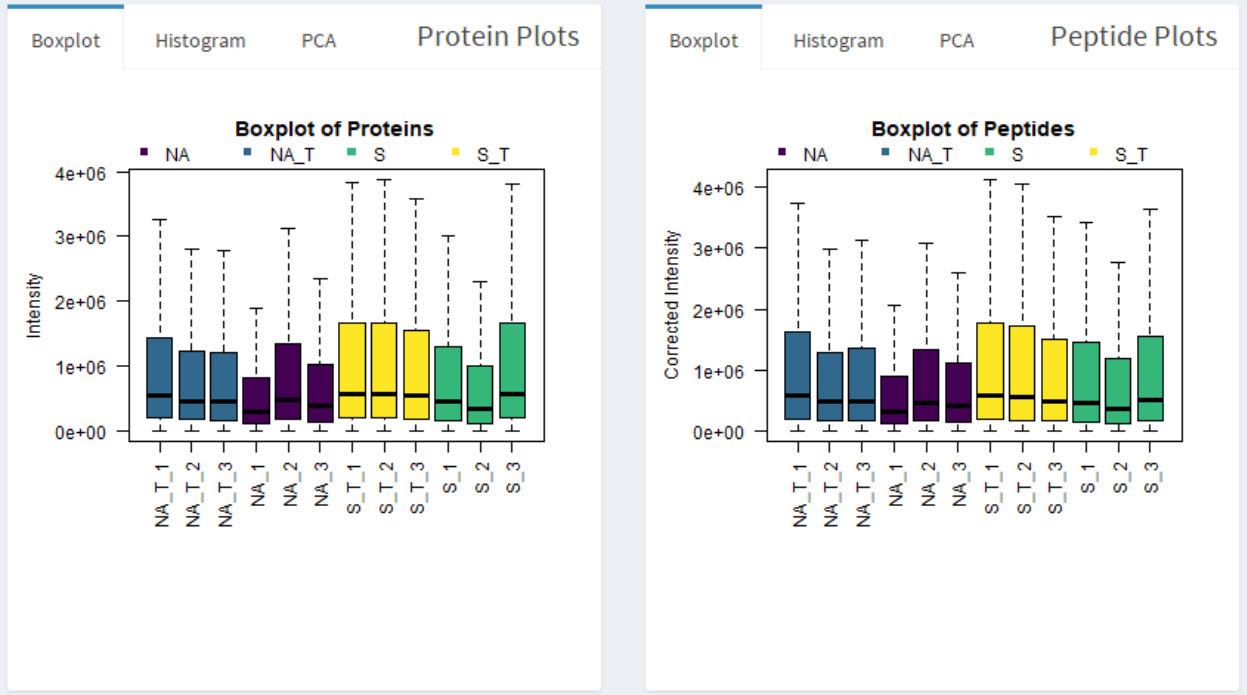

Figure 3 Intensity distribution by sample for peptide and protein data

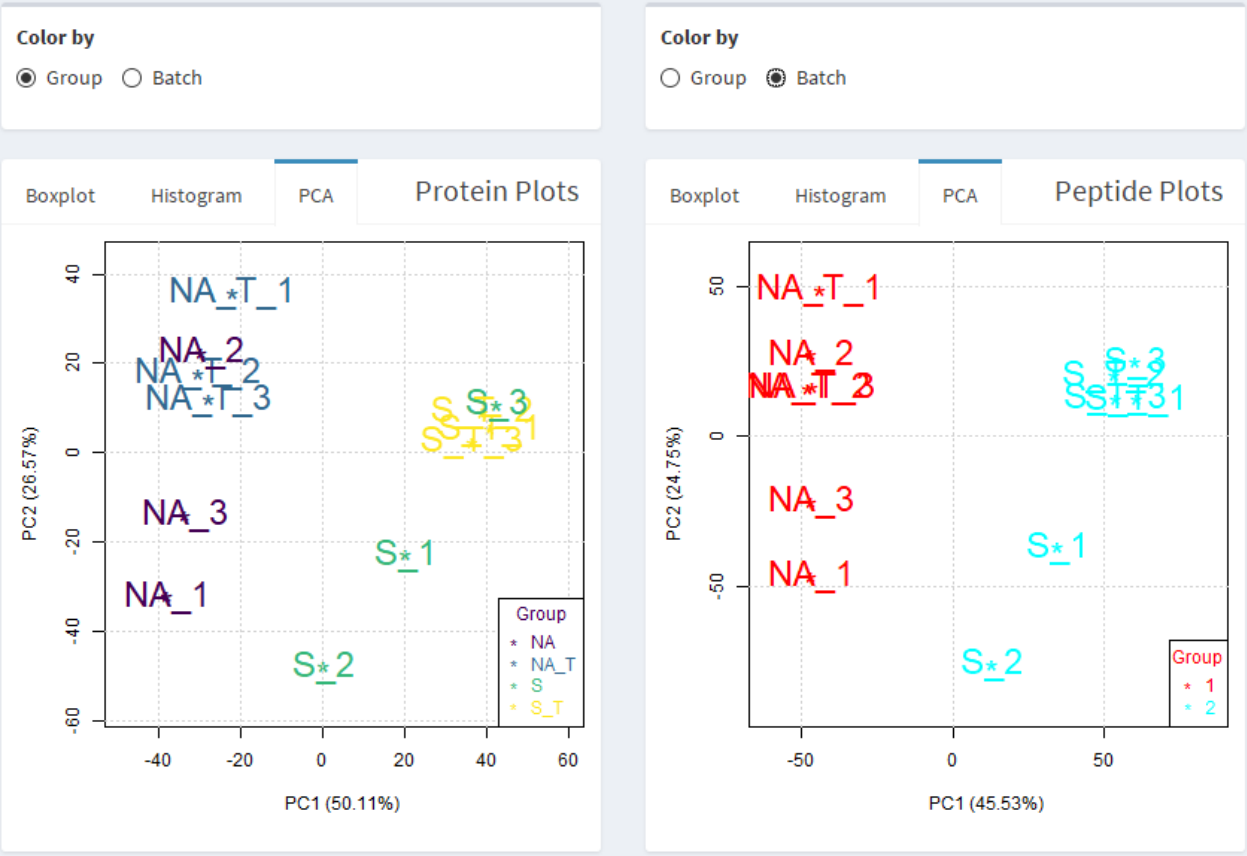

Figure 4 PCA plots of peptides and proteins, color-coded by group or batch.

In addition, the user can the minimum number of samples with measurements in Y groups and the number of Y groups, meaning that only proteins are used that were measured in at least X samples in Y groups, where X is the number specified in “Minimum number of samples with measurements in Y groups” and Y is the number specified in “Number of Y groups”. For example, if “Minimum number of samples with measurements in Y groups” is set to 2 and “Number of Y groups” is set to 3, only proteins that were measured in at least 2 samples in 3 of the groups will be used downstream.