Protein Groups Output (MaxQuant or MQ)

Description of the worksheets:
Raw data – Unprocessed output from the MQ software.
Result – Includes only the Majority protein IDs and relevant data columns.

Background information:

Label-free quantification based on extracted ion chromatogram (XIC): quantification based on peptide-ion intensity (MS) and subsequent identification (MS/MS).
Note: Label-free: scale according to the total protein amount

Posterior Error Probability (PEP) probability of a peptide being wrongly identified
Protein score = product of peptide PEPs (one for each sequence)

Match between runs
Peptides, which are present in several samples, but not identified via MS/MS in all of them: can still be identified via matching between runs—boosts number of identifications.

Normalization
Normalization is used in the MaxQuant LFQ algorithm for label-free quantification.

Firstly, the intensities for a peptide are summed up over the factions of one sample with introducing the normalization factors as unknown variables. These normalization factors are then calculated by an optimization approach, where the overall proteome change should be kept minimal. This is based on the assumption that most proteins do not or only minimally change between conditions, to have a constant baseline.

Protein Groups: Proteins, which cannot be unambiguously identified by unique peptides (but have only shared peptides) are grouped in one protein group and quantified together, e.g. if all detected peptides of protein A also belong to protein B, A and B form one protein group.

Razor peptides: Peptides, which are shared between protein groups (e.g. protein X and protein Y are unambiguously identified via unique peptides, but nevertheless share a common peptide). The razor peptide will be assigned to the protein group with the largest number of total peptide identified.

Majority protein IDs These are the IDs of those proteins that have at least half of the peptides that the leading protein has.

Protein names Name(s) of protein(s) contained within the group.

Gene names Name(s) of the gene(s) associated to the protein(s) contained within the group.

Number of proteins Number of proteins contained within the group.

Peptides The total number of peptide sequences associated with the protein group (i.e. for all the proteins in the group).

Razor + unique peptides The total number of razor + unique peptides associated with the protein group (i.e. these peptides are shared with another protein group).

Unique peptides The total number of unique peptides associated with the protein group (i.e. these peptides are not shared with another protein group).

Sequence coverage [%] Percentage of the sequence that is covered by the identified peptides of the best protein sequence contained in the group.

Q-value This is the ratio of reverse to forward protein groups.

Score Protein score which is derived from peptide posterior error probabilities.

$$Sum\ PEP\ Score = \sum_i -\log_{10}(PEP_{best\ peptide,i})$$

Identification type Indicates whether this experiment was identified by MS/MS or only by matching between runs.

Intensity Summed up eXtracted Ion Current (XIC) of all isotopic clusters associated with the identified AA sequence. In other words, intensities are the sums of all individual peptide intensities belonging to a particular protein group. Unique and razor peptide intensities are used as default.

iBAQ (intensity Based Absolute Quantification): these are the (raw) intensities divided by the number of theoretical peptides. Thus, iBAQ values are proportional to the molar quantities of the proteins. The iBAQ algorithm can roughly estimate the relative abundance of the proteins within each sample.

LFQ intensity: Normalized intensity

MS/MS count: Peptide spectrum matches

Only identified by site When marked with '+', this particular protein group was identified only by a modification site.

Reverse When marked with '+', this particular protein group contains no protein, made up of at least 50% of the peptides of the leading protein, with a peptide derived from the reversed part of the decoy database. These should be removed for further data analysis. The 50% rule is in place to prevent spurious protein hits to erroneously flag the protein group as reverse.

Potential contaminant When marked with '+', this particular protein group was found to be a commonly occurring contaminant. These should be removed for further data analysis.