

Summary of reanalysed MaxQuant output exported into Expression Atlas

PRIDE dataset identifier: PXD012203
PRIDE dataset URL: <https://www.ebi.ac.uk/pride/archive/projects/PXD012203>
Lab Head: SK, Sze
E-mail: sksze@ntu.edu.sg
Affiliation: Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551, Singapore.
Original dataset submitter: Adav, Sunil
E-mail: ssadav@ntu.edu.sg
PubMed ID: 30691479

Quantification method: iTRAQ
Search database: Human SwissProt Proteome (UniProt, January 2020. 20,367 sequences)
Contaminant database: MaxQuant contaminants database (conf/contaminants.fasta)
Analysis software: MaxQuant v1.6.3.4
Operating system: Red Hat Enterprise Linux Server
Differential analysis: MSstatsTMT v1.4.6

Experimental design

Labelling method: iTRAQ, 4-plex
Design: 3 replicates (AD_A, AD_B, AD_C)
... Non-demented brain samples (controls).
... label 114 (channel name: 1.AD)
... label 116 (channel name: 3.AD)
... Alzheimer's disease brain samples (cases).
... label 115 (channel name: 2.AD)
... label 117 (channel name: 4.AD)

Sample	Replicate	iTRAQ.channel	Condition
1.AD	A	114	Early onset Non-demented control
1.AD	B	114	Early onset Non-demented control
1.AD	C	114	Early onset Non-demented control
2.AD	A	115	Early onset Alzheimer's disease
2.AD	B	115	Early onset Alzheimer's disease
2.AD	C	115	Early onset Alzheimer's disease
3.AD	A	116	Late onset Non-demented control
3.AD	B	116	Late onset Non-demented control
3.AD	C	116	Late onset Non-demented control
4.AD	A	117	Late onset Alzheimer's disease
4.AD	B	117	Late onset Alzheimer's disease
4.AD	C	117	Late onset Alzheimer's disease

Summary table

MaxQuant output before and after processing.

The submitted original ‘raw’ files were run through MaxQuant; the output (pre-processed) intensities were then normalised, proteins mapped to Ensembl gene IDs and filtered results (post-processed) were uploaded to Expression Atlas.

Number of differential expressed proteins numbers. . . .

	Pre-processed	Post-processed*
Number of samples	12	12
Number of potential contaminants •	18	0
Number of reverse decoys ^	52	0
Number of identified proteins †	5378	2477
Total number of mapped peptides ^a	28129	24896
Protein groups mapped to unique gene id ^c	NA	3493
.... Number of mapped unique peptides §	25946	23063

* Data show in Expression Atlas.

• The total number of protein groups found to be a commonly occurring contaminant.

^ The total number of protein groups with a peptide derived from the reversed part of the decoy database.

† The total number of non-isoform SwissProt proteins within the protein group, to which at least 2 or more peptides from each sample are mapped to.

^a Sum of peptides that are mapped across all protein groups.

^c The total number of protein groups which are mapped to an unique Ensembl Gene ID.

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

Post-processing filters applied:

- (i) Remove reverse decoys.
- (ii) Remove potential contaminants.
- (iii) Include protein groups to which 2 or more unique peptides are mapped.
- (iv) Include protein groups wherein all protein IDs within are mapped to an unique Ensembl Gene ID.
- (v) Include protein groups wherein at least 50% sample replicates have non-zero intensity values.

Normalisation method:

Fraction Of Total (FOT): Each protein intensity value is scaled to the total amount of signal in a given MS run (column) and transformed to parts per billion (ppb)

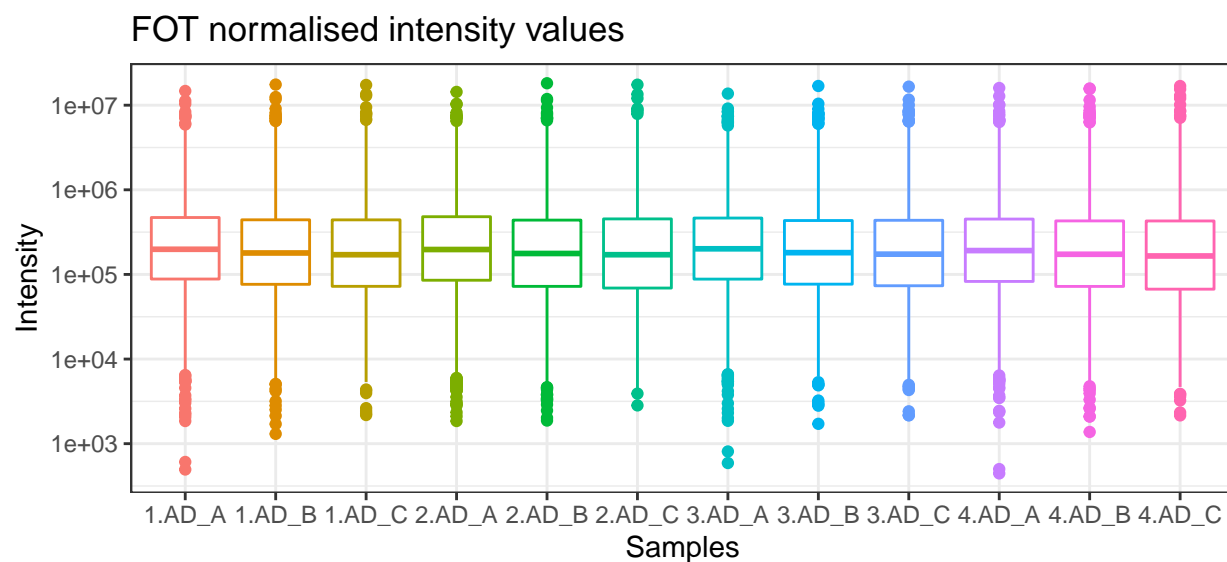


Figure 1. Boxplots with distribution of intensity values for each sample after FOT normalisation. Sample name description: **AD**: Alzheimer's Disease; A, B and C represent replicates.

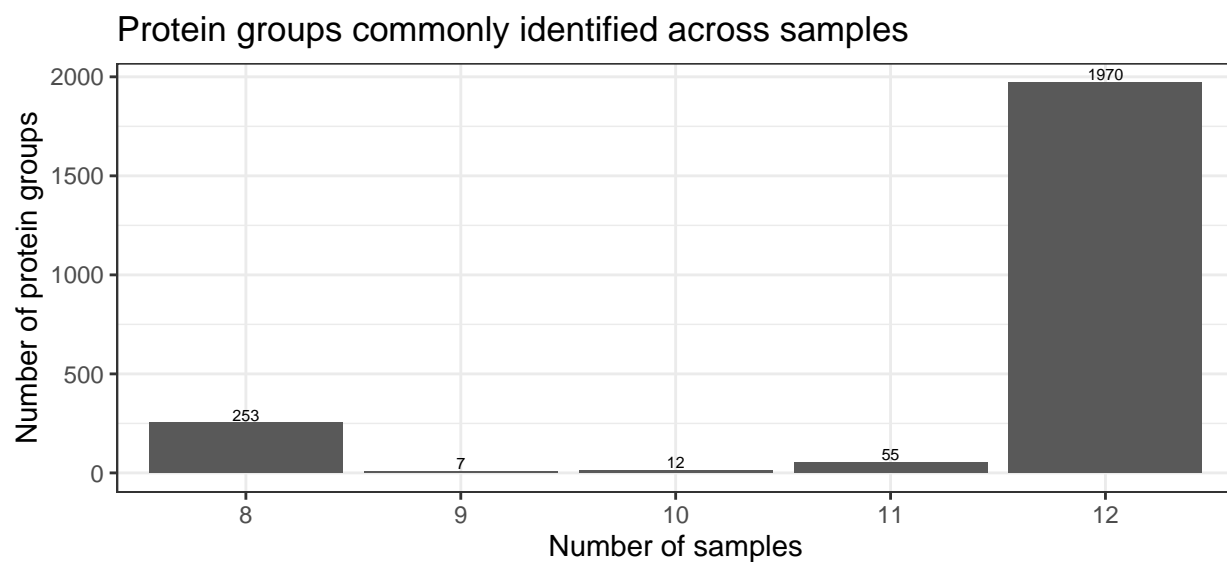


Figure 2. Protein overlap. Indicates the number of protein groups that were identified across different samples. Protein groups were counted as present in a sample when the sample had registered intensity.

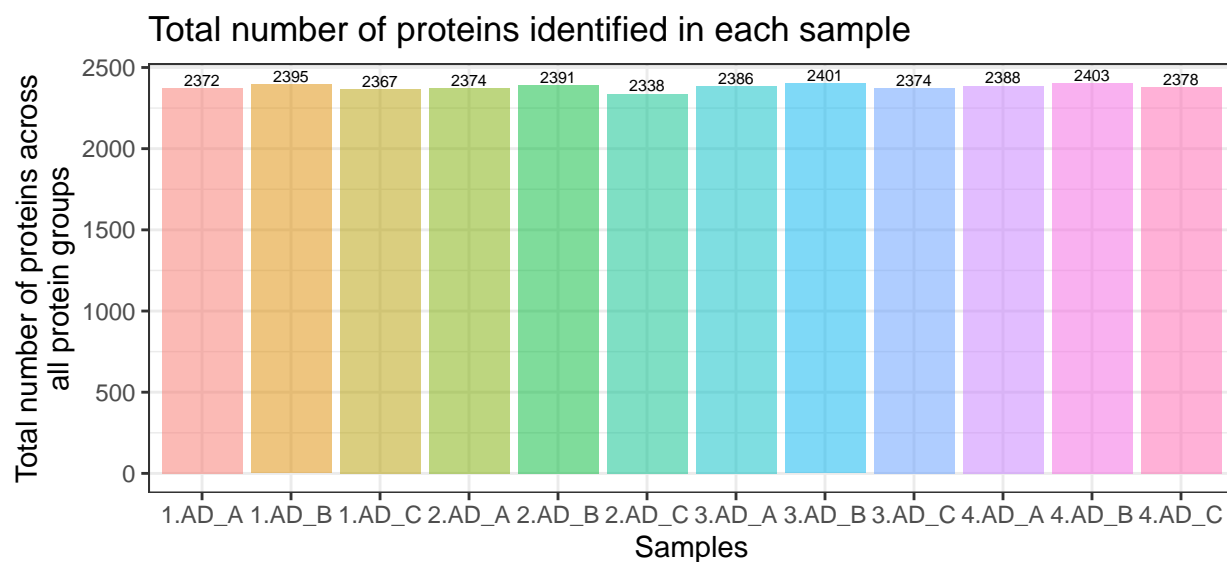


Figure 3. Protein counts in each sample. The total number of proteins (SwissProt non-isoforms) from all protein groups to which at least 2 or more unique peptides from each sample are mapped to.

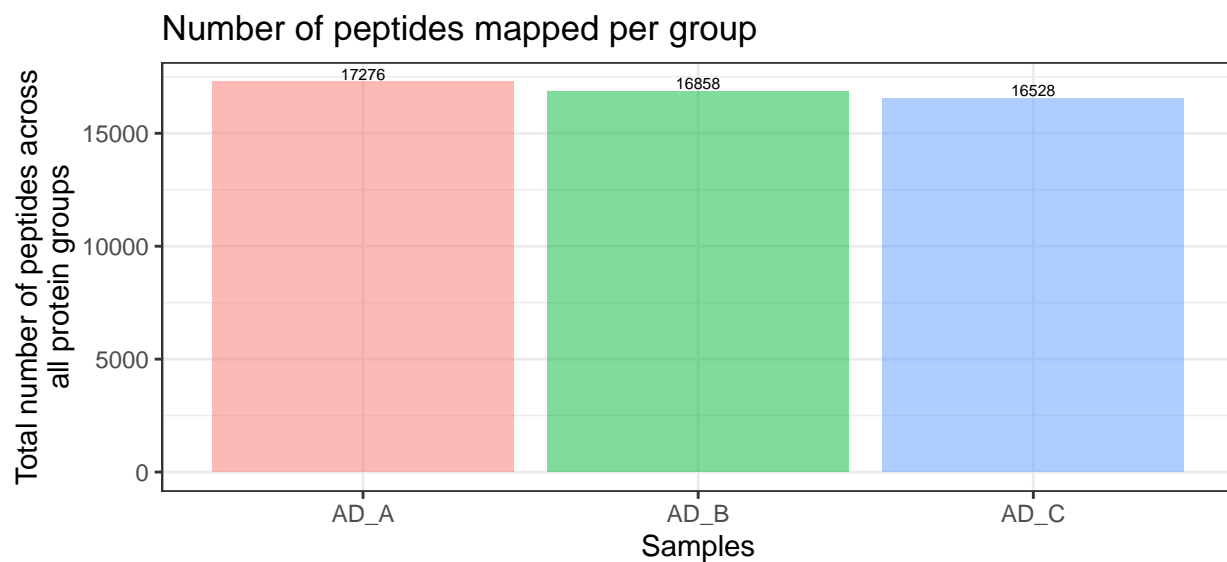


Figure 4. Peptide counts in each sample. The total number of peptides that are mapped across all protein groups from each sample.

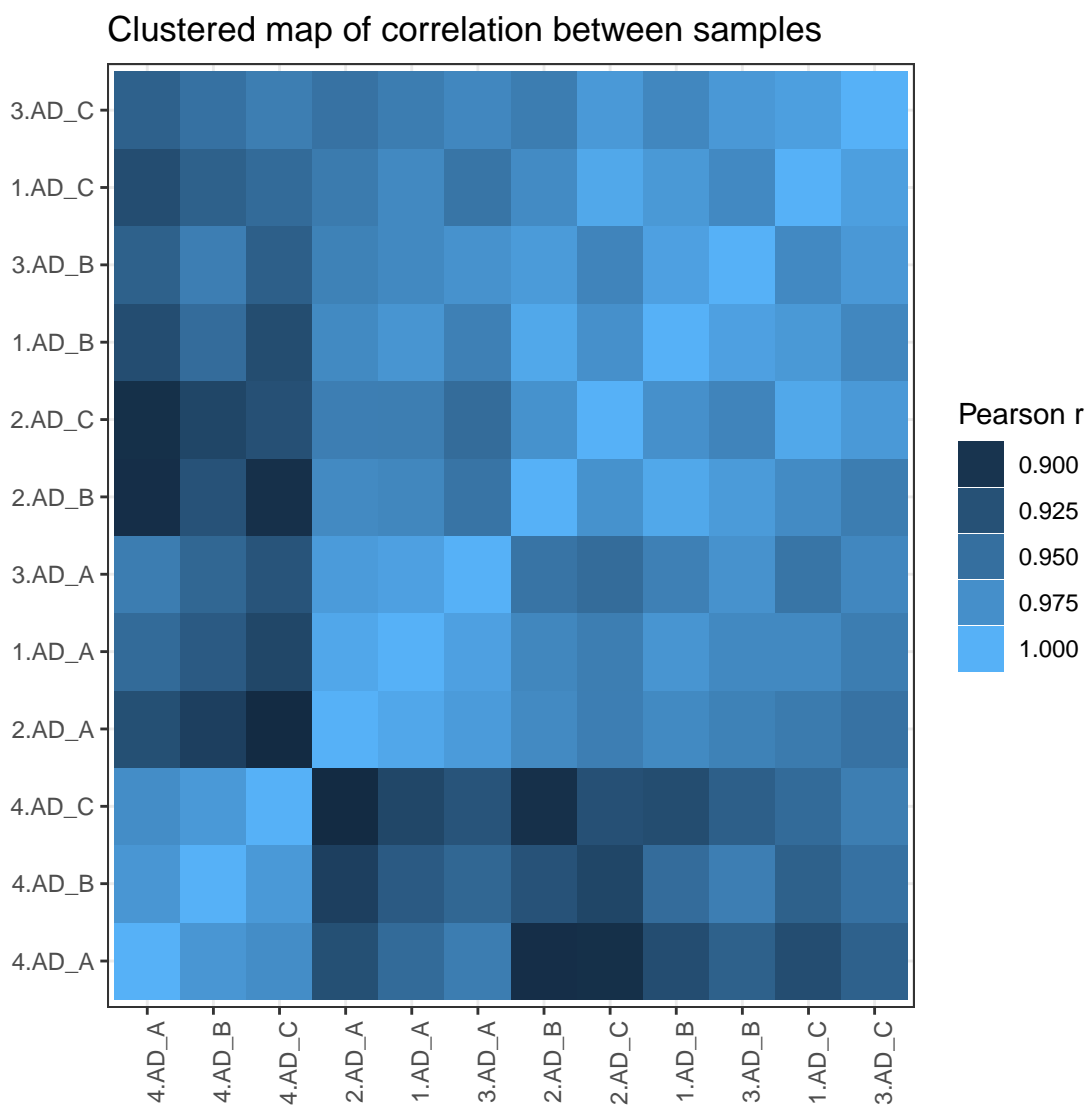


Figure 5. Correlation between samples with cases and controls. The pairwise Pearson correlation was calculated between normalised intensities of each sample and clustered hierarchically.

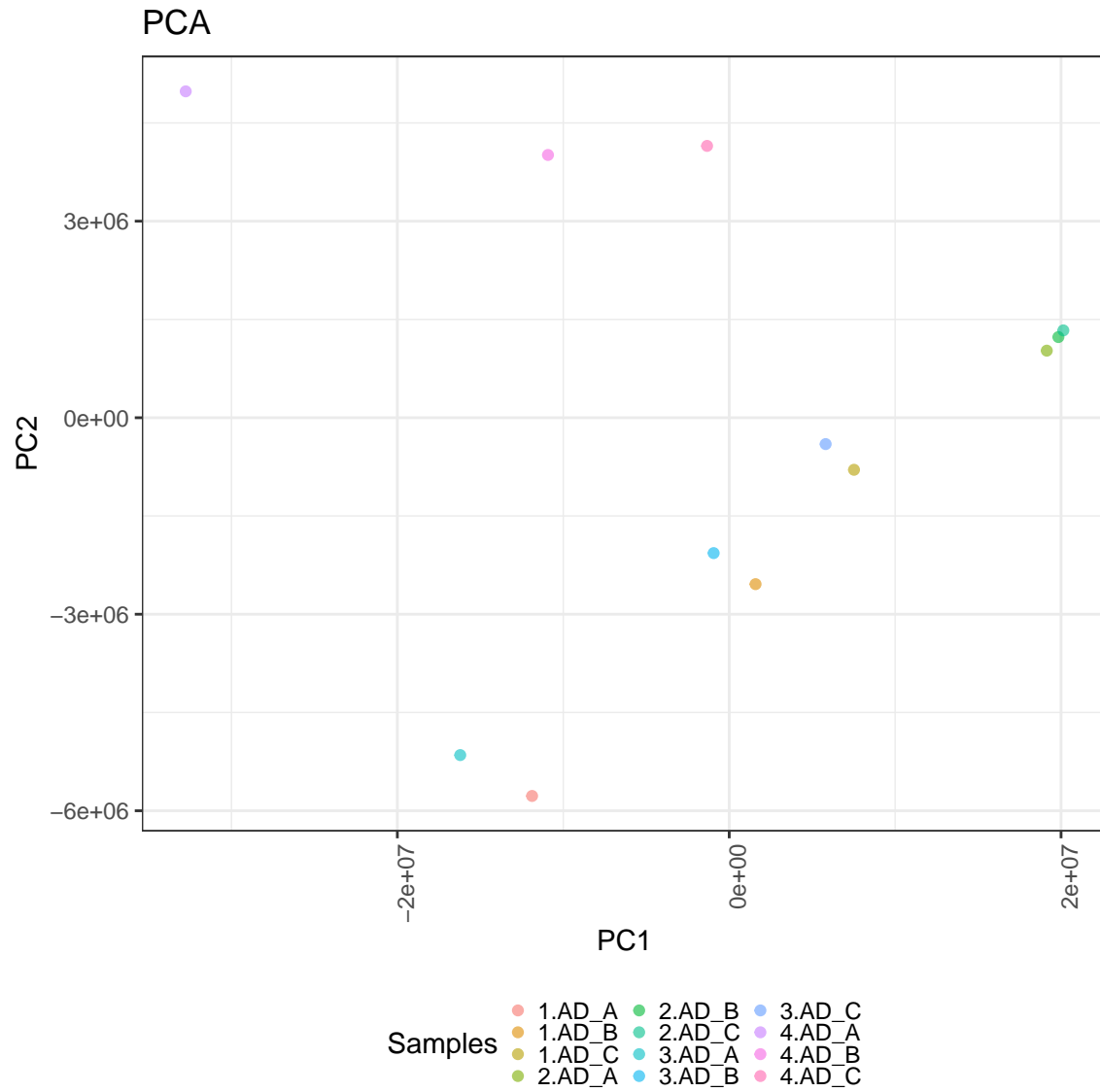


Figure 6. PCA plot. The Reporter corrected intensity values of cases and controls are used to compute principle components.

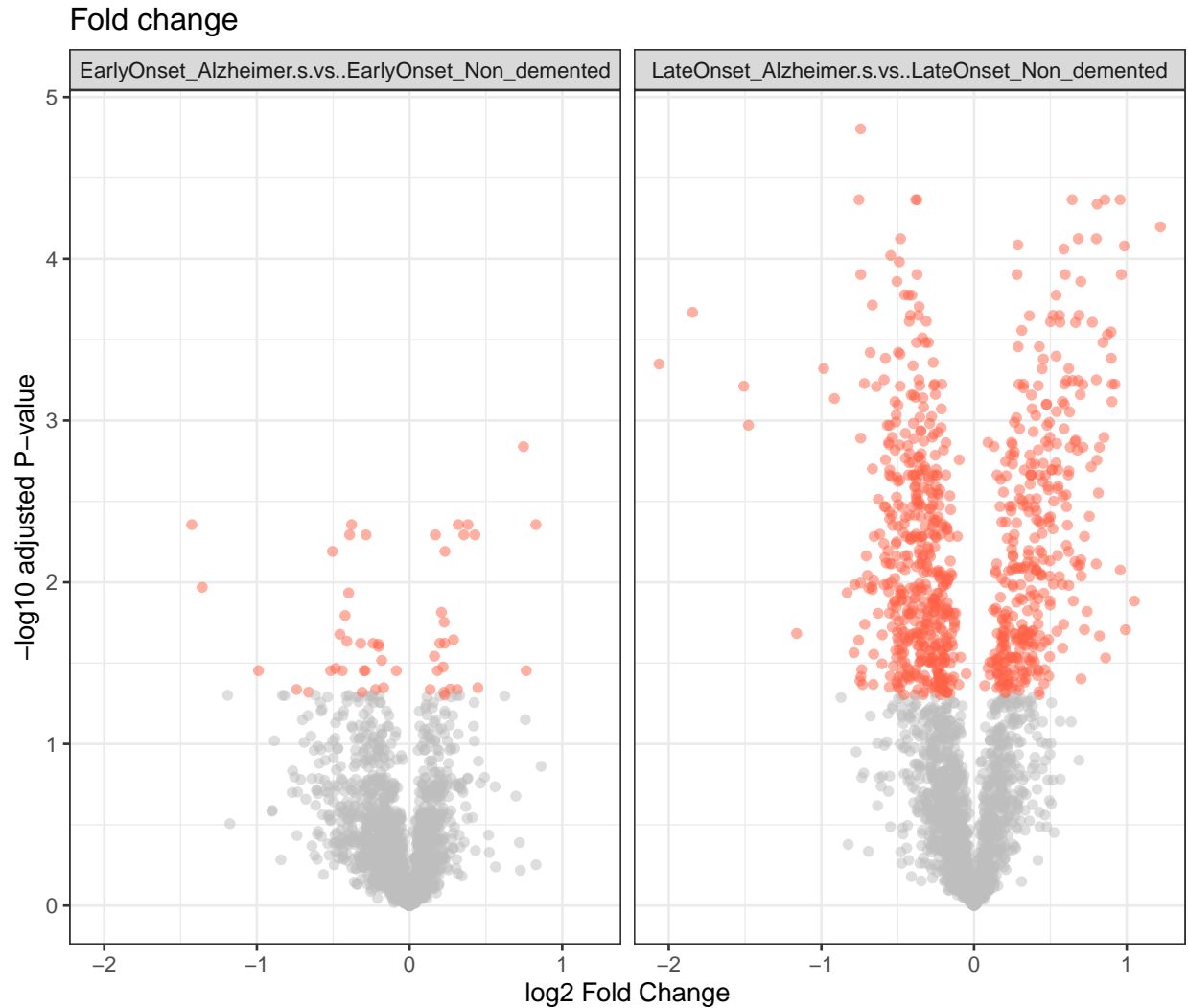


Figure 7. Differential protein expression. Proteins which show statistically significant differential expression ($P\text{-value} \leq 0.05$) are coloured red.

Glossary

The descriptions of the terms used in here are taken from MaxQuant documentation.

Reverse decoy: 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 are 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: This particular protein group was found to be a commonly occurring contaminant. These are removed for further data analysis.

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