Summary of reanalysed MaxQuant output exported into Expression Atlas

PRIDE dataset identifier:	PXD016958
PRIDE dataset URL:	https://www.ebi.ac.uk/pride/archive/projects/PXD016958
Lab Head:	Knepper, Mark
E-mail:	knepperm@nhlbi.nih.gov
Affiliation:	Epithelial Systems Biology Laboratory, Systems Biology Center, NHLBI, NIH
Original dataset submitter:	Limbutara, Kavee
E-mail:	kavee.limbutara@nih.gov
PubMed ID:	32358040
Quantification method:	Label-free (baseline)
Search database:	Rat Reference Proteome with isoforms (Rattus norvegicus, UniProt, Nov 2021. 31,562 sequences)
Contaminant database:	MaxQuant contaminants database (conf/contaminants.fasta)
Analysis software:	MaxQuant v1.6.3.4
Operating system:	Red Hat Enterprise Linux Server

Summary table

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

	Pre-processed	Post-processed*
Number of samples	44	44
Number of potential contaminants •	95	0
Number of reverse decoys^	113	0
Number of identified proteins†	4224	3974
Total number of mapped peptides[a]	105804	87149
.... Number of mapped unique peptides§	NA	72057
Protein groups mapped to unique gene id¢	84447	6132

* 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 th 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.
¢ 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.

Normalisation method:

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

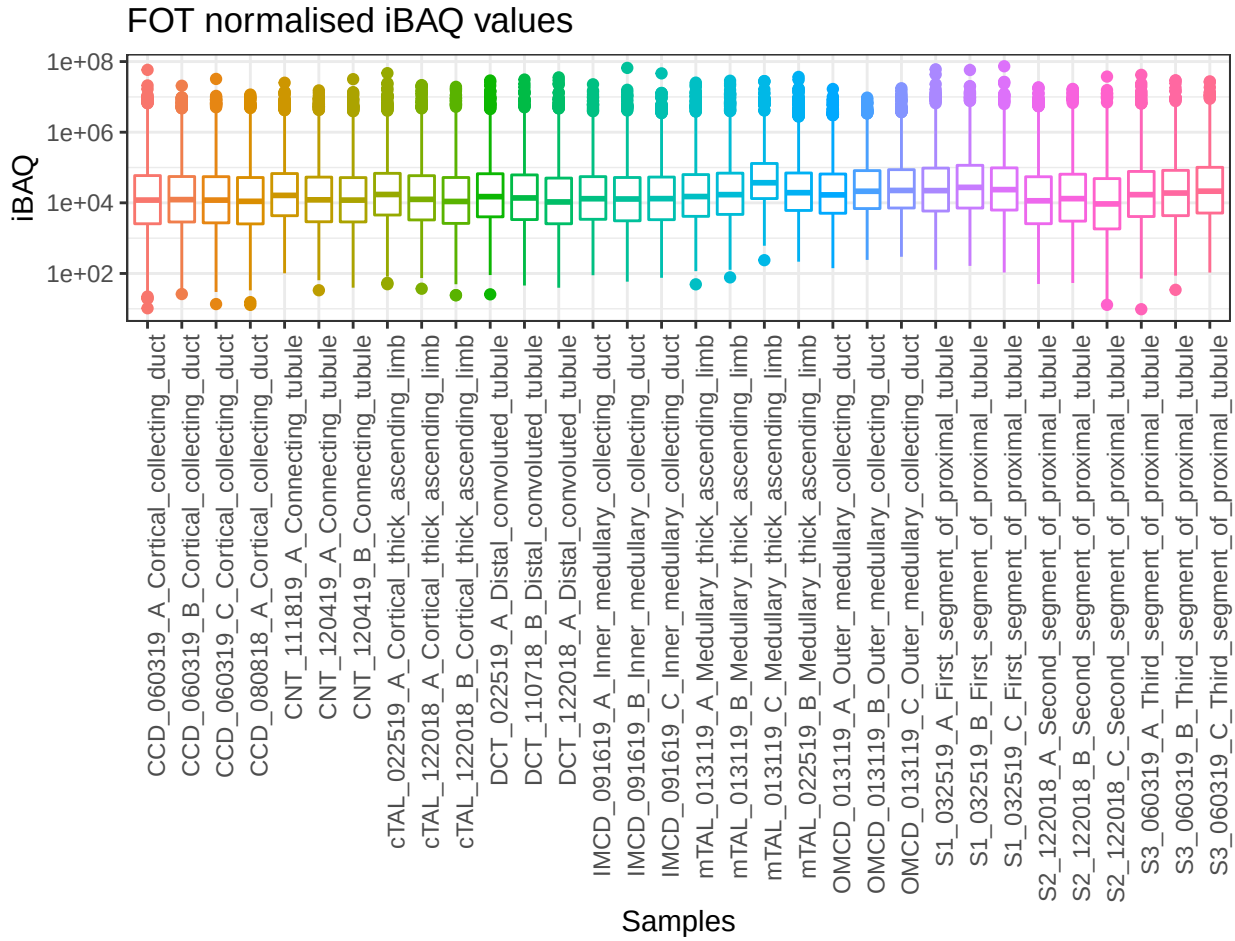

Figure 1. Boxplots with distribution of iBAQ values for each sample after FOT normalisation.

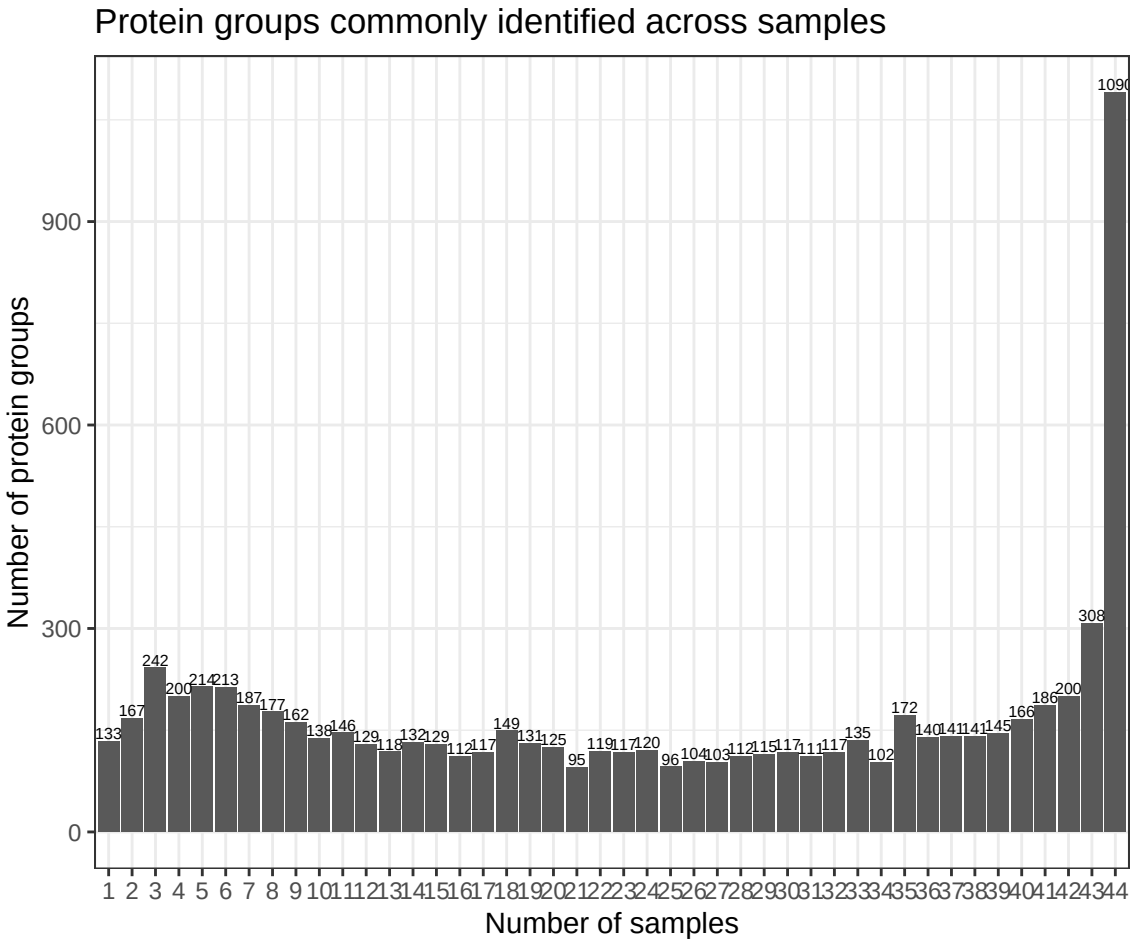

Figure 2. Protein overlap. Indicates the number of protein groups that were identified across different samples. Protein groups were counted as present in a sample when the sample had registered intensity.

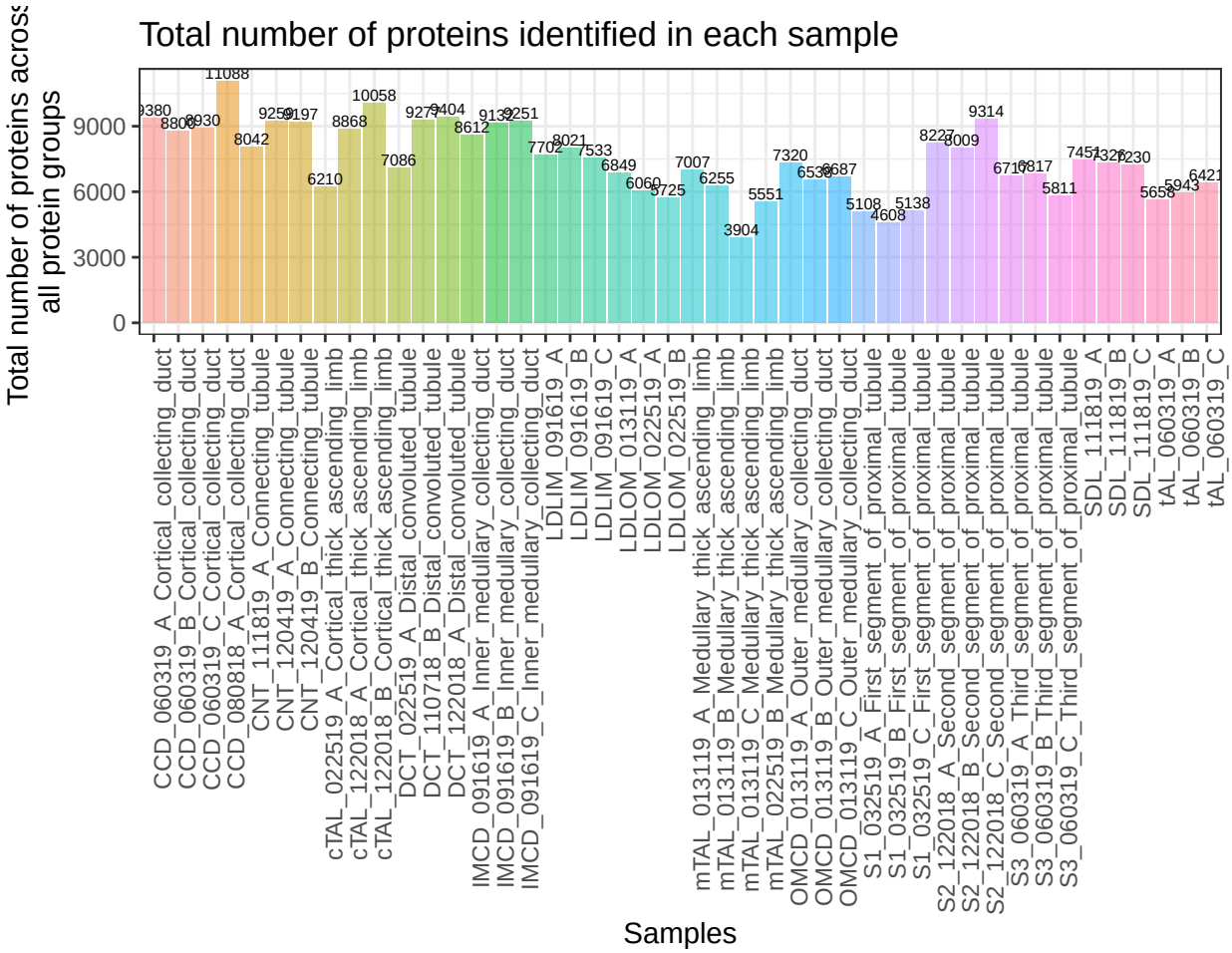

Figure 3. Protein counts in each sample. The total number of proteins (SwissProt non-isoforms) from all protein groups to which at least 2 or more unique peptides from each sample are mapped to.

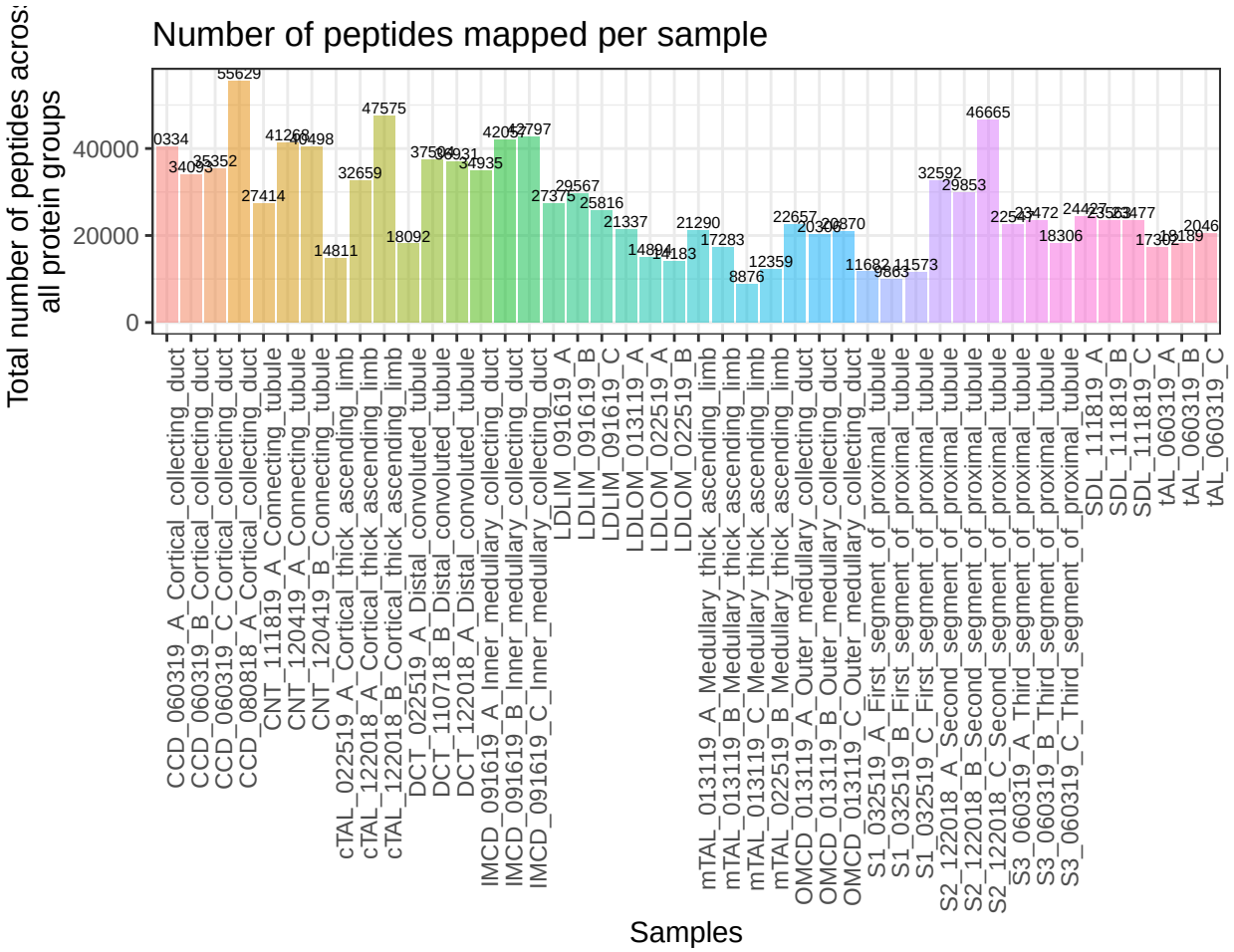

Figure 4. Peptide counts in each sample. The total number of peptides that are mapped across all protein groups from each sample.

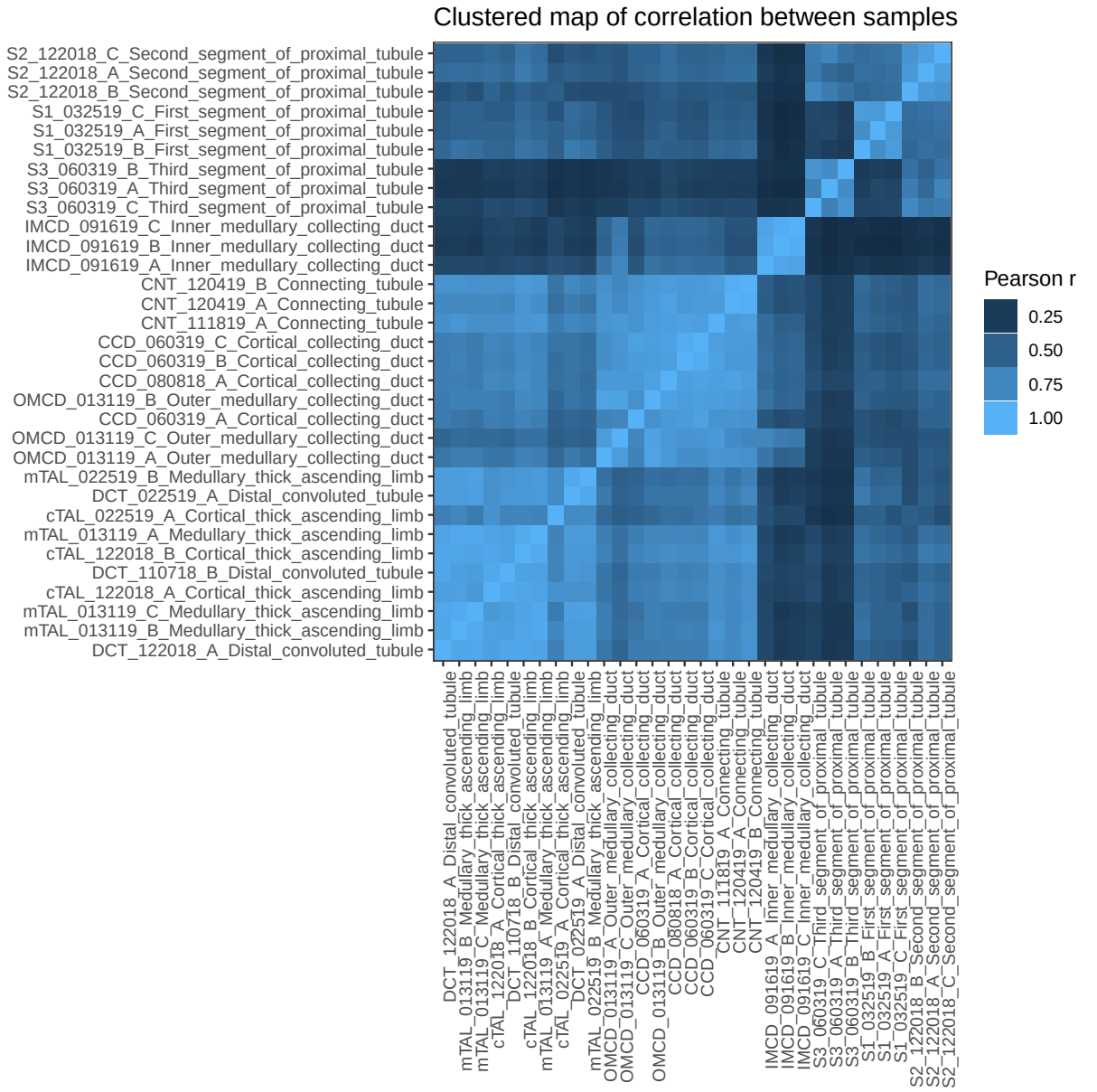

Figure 5. Correlation between samples. The pairwise Pearson correlation was calculated between normalised intensities (iBAQs) of each sample and clustered heirarchically.

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).