FragPipe-Analyst report

05 March, 2025

Method details

The raw data files were analyzed using TMT workflows in FragPipe to obtain protein quantification.

Statistical analysis was performed using an in-house generated R script based on the ProteinGroup.txt file. Protein-wise linear models combined with empirical Bayes statistics were used for the differential expression analyses. The *limma* package from R Bioconductor was used to generate a list of differentially expressed proteins for each pair-wise comparison. A cutoff of the *adjusted p-value* of 0.05 (`params$fdr_correction` method) along with a |log2 fold change| of 1 has been applied to determine differentially expressed proteins in each pairwise comparison.

Quick summary of parameters used:

- Normalization = none
- Imputation = none
- Tested comparisons = Tumor_vs_Normal
- Adjusted *p-value* cutoff <= 0.05
- Log2 fold change cutoff >= 1

Results

Result output contains features of which *10671* features were reproducibly quantified.

1174 proteins differ significantly between samples.

## Exploratory Analysis (QC Plots)

Principle Component Analysis (PCA) plot

```
## Warning: Use of `pca_df[[indicate[1]]]` is discouraged. Use
## `.data[[indicate[1]]]` instead.
```

```
## Warning: Use of `pca_df[[indicate[2]]]` is discouraged. Use
## `.data[[indicate[2]]]` instead.
```
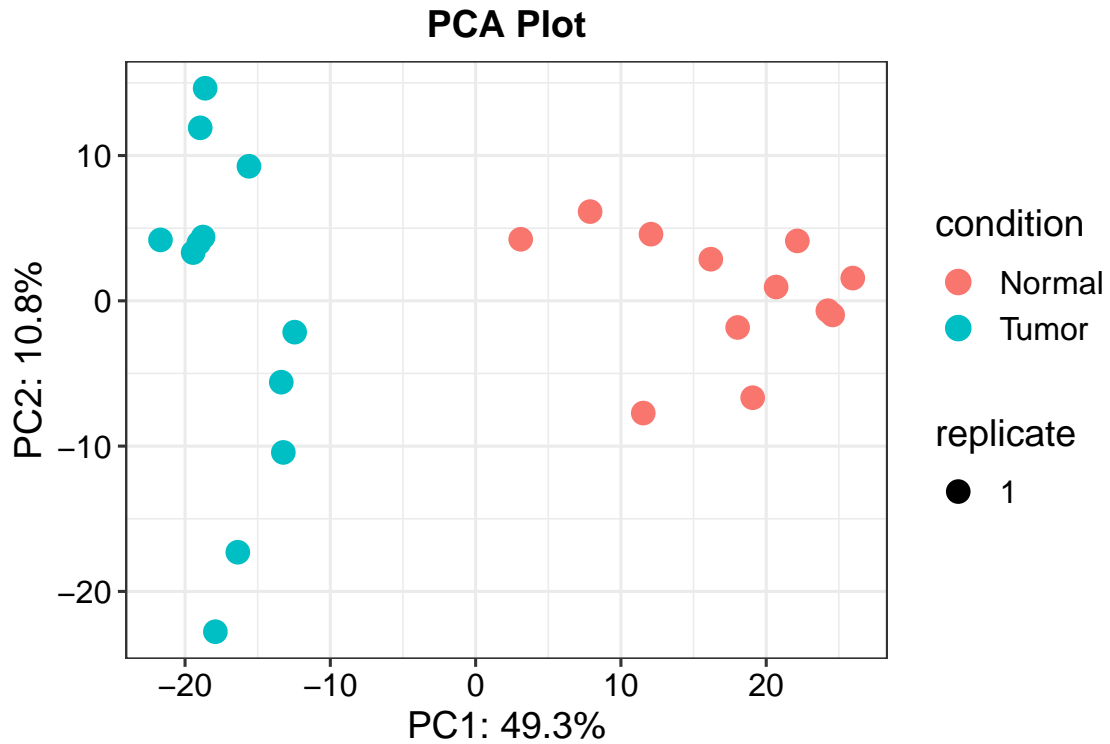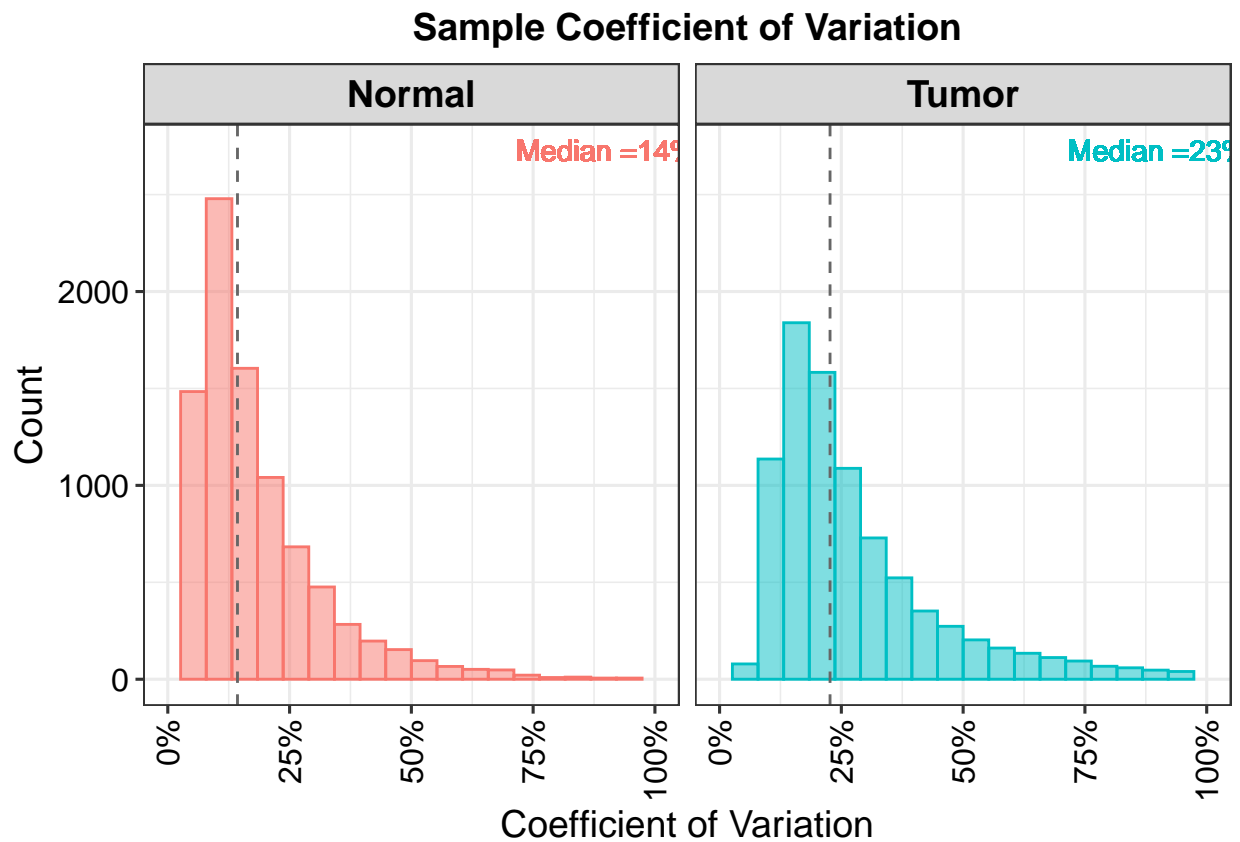
Sample Correlation matrix

```
## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## i Please use `all_of()` or `any_of()` instead.
##    # Was:
##    data %>% select(indicate)
##
##    # Now:
##    data %>% select(all_of(indicate))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```
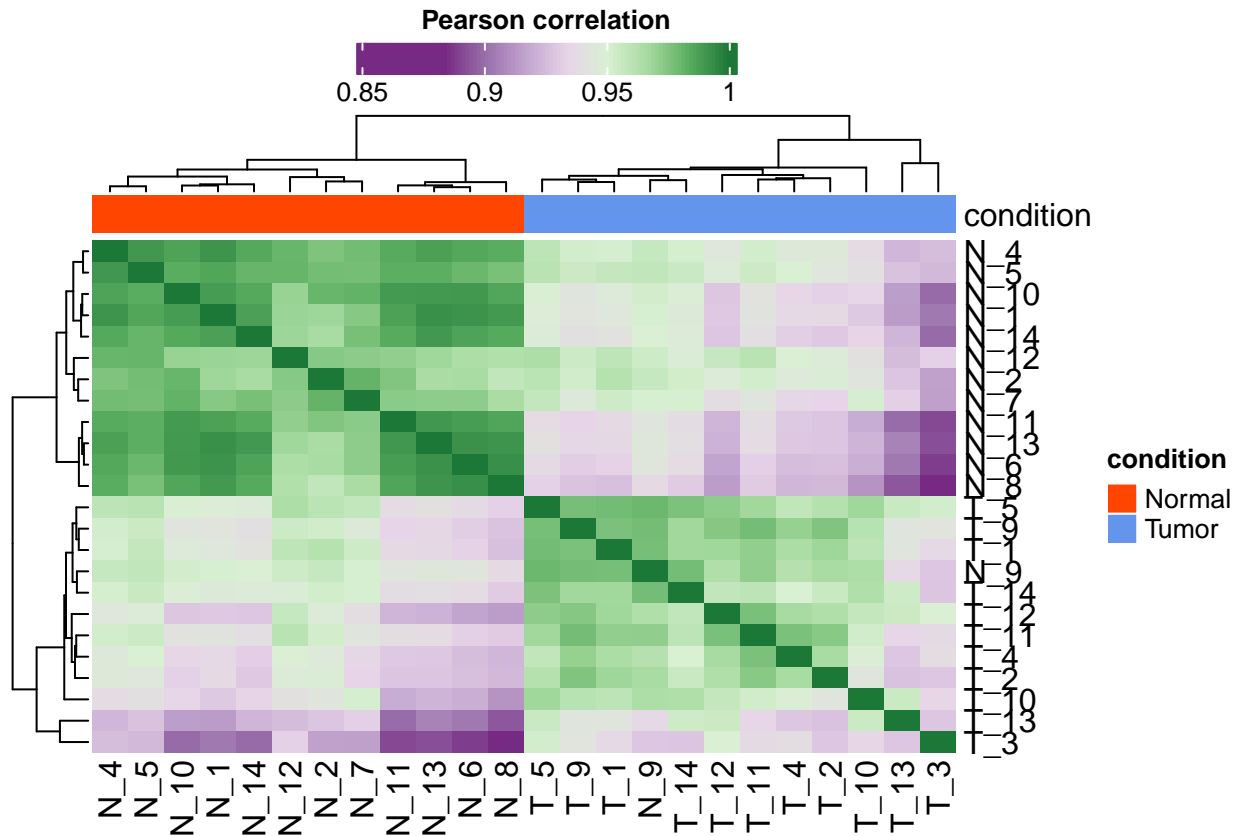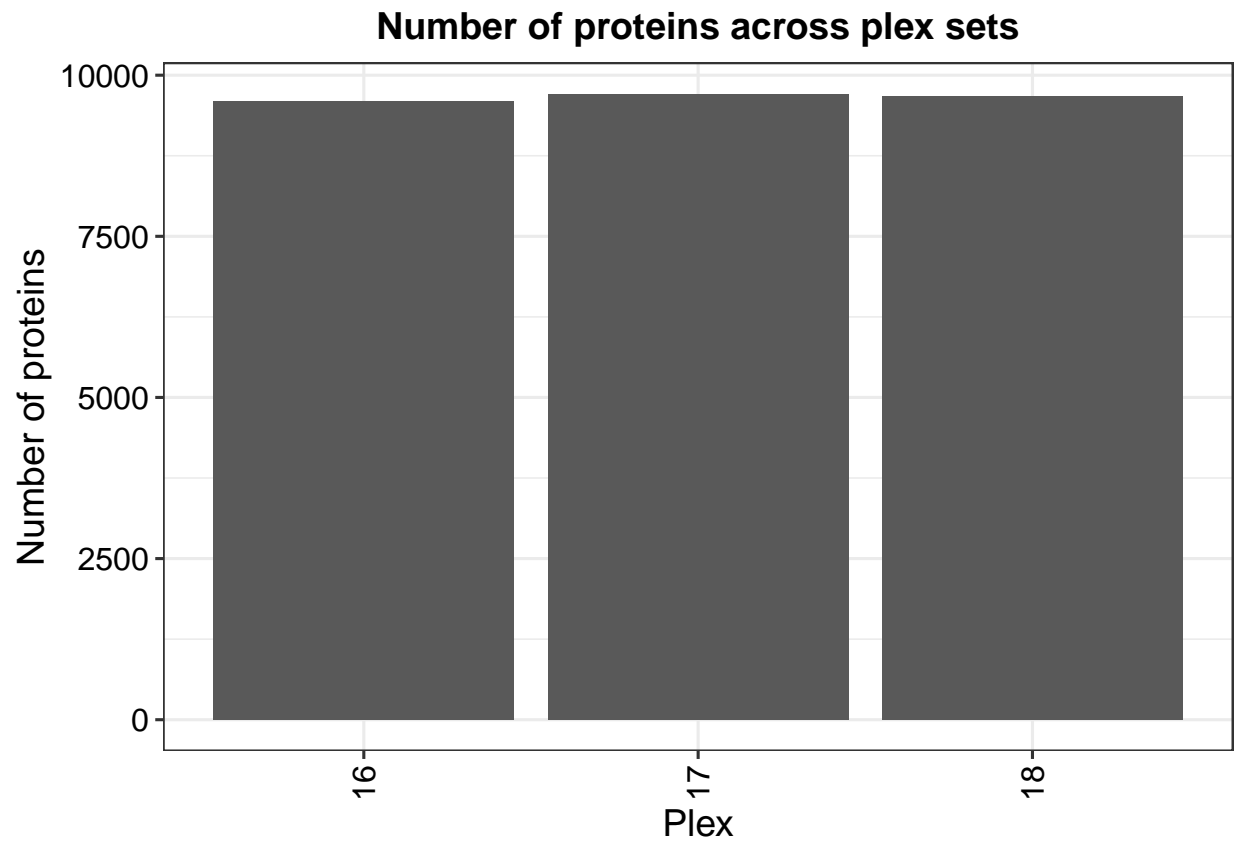
Sample Coefficient of variation (CVs)

```
## `summarise()` has grouped output by 'rowname'. You can override using the
## `.groups` argument.
```
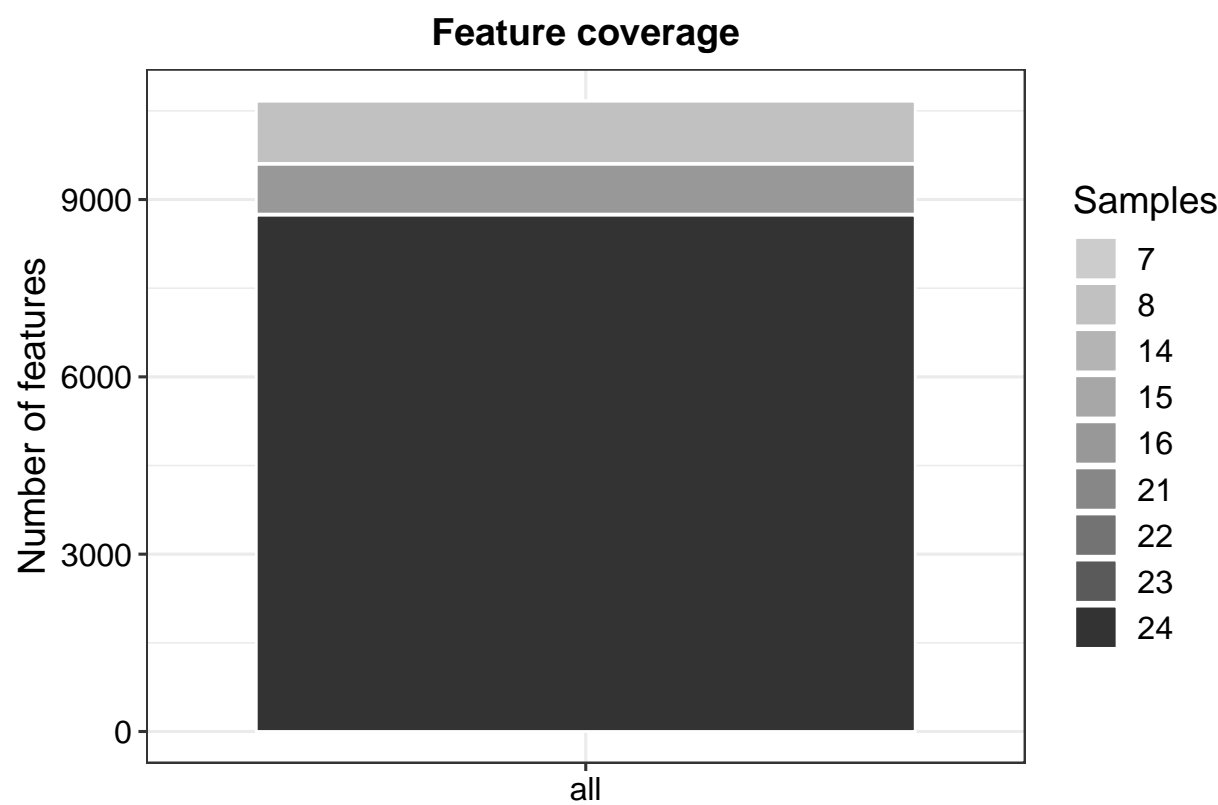

Sample Coefficient of Variation

Proteomics Experiment Summary

Protein quantified per sample (after pre-processing).

## Number of proteins across plex sets

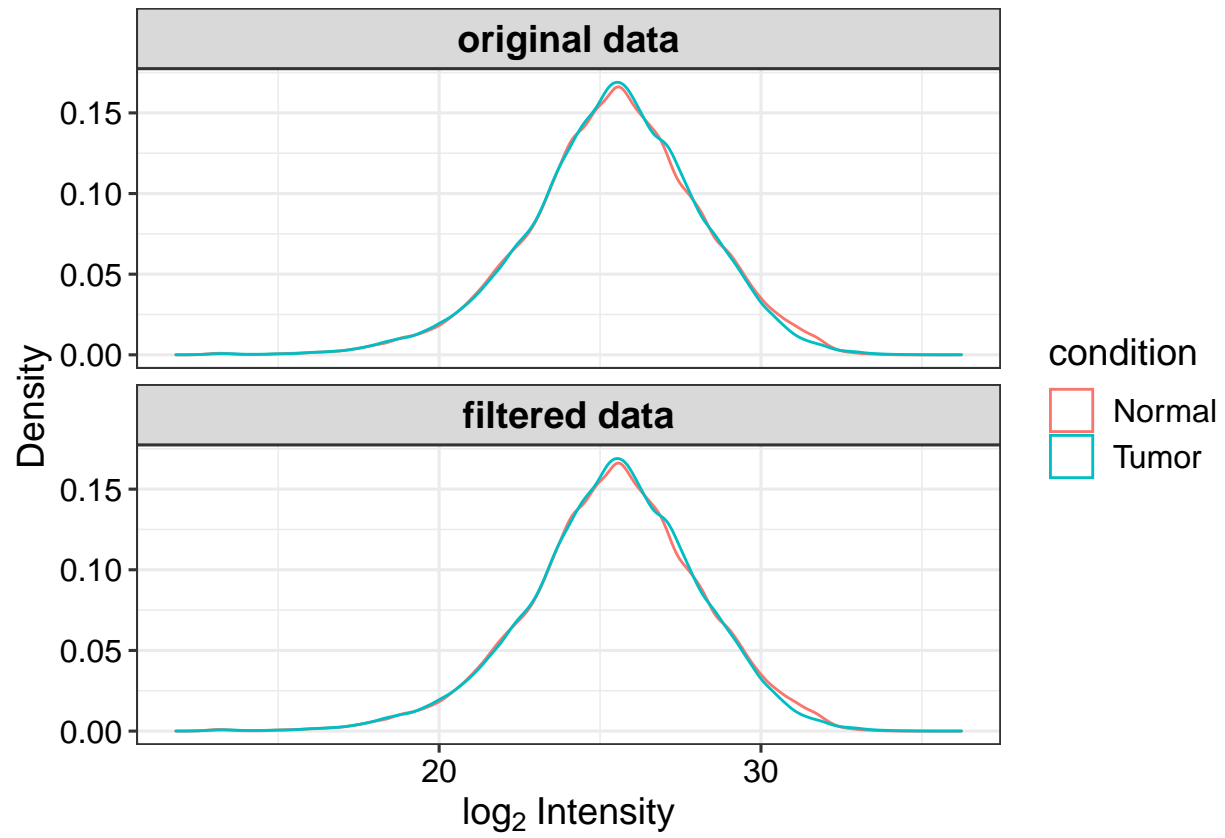Protein overlap in all samples.

Feature coverage

Missing Value handling

Missing value heatmap A heatmap for proteins with missing value in each dataset. Each row represent a protein with missing value in one or more replicate. Each replicate is clustered based on presence of missing values in the sample.
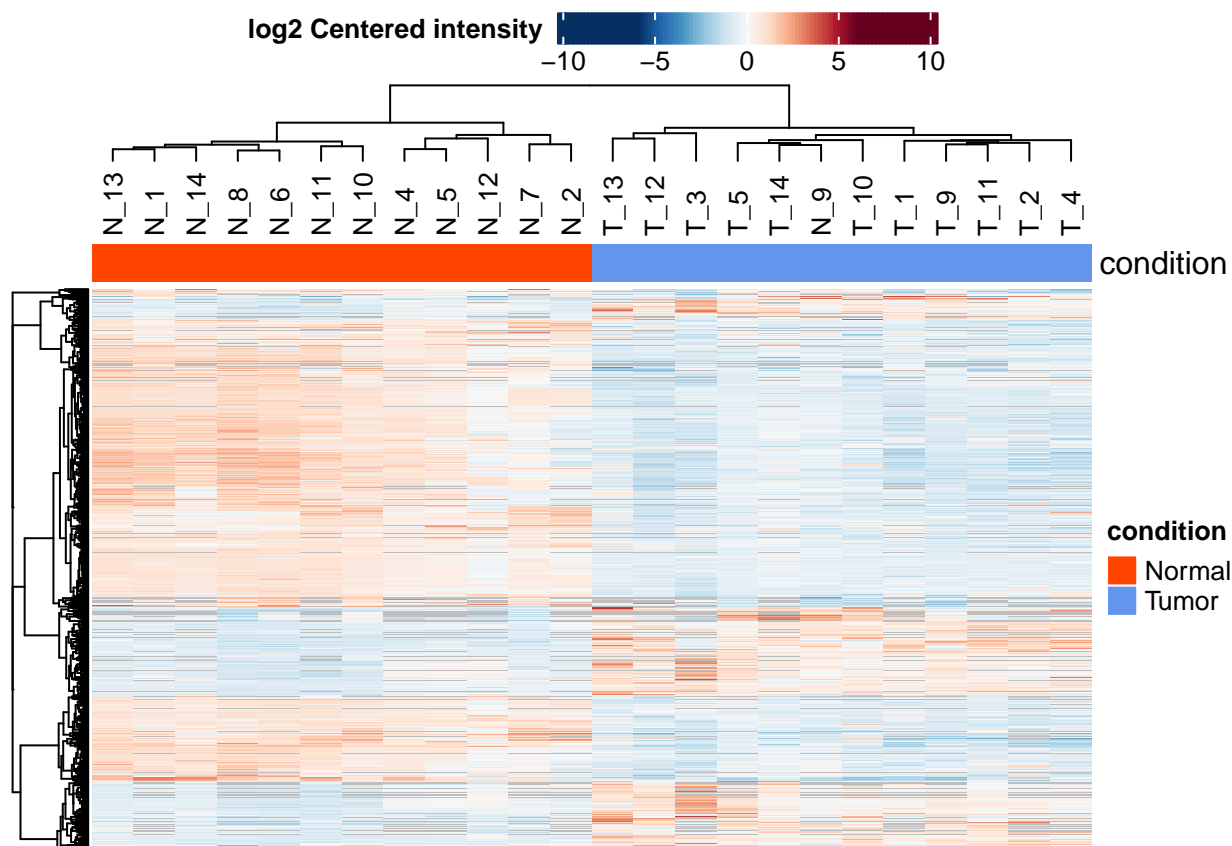

Missing values pattern (1936 proteins)
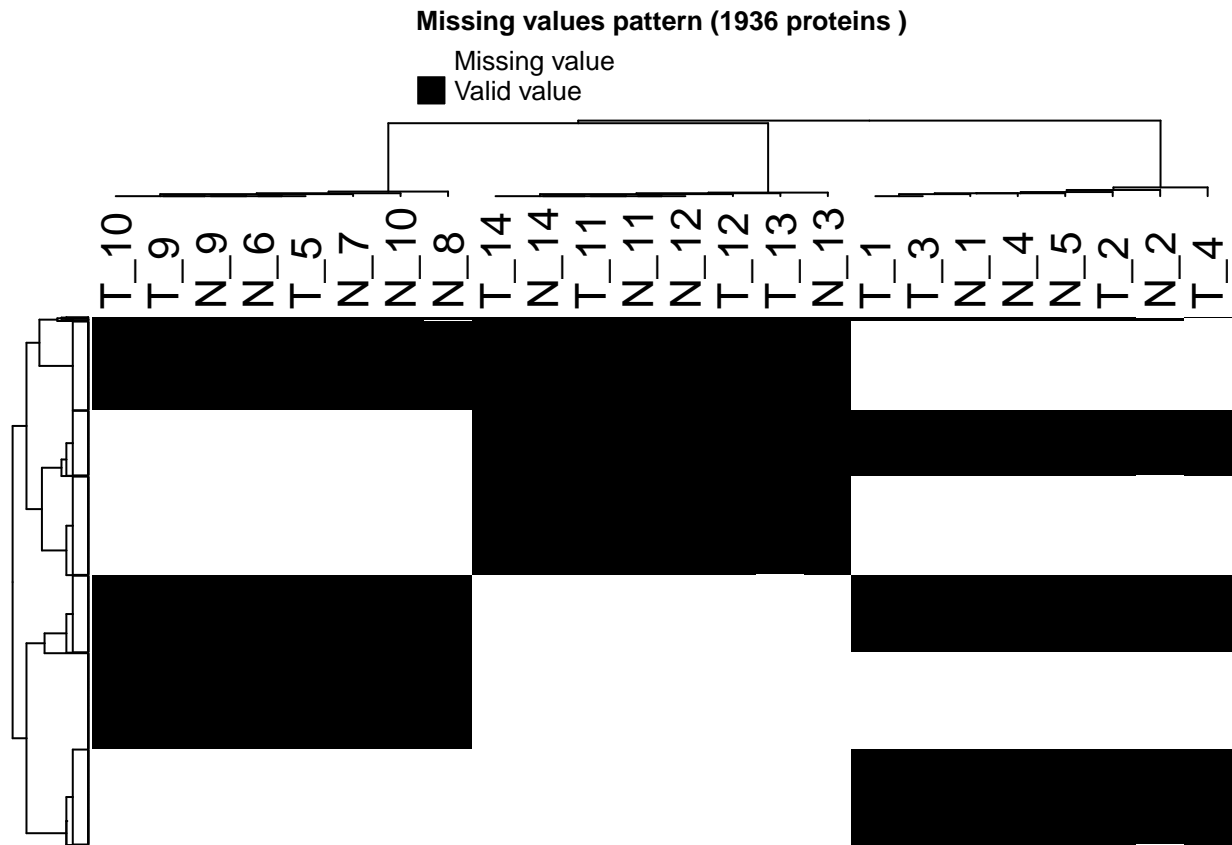
Missing value distribution   Protein expression distribution before and after imputation. The plot showing the effect of imputation on protein expression distribution.
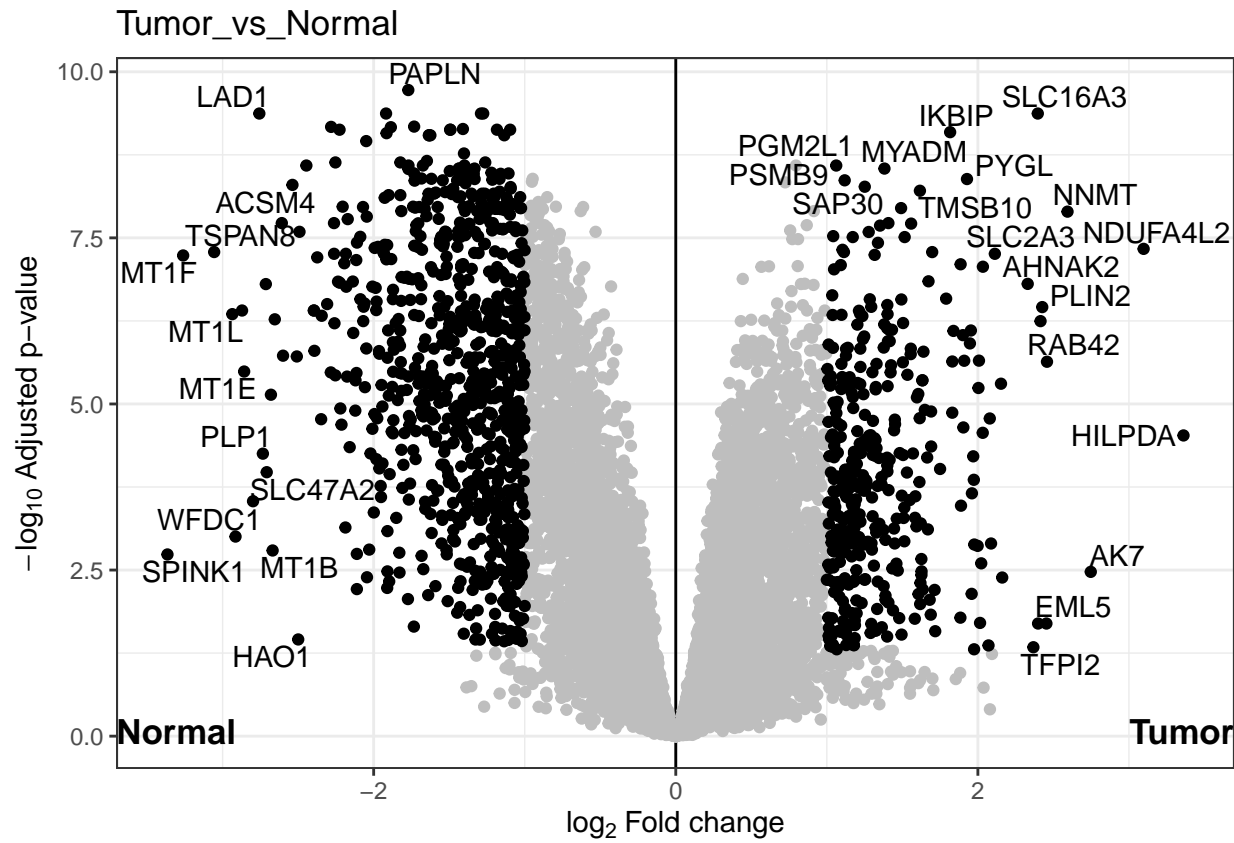
Differential Expression Analysis (Results Plots)

Heatmap   A plot representing an overview of expression of all significant (differentially expressed) proteins (rows) in all samples (columns).

Tumor_vs_Normal

Volcano Plots