Phospho-Analyst(phosphosite) report

14 February, 2022

Method details

The raw data files were analyzed using MaxQuant to obtain phosphosite identifications and their respective label-free quantification values using in-house standard parameters. Of note, the data were normalization based on the assumption that the majority of phosphosites do not change between the different conditions. Statistical analysis was performed using an in-house generated R script based on the Phospho(STY)Sites.txt file. First, contaminant phosphosites, reverse sequences and phosphosites identified "only by site" were filtered out. In addition, phosphosites that localization probability <0.75 have been removed as well. The phosphosite data was converted to log2 scale, samples were grouped by conditions and missing values were imputed using the 'Missing not At Random' (MNAR) method, which uses random draws from a left-shifted Gaussian distribution of 1.8 StDev (standard deviation) apart with a width of 0.3. 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 phosphosites for each pair-wise comparison. A cutoff of the *adjusted p-value* of 0.05 (Benjamini-Hochberg method) along with a |log2 fold change| of 1 has been applied to determine significantly regulated phosphosites in each pairwise comparison.

Quick summary of parameters used:

- Tested pairwise comparisons = Day0_vs_Day2, Day0_vs_Day7, Day2_vs_Day7

- Adjusted *p-value* cutoff $<= 0.05$
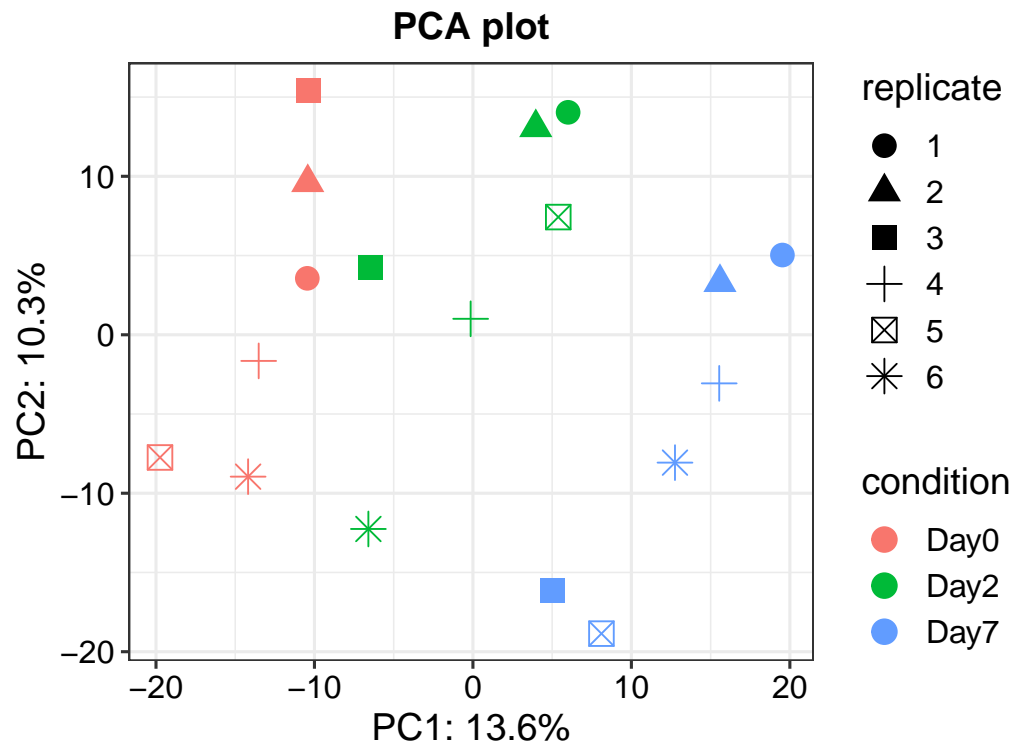
- Log fold change cutoff $>= 1$

Results

MaxQuant result output contains phosphosites of which *1856* phosphosites were reproducibly quantified.

137 phosphosites differ significantly between samples.

Exploratory Analysis (QC Plots)

**Principle Component Analysis (PCA) plot**   A plot used the PCA method to emphasize the variation.

Sample Correlation matrix   Correlation plot is similar to a heatmap, to visualise the relationship among different samples. The darker the stronger relevance between each sample.
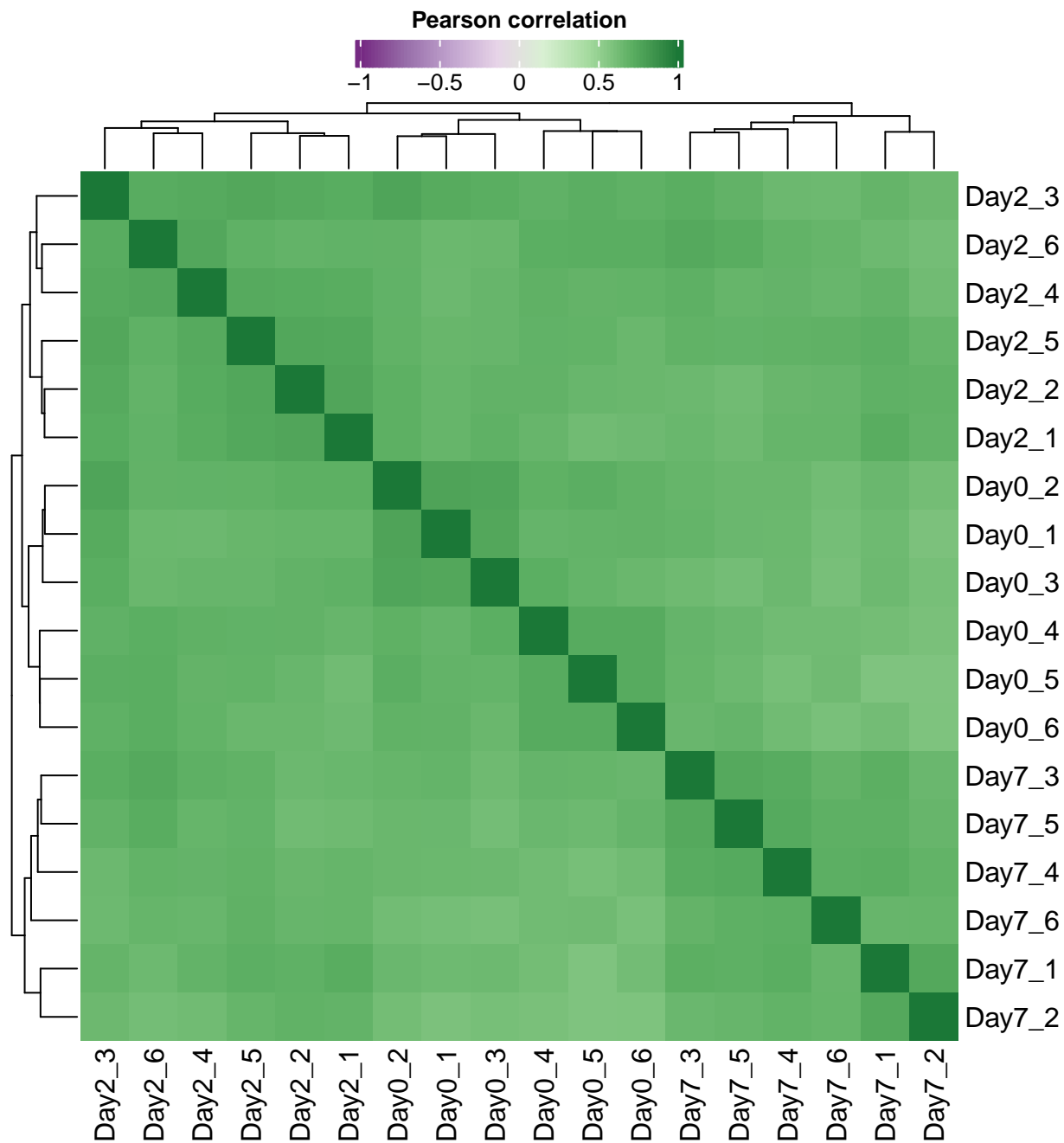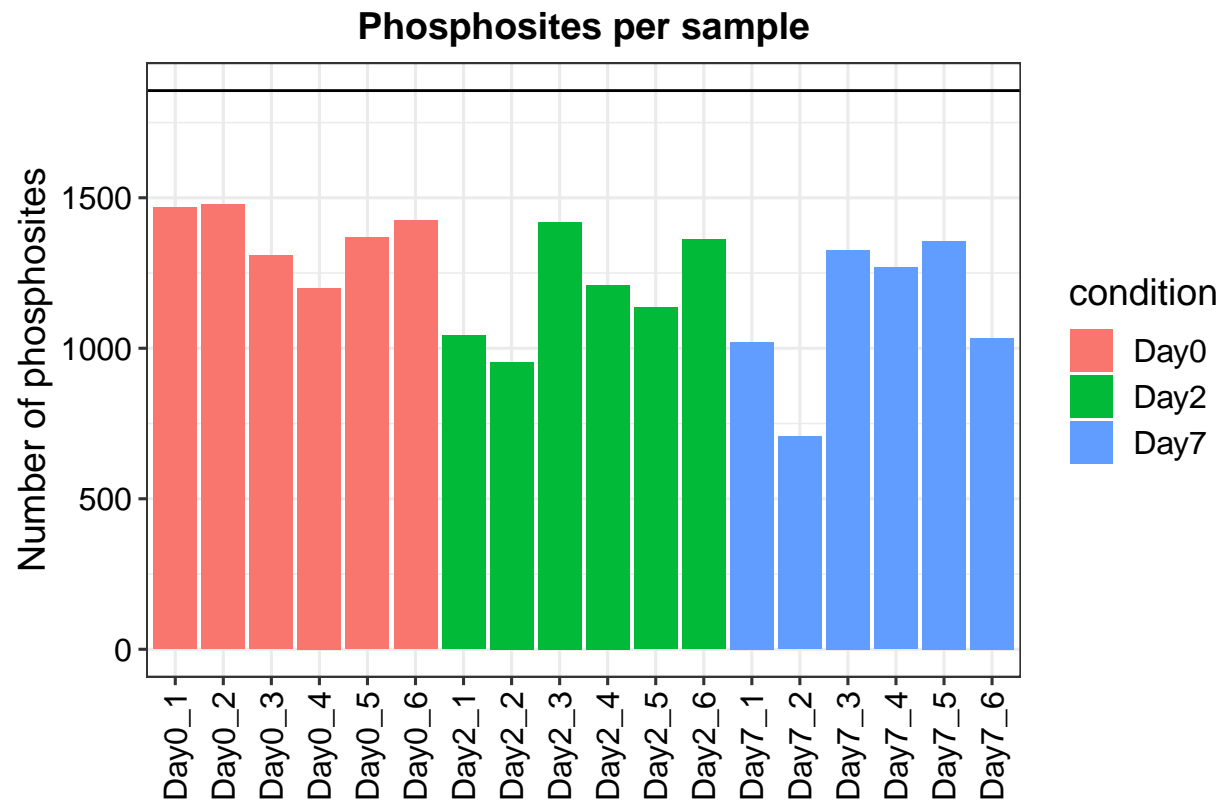
Sample Coefficient of variation (CVs) Also called relative standard deviation (RSD), it is present as a histogram plot, and illustrates the degree of variation relative to the overall mean.
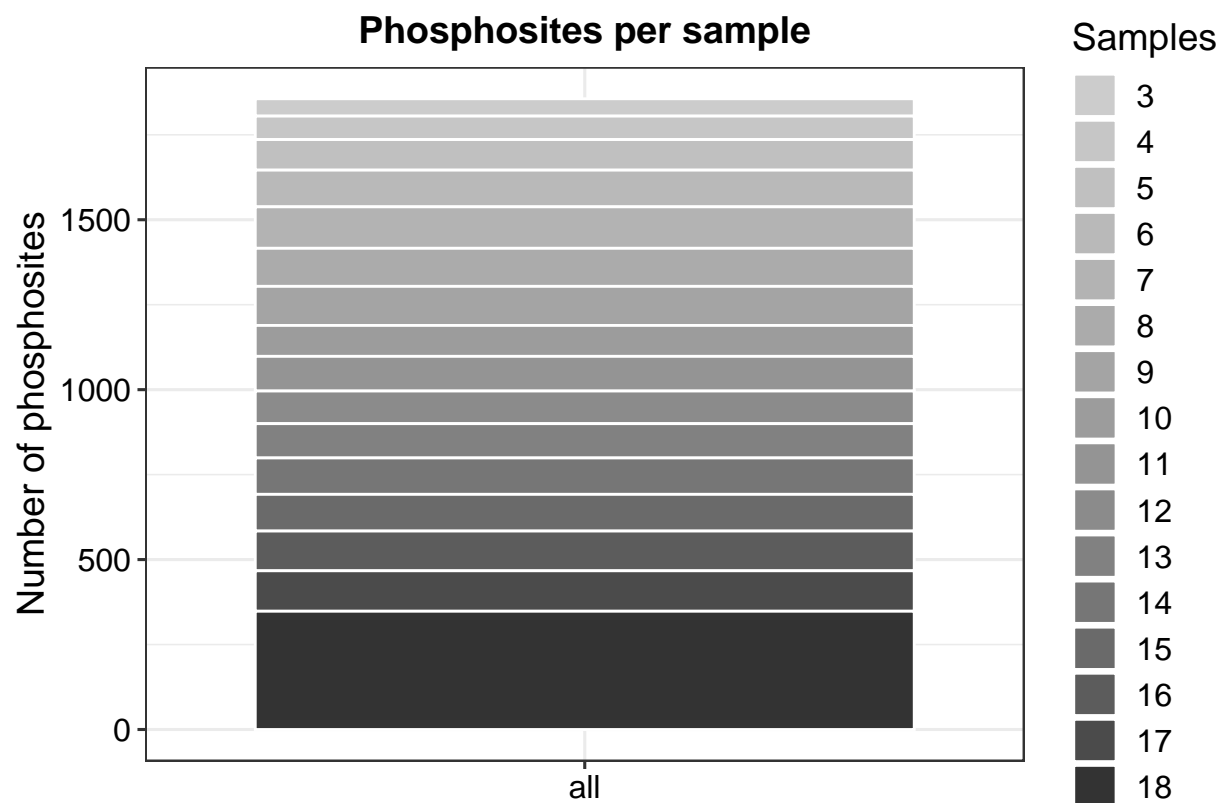
Sample Coefficient of Variation
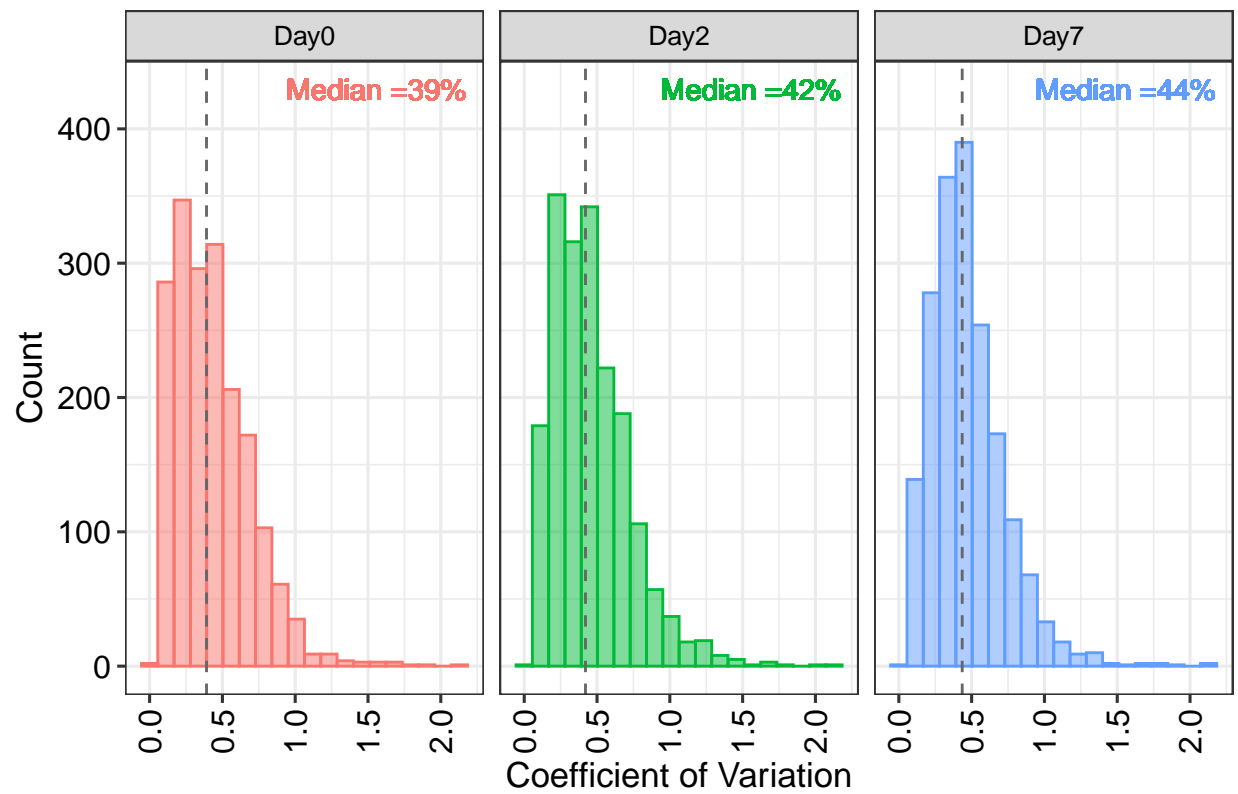
Proteomics Experiment Summary

Phosphosite quantified per sample (after pre-processing).

Phosphosites per sample

Phosphosite overlap in all samples.
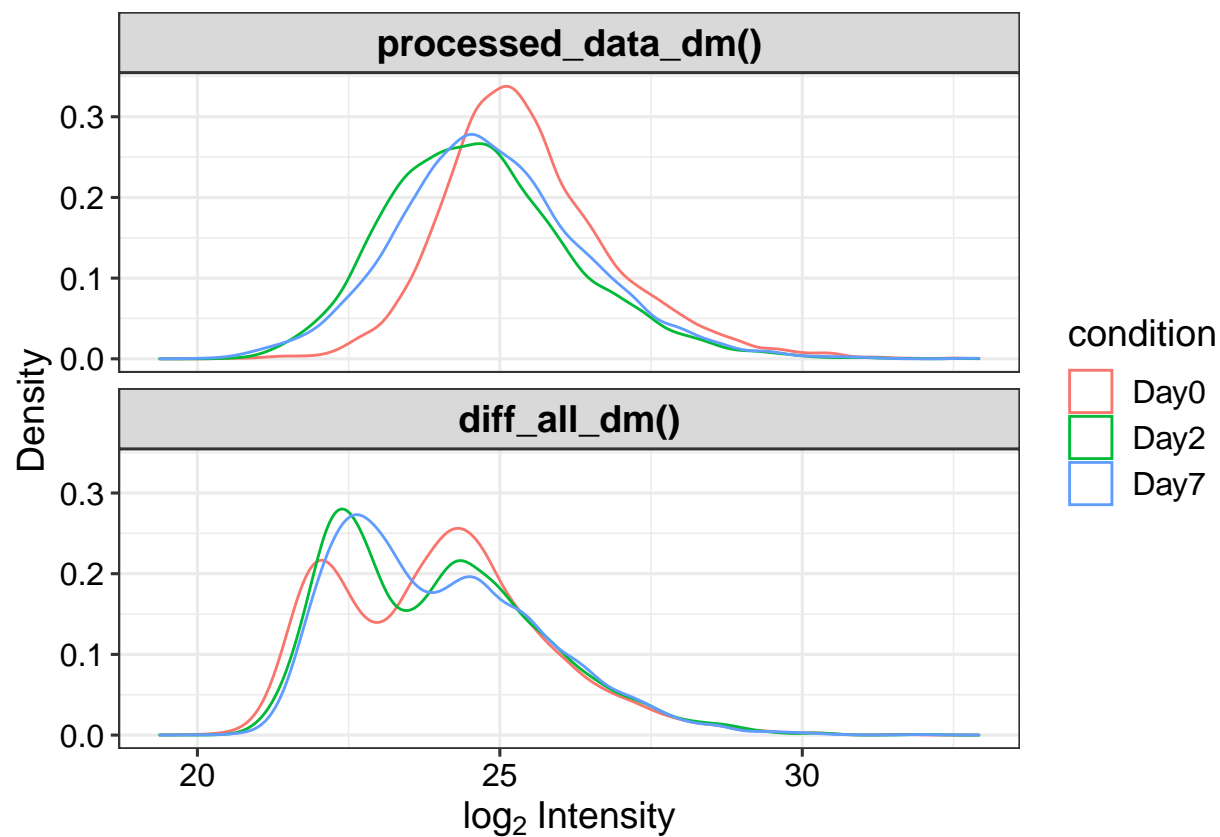
Phosphosites per sample

Missing Value handling

Missing value heatmap A heatmap for phosphosites with missing value in each dataset. Each row represent a phosphosite with missing value in one or more replicate. Each replicate is clustered based on presence of missing values in the sample.
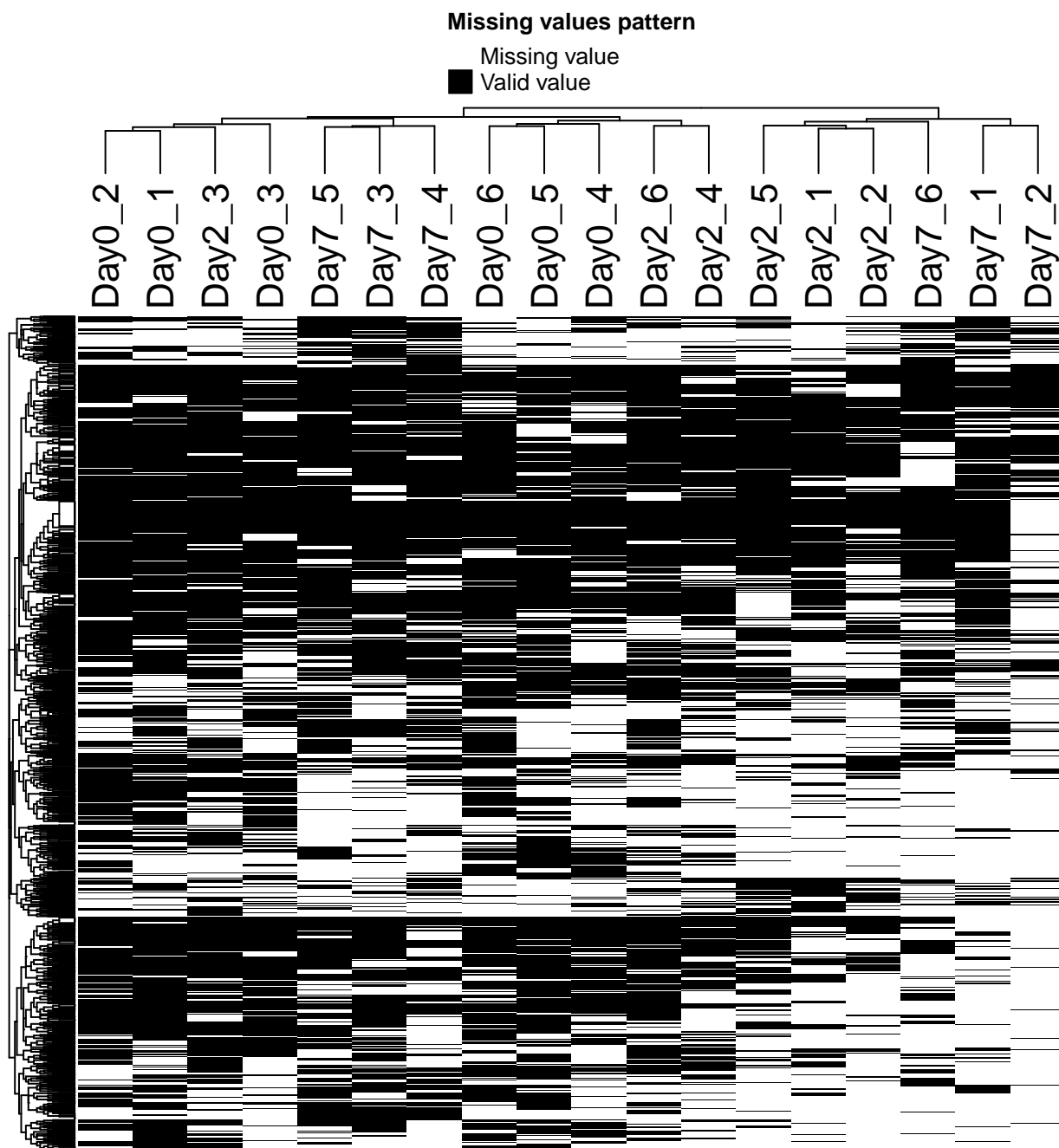
Missing value distribution   Phosphosite expression distribution before and after imputation. The plot showing the effect of imputation on phosphosite expression distribution.

Differential Expression Analysis (Results Plots)

Heatmap A plot representing an overview of expression of all significant (differencially expressed) phosphosites (rows) in all samples (columns).
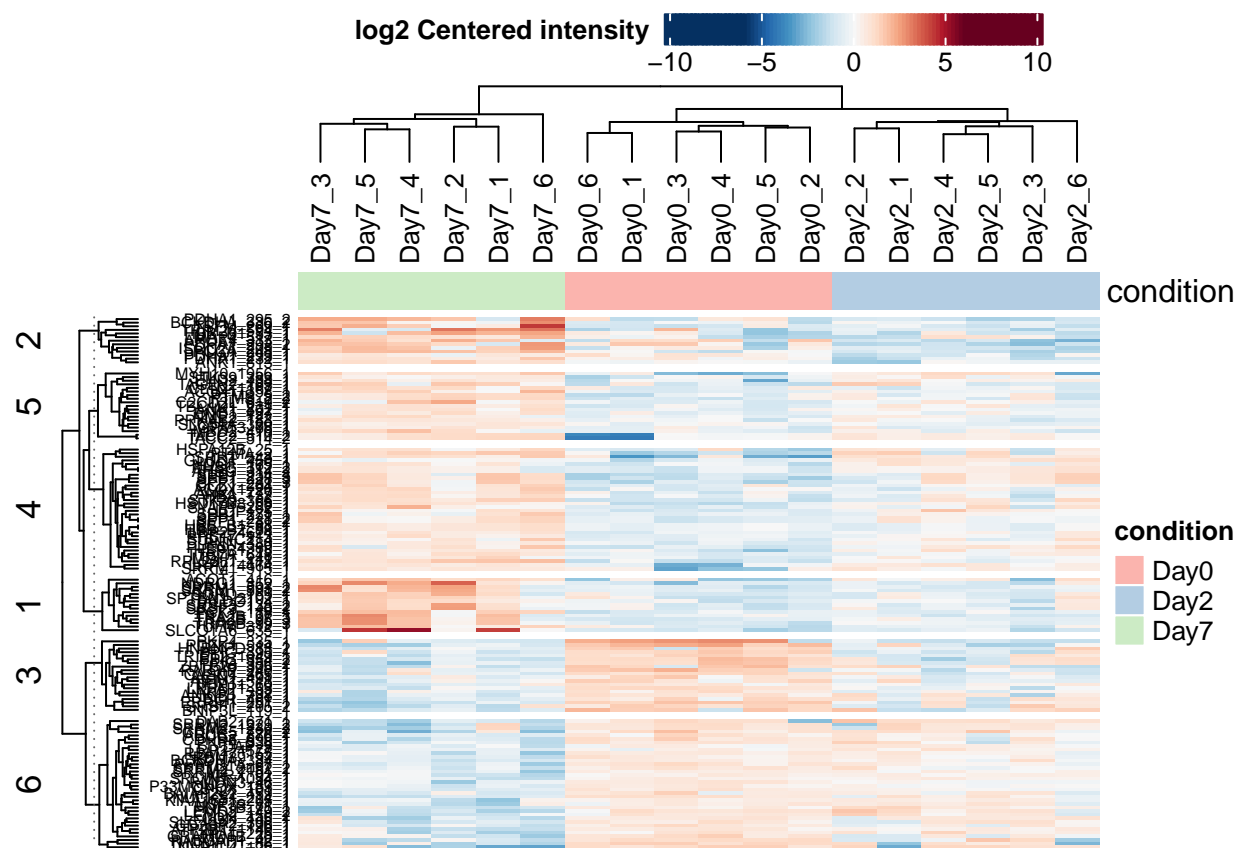
Volcano Plots Plots illustrate statistically significant (P value) versus change magnitude (fold change) in different contrasts.

Day0_vs_Day2

Day0_vs_Day7

Day2_vs_Day7