

FGCZ Two-Group Analysis for: p2084 Statistics for a Quantitative Protein Matrix

Functional Genomics Center Zuerich

19 July 2016

Input Matrix

Experiment is called: OID2264

Numbers

The protein matrix is read like this:

All are QUANTIFIED, as filtering should be done before uploading the matrix file. Zero values will eventually be imputed depending on the argument.

The total number of proteins in this experiment is: 3516

Condition	# samples
KO	6
WT	6

The Groups

Here the files in each group:

Condition	Raw.file
KO	20160217_04_KO_3
KO	20160217_05_KO_2
KO	20160217_08_KO_5
KO	20160217_13_KO_1_rep
KO	20160217_14_KO_4
KO	20160217_15_KO_6
WT	20160217_02_WT_4_rep2
WT	20160217_03_WT_2_rep
WT	20160217_07_WT_1
WT	20160217_09_WT_5
WT	20160217_10_WT_3_rep
WT	20160217_12_WT_6_rep

Proteins Used for Quantitation

The input matrix has the following structure.

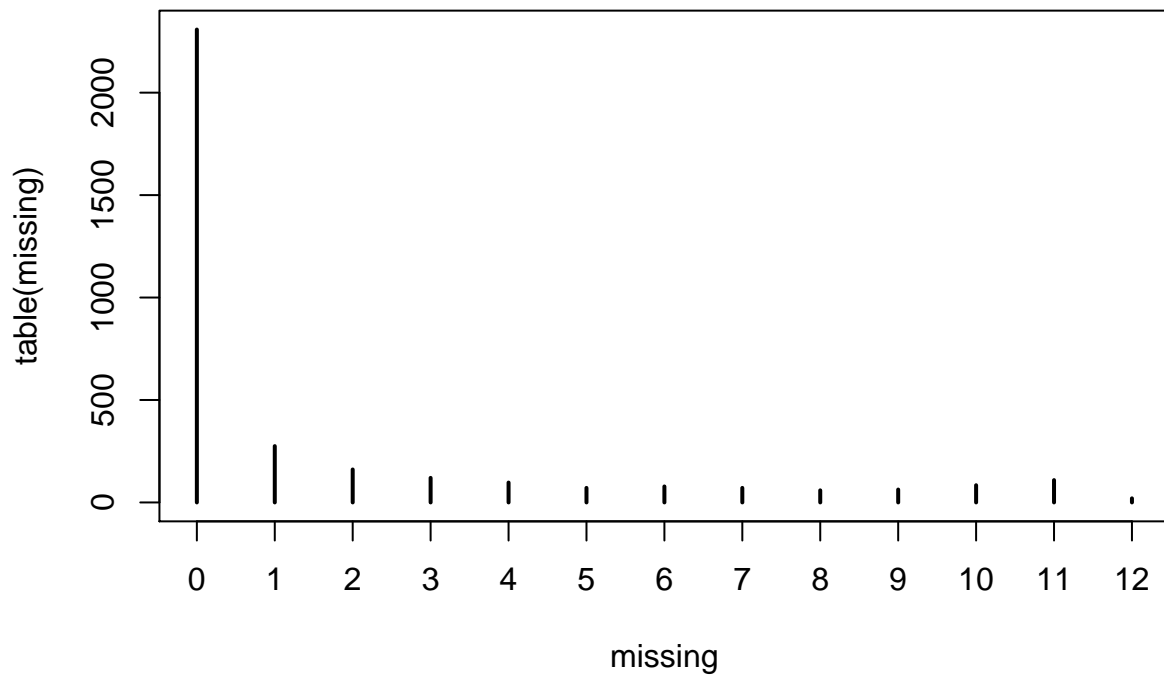


Figure 1: nr of missing values per protein

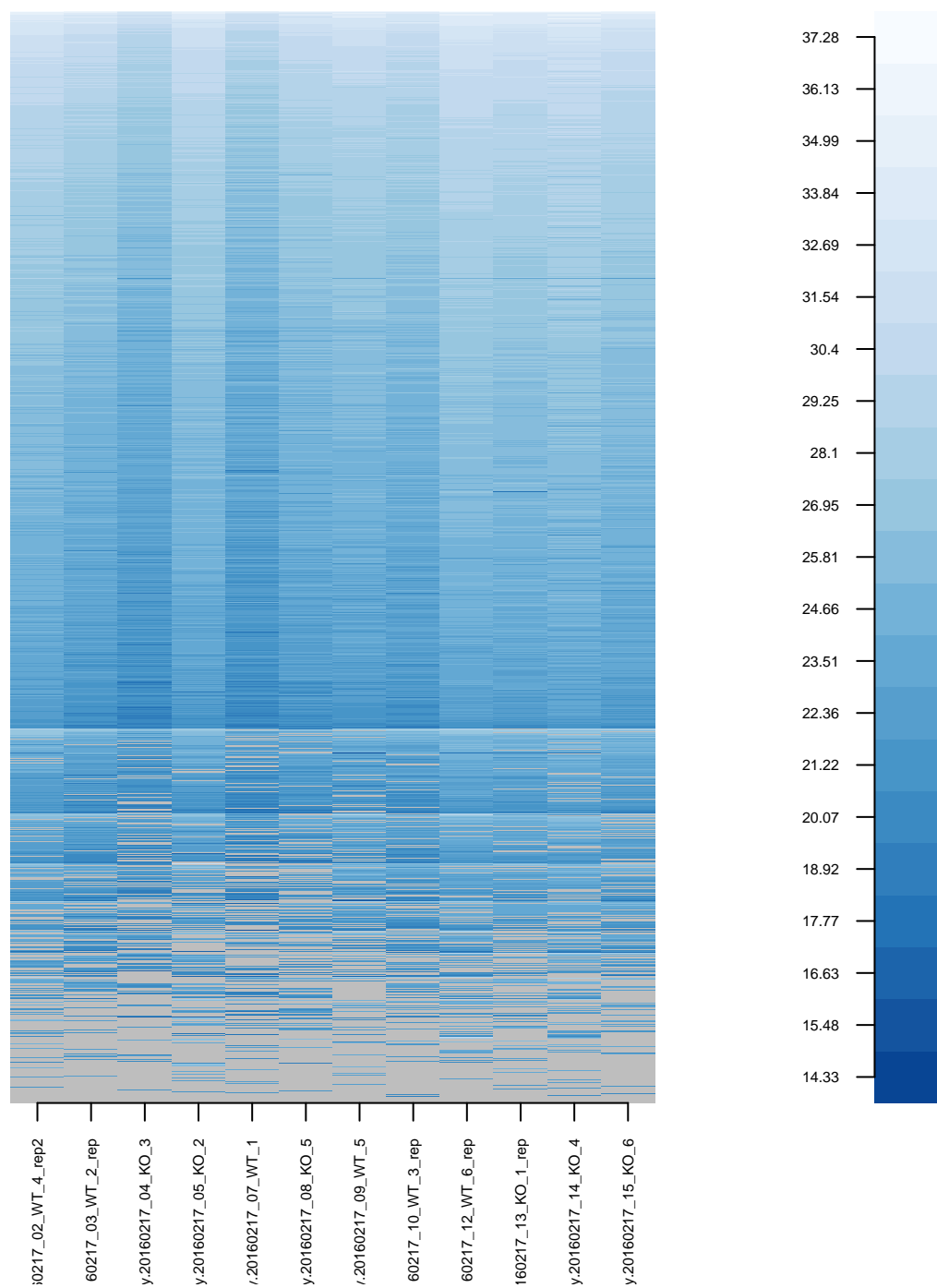


Figure 2: Heatmap for quantifiable proteins sorted by missigness and intensity (log2). Gray - missing values

Distribution of intensities

Shown in Figure @ref(distributionRaw) are the un-normalized values while in Figure @ref(normalized) are the z transformed values (subtracted median and divided by variance).

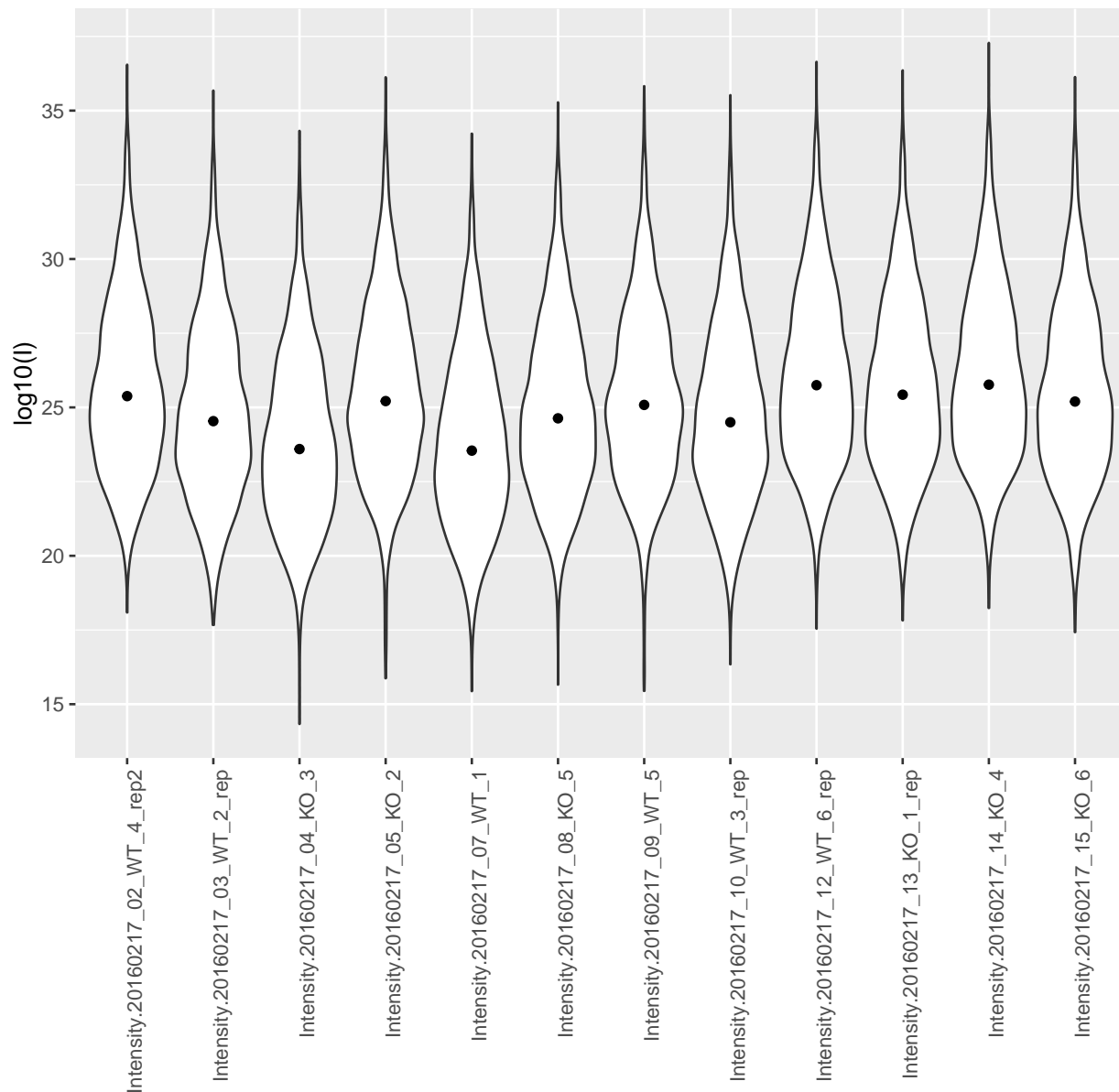


Figure 3: Density plot for quantifiable proteins (log2 transformed)

The scaling factors are visualized in Figure @ref(scaling). It shows with what factor the individually samples are normalized. Large differences in the scaling factors are critical.

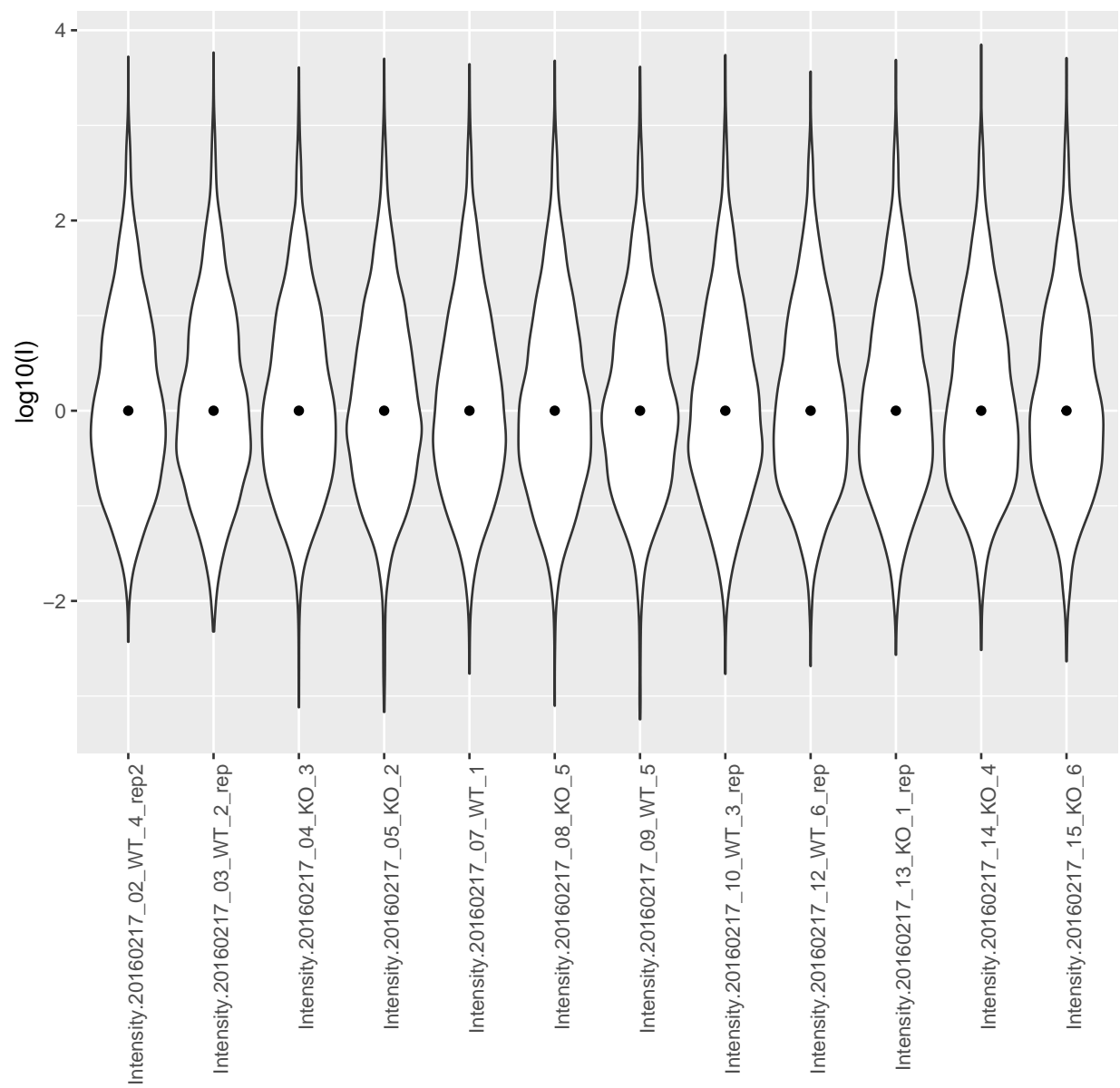
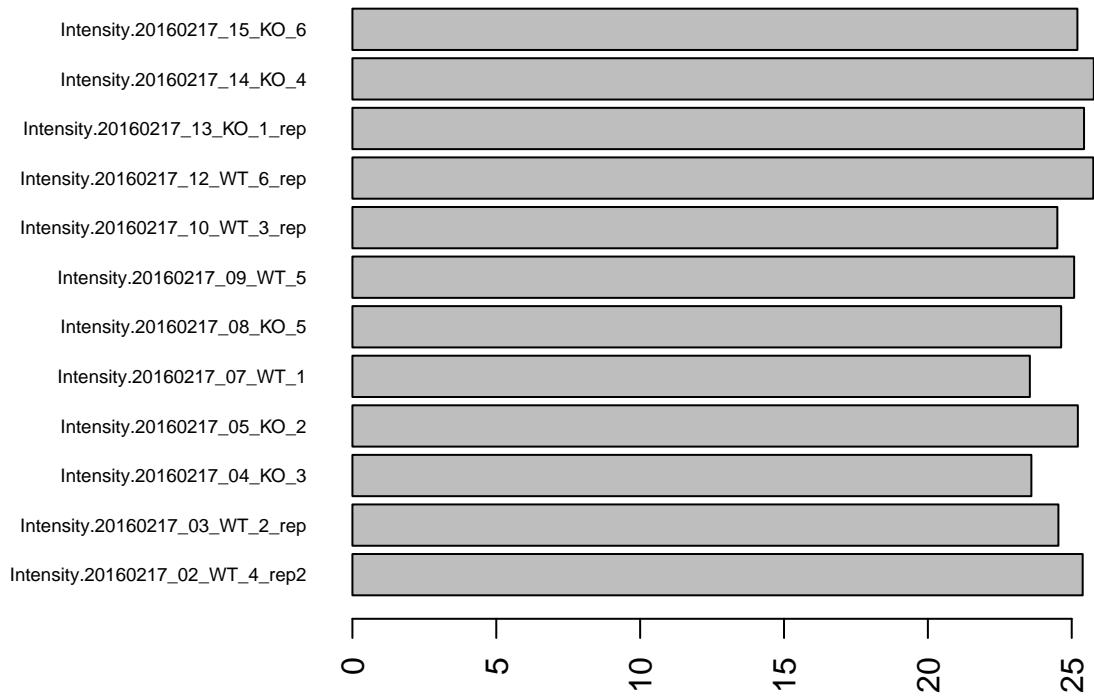


Figure 4: Density plot for normalized values (log2)

median correction



standard deviation correction

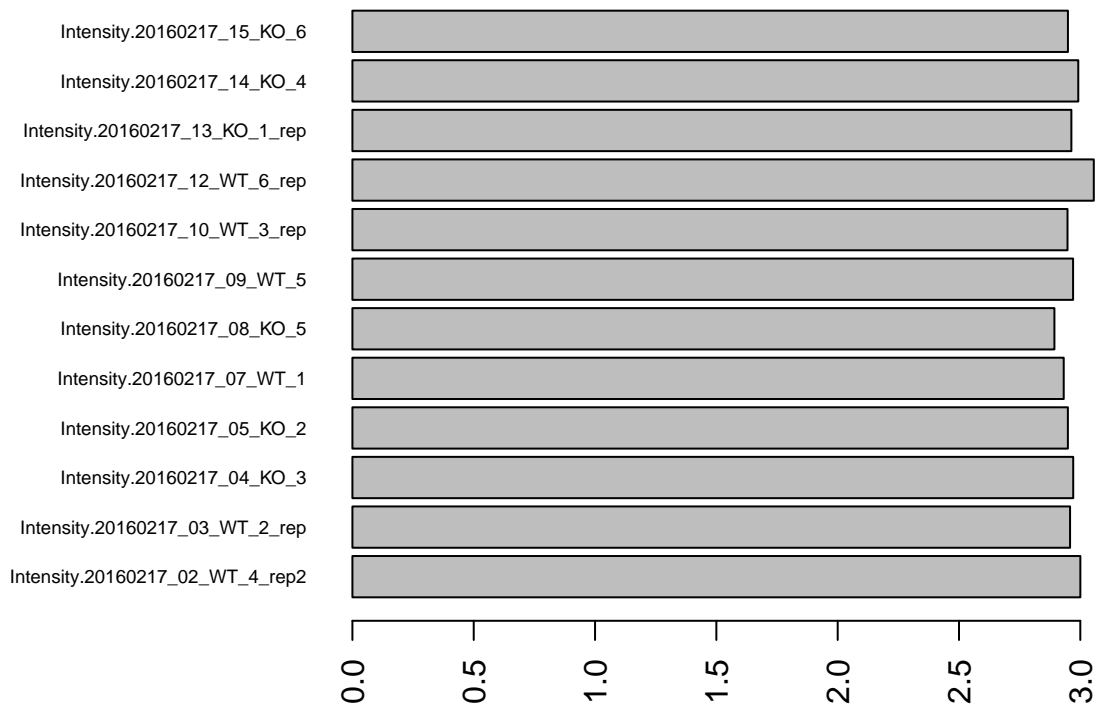


Figure 5: Applied scaling factors for normalization (calculated using median normalization).

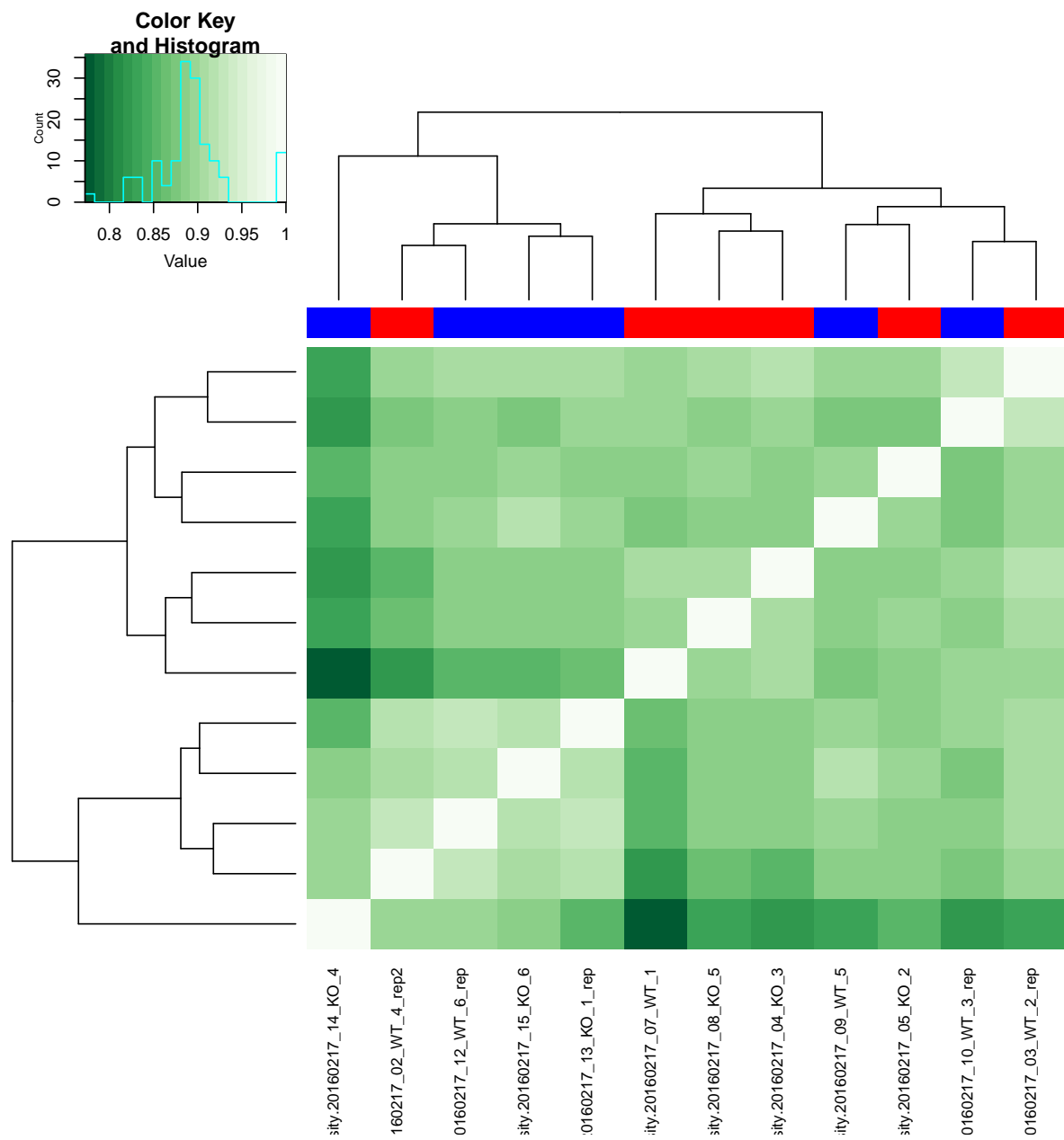
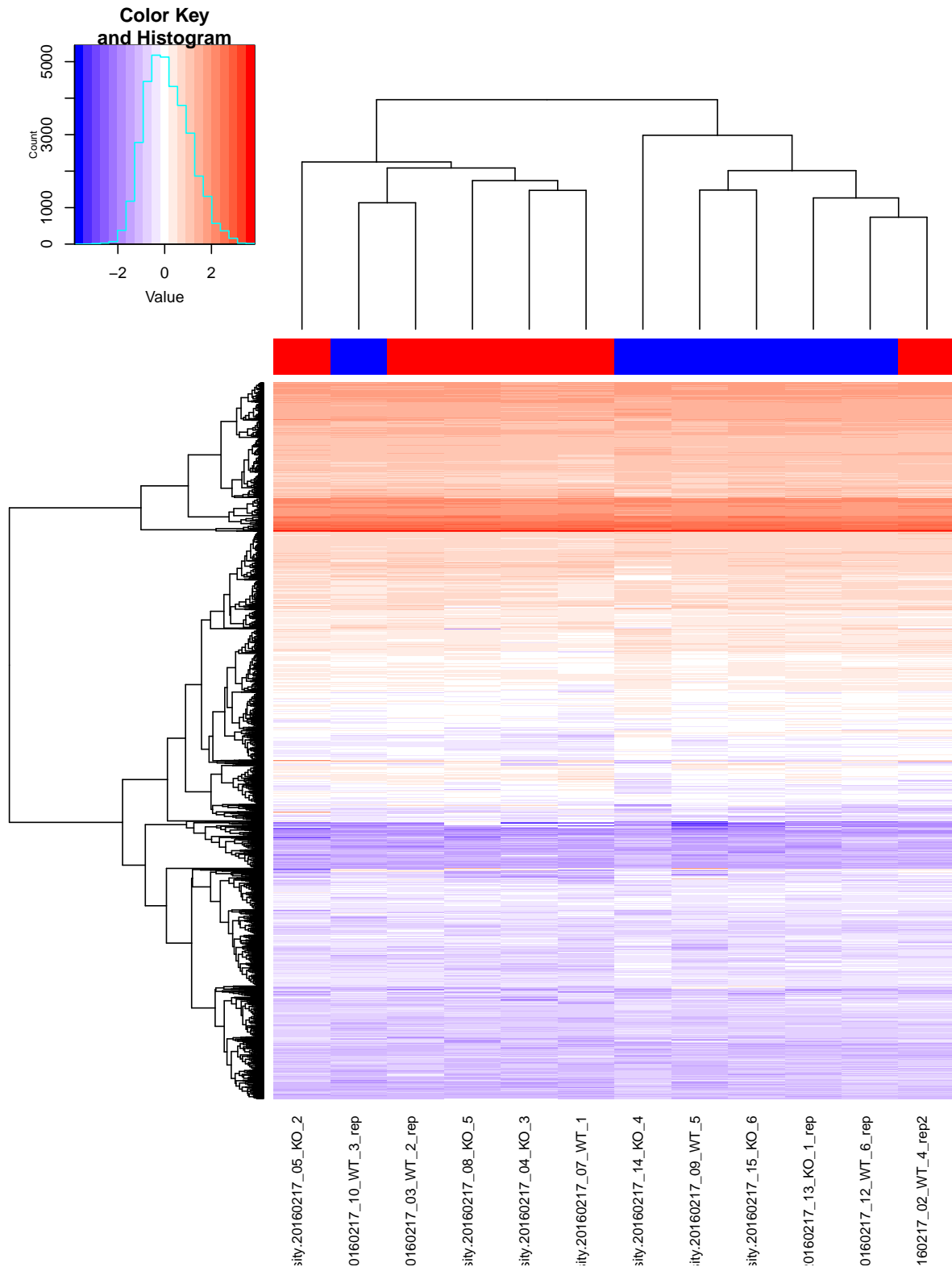


Figure 6: Heatmap of correlations (spearman) between samples.

Heatmaps and Clustering for Proteins

In Figure @ref(heatmapData) we show how samples are clustering depending on their protein expression profiles. Also proteins are clustered according their profiles within the different samples.



FGCZ - Two Group Analysis