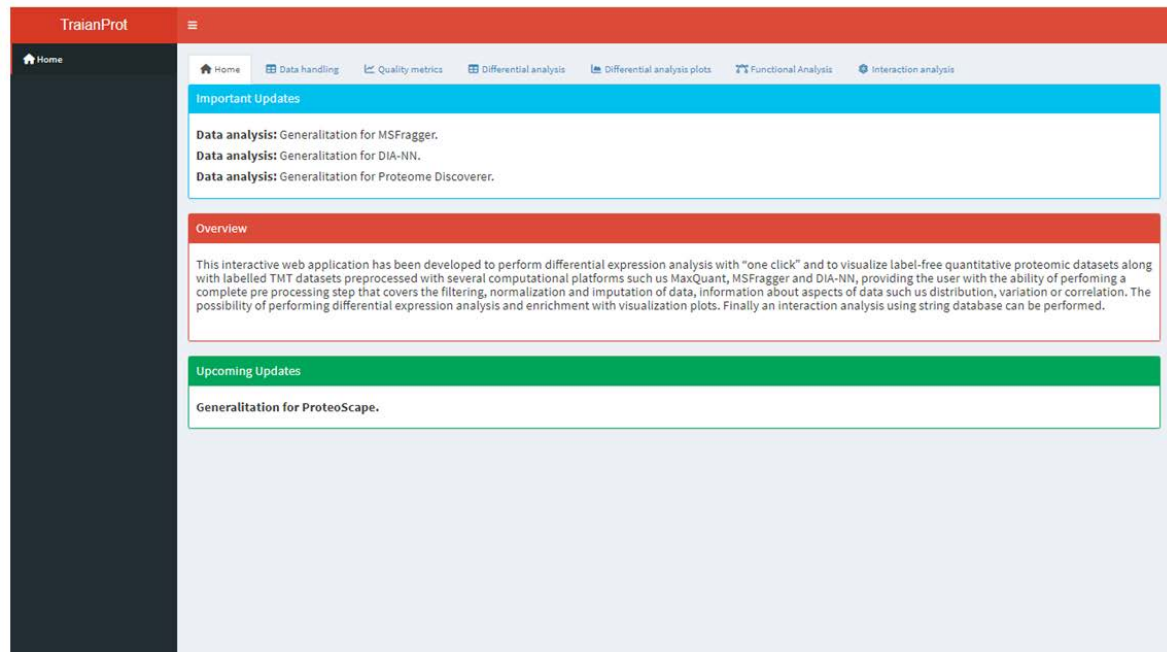


TraianProt is a software platform focused on the down streaming analysis, achieving the differential expression and visualization of shotgun label free/labelled proteome datasets. TraianProt supports several formats from different proteomic computational platforms such as MaxQuant, MSFragger, DIA-NN and Proteome Discoverer in both Data Dependent and Data Independent Acquisition mode. Among its functionalities a preprocessing, differential analysis, functional analysis and protein interaction analysis step can be underlined along with the visualization of the previous steps.

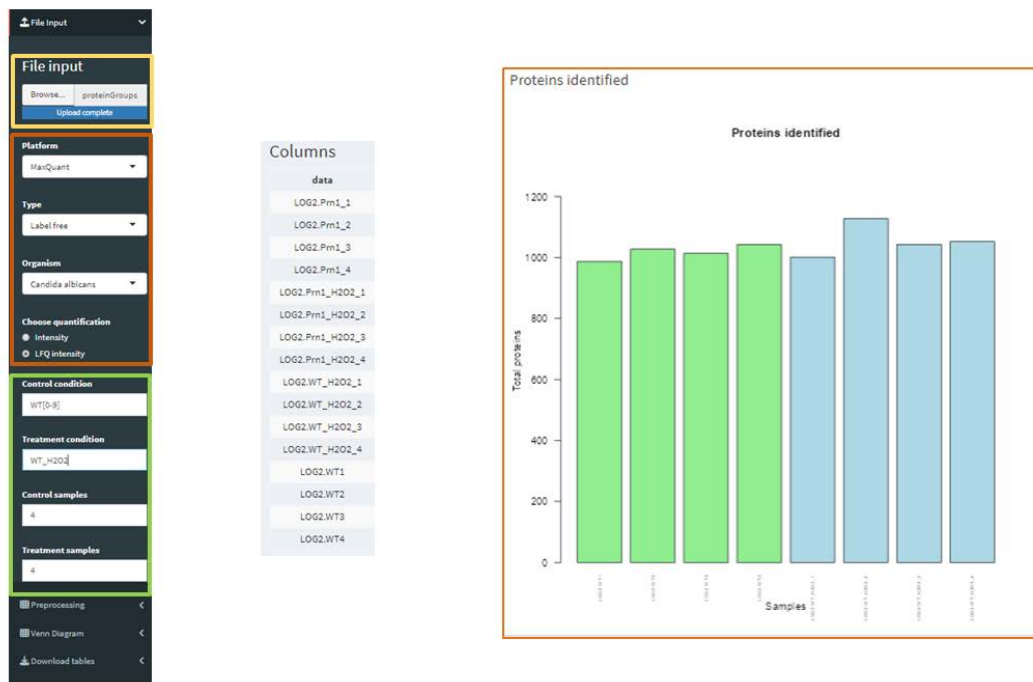


The software platform is divided into six sections:

The **Preprocessing** section is focused on data preprocessing:

In the first step the dataset is loaded in the “File input” button. The selection of the computational platform that yielded the dataset loaded is needed. The program works with the UniProt database format and the *Candida albicans* CGD database format for the protein identifiers in the dataset. The option “Intensity” or “LFQ intensity” can be chosen depending on the type of intensity we want to work with.

The software has a peculiar method for the selection of samples for each condition. Clicking the button “Show columns” will show all the sample intensity columns ‘name. This will allow us to establish a regular expression that matches all control and treatment sample names. By this approach we get rid of the typical experimental annotation file whose confection is laborious. Finally, it is necessary to set the number of samples for each condition.

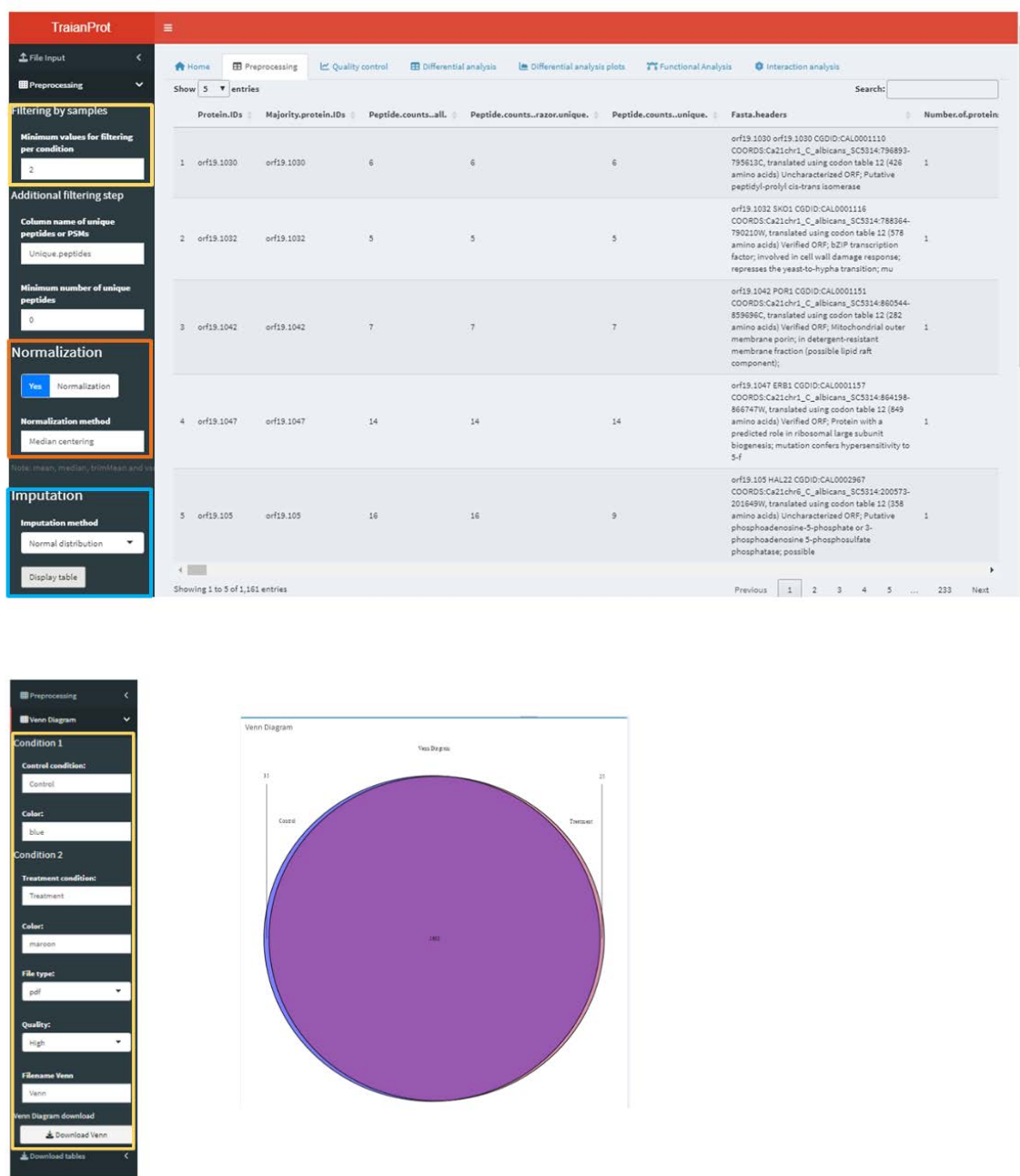


Next the pre preprocessing steps are performed:

- **Filtering:** the aim of this functionality is to get rid of those proteins which has not been quantified in a certain number of samples between both conditions in the experiment. To perform this action a minimum value for filtering per condition is set, in the screenshot a value equal to 2 is set. An additional filtering step according to unique peptides can be made.
- **Normalization:** several types of normalization are included such us mean, median, trimMean and vs.
- **Imputation:** the imputation step is executed choosing between two methods: according to a normal distribution and through the K-Nearest Neighbors algorithm.

Once the previous actions are made, the “Display table” button will show the resulting dataset where common proteins and presence and absence proteins can be seen and downloaded.

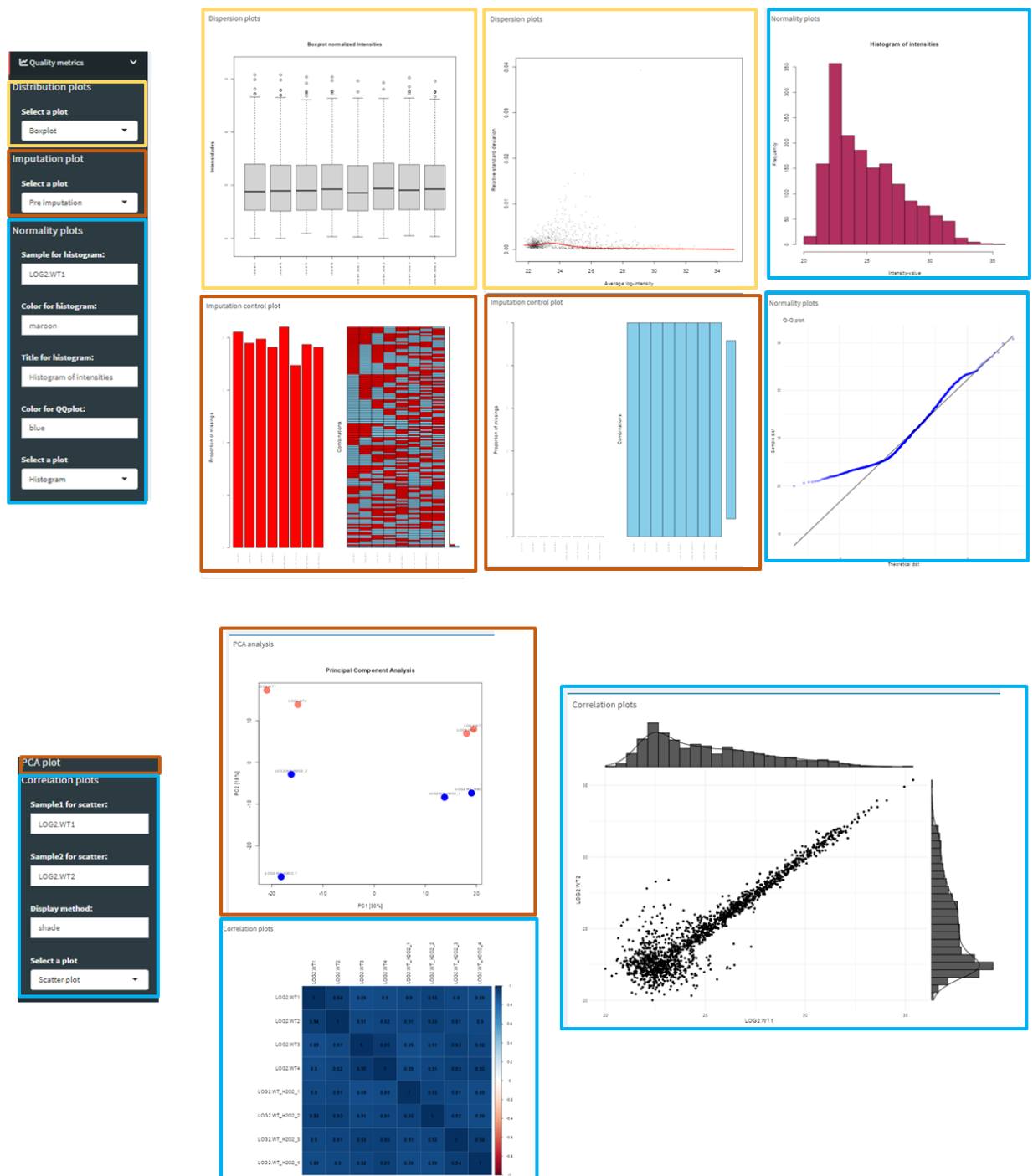
It is necessary to point out the representation of a Venn Diagram in this section, with several options for labeling of conditions, color representation and format (.tiff, .png, .jpeg).



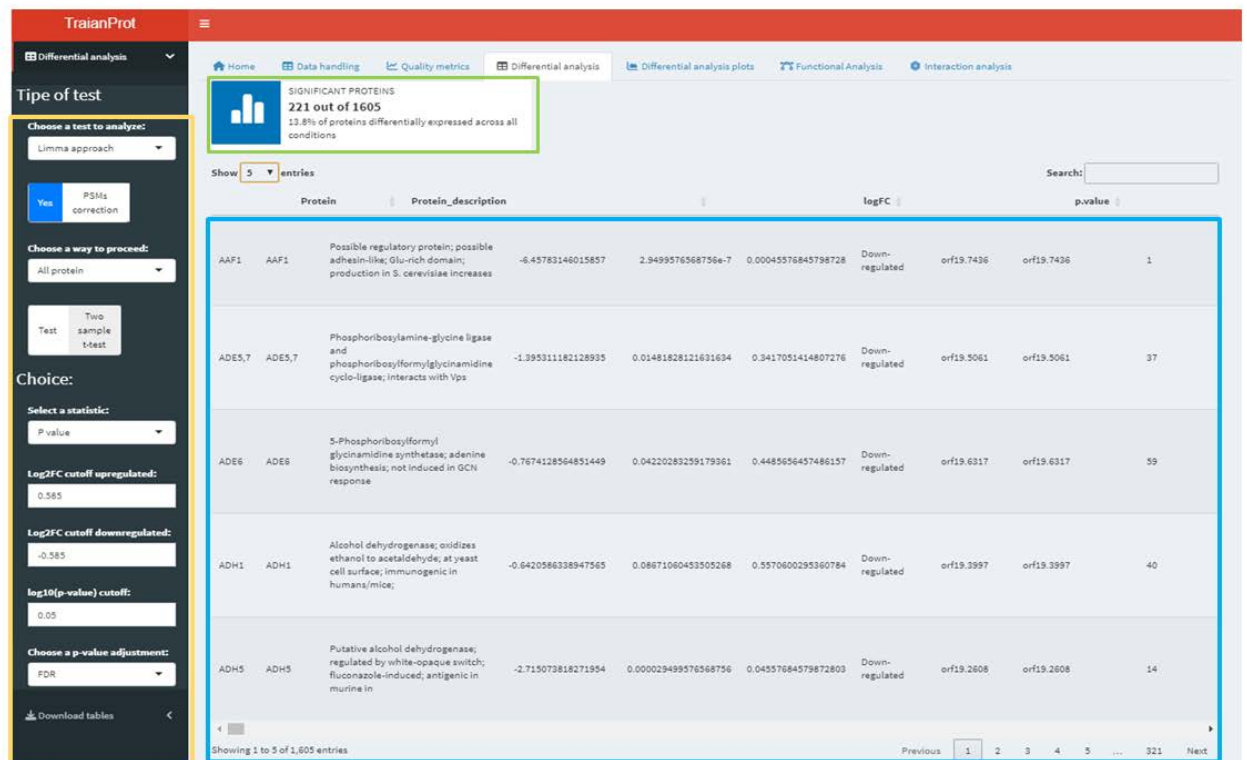
The **Quality metrics** section covers a group of plots that describe the data nature (distribution, dispersion, missing values proportion in our data...) Inside this section we can highlight the following sections:

- Dispersion plots: includes a boxplot and dispersion plot.
- Imputation plots: includes a representation of the amount of missing values in the data before and after performing imputation.
- Normality plots: covers a set of plots whose purpose is to the representation of data's distribution, including histogram of proteins abundances and a Q-Q plot.
- PCA plot: plot with Principal Component Analysis.
- Correlation plots: includes a Scatter plot and correlation plot.

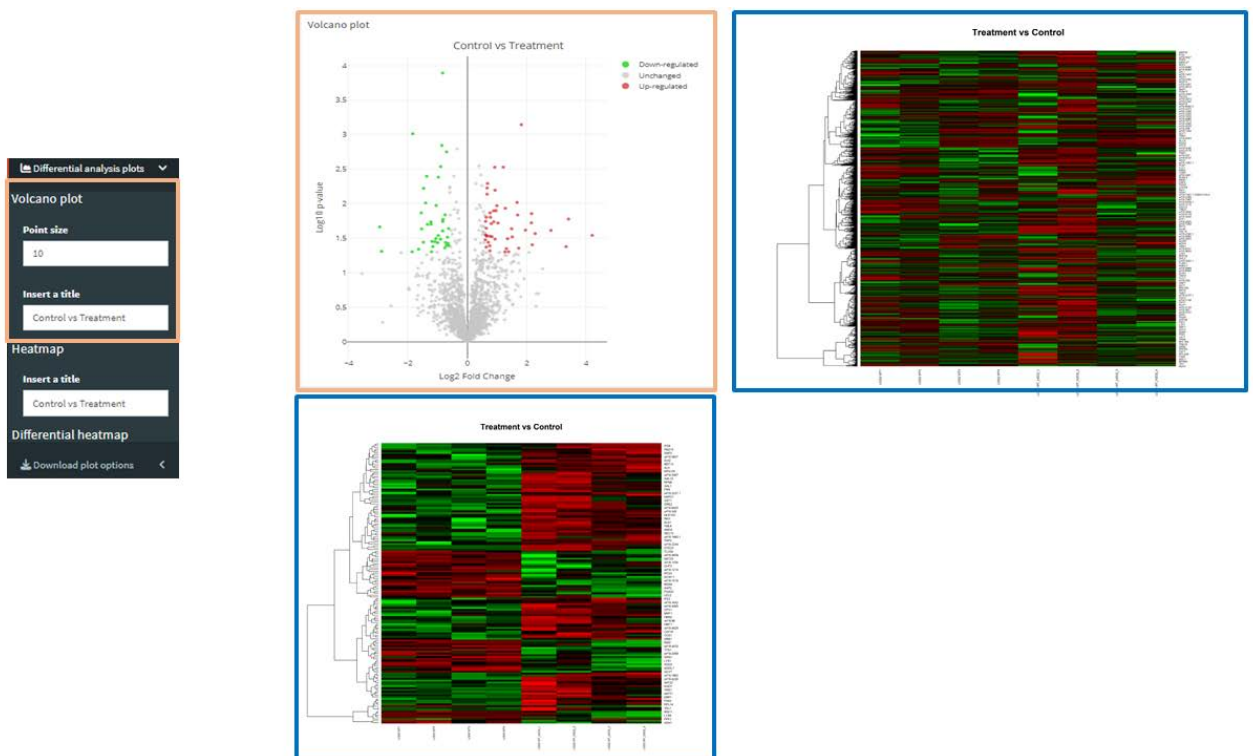
All the previous plot can be downloaded in a paper ready format (.tiff).



The **differential analysis** section allows to perform a differential expression analysis using a two sample T test (from limma package, recommended for dealing with a big number of samples or base R). a paired test specification can be made. A variance correction for each protein according to PSMs identified can be made if the limma package is selected. Finally, several parameters such us Fold change, p value or q value threshold are set. The option for downloading the resulting table is shown below.



The **differential analysis plot** section covers several plots such as volcano plot, heatmap and a heatmap with those proteins that exhibit significant abundant changes. The aim of this section is the representation of those proteins that exhibit significant changes in their relative abundance. The possibility of downloading the plots is shown below.



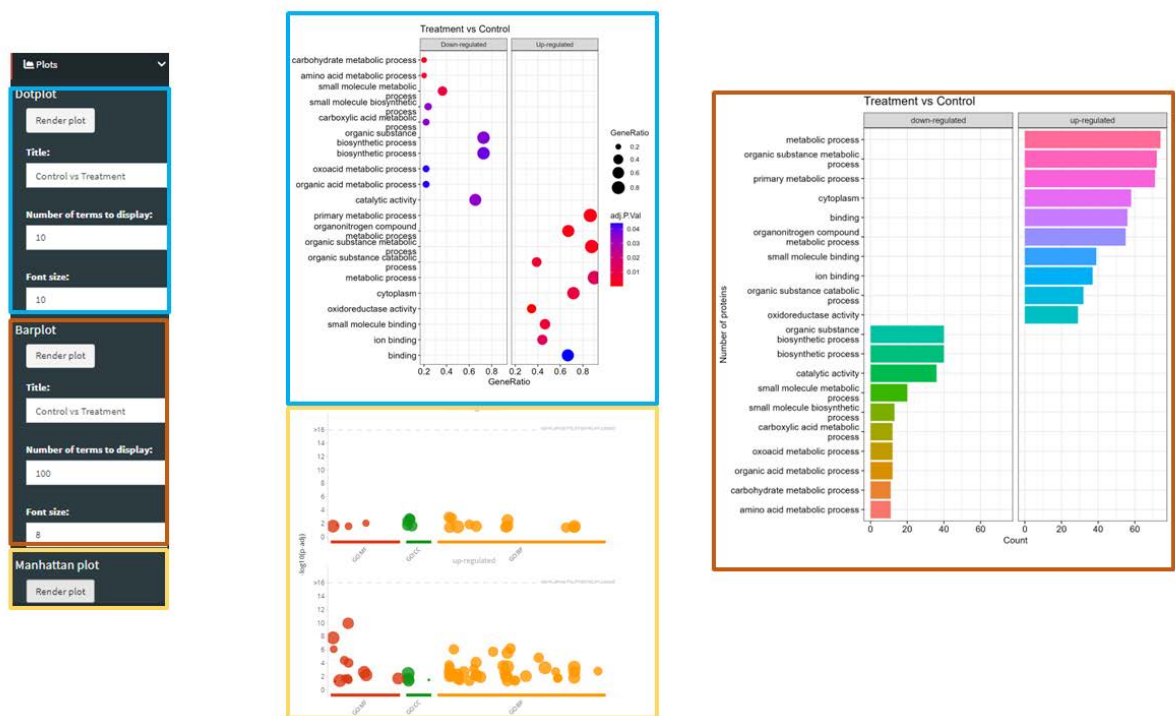
The **functional analysis** section carries out a functional analysis for those proteins that exhibit significant changes in their relative abundance, the organism is specified, and a p value threshold is set.

The organism text input needs the correct format for the organism of interest, complete list in the following link: <https://biit.cs.ut.ee/gprofiler/page/organism-list>

Several plots that illustrate the different biological processes for which our dataset is enriched are included. A Dotplot, Barplot and Manhattan plot are included.

Download options for the resulting table of protein and plots are included in the section below.

Cluster	Category	ID	Description	p-value	adj.PVal	query_size	Count	term_size	effective_domain_size	geneID
GO:0005975	down-regulated	GO:0005975	carbohydrate metabolic process	0.001151925427032523	0.004789584670293123	55	11	204	5922	FUN31/G
GO:0006520	down-regulated	GO:0006520	amino acid metabolic process	0.001680682543951494	0.006197005250726103	55	11	212	5922	GLN1/M
GO:0044281	down-regulated	GO:0044281	small molecule metabolic process	0.0032439003402394567	0.008542270896109027	55	20	748	5922	COQ5/EC
GO:0016052	down-regulated	GO:0016052	carbohydrate catabolic process	0.01415460916137887	0.02430899238584631	55	6	64	5922	GP42/S
GO:0044283	down-regulated	GO:0044283	small molecule biosynthetic process	0.02610363474063084	0.03435961183411029	55	13	393	5922	COQ5/S
GO:0019752	down-regulated	GO:0019752	carboxylic acid metabolic process	0.02653083951747756	0.03435961183411029	55	12	337	5922	EC1/GP1
GO:1901576	down-regulated	GO:1901576	organic substance biosynthetic process	0.02881325105176357	0.03671365859821487	55	40	2723	5922	AAF1/CO
GO:0009058	down-regulated	GO:0009058	biosynthetic process	0.03195203676099957	0.03911600113749465	55	40	2733	5922	AAF1/CO
GO:0043436	down-regulated	GO:0043436	oxoacid metabolic process	0.03635060408740412	0.0428611600433571	55	12	348	5922	EC1/GP1
GO:0006082	down-regulated	GO:0006082	organic acid metabolic process	0.03738269697599395	0.04342989795740474	55	12	349	5922	EC1/GP1



Finally, the **interaction analysis** section performs a protein interaction analysis that exhibited significant changes in their relative abundance, both networks of proteins up-regulated and down-regulated along with some metrics are shown in this section.

