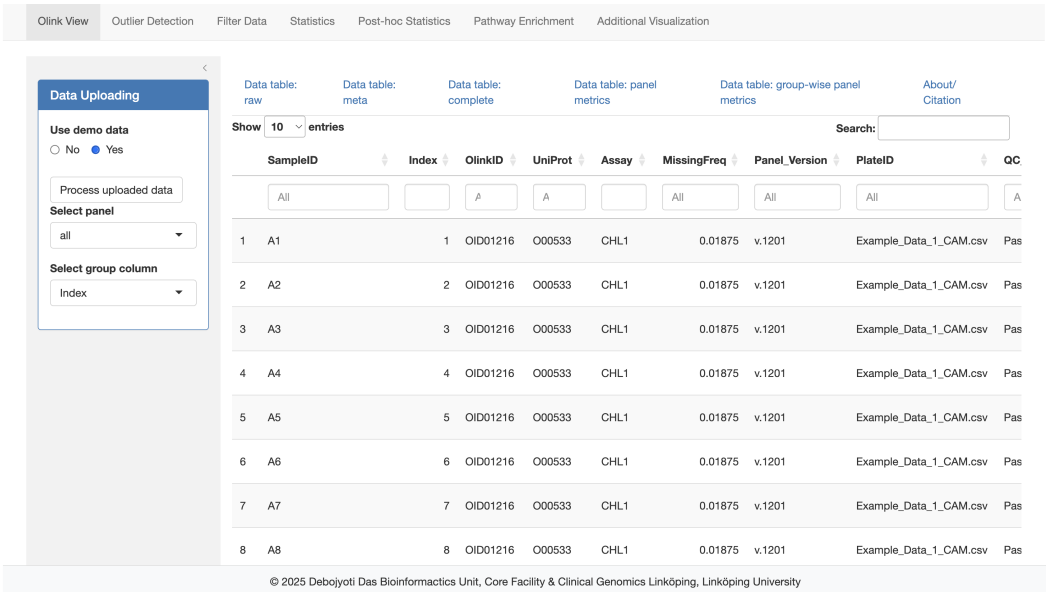
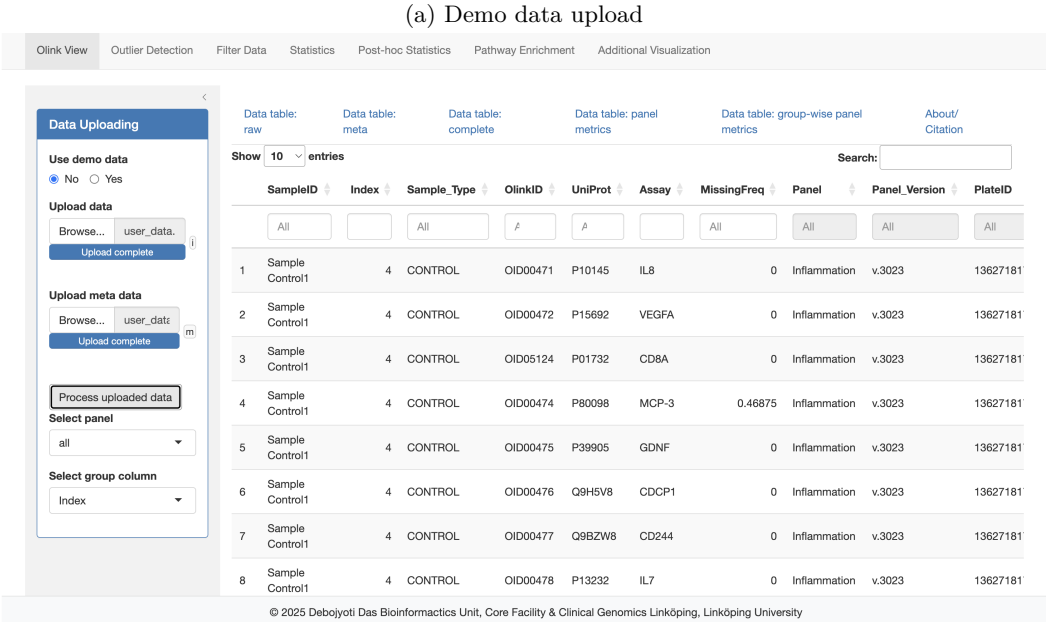
completeOlink - Manual

Here is an easy, step-wise guide to the use of completeOlink webserver for **OLINK** proteomics data analysis

A. Data upload:



(a) Demo data upload

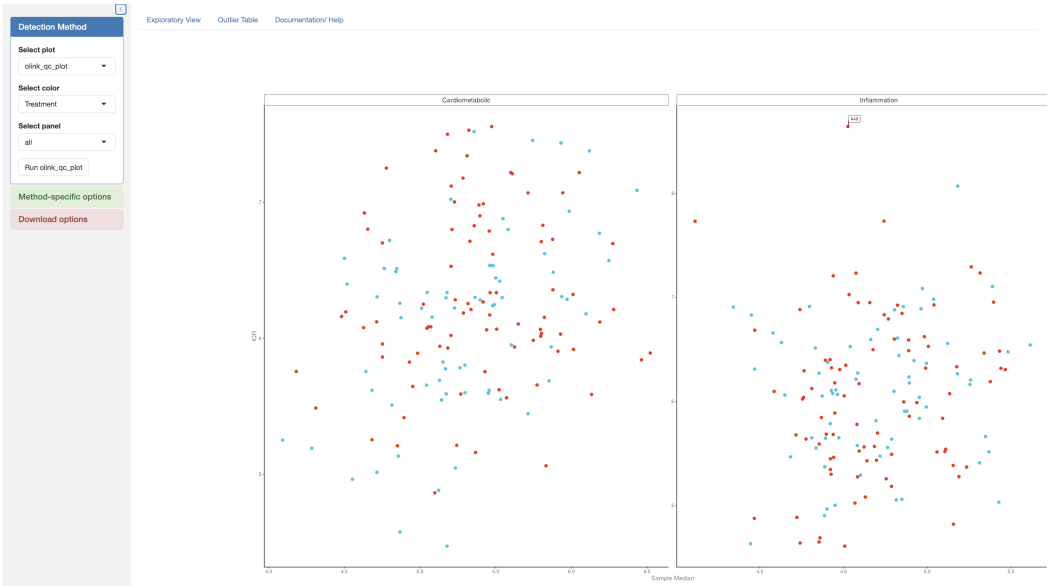


(b) User data upload

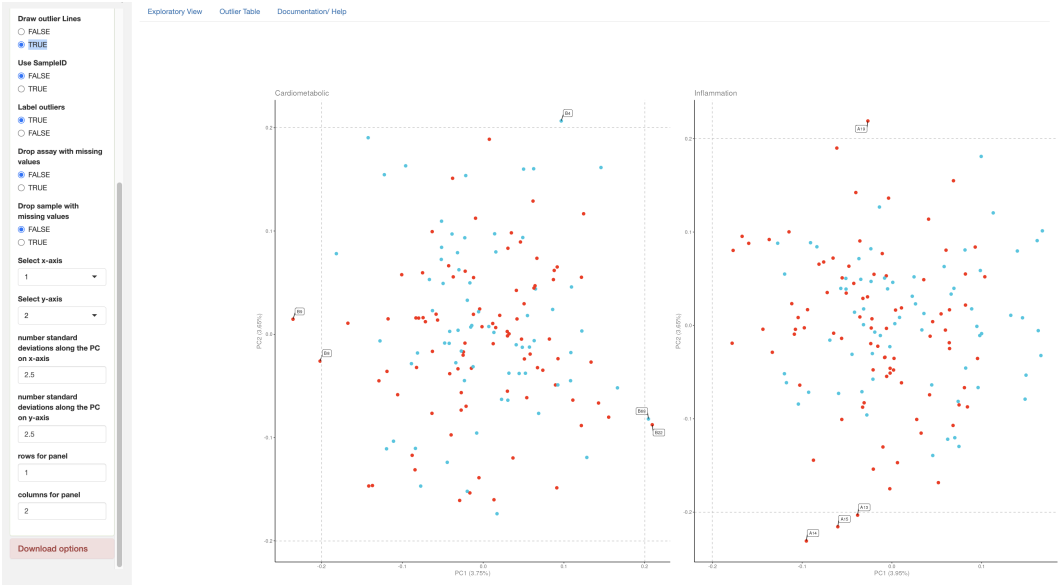
Figure 1: Olink View tab

The **Olink View** tab allows data upload. User can optionally use the demo data by selecting **"Yes"** for "Use demo data". Figure 1 depicts both scenarios.

User needs to upload the csv file (the CSV file is characterized by the presence of ";" as a delimiter instead of comma) and a sample manifest file (regular CSV file) where information about each unique "SampleID" in the npx raw data file is available. SampleID (s) are required to be character strings.



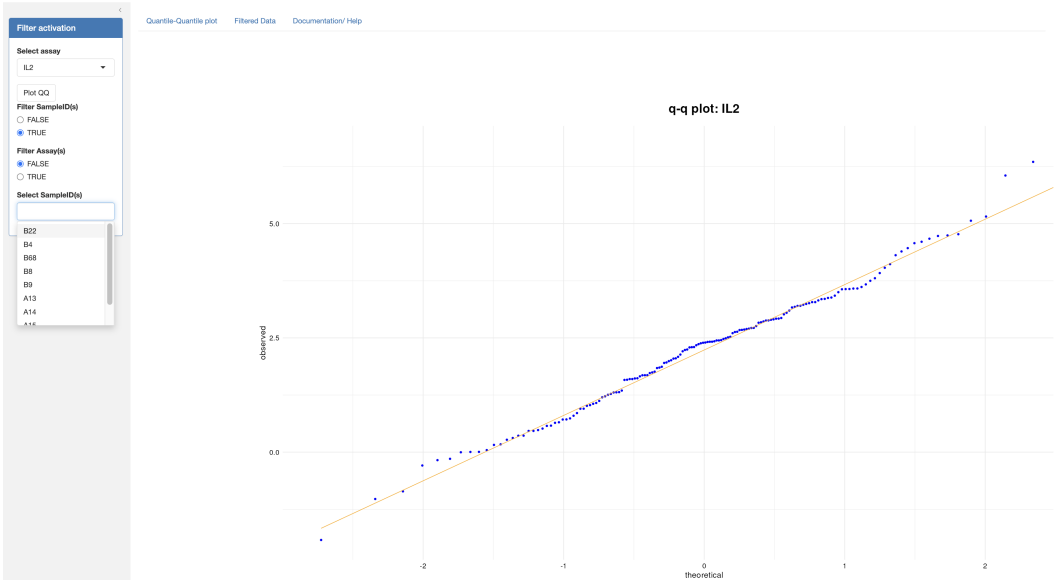
(a) method: olink_qc_plot



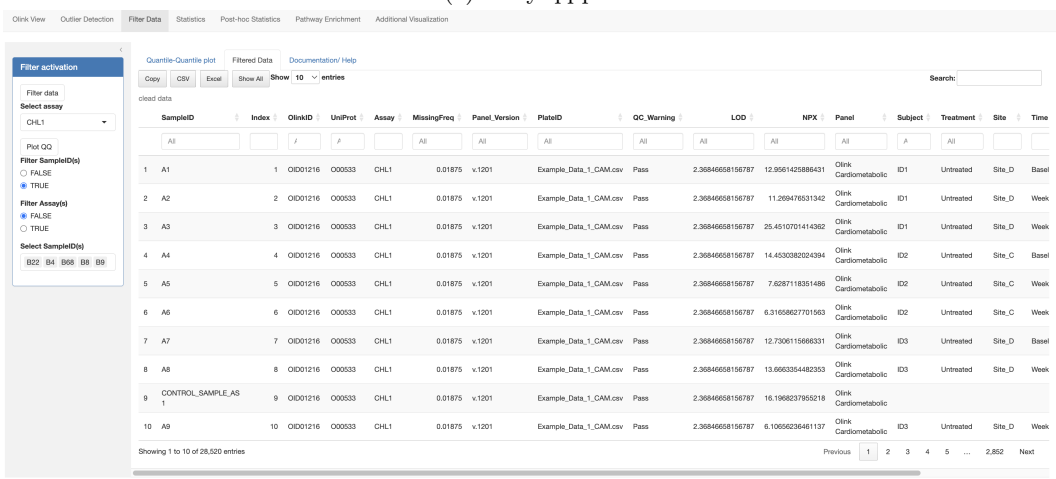
(b) method: olink_pca_plot

Figure 2: Outlier Detection tab (demo data; color: Treatment)

B. Outlier detection: Three outlier detection methods are available with control over parameters (optional) for accurate detection. The user can choose any of the three methods to find outliers or this step can be entirely skipped. See Figure 2. Note: If this step is used the outliers that are found in the most recent run/execution, are propagated to the next analysis step. Although, the user can choose to use the original data instead of the filtered data at every subsequent step. Results of **olink_umap_plot** can change in each run because of the inherent stochasticity of UMAP method.



(a) assay qq plot

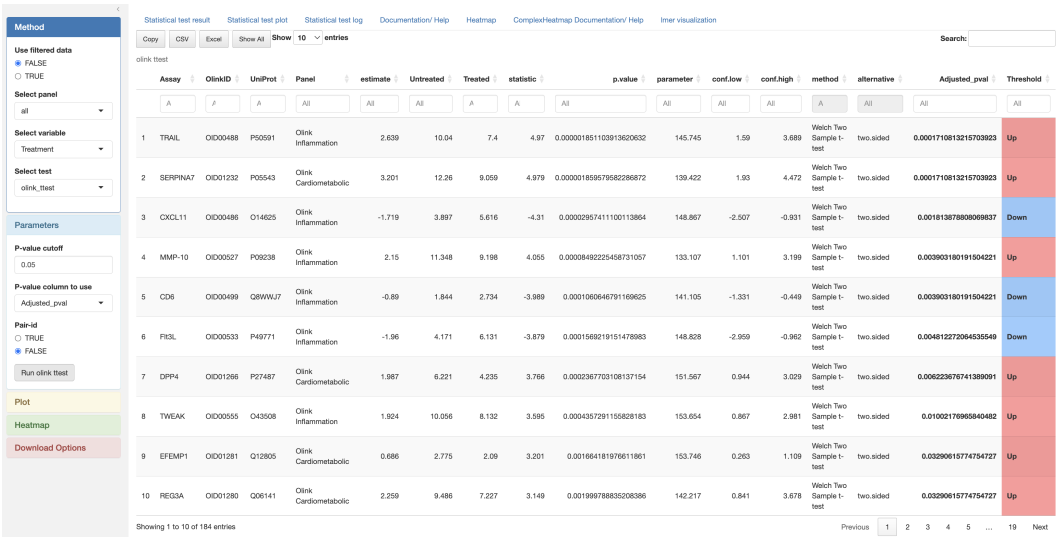


(b) sample filtered data

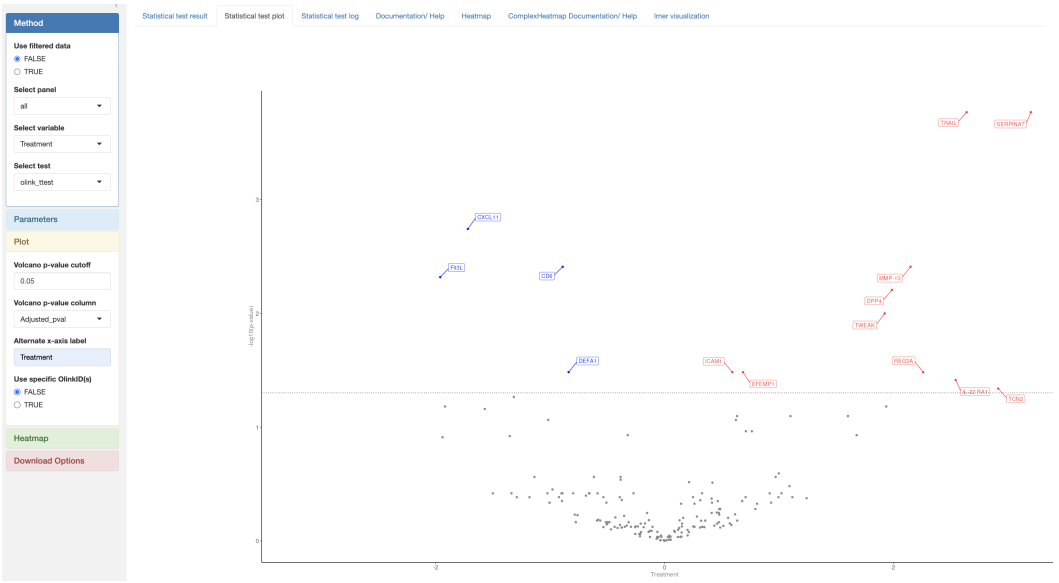
Figure 3: Filter Data tab (demo data)

C. Filter Data:

While the filtering step is optional, the user has the choice of removing samples (outliers from outlier detection step B) as well as assays (having non-parametric or non-standard distribution) before further analysis. See Figure 3. Qualitative feature of each assay is captured using quantile-quantile plot, a standard way to detect if there are departures from normal distribution (non-parametric distribution).



(a) t-test result



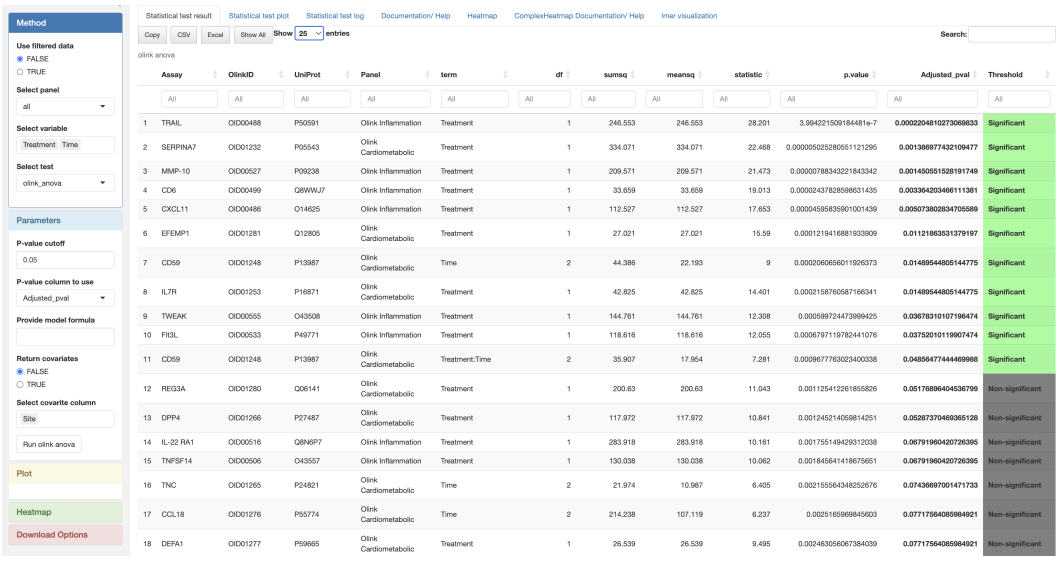
(b) t-test plot

Figure 4: Statistics tab (two-level variable)

D. Statistics:

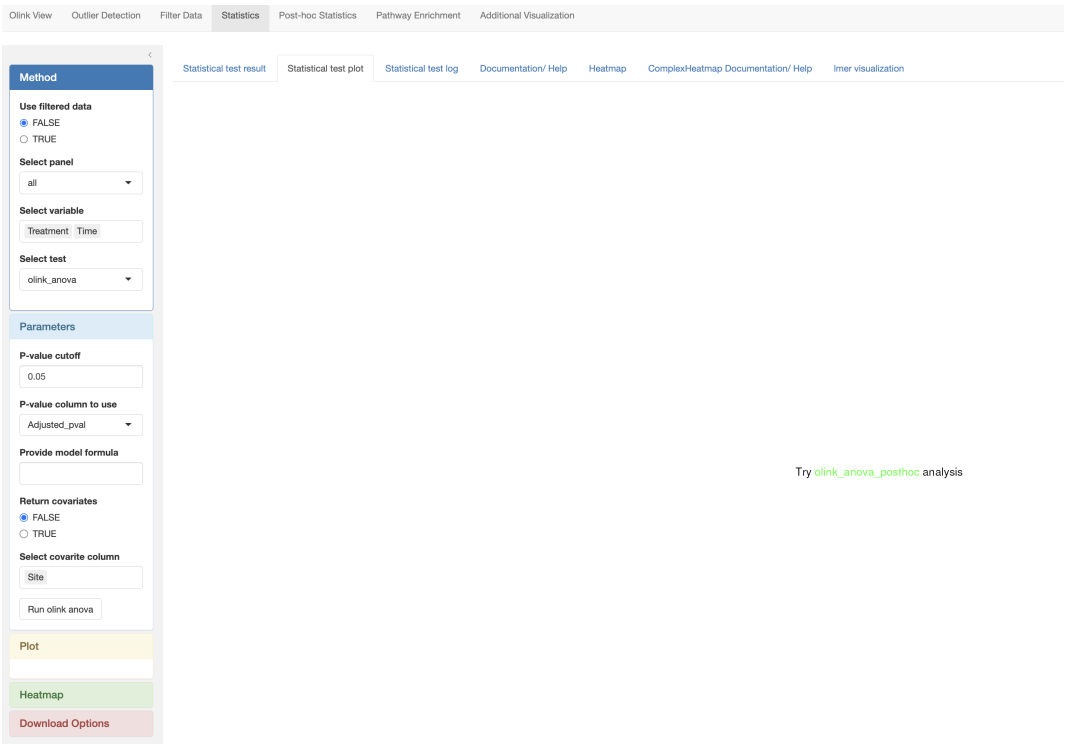
Statistical analysis methods offered in the application can be broadly grouped as "parametric" and "non-parametric" statistics. Parametric statistical tests can be further grouped as tests where "Outcome factor with exactly two levels" and "Outcome factor with greater than two levels". The former comprises of t-test (olink_ttest) [See Figure 4.] whereas the latter includes analysis of variance (anova; olink_anova)[See Figure 5.] and linear mixed effects model (lmer; olink_lmer).

The corresponding "non-parametric" test for "Outcome factor with exactly two levels" is the Mann-Whitney U test (olink_wilcox), while for "Outcome factor with greater than two levels" the tests are Kruskal-Wallis test/ Friedman test (olink_one_non_parametric) and two-way ordinal analysis of variance (olink_ordinalRegression). Both olink_ttest and olink_wilcox results can be used to generate volcano plots corresponding to the test outcome. In addition to statistical tests, this section allows user to generate heatmaps for a choice of grouping variable. Clustering can be invoked both on rows and/or columns, and



	Assay	OlinkID	UniProt	Panel	term	df	sumsq	meansq	statistic	p.value	Adjusted_pval	Threshold
1	TRAIL	OID00488	P50591	Olink Inflammation	Treatment	1	246.553	246.553	28.201	3.994221509184481e-7	0.000220481027306983 3	Significant
2	SERPINA7	OID01232	P05543	Olink Cardiometabolic	Treatment	1	334.071	334.071	22.468	0.000005025280551121295	0.001386977432109477	Significant
3	MMP-10	OID00527	P09238	Olink Inflammation	Treatment	1	209.571	209.571	21.473	0.000007883432218433 42	0.001450551528191749	Significant
4	CD6	OID00499	Q8WWJ7	Olink Inflammation	Treatment	1	33.659	33.659	19.013	0.000024378285986314 35	0.003364203466111381	Significant
5	CXCL11	OID00486	O14625	Olink Inflammation	Treatment	1	112.527	112.527	17.653	0.000045958359010014 39	0.005073802834705589	Significant
6	EFEMP1	OID01281	Q12805	Olink Cardiometabolic	Treatment	1	27.021	27.021	15.59	0.000121941688193390 9	0.011218635313791 97	Significant
7	CD59	OID01248	P13987	Olink Cardiometabolic	Time	2	44.386	22.193	9	0.000206065601192637 3	0.014895448051447 75	Significant
8	IL7R	OID01253	P16871	Olink Cardiometabolic	Treatment	1	42.825	42.825	14.401	0.000215876058716634 1	0.014895448051447 75	Significant
9	TWEAK	OID00555	O43508	Olink Inflammation	Treatment	1	144.761	144.761	12.308	0.000599724473999425	0.036783101071964 74	Significant
10	Fit3L	OID00533	P49771	Olink Inflammation	Treatment	1	118.616	118.616	12.055	0.000679711978244107 6	0.037520101199074 74	Significant
11	CD69	OID01248	P13987	Olink Cardiometabolic	Treatment:Time	2	35.907	17.954	7.281	0.000967776302340033 8	0.048564774444698 8	Significant
12	REG3A	OID01280	Q06141	Olink Cardiometabolic	Treatment	1	200.63	200.63	11.043	0.001125412261855826	0.051768964045367 99	Non-significant
13	DPP4	OID01266	P27487	Olink Cardiometabolic	Treatment	1	117.972	117.972	10.841	0.001245214059814251	0.052873704693651 28	Non-significant
14	IL-22 RA1	OID00516	Q8N6P7	Olink Inflammation	Treatment	1	283.918	283.918	10.161	0.001755149429312038	0.067919604207263 95	Non-significant
15	TNFSF14	OID00506	O43557	Olink Inflammation	Treatment	1	130.038	130.038	10.062	0.001845641418675651	0.067919604207263 95	Non-significant
16	TNC	OID01265	P24821	Olink Cardiometabolic	Time	2	21.974	10.987	6.405	0.002155564348252676	0.074366970014717 33	Non-significant
17	CCL18	OID01276	P55774	Olink Cardiometabolic	Time	2	214.238	107.119	6.237	0.00251659698456 03	0.077175640859849 21	Non-significant
18	DEFA1	OID01277	P59665	Olink Cardiometabolic	Treatment	1	26.539	26.539	9.495	0.002463056067384039	0.077175640859849 21	Non-significant

(a) ANOVA result

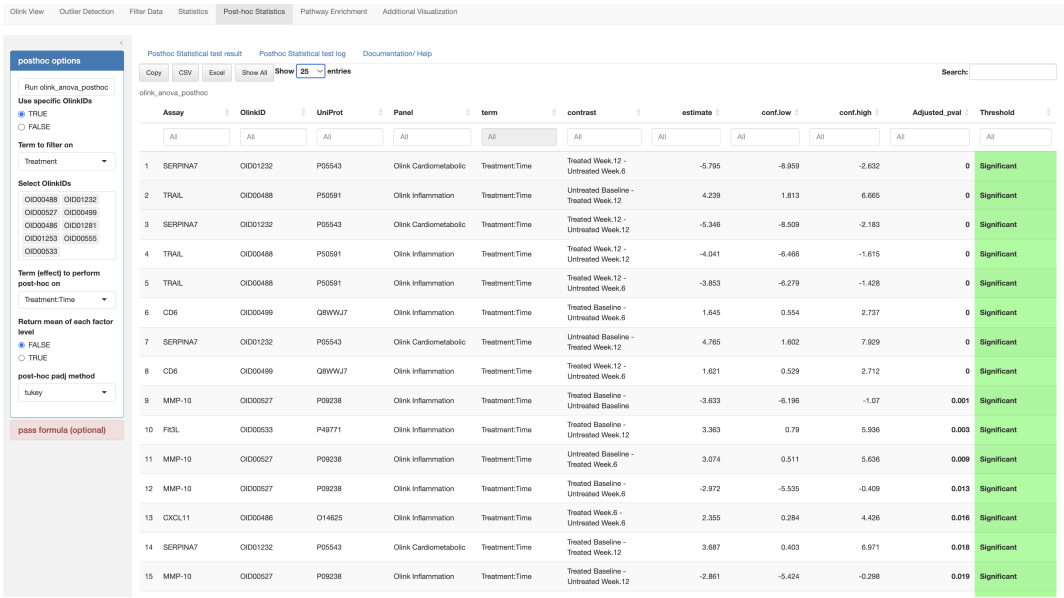


(b) ANOVA plot

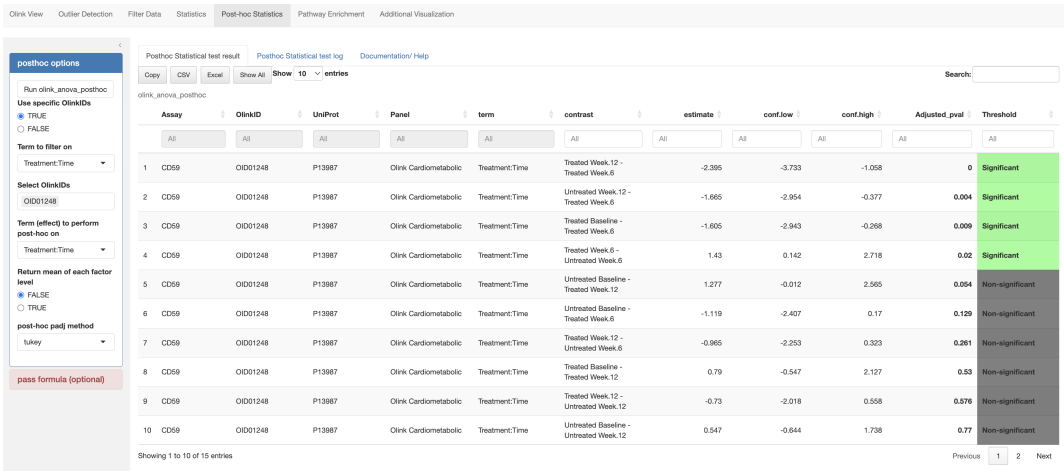
Figure 5: Statistics tab (>two-level variable)

the user can choose to use only significant assays obtained in the most recently run/executed statistical test.

E. Post-hoc Statistics: All statistical tests in the group “Outcome factor with greater than two levels”[See Figure 6] irrespective of distribution, require post hoc analysis before meaningful results are generated. All active (chosen) functions have their corresponding usage information available in a tab labelled “Documentation/ Help”.



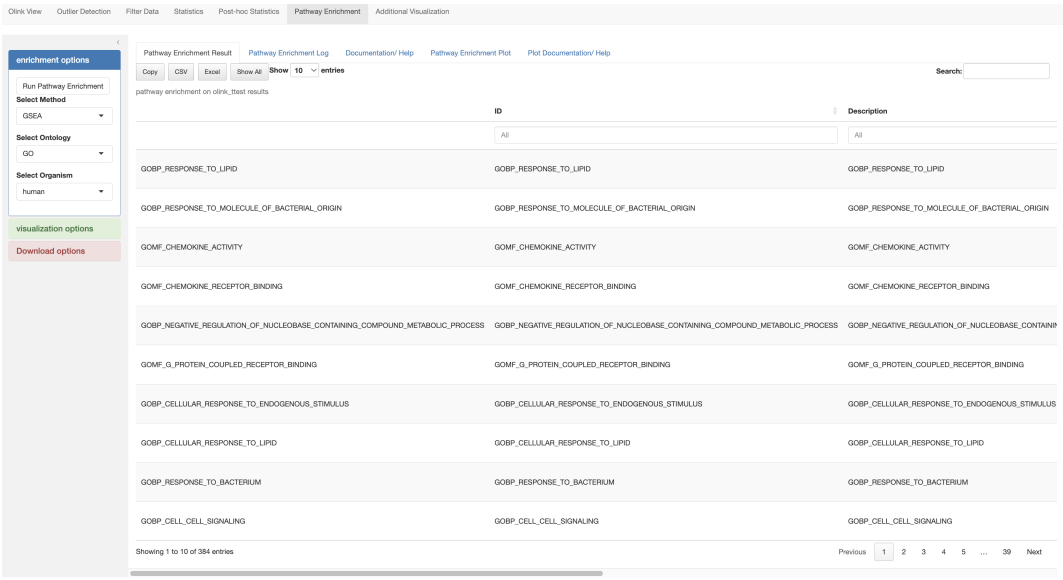
(a) different effect & filter term



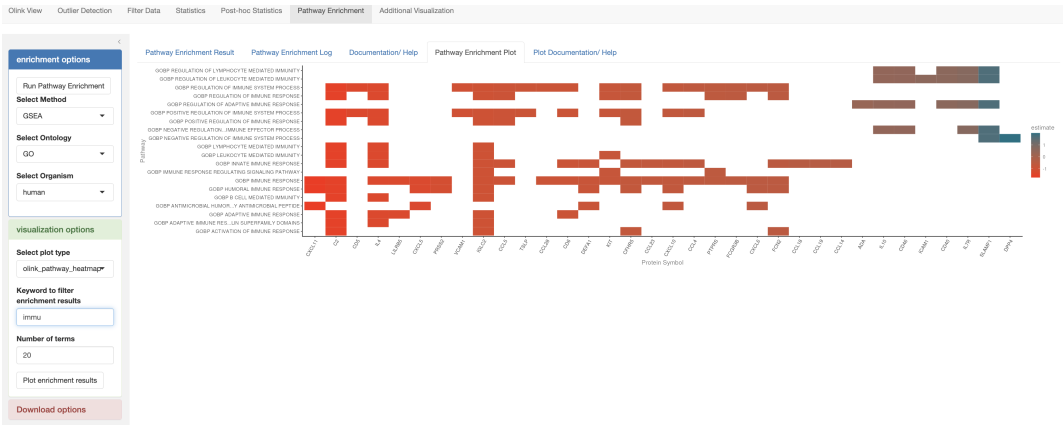
(b) same effect & filter term

Figure 6: Post-hoc Statistics tab (>two-level variable)

F. Pathway Enrichment: Following t-test or Mann-Whitney U test or the appropriate post-hoc analysis, the user has the option of performing pathway enrichment using over-representation analysis (ORA) or gene set enrichment analysis (GSEA)[See Figure 8].



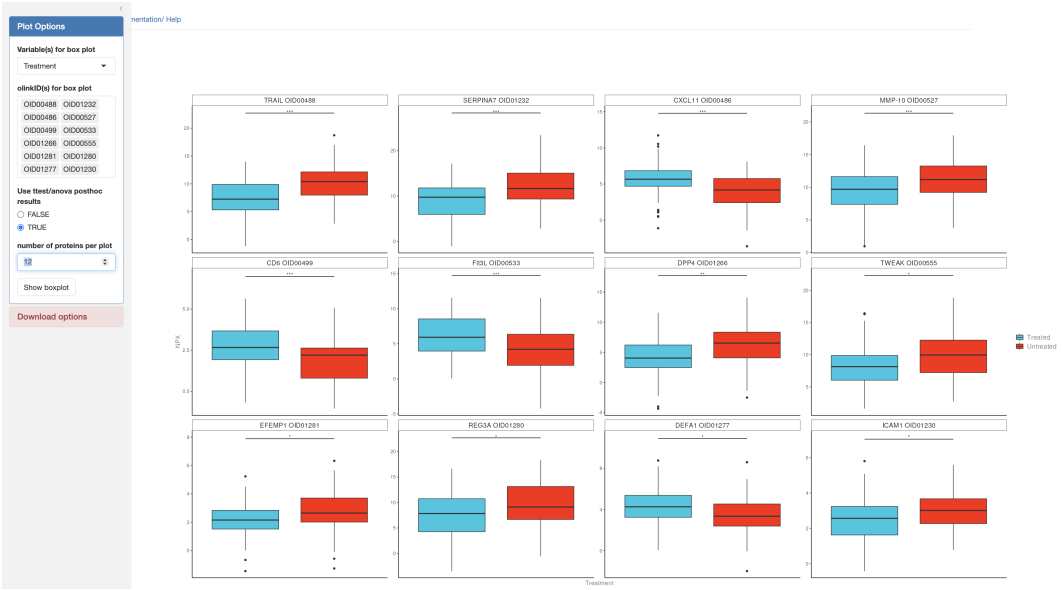
(a) Gene Set Enrichment Results



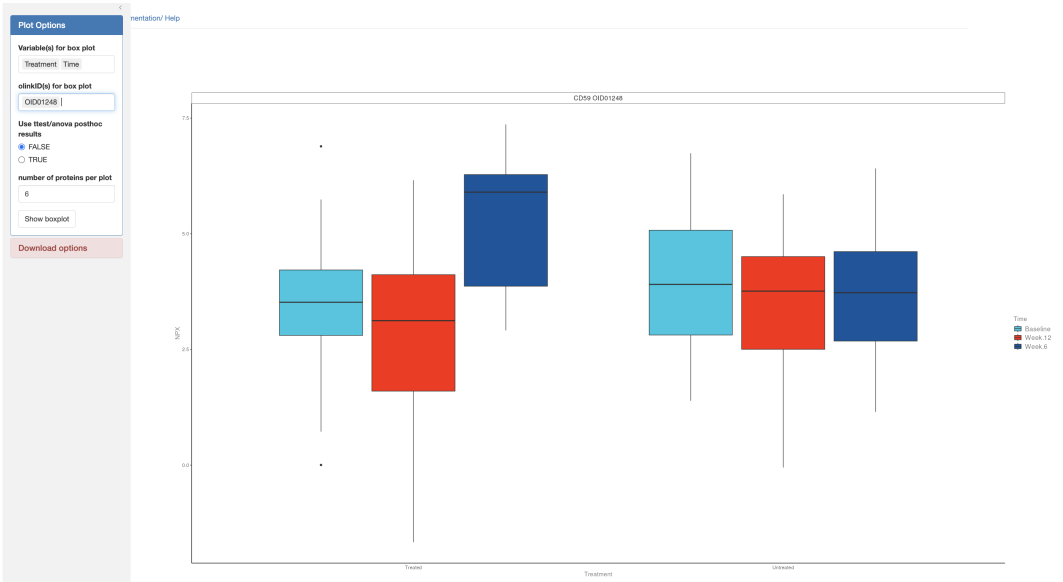
(b) Gene Set Enrichment Heatmap

Figure 7: Pathway Enrichment tab (demo data t-test results)

G. Additional Visualization: Lastly, significant assays can be visualized using boxplots and the p-value significance added to the plot in case t-test were previously performed. Download handlers have been implemented for easy download of all result tables. Figures generated during analysis can be customized for size and downloaded in three file formats.



(a) Boxplot with t-test results



(b) Boxplot with ANOVA results

Figure 8: Additional Visualization tab (demo data)