How to analyze Olink data

PREFACE

This document introduces statistical analysis of data from Olink® panels in Qlucore Omics Explorer (QOE). It is assumed that the user is familiar with the basic functionality of QOE.

In addition to the statistical tests that are highlighted below other functionality such as plots and visualizations are also relevant.

Content

1. OLINK® PANELS

Olink® panels are used for a multitude of purposes. Qlucore Omics Explorer is used for data visualization, exploration and analysis of many different types of data, and can also be used to analyze data from Olink® panels.

2. DATA PREPARATION

You can use the Olink software to export NPX data to .csv files or in some cases to .parquet files.

Csv files can be imported into QOE, and a parquet file can after conversion to a csv file also be imported. The data shall be exported from the Olink software as normalized data.

Files exported from Olink® NPX Signature software and Olink® NPX Map software (as an example Olink® Reveal data) are normally .csv files in a vector (column) format. The NPX files exported by the Olink NPX Manager, for Olink® Target 48 / 96 panels, are normally .csv files in a matrix format, but can also be in vector format.

NPX data from Olink® High-plex panels, like Explore HT/3K, may be exported as .parquet files. After export you then convert the .parquet file to a .csv file. There are several web sites providing free tools to convert a .parquet file to a .csv file.

Note that you can import tab separated txt files and csv files. If you have Microsoft Excel file (.xlsx) then you first save the data as a tab separated txt file or a csv file. If you have several sheets in the Microsoft Excel file, then you save each sheet as a tab separated txt file or csv file.

Data can be organized in several ways. Both as a matrix and as vectors (columns) where data for all samples are presented for each variable after each other. This format depends on which software you use and how data is exported from the Olink software.

202501_test	Olink NPX Manager xxx		
NPX data			
Panel	Olink YYY	…	Olink YYY
Assay	Assay name_1	…	Assay name_n
Uniprot ID	P42785	…	P00915
OlinkID	OID01272	…	OID01223
Sample_1	0,964551	…	-2,17105
Sample_2	0,43123	…	-1,98125
….	1,56789	…	0,96531
Sample_n	-2,04536	…	0,43289

***Figure**: Example of matrix output from Olink NPX Manager software (not all columns shown)*

SampleID	SampleType	OlinkID	UniProt	LOD	NPX
Sample_01	SAMPLE	OID50263	A1E959	0,2178	-0.1219
Sample_02	SAMPLE_CONTROL	OID50263	A1E959	0,2178	-0.0869
Sample_01	SAMPLE	OID50375	Q9Y2D5	0.9848	0.1711
Sample_02	SAMPLE	OID50375	Q9Y2D5	0.9848	-0.1490
…	…	…	…	…	…

***Figure**: Example of vector (column) output from Olink NPX Signature software (Reveal) (not all columns shown)*

3. EXAMINE THE DATA BEFORE IMPORT

Before importing, examine the data.

3.1. SAMPLES

Make sure to identify the name of the unique sample identifier. That should be selected during the import.

Note: Replicates are handled after import.

3.2. VARIABLES

If you have replicates among the variables it is managed after import by suing the collapse functionality (in the Data tab).

Note: Sometimes variable names need correction depending on the planned downstream analysis. A variable may have hyphens in the name, like IL-8 (interleukin 8), can also be coded as IL8 or alt CXCL8. You may want to remove hyphens to work with gene symbols in databases like pathways from MSigDB (Molecular Signatures Database) from Broad Institute.

4. IMPORT

Start QOE, select the file menu and then “Import with Wizard…”. The Wizard and steps are described in detail in the reference manual and in several videos that can be viewed at qlucore.com . You answer questions in the steps and point out where you have the data. The Olink data is of type “Other” (What type of data is this) and you select either Matrix or Vector depending on the format of your Olink data (How is the data organized). (Tip: Search for “Wizard” on https://qlucore.com/videos). After import, save the imported data as a gedata file, the the file menu and “Save as …”.

Tip: you can go back and forward in the wizard and the selections are highlighted with colors.

5. GET AND OVERVIEW OF THE IMPORTED DATA SET

5.1. CELL VALUES

For smaller Olink® panels with a limited number of variables, like 96, 48 etc., you can easily get an overview of the data using a Table plot, available under the Methods. The data imported should already have been normalized, and log transformed.

A heatmap will also give a good overview. You can then order the samples according to a Sample Annotation, and you can order the variables too, perhaps using hierarchical clustering to get an easy overview of correlated variables in a dendrogram.

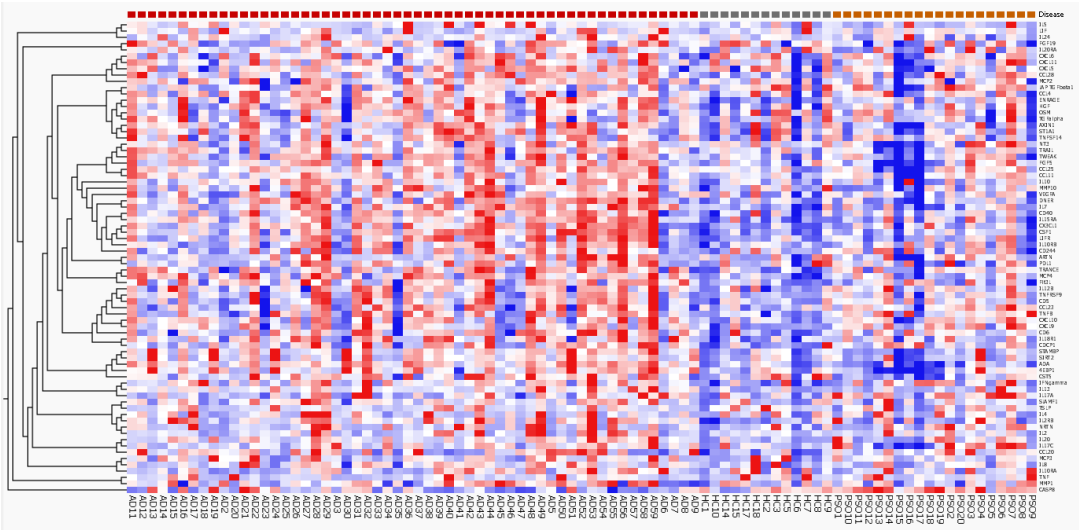

***Figure**: Heatmap in Qlucore Omics Explorer with data from an Olink® panel*

You can also select to have a look at the data in a histogram, to look at the distribution of the data, to understand if the data set can be considered to be roughly normal distributed, or not.

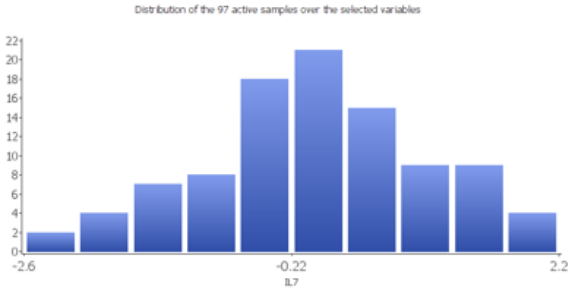

***Figure**: Histogram with 10 bins to get an overview of distribution*

5.2. SAMPLES AND THEIR ANNOTATIONS

If you imported sample annotations with the data, or if you have imported sample annotations to add them to the data set afterwards, using the Import support of Sample Annotations, it is now time to examine the annotations.

5.3. VARIABLES AND THEIR ANNOTATIONS

In the “Variables” tab you can inspect the variables. If there are several available annotations, you can make these columns visible with the “Columns” icon. If the variables are not unique you can in the “Data” tab select to collapse variables (calculation can be based on median, mean value etc.). In the “Data” tab you can also select to change the identifier if several identifiers are available in the dataset.

6. PRE-FILTER, TRANSFORMATIONS, NORMALIZATION AND MISSING VALUES

In the Data tab there is a rich suite of functionality to let you pre-process and manage the data.

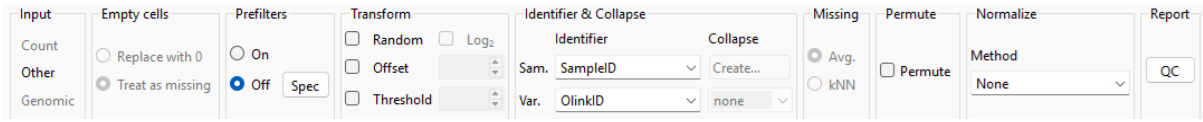

***Figure**: The Data tab*

Examples are:

- Manage empty cells in the data set. How should they be treated.
- Use prefilters to remove variables with low quality, see example below.
- Add offset or a Threshold to make it possible to take the logarithm of data.
- Normalize data (if it has not already been normalized before import)

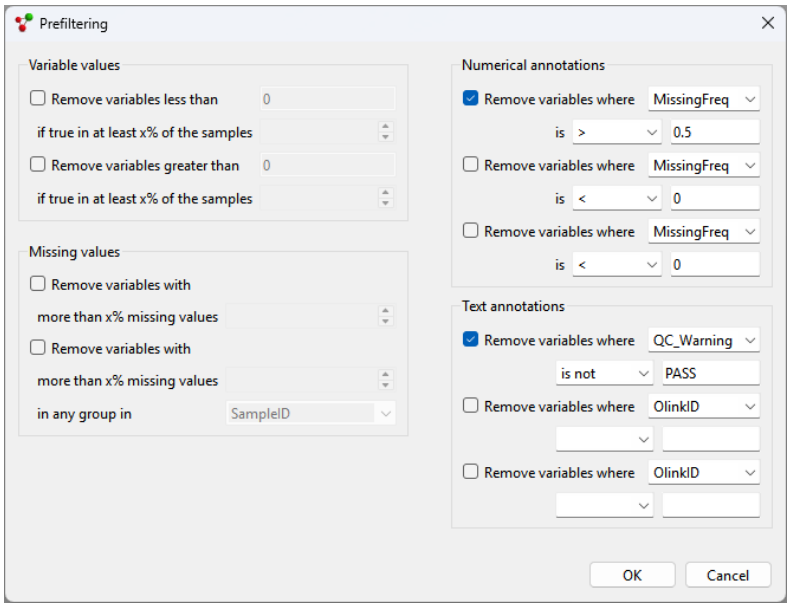

***Figure**: Prefiltering*

7. MANAGE SAMPLE REPLICATES

It is possible to collapse sample replicates. This is done from the Data tab. Select the Create button in the ***Identifier & Collapse*** group box. It is required that there is sample annotation outlining the replicates. An example of how the sample annotation can look is shown in the table below. In the dropdown you would then select to Collapse on the annotation “Samples” and will get a new data set with half the number of samples. You can select the collapse algorithm.

Sample ID	Samples	Prot 1	…	Prot_n
Sample_01_1	Sample_01	0,1		2,1
Sample_01_2	Sample_01	2,4		9,6
Sample_02_1	Sample_02	7,3		1,9
Sample_02_02	Sample_02	5,3		3,4
…				

***Figure**: Replicates*

8. DATA ANALYSIS

8.1. A TWO GROUP STATISTICAL TEST

A two-group comparison (t-test) can be used to test the significance of the difference between two conditions. The

standard t-test assumes that the input values are normally distributed and the variance between conditions are comparable.

As an alternative one can use the Mann Whitney U Test (Wilcoxon Rank Sum Test) when data is not normally distributed. Below is an overview of the different options.

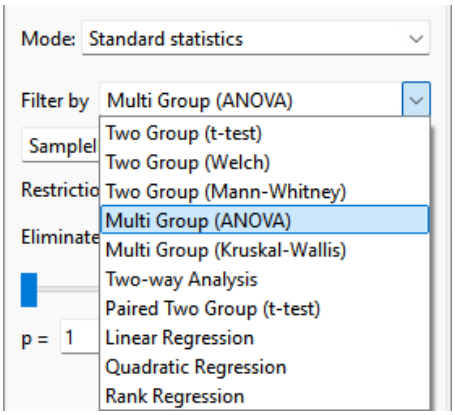

***Figure**:Standard tests*

If you would like to eliminate the influence from covariates, you can do so. You can select the sample annotations that you would like to eliminate on. Here an example with three annotations treated as covariates and eliminated; BMI, Age and Gender below in the t-test between atopic dermatitis (AD) and Healthy samples.

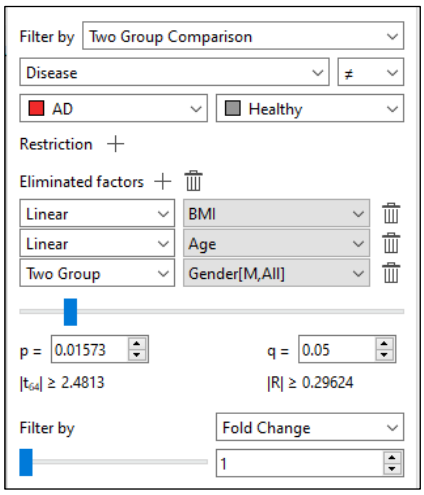

***Figure**: t-test between AD and Healthy sample, elimination on covariates. The threshold for the q-value set to 0,05 (Benjamini-Hochberg)*

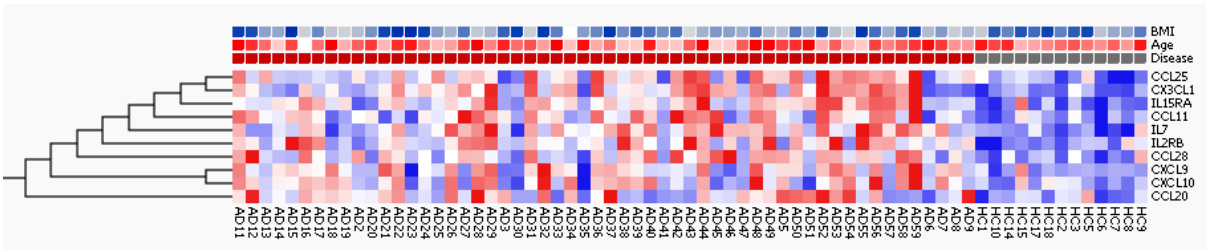

***Figure**: Heatmap after the test above.*

8.2. A REGRESSION TEST

Let´s assume that you have another annotation “BMI” with numerical values. In this case you can select to run a linear regression test. If you have an annotation with ordered, but not numerical values, you can instead select to run a rank regression test.

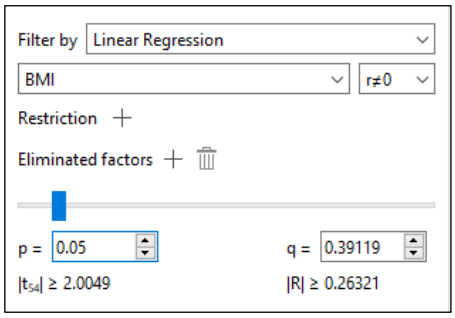

***Figure**: Linear regression test using BMI (numerical value)*

You can look at each of the remaining variables look at the regression pattern and trend lines in Scatter plots. You select Scatter Sample Plot, with BMI (numerical value) on the X-axis and variables on the Y axis.

Ds

	VariableID	R-statistic
1	CASP8	0,284574
2	CXCL5	0,287155
3	DNER	-0,274189
4	FGF5	-0,374872
5	IL17C	-0,443196
6	LAP TGFbeta1	-0,273067
7	NT3	-0,343557
8	TWEAK	-0,314403

***Figure**: The correlation values in the R-statistic column*

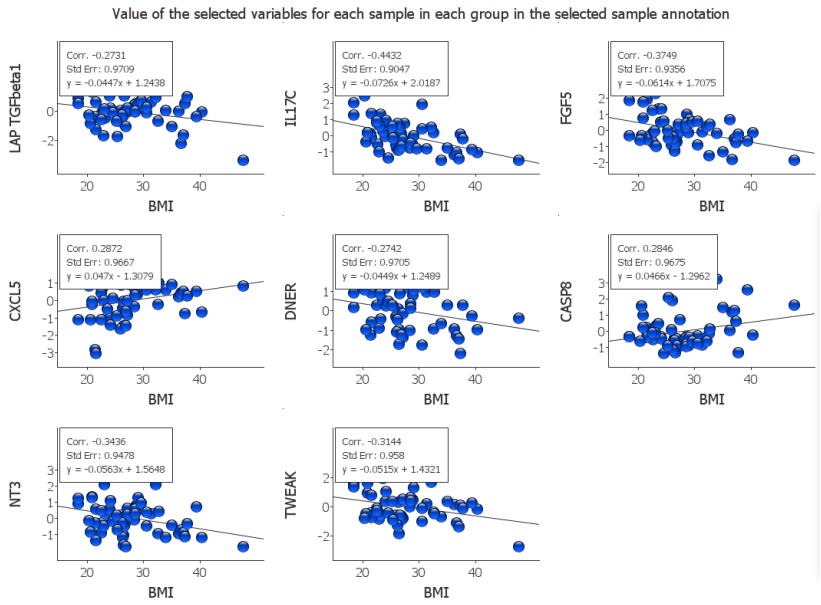

***Figure**: Scatter plots for variables after the linear regression test*

8.3. REFERENCES

One of the data example in this document comes from supplementary data in the article “The atopic dermatitis blood signature is characterized by increases in inflammatory and cardiovascular risk proteins”, published in August 18th, 2017 in Scientific Reports, DOI:10.1038/s41598-017-09207-z.

9. ACKNOWLEDGEMENTS

Olink® is a registered trademark of Olink Proteomics AB.

Excel is a trademark of Microsoft Corporation.

10. DISCLAIMER

The contents of this document are subject to revision without notice due to continuous progress in methodology, design, and manufacturing.

Qlucore shall have no liability for any error or damages of any kind resulting from the use of this document.

Qlucore Omics Explorer is only intended for research purposes.