QLUCORE

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 for Olink® data analysis.

Contents

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

Use the Olink® NPX Signature software from Olink Proteomics to export normalized data from the Olink® panels. Export can be done to Microsoft Excel (.xlsx) files or text (.csv) files for use in other programs, like Qlucore Omics Explorer. If you select to export several panels, you may get a Microsoft Excel file with several sheets, where there is one panel per sheet.

Note that you can import tab separated txt files and csv files into Qlucore Omics Explorer. If you have Microsoft Excel file (.xlsx) then you 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.

3. DATA ORGANIZATION

The exported data probably needs to be re-organized before imported into QOE.

It is likely that your data is organized with the sample identifier and additional sample annotations as columns. The variables and the corresponding variable annotations are then the rows.

To import the data into Qlucore Omics Explorer you would need to have the Sample identifier as the leftmost column.

You will likely have the cell values in the columns to the left in the file exported, and the corresponding sample annotations to the right. You then just swap the columns in Excel, so that you get the sample annotations to the left, and the actual cell values to the right.
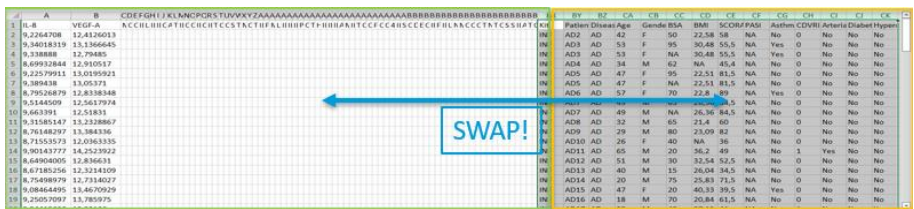

Figure: Re-arrange data in the Excel file

Note: Qlucore Omics Explorer can also handle import of transposed data, where the samples are the rows, and the variables the columns.

4. EXAMINE THE DATA BEFORE IMPORT

Before import into Qlucore Omics Explorer, first examine the data.

4.1. SAMPLES

Are there any replicates among the samples?

Is the main sample identifier unique or not? Let´s say that you have an identifier "Patient.ID" that is not unique. Then you need to decide how you should handle this – keep it as not unique or collapse to make it unique.

What identifier shall be used for the samples?

The sample identifier shall be in the leftmost column in the Excel file. If replicates are kept, then create a new unique sample identifier to the left, like "SampleID".

4.2. VARIABLES

Does the variable identifier have a header?

If there is no header, then add a column for header (like "VariableID")

Are there any replicates among the variables?

If you have replicates in the variables, then you can collapse after import into Qlucore Omics Explorer.

Please also check the variable names. 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.

4.3. GENERAL

To make it easier to read, add empty row below the variableIDs to get the sample annotations on a row of their own.

Also, make sure that the last row in the matrix is a row with cell data, not summaries etc. Remove rows below the last row with cell data.

After examination and adjustments, then save the sheet as a tab separated txt file or a csv file.

5. IMPORT FILE INTO QLUCORE OMICS EXPLORER

QOE provides import of data and annotations through Wizards. Start QOE, select the file menu and the Import Wizard. The Wizard and steps are described in detail in the reference manual and in several videos that can be viewed at qlucore.com (Tip: Search for "Wizard" on https://qlucore.com/videos). After import, save the imported data as a gedata file in QOE.

6. EXAMINE THE IMPORTED DATA SET

6.1. EXAMINE THE DATA MATRIX AND CELL VALUES

The Olink® panels are for a limited number of variables, like 96, 48 etc., so you can easily get an overview of the data using a Table plot, available under the Methods. The data should be normalized, and log transformed already. There could be negative values due to the log transformation.

A heatmap will also give a good overview. You can then select to 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.
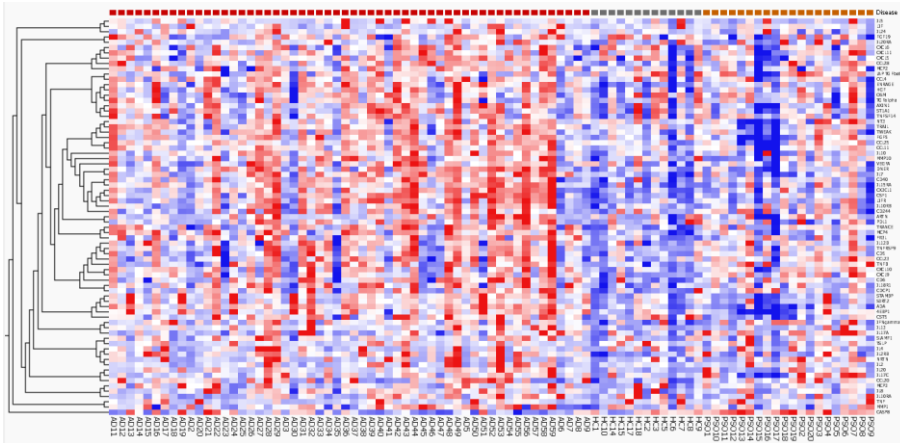
Figure: *Heatmap in Qlucore Omics Explorer with data from an Olink® panel*

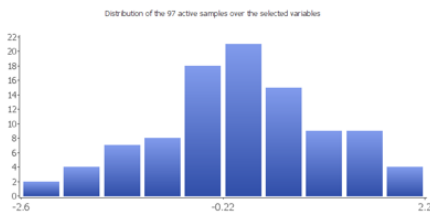You can also select to have a look at the data in a histogram, to look at the distribution.

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

6.2. EXAMINE THE SAMPLES AND THEIR ANNOTATIONS

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

It may be that you have missing information for some annotations. As an example, you could have an annotation "Age" with numerical values, but then find out that there is one group called "N/A" which contains the samples for which no Age information was available.

6.3. EXAMINE THE VARIABLES AND THEIR ANNOTATIONS

Now you can go to the "Variables" tab and have a look at 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.

7. PLOT OVERVIEW

The Olink® panels have a limited number of variables, and a heatmap often gives a good overview of the data. You can change the plot to a heatmap in the Methods Tab. You can also select how you want to order the samples, perhaps based on a Sample annotation that you intend to use for statistical analysis. You can also order the variables, one idea is to order them using hierarchical clustering, to better see correlations between variables.

8. STATISTICAL TESTS

8.1. A TWO GROUP 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, and the number of samples are small, see the "Extended Statistics" in the inbuilt Qlucore manual. Another alternative is the limma t-test from the LIMMA R package, that can be called from inside Qlucore Omics Explorer through a predefined wrapper.

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, like BMI, Age and Gender below in the t-test between atopic dermatitis (AD) and Healthy samples.
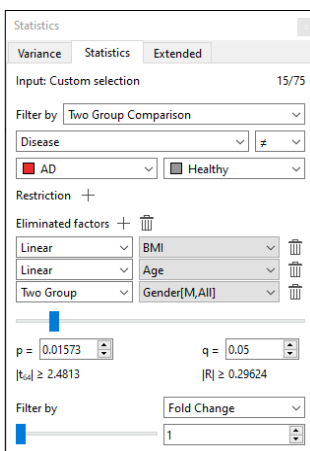
Figure: *t-test between AD and Healthy sample, elimination on covariates. The threshold for the q-value set to 0,05 (Benjamini-Hochberg)*
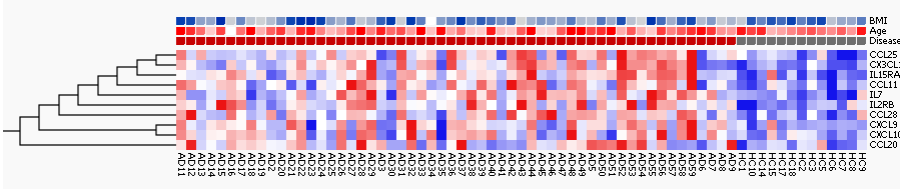
Figure: *Heatmap after the test above.*

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.
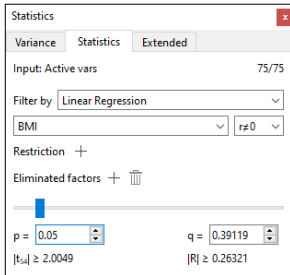

Figure: *Linear regression test using BMI (numerical value). The threshold for the p-value set to 0,05 (Benjamini-Hochberg)*

You can for 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.
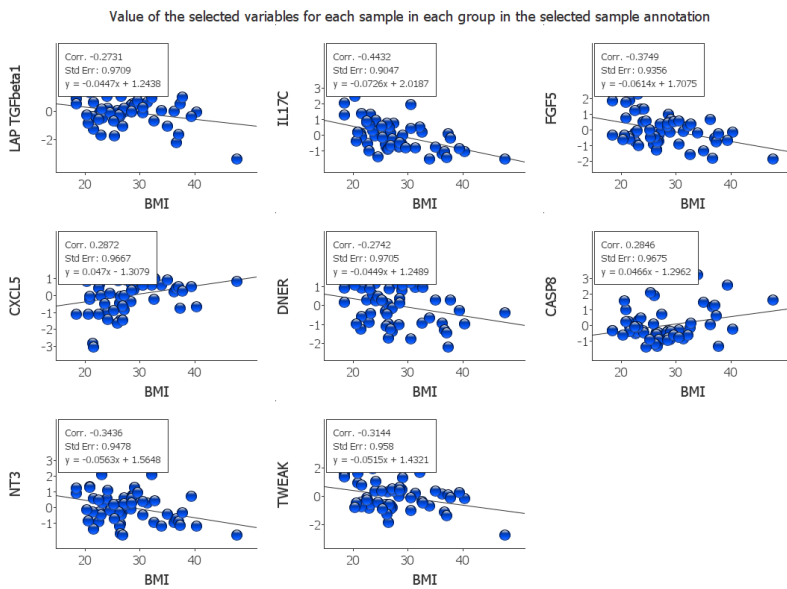
Value of the selected variables for each sample in each group in the selected sample annotation

Figure: Scatter plots for variables after the linear regression test

Figure: The correlation values in the R-statistic column

9. REFERENCES

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.

10. ACKNOWLEDGEMENTS

Olink® is a registered trademark of Olink Proteomics AB. Qlucore AB is not affiliated with Olink Proteomics AB.

11. 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.