

# Qlucore Omics Explorer 3.3 feature overview

## INTRODUCTION

Qlucore Omics Explorer (QOE) is developed to support the user with fast, simple and visual analysis of measured data considering publicly available information such as gene ontologies, pathways and other system biology information to maximize the output of the analysis.

You reach all key functionality with one or two mouse clicks and the results of your actions are always presented to you in real time by a visual update. The visual approach makes it easy to publish results as well as working in teams.

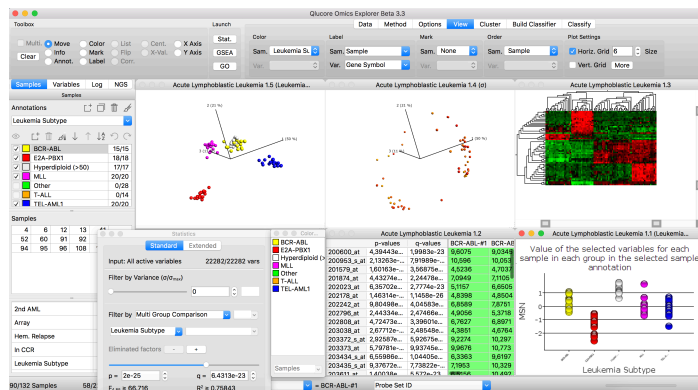
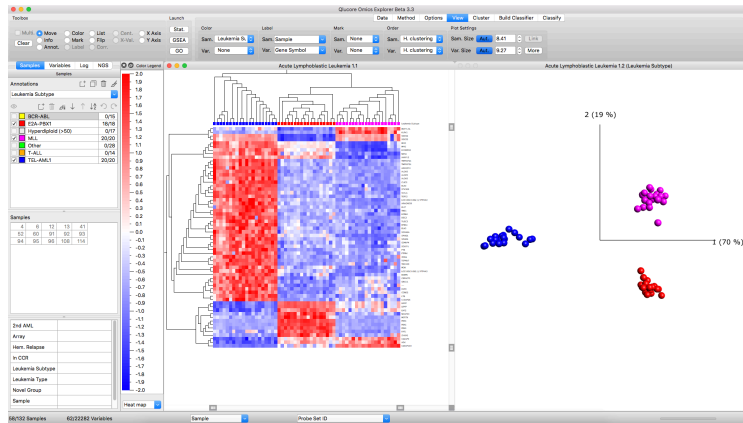
Qlucore Omics Explorer ships in a base module with an option to add a NGS module with extensive functionality for NGS data analysis. Detailed information about the NGS module features is presented in the NGS Module feature overview document.

Qlucore Omics Explorer is tailored for creative analysis with a focus on instant results and effective visualizations.

- QOE works in full real time with both 2D and 3D presentations of all data. All plots are truly interactive. The user is encouraged to explore the data by changing filters and parameters dynamically.
- QOE uniquely combines powerful statistical analysis with instant visualization. Most actions are controlled with only one mouse click.
- QOE provides simple workflows for mRNA, miRNA data and DNA Methylated data with direct import and normalization of Agilent, Affymetrix data and aligned BAM files for RNA-seq data. The NGS module supports a wide range of options for NGS data.
- The integrated Gene Set Enrichment Analysis (GSEA) workbench allows a straightforward analysis of the biological context (pathways, ontology categories or any other relevant set of genes).
- Classifiers can be constructed using any of the following methods: Support Vector Machines, Random Forest and kNN.
- QOE includes an open interface to R.
- QOE support hierarchical clustering, K-means clustering, heatmaps with dendograms and Dynamic Principal Component Analysis (PCA).
- A direct link to Gene Expression Omnibus (GEO) enables one button data downloads and easy comparison of findings with published material and with the GO Browser you can quickly search in ontologies.
- QOE only requires a normal computer to handle huge data sets (more than 100 million entries).

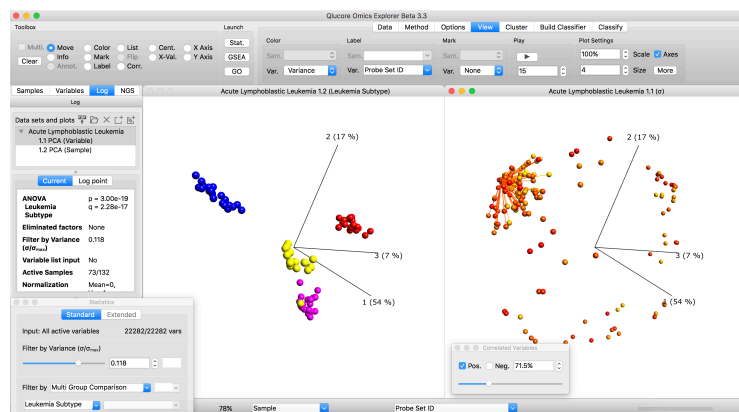
## MAIN FUNCTIONALITY

- Analyze and explore data set by a combination of visualizations and intuitive filters.
- Do statistical analysis using a wide range of built in tests, such as ANOVA, as well as through the open interface to R. Generate results with false discovery rates(q-value), fold change and p-values.
- Perform hierarchical clustering and generate dynamic heatmap plots.
- Analyze RNA-seq data both in the Genome browser and a PCA plot in a synchronized view
- Fine tune and generate results using any combination of scatter plots, box plots and line plots.
- Instantly create Principal Component Analysis PCA plots of large data sets and confirm the information content by using the Qlucore unique functionality Projection Score.
- Use K-means++ clustering
- Use any of several methods, Hierarchical clustering, PCA, clustering, ISOMAP and or graphs for visual data exploration.
- Build classifiers and classify new samples.
- Do functional analysis in the context of public available gene sets such as pathways and so on using GSEA.
- Download data from Gene Expression Omnibus (GEO) to compare your own results with published material.
- Remove unwanted dependencies such as artifacts and outliers. Manage batch effects.
- Benefit from the streamlined workflows for Affymetrix gene expression microarray and Agilent miRNA and mRNA data as well as direct import of aligned BAM files with RNA-seq data for digital gene expression analysis.
- Keep track of your work with powerful global log and restore function.



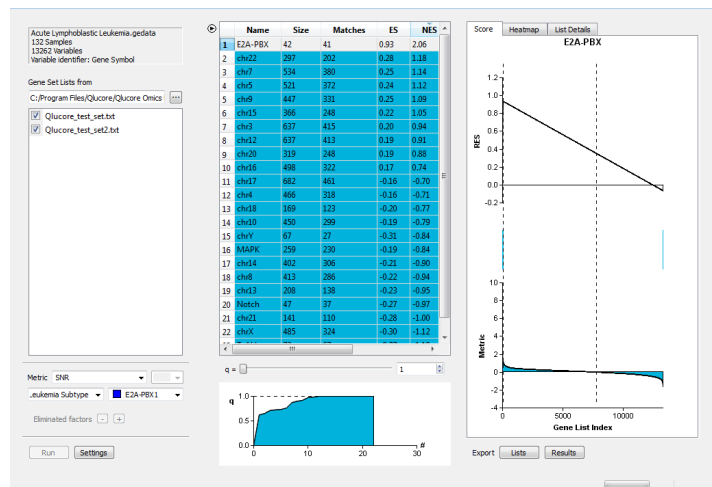
## OUTPUT

- High quality 2-D and 3-D graphics.
- 11 plot types: Heatmap, sample PCA, variable PCA, bar plot, sample scatter plot, variable scatter plot, box plot, line plot, histogram, Kaplan-Meier and ROC curves.
- Data table view.
- GSEA results; enrichment plots as well as leading edge heatmaps and result lists.
- Flexible ordering in heatmap. Order according to hierarchical clustering or an annotation or a statistical value.
- Present multiple annotations in a heatmap.
- Plot both samples and variables.
- Open multiple data sets at one time.
- Synchronized plots (as many as you like). Synchronized plots are updated simultaneously
- Variable lists with p-values, fold change and FDR (q-values) values including a complete description of how the list was generated.
- Plot arbitrary principal components.
- Color the samples and the variables through different methods.
- Color the variables according to any list.
- Label the samples and the variables through different methods.
- Present multiple scatter plots in one view.
- Color legend window. Explains the colors and scales in the active plot. Can be exported with a plot.
- Classifiers



## GSEA WORKBENCH

- Start the analysis with one key press. All plots and lists directly available.
- Work with publicly available gene sets or work with your own sets.
- Filter results on q-value with slider.
- Select ranking criteria from a broad range (SNR, two group comparison, multi group).
- Export selected lists as variable lists to be used in OE main window.
- Export plots and lists for publication and further analysis



## GO BROWSER

- Use any ontology that you prefer.
- Excellent overview by both tree and flat view of results.
- Very fast search.
- Export lists as variable lists to be used in OE main window.

## EDITING OF DATA

- Interactive editing of sample annotations.
- Interactive editing of variable lists.
- Variable collapse. Select any variable annotation and collapse<sup>1</sup> the data on this annotation.

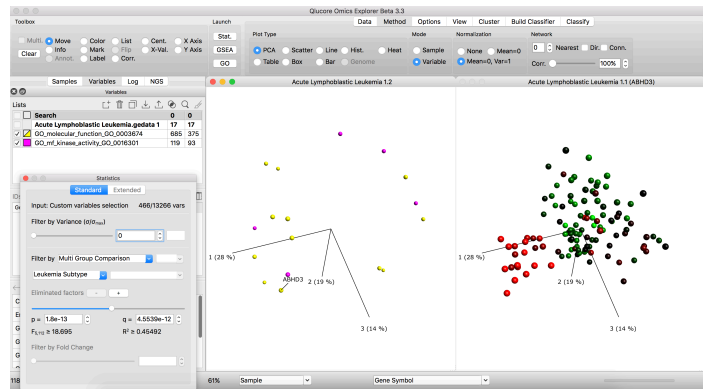
<sup>1</sup> Combine information from one or two variables to a new variable. Example: two measured variables match one Gene and the study should be conducted on gene level.

## SELECTIONS

- Work with subsets of samples and variables.
- Select samples based on clinical variables and other annotations.
- Select variables based on variance, F-test(ANOVA), t-test, rank correlation, correlation coefficients, Fold change, annotation searches and imported variable lists (such as pathways).
- Select variables based on statistical methods from the open interface. See below.
- Select variables based on linear or quadratic regression
- Study part of data set based on imported variable lists (such as pathways) and or combinations of lists.

## VARIABLE LISTS

- Automatic variable list for all active variables.
- Automatic variable list for Search results.
- Set operations on variable lists.
- Color any variable plot according to any selection of variable lists.
- Save variable lists including information about how they was created, this helps in creating good result traceability



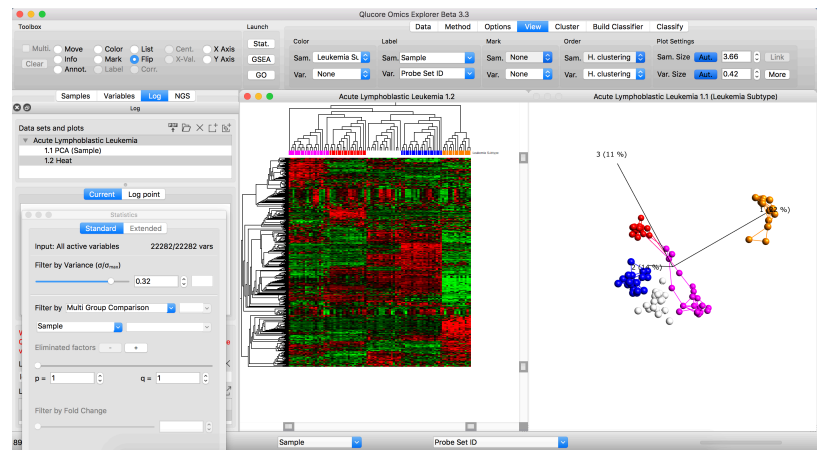
## VERIFICATION

- Projection score to understand how much information that is captured in a PCA plot.
- Get direct feedback on p-values and q-values during variable selection.
- Verify results by redoing the analysis with permuted sample annotations or with random numbers.
- Verify results through remove-one-at-a-time cross validation or several at the time cross validation.
- Continuous update on captured variance in PCA plot.

- Generate histograms to check variable distributions.
- Generate box plots to visualize results.

## CLUSTERS AND NETWORKS

- Visualize sample clusters by connecting each sample with its nearest neighbors.
- Visualize variable clusters by connecting correlated variables.
- Create clusters using K-means++.
- Perform hierarchical clustering in the heatmap plot



## CLASSIFICATION

- Build classifiers with Support Vector Machines, Random Trees or kNN.
- Validate the classifier either on the internal data set with cross validation scheme or an external data set.
- Classify new samples based on the build classifier.

## BATCH CORRECTIONS

- Correct multiple batch effects using in-built methods.

## OPEN INTERFACE

- Interface to R
- Expands the available statistical tests.

- Works for two-group, paired tests and multi groups, with or without eliminated factors.
- Example on supported methods are: Limma, Wilcoxon, Welch,...

## IMPORT

- Affymetrix .cel files and .chp files. Including normalization and generation of QC-report.
- Agilent Text Files (\*.txt) (from Feature Extraction Software). OE imports and normalizes miRNA data, mRNA data and DNA methylated data. Both single color and two color arrays can be handled.
- Agilent GeneView files (\*.txt) for miRNA data.
- Aligned BAM files with RNA-seq data.
- Flexible import wizard for import of (\*.txt, \*.csv, \*.tsv) files.
- Data files of .gedata format (normal or transposed).
- Annotation files, both samples and variables (\*.txt, \*.csv).
- Basic file formats such as
  - Data files of .txt format (one variable identifier column and one sample identifier row). Called Simple text files.
  - Compact Text files (\*.csv)
- Data (GDS and GSE) from Gene Expression Omnibus (GEO). Direct download.
- GEO soft files (\*.soft and \*.soft.gz).
- GEO Series Matrix (\*.txt and \*.txt.gz).
- Gene set files for the GSEA Workbench (\*.txt and \*.gmt).
- Ontology files (\*.obo, \*.obo.xml, obo-xeml.gz, obo-xml).
- Variable lists
- Direct NetAffx import
- Affymetrix CHP and ARR files
- Log files

- Quality control plot files
- Created classifiers
- Data files of .cedata format

#### EXPORT

- Still images (plots) including a legend plot. Select resolution and the plot is exported.
- Data files
- Variable lists (with annotations and data if so preferred)
- Videos
- Log files
- Correlation and co-variance matrixes
- QC-report
- PCA loadings
- PCA plot coordinates
- Classifiers

#### OTHER

- Missing value reconstruction (two versions)
- Variable normalization
- Multi-dimensional rescaling
- Isomap
- Take logarithm of data
- Simplified Affymetrix annotations





#### RESEARCH PURPOSE ONLY

Qlucore Omics Explorer is only intended for research purposes.

#### 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.

#### TRADEMARK LIST

NetAffx is a trademark of Affymetrix

#### CREDITS

GSEA: Subramanian, Tamayo, et al. 2005 Proc Natl Acad Sci U S A 102(43):15545-50

The Gene Ontology Consortium. "Gene ontology: tool for the unification of biology." Nat. Genet.. May 2000;25(1):25-9.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.