Qlucore Omics Explorer 3.9 feature overview

INTRODUCTION

Qlucore Omics Explorer (QOE) is developed to enable fast, easy-to use and visual analysis of measured data from a very wide range of sources and instruments transforming data into insights considering publicly available information such as TCGA data, gene ontologies, pathways and other system biology information to maximize the results and impact.

You reach all key functionality with one or two mouse clicks and the results of actions are always presented 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 an 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.** Available with native versions for Windows and Mac OS.

The program integrates well with work-flows and through the direct import and export of R data frames as well as the Python based template functionality it is possible to control the program through a script and configure well defined analysis steps.

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

o 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.

o QOE uniquely combines powerful statistical analysis with instant visualization. Most actions are controlled with only one mouse click.

o QOE provides simple workflows for mRNA (including single cell), miRNA data, protein, metabolomics and DNA Methylated data with direct import and normalization of 10X Genomics, Agilent, aligned BAM files for RNA-seq data and also array data for backward compatibility. The NGS module supports a wide range of options for NGS data including Fusion gene analysis, Chip-seq and ATAC-seq data.

o The powerful statistical framework supports a wide range of tests including, but not limited to, t-tests, Welch, ANOVA, paired tests, Two-way ANOVA, linear/quadratic/rank regression and a suite of nonparametric tests. The tests can be further customized using eliminated factors and restrictions.

o Script based execution is possible through templates that run in Python. The user can develop own templates. Integrating with a larger workflow is easy through templates and Python.

o   The integrated Biomarker Workbench is optimized for experiments and studies in the areas of drug development and biomarker discovery.

o   Survival analysis with Kaplan-Meier plots and Cox proportional hazards model.

o   QOE includes an open interface to R for statistical methods. R data files can be opened directly.

o   Analysis of single cell RNA-seq data is supported by features such as subsampling and UMAP and t-SNE plots.

o   The integrated Gene Set Enrichment Analysis (GSEA) workbench allows a straightforward analysis of the biological context (pathways, ontology categories or any other relevant sets of genes).

o   Classifiers can be constructed using any of the following methods: Support Vector Machines, Boosted trees, Random Forest and kNN. In combination with Qlucore Insights, developed predictors can be used for companion diagnostics (CDx)

o   QOE support hierarchical clustering, K-means clustering, heatmaps with dendrograms and Dynamic Principal Component Analysis (PCA).

o   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.

o   A Template tailored for TCGA data analysis is included.

o   QOE only requires a normal computer to handle huge data sets (more than 100 million entries).

MAIN FUNCTIONALITY

o   Analyze and explore data set by a combination of visualizations and intuitive filters.
o   Do statistical analysis using a wide range of built-in tests, such as ANOVA and nonparametric tests, as well as through the open interface to R.
o   Generate results with false discovery rates(q-value), fold change, difference and p-values.

o   Use the Biomarker workbench to run many statistical tests in batch mode.
o   Import data and clinical annotations with easy to use Wizards.
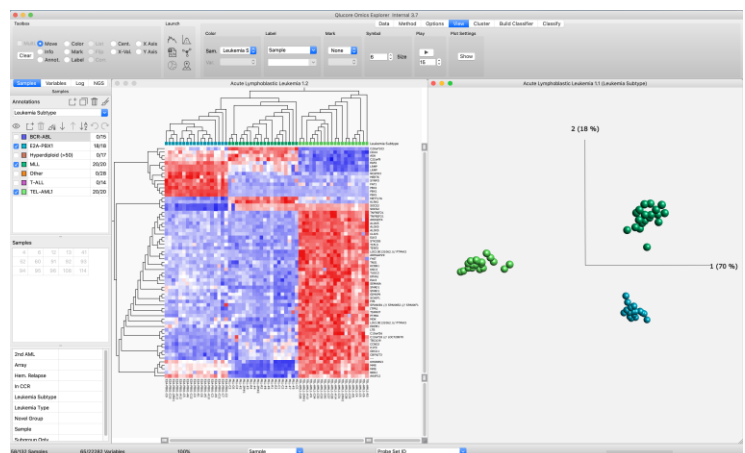
- o Pre-filter to remove unwanted variables.
- o Perform hierarchical clustering and generate dynamic heatmap plots.
- o Analyze RNA-seq data both in the Genome browser and a PCA plot in a synchronized view. Analyze single cell RNA-seq data using K-means++ clustering and t-SNE plots.
- o Fine tune and generate results using any combination of scatter, violin, venn, volcano, line and more plots.
- o Instantly create Principal Component Analysis PCA plots of large data sets and confirm the information content by using the Qlucore unique functionality Projection Score. Maximize the Projection Score automatically by letting the program select the optimal variance filter settings.
- o Use unsupervised clustering based on K-means++ clustering and silhouette values.
- o Use any of several methods, Hierarchical clustering, PCA, clustering, ISOMAP and or graphs for visual data exploration.
- o Build classifiers and classify new samples.
- o Do functional analysis in the context of public available gene sets such as pathways and so on using GSEA.
- o Download data from Gene Expression Omnibus (GEO) and TCGA to compare your own results with published material.
- o Remove unwanted dependencies such as artifacts and outliers. Manage batch effects.
- o Benefit from the streamlined workflows for aligned BAM files with RNA-seq data and 10X generated H5 files.
- o Keep track of your work with powerful global log and restore function.
- o Estimate data "intrinsic" dimensionality with the help of Scree and Talus plots
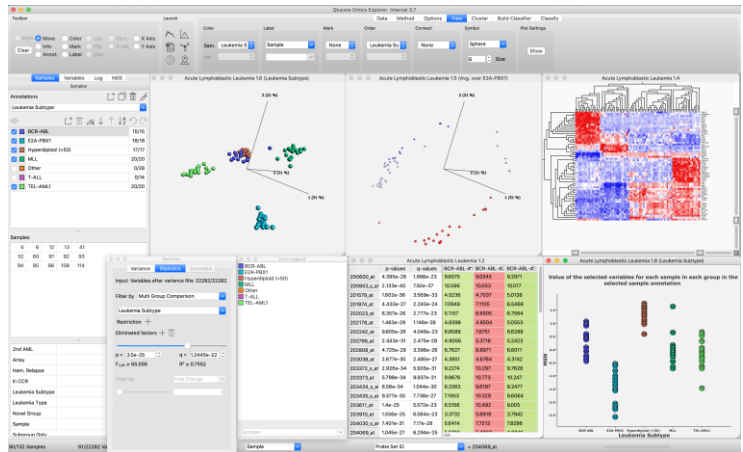
OUTPUT

- o High quality 2-D and 3-D graphics.
- o 17 plot types: Heatmap, sample PCA, variable PCA, bar plot, sample scatter plot, variable scatter plot, box plot, violin plot, line plot, pie chart, histogram, volcano plot, Venn diagram, UMAP plot, t-SNE plot, Talus plot, Scree plot and Kaplan-Meier.

- o Data table view.
- o GSEA results; enrichment plots as well as leading edge heatmaps and result lists.
- o Flexible ordering in heatmap. Order according to hierarchical clustering or an annotation or a statistical value.
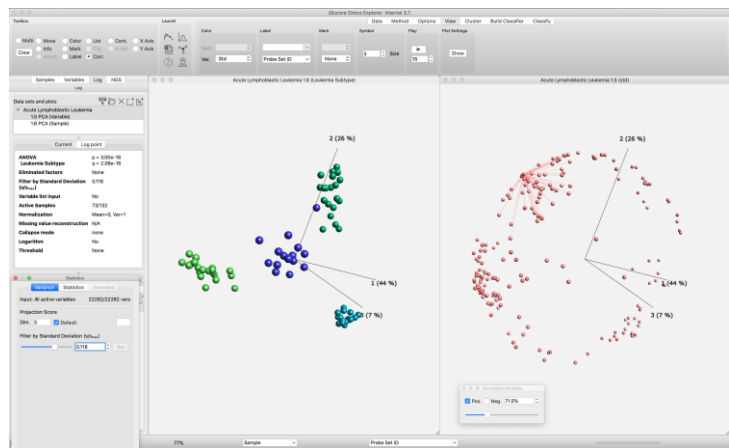
- o Present multiple annotations in a heatmap.
- o Plot both samples and variables.
- o Open multiple data sets at one time.
- o Synchronized plots (as many as you like). Synchronized plots are updated simultaneously.
- o Variable lists with p-values, fold change and FDR (q-values) values including a complete description of how the list was generated.
- o Plot arbitrary principal components.
- o Color the samples and the variables through different methods.
- o Show and visualize sample annotations with both colors and symbols.
- o Color the variables according to any list.
- o Label the samples and the variables through different methods.
- o Present multiple scatter plots in one view.
- o Color legend window. Explains the colors and scales in the active plot. Can be exported with a plot.
- o Classifiers

TEMPLATES

- o Execute templates to perform a scripted sequence of events in the program.
- o Use built in templates or develop new ones.
- o Combine QOE with other steps in a larger work-flow with the Python based templates
- o Use templates to make it easier to get started
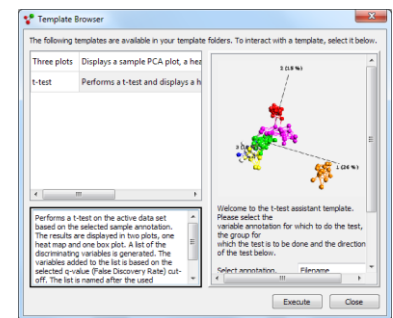
GSEA WORKBENCH

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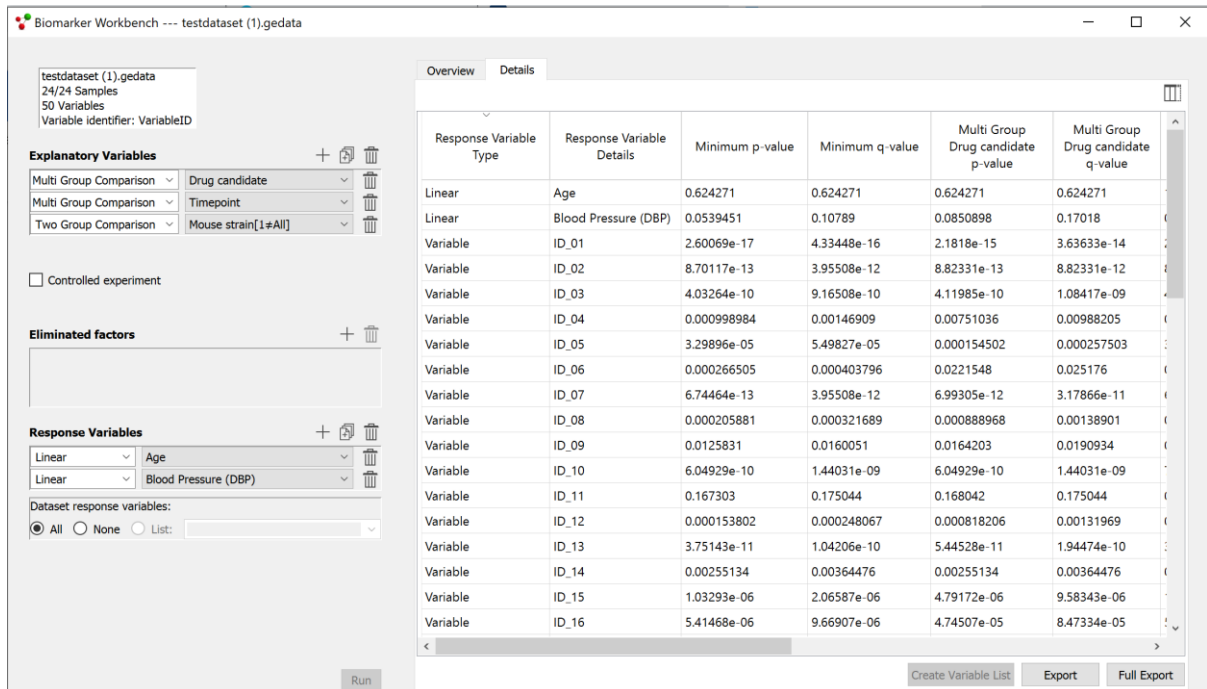- o Export plots and lists for publication and further analysis
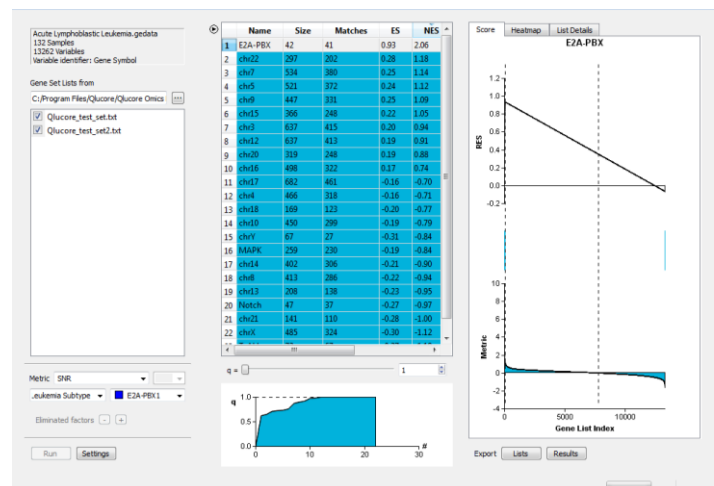
BIOMARKER WORKBENCH

- o The integrated Biomarker Workbench is optimized for experiments and studies in the areas of drug development and biomarker discovery.
- o Easy set-up of a suite of different statistical tests to run in batch mode with the objective to select effective compounds or other relevant signals.
- o Drug candidate testing
- o Large scale clinical annotation analysis by setting up Response variables. The Response variable option is tailored for biomarker discovery and assists in finding correlation

between sample annotations – an excellent way to focus on key annotations when working with large amounts of clinical data.

o   Linear models and logistic regression as part of flexible Generalized Linear Model framework.

o   Detailed output of results in comprehensive tables

o   Export of lists and tests for continued exploration in QOE

GO BROWSER

o   Use any ontology that you prefer.

o   Excellent overview by both tree and flat view of results.

o   Very fast search.

o   Export lists as variable lists to be used in OE main window.

NGS MODULE (OPTIONAL ADD ON)

o   Genome Browser

o   Gene Fusion Workbench including circular plots

o   Variant calling

o   Extensive filtering

o   RNA-seq expression analysis

o   Chip-seq peak detection and peak count data analysis

o   ATAC-seq peak detection and peak count data analysis

o   Combined analysis, i.e. run powerful statistics and machine learning on count data matrix and combine with detailed view in the genome browser.

Detailed information about the NGS module features is presented in the NGS Module feature overview document.

EDITING OF DATA

o   Interactive editing of sample annotations.

- o Subsampling
- o Interactive editing of variable lists.
- o Variable collapse. Select any variable annotation and collapse (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.) the data on this annotation.
- o Pre-filter for removal of variables below certain thresholds

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

VARIABLE LISTS
- o Automatic variable list for all active variables.
- o Automatic variable list for Search results.
- o Set operations on variable lists.
- o Convert gene identifiers using Broad Institute CHIP files.
- o Color any variable plot according to any selection of variable lists.
- o Save variable lists including information about how they were created, this helps in creating good result traceability.
- o Export Tukey statistics with a variable list.
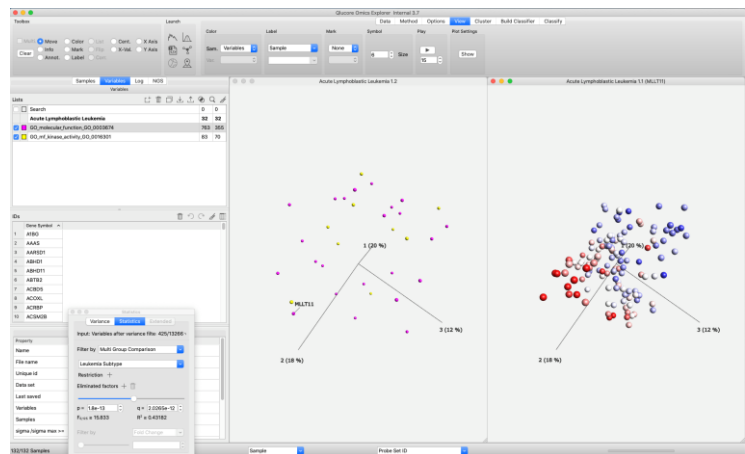
VERIFICATION
- o Projection score to understand how much information that is captured in a PCA plot.
- o Get direct feedback on p-values and q-values during variable selection.
- o Verify results by redoing the analysis with permuted sample annotations or with random numbers.
- o Verify results through remove-one-at-a-time cross validation or several at the time cross validation.
- o Continuous update on captured variance in PCA plot.
- o Generate histograms to check variable distributions.
- o Generate box plots to visualize results. Select from a range of significance levels.

CLUSTERS AND NETWORKS

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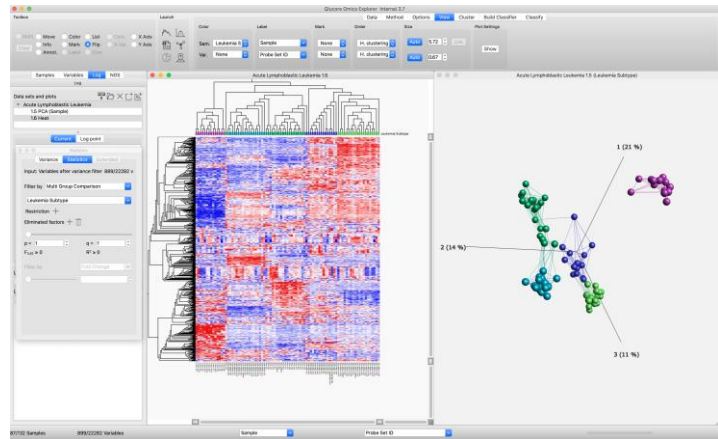
CLASSIFICATION

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

BATCH CORRECTIONS

o Correct multiple batch effects using in-built methods.

OPEN INTERFACE

o Interface to R for statistical tests
o Expands the available statistical tests.
o Works for two-group, paired tests and multi groups, with or without eliminated factors.
o Example on supported methods are: Limma, Mann-Whitney, Kruskal-Wallis, Welch,...

SURVIVAL ANALYSIS

o Kaplan Meier plots
o Median survival
o Hazard ratios (Cox)

IMPORT

o Aligned BAM files with RNA-seq data.
o Single cell RNA-seq data import of 10X Genomics data
o Affymetrix .cel files and .chp files. Including normalization and generation of QC-report.
o 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.
o Salmon file import (quant.sf)
o Agilent GeneView files (*.txt) for miRNA data.
o Flexible import wizard for import of (*.txt, *.csv,*.tsv) files both for data and annotations. Including options for count data.
o Data files of .gedata format (normal or transposed).
o Annotation files, both samples and variables (*.txt, *.csv).

- o Basic file formats such as
  - o Data files of .txt format (one variable identifier column and one sample identifier row). Called Simple text files.
  - o Compact Text files (*.csv)
- o Data (GDS and GSE) from Gene Expression Omnibus (GEO). Direct download.
- o GEO soft files (*.soft and *.soft.gz).
- o GEO Series Matrix (*.txt and *.txt.gz).
- o Gene set files for the GSEA Workbench (*.txt and *.gmt).
- o Broad CHIP files for identifier conversions
- o Ontology files (*.obo, *.obo.xml, obo-xeml.gz, obo-xml).
- o Raw count files
- o Variable lists
- o Direct NetAffx import. Affymetrix CHP and ARR files
- o Log files
- o Quality control plot files
- o Created classifiers

EXPORT
- o Still images (plots) including a legend plot. Select resolution and the plot is exported.
- o Copy to clipboard
- o Data files
- o Variable lists (with annotations and data if so preferred)
- o Videos
- o Pdf report with all open plots
- o Log files
- o Correlation and co-variance matrixes
- o QC-report
- o PCA loadings
- o PCA plot coordinates
- o Classifiers
- o Hazard Ratios
- o In-depth statistical test tables from Biomarker workbench.

OTHER
- o Missing value reconstruction (two versions)
- o Variable normalization
- o Multi-dimensional rescaling
- o Isomap
- o Take logarithm of data and set thresholds
- o 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 and Affymetrix are trademarks of ThermoFisher

CREDITS

- o GSEA: Subramanian, Tamayo, et al. 2005 Proc Natl Acad Sci U S A 102(43):15545-50
- o The Gene Ontology Consortium. "Gene ontology: tool for the unification of biology." Nat. Genet.. May 2000;25(1):25-9.
- o R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.