Understand the biology behind your Qlucore Omics Explorer data

Introduction

This is a guide for biologists who are using Qlucore Omics Explorer (QOE) visualization and statistical analysis tools and who want to gain additional insight into the underlying biology that drives your omics. The combination of Qlucore Omics Explorer with Clarivate's MetaCore™ allows you to analyze huge amounts of complex omics data and rationalize the biological processes behind that data in one easy workflow.

Qlucore Omics Explorer allows you to quickly explore, analyze and visualize large complex datasets to find discriminating variables (genes, proteins, metabolites etc) and to identify potentially new, interesting groups in the datasets. The built-in analytics allow you to quickly perform statistical tests such as t-test, ANOVA and regression analyses and identify discriminating variables. The results are visualized in plots that are updated in real time. QOE supports many different types of data. For gene expression analysis, data is typically loaded as RNA-seq BAM-files, pre-normalized txt or csv files. It is also possible to download and analyze public datasets from, for example, GEO and TCGA.

MetaCore is a molecular interactions database that allows you to upload gene lists from QOE and understand the biological pathways and processes that are driving your disease or treatment-response of interest. Various analysis workflows will guide you through mapping your data to manually curated interactions, signalling pathways, process networks and other ontologies to identify disrupted biological process in your studies. In addition, search across the wealth of information in MetaCore for relationships between genes, diseases, pathways, and other biological topics to understand disease mechanisms and identify potential targets and biomarkers for further investigation.
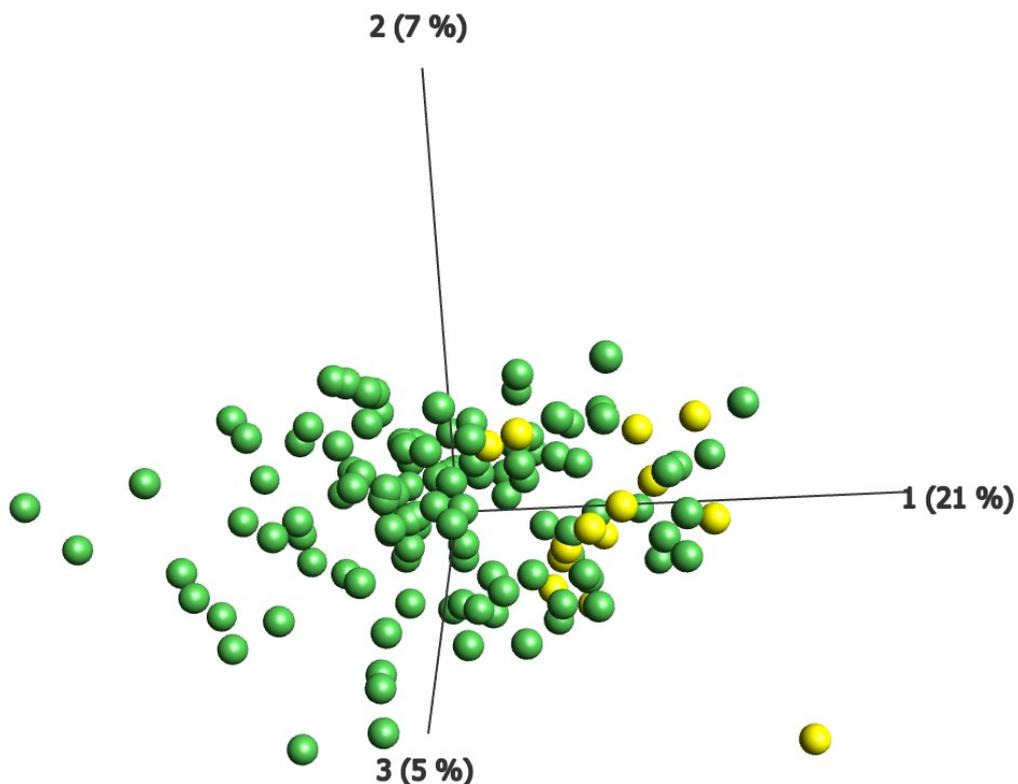
Use Case: Acute Lymphoblastic Leukemia

As an example of how QOE and MetaCore can be combined, we will analyze a gene expression dataset from pediatric acute lymphoblastic leukemia (ALL) patients, described by Ross et al., "Classification of pediatric acute lymphoblastic leukemia by gene expression profiling". *Blood* (2003), 102: 2951-9 (**https://www.stjuderesearch.org/site/data/ALL3/**).

1. Using QOE to refine the data to a set of genes most likely to be driving the disease.

2. Exporting the gene list and expression fold-change so it can be imported into MetaCore.

3. Using MetaCore enrichment analysis to identify the disrupted pathways and biological processes in the pediatric ALL patients in this experiment.

Basic Statistical analysis in Qlucore Omics Explorer

See the **Additional Support** section below for how to upload different types of data into QOE.

Once the data set is loaded, it is instantly visualized in a sample principal component analysis (PCA) plot:

132 samples from patients with Acute Lymphoblastic Leukemia (ALL) belonging to two subgroups – T-ALL (yellow) and B-ALL (green) visualized in a sample PCA plot. No filtering is done – the plot is based on all 22282 gene probes. The sample coloring was based on the annotations from the data set.

To separate the groups, a t-test (Two-group comparison) is applied, and the filters are set to a false discovery rate (FDR q-value) cutoff of 0.0019 and a Fold-change cutoff of 2,75.

This results in 250 highly discriminating genes, as shown in the plots below.
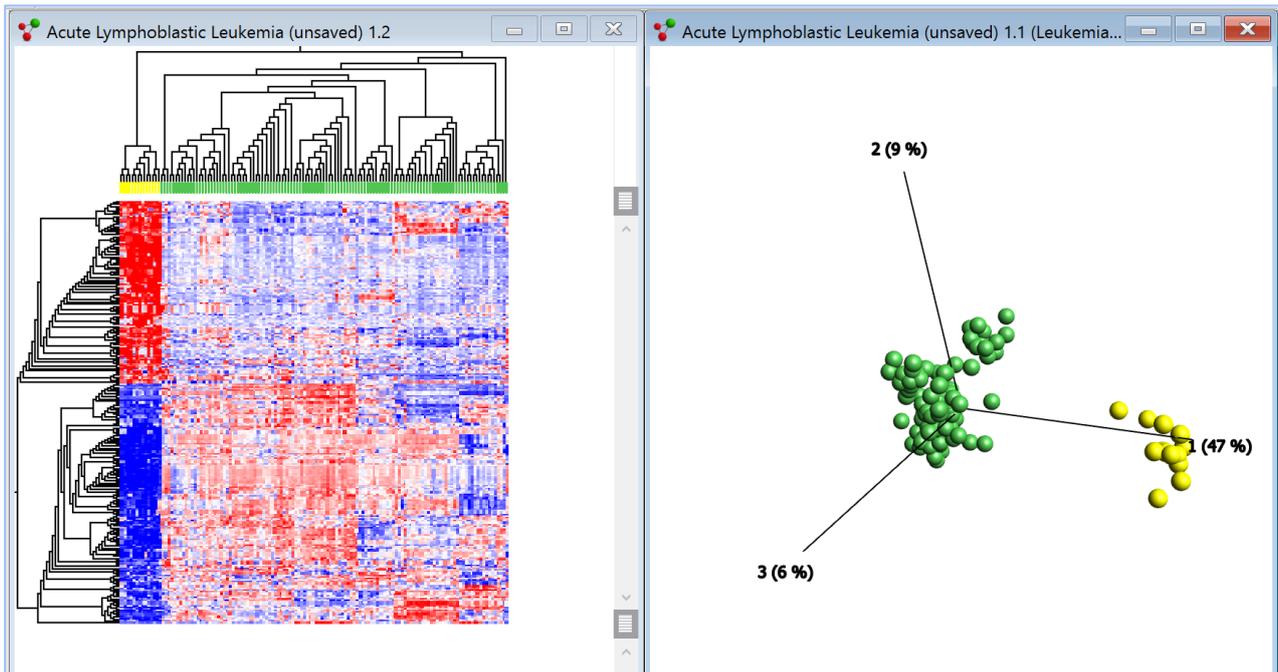
The PCA plot and heatmap show a clear discrimination between genes that are over/under expressed in B-ALL versus T-ALL.

Exporting omics data for analysis in MetaCore

Having refined the data to a list of 250 highly discriminating genes, they can now be easily exported to MetaCore by right clicking in one of the plots, selecting Copy to clipboard > MetaCore symbols. The probe IDs and log expression fold-change is exported. This information should then be pasted into Excel and saved as a tab delimited text file.
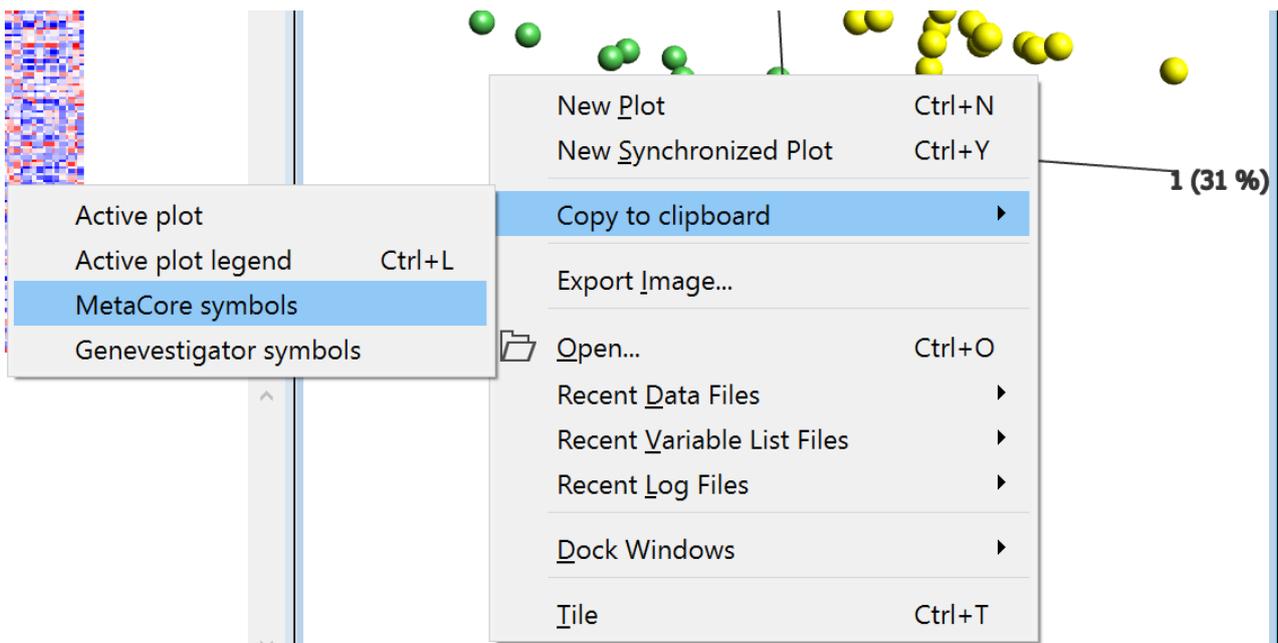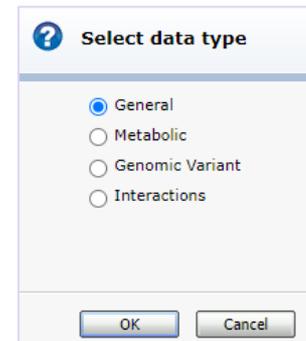
Uploading omics data to MetaCore

Below is a quick reference guide to uploading the data:

1. Click "Upload Data" in the top left of your screen.
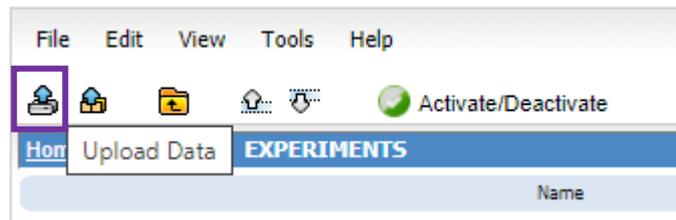
2. Select data type, General, to upload gene or proteins.

3. Select your file to upload in step 1 of the Data Analysis Wizard, then click "Next".
4. In Step 2, check the details are correct and select the column headers, then click "Next"
5. Choose the species in Step 3 > Next

6. A status bar will let you know when your data is ready to analyze
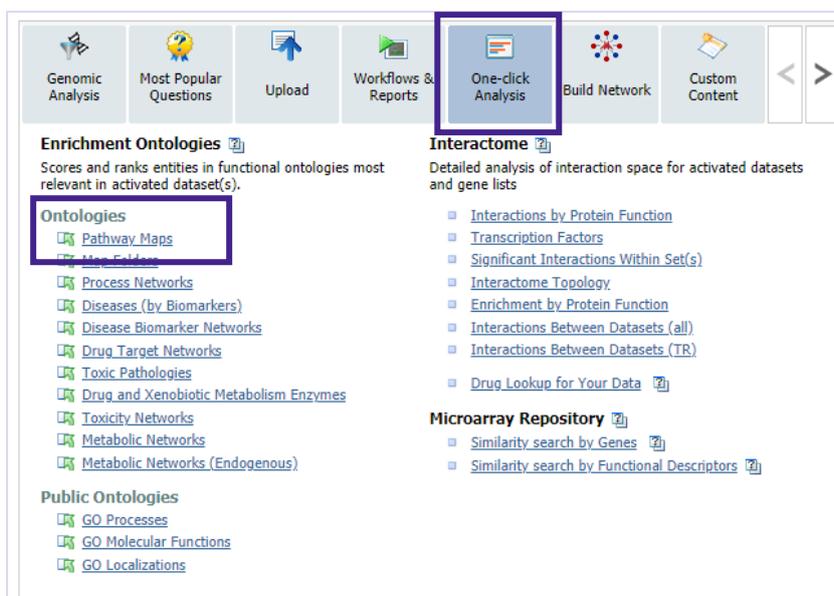7. Click the Experiment Name to activate your dataset in the Data Manager window

Metacore pathway enrichment

So far, we have refined the data using QOE and uploaded it to MetaCore for analysis. In this example, we'll use enrichment analysis to explore the most relevant pathways, processes, targets, biomarkers, and molecular functions in the dataset. The following steps will guide you through pathway map enrichment.

1. Make sure your dataset is in the "Active Data" window in the lower left panel of your screen

Name	Type	Date	
Home ▸ **Active Data**			
⬆ ... [..] Active Data			
🔲 Gene list 250 genes sep T-ALL ... ✅	GX	06/23/2021 04:17:51	
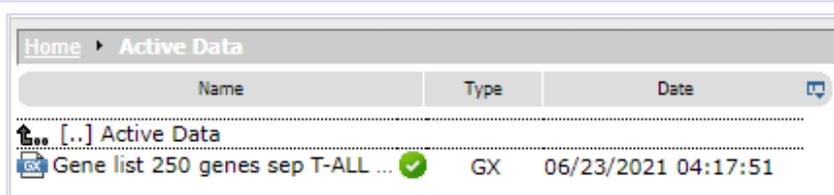
2. From the control panel on the upper right, select > One-Click Analysis, then > Pathway Maps

Genomic Analysis | Most Popular Questions | Upload | Workflows & Reports | **One-click Analysis** | Build Network | Custom Content | ‹ ›

Enrichment Ontologies ⓘ
Scores and ranks entities in functional ontologies most relevant in activated dataset(s).

Ontologies
- Pathway Maps
- Map Folders
- Process Networks
- Diseases (by Biomarkers)
- Disease Biomarker Networks
- Drug Target Networks
- Toxic Pathologies
- Drug and Xenobiotic Metabolism Enzymes
- Toxicity Networks
- Metabolic Networks
- Metabolic Networks (Endogenous)

Public Ontologies
- GO Processes
- GO Molecular Functions
- GO Localizations

Interactome ⓘ
Detailed analysis of interaction space for activated datasets and gene lists

- Interactions by Protein Function
- Transcription Factors
- Significant Interactions Within Set(s)
- Interactome Topology
- Enrichment by Protein Function
- Interactions Between Datasets (all)
- Interactions Between Datasets (TR)
- Drug Lookup for Your Data ⓘ

Microarray Repository
- Similarity search by Genes ⓘ
- Similarity search by Functional Descriptors ⓘ

Interpreting the pathway map enrichment results

The pediatric ALL data has been mapped onto MetaCore's high quality, manually curated pathway maps.

| #	Maps	0 2 4 6 8 10 -log(pValue)	pValue ↑	FDR	Ratio
1	Immune response_NF-AT in immune response		3.040e-14	2.696e-11	13/51
2	Immune response_Induction of the antigen presentation machinery by IFN-gamma		1.443e-12	6.399e-10	12/53
3	Breakdown of CD4+ T cell peripheral tolerance in type 1 diabetes mellitus		3.408e-10	9.348e-8	10/49
4	Rheumatoid arthritis (general schema)		4.216e-10	9.348e-8	10/50
5	Maturation and migration of dendritic cells in skin sensitization		1.343e-9	2.383e-7	9/41
6	Immune response_Antigen presentation by MHC class II		2.181e-9	3.225e-7	13/118
7	COVID-19: immune dysregulation		3.450e-9	4.295e-7	12/100
8	SLE genetic marker-specific pathways in T cells		3.874e-9	4.295e-7	12/101
9	Inflammatory response in dry eye		8.114e-9	7.997e-7	8/35
10	B cell signaling in hematological malignancies		3.384e-8	3.001e-6	10/77
11	Signal transduction_Erk Interactions: Inhibition of Erk		1.573e-7	1.269e-5	7/34
12	Transition of Monoclonal gammopathy of undetermined significance to active multiple myeloma (schema)		2.060e-7	1.523e-5	6/22

The overlap between the dataset and pathway maps is calculated using a hypergeometric mean, and the P value of the mean overlap is used to sort the maps with the most significant overlap at the top of the list.

A quick scan of the map names shows highly enriched maps related to T- and B-Cell immune responses, which makes sense given the dataset came from ALL patients, and the disruptions occurring in their lymphocytes.

To further refine the significance of the overlap, a False Discovery Rate (FDR) correction for the *P* value is also provided for each map.

The ratio indicates the number of genes in the experimental dataset over the total number of genes on the map. Click on this number to see a list of "Network Objects" (the proteins that correspond to your gene list) and the over/under expression values.

Click on the map name to see the map itself. For this example, the "B-cell signalling in hematologic malignancies" map was selected to demonstrate the visualization of the experimental pediatric ALL data in the context of this disease map:
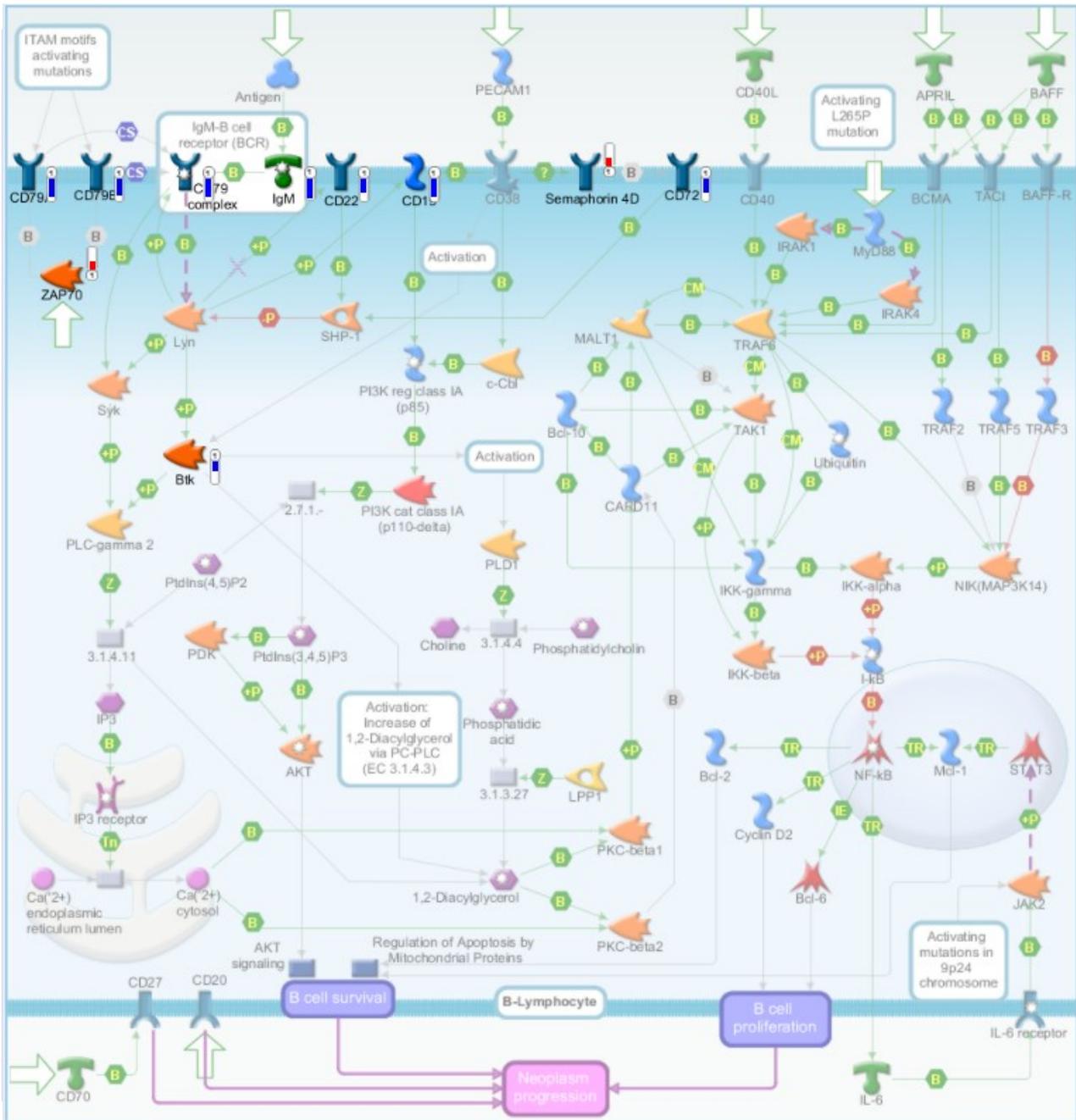
This map shows a series of B-cell surface receptors have been under or over-expressed in the experimental dataset, displayed as thermometers where the amount of blue/red indicates the level of under/over-expression, activating a signal cascade that results in B-cell survival and progression of the neoplasm.

In Summary

The initial analysis with Qlucore Omics Explorer was able to differentiate B-Cell and T-Cell lymphomas in the dataset. From the differences in expression between the B-cell and T-cell lymphomas, we can hypothesize that the cell surface receptors shown above are driving the B-cell lymphoma in this subset of patients.

What next?

Having used QOE with MetaCore to rationalize the mechanistic drivers behind the ALL in these pediatric patients, you could go on to use MetaCore to explore possible biomarkers to differentiate the B- and T-Cell subtypes using the biomarker enrichment ontologies; or hypothesize druggable targets using the target enrichment.

The example above is for analysis of differential gene expression data from microarrays. The following datatypes can also be exported from QOE for analysis in MetaCore; and note that data from multiple different experiments on the same population can be combined in MetaCore with the Compare Experiments workflow.

- Gene expression (RNA-Seq and Array)
- Single-cell RNA-Seq
- Proteomics
- Metabolomics
- miRNA data
- qPCR data
- NGS module: RNA-Seq, DNA-Seq, ChIP-Seq and ATAC-Seq data (in QOE export to .txt file for analysis in MetaCore)

Additional Support

Learn more about how to explore, analyze and visualize your data with Qlucore Omics Explorer from the Video tutorials in the Help menu or join a webinar or hands-on training from **here**

Or contact Qlucore directly from **here**

Explore MetaCore's library of training resources to understand enrichment, interactome and network analysis:

Or use the "Contact us" link shown above to contact Clarivate Support.