



Qlucore webinar

Bulk RNAseq data analysis

Yana Stackpole, PhD

Qlucore



QluCore Webinars and courses at
<http://www.qlucore.com/webinar>

Making a difference by transforming complexity into clarity

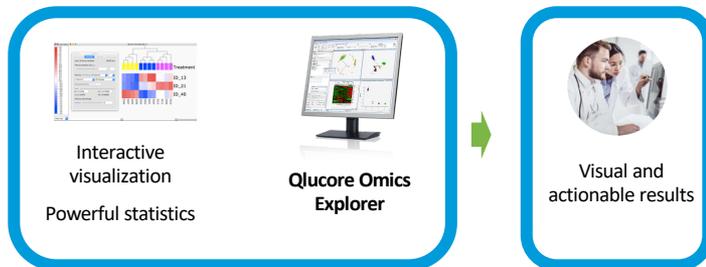


Multi Omics Data Analysis Software

- Since 2007
- Customers in 25 countries
- 1000+ scientific articles
- >2500 times faster

Precision & Companion Diagnostic Software

- AI-based machine learning classifiers
- Leukemia, Lung cancer and Bladder cancer classifiers for gene expression and fusions.
- **Qlucore Insights (RUO) & Qlucore Diagnostics (IVDR)** software.



Data types

Examples

- Gene Expression (RNA-seq & array)
- Single cell RNA-seq
- DNA Methylation
- Proteomics
- Metabolomics
- Protein array data
- miRNA data
- qPCR data
- Flow cytometry data
- NGS module: RNA-seq, DNA-seq, CHIP-seq and ATAC-seq data

Any multivariate data



Supported File formats

- RNA-seq (aligned BAM files & gtf file, quant.sf files)
- Thermo Fisher/Affymetrix GeneChip compatible (.cel files)
- Agilent txt files
- GEO soft files
- TCGA Data
- 10X Genomics data
- Wizard (*.txt, *.csv)
- NGS module: fasta, vcf, gtf, bed and cytoband files



Aligned BAM files

Preparation

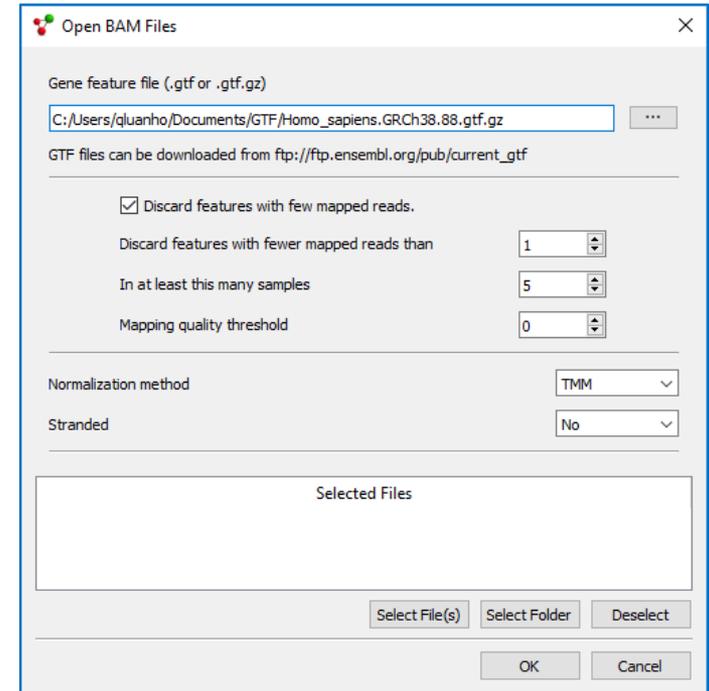
- Enter the path to the Reference Genome in File->Open BAM files
- Define optional pre-filtration
- Select normalization method

Import

- Select individual BAM files or folder
- The BAM files will be counted, normalized using the selected method and log transformed. Data set will be opened up in a PCA plot

Import additional annotations

- See "Add annotations"



Note: The BAM files need to be aligned and sorted on coordinate. The reference genome must be the same that was used used for the alignment.

Raw count based RNA-seq matrix

- Start "Open with Wizard..."
- Select radio button for Raw count data
- Select normalization method
- Select where to find feature length info (if available)
- Select data, variable and sample annotations
- Finish import

 Data Import Wizard

Is this a raw count matrix?

Yes No

File Window License Help

-  New
-  Open... Ctrl+O
-  Open with Wizard... →
-  Open BAM Files (RNA seq)...
-  Open quant.sf Files...
-  Open R Data Frame...

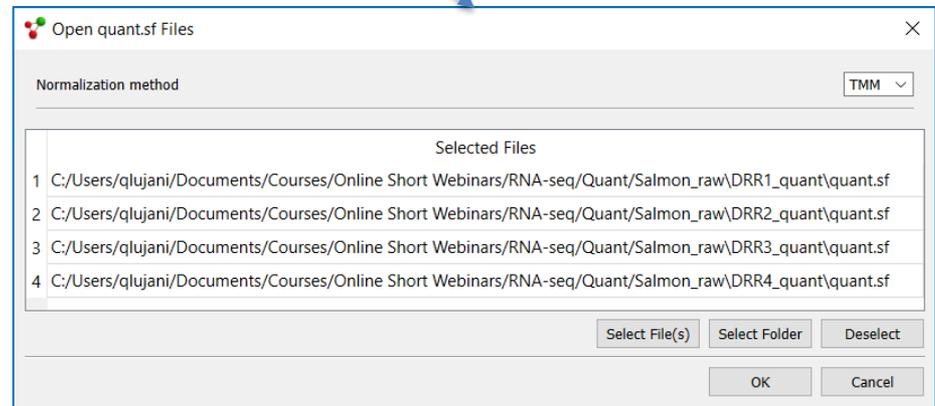
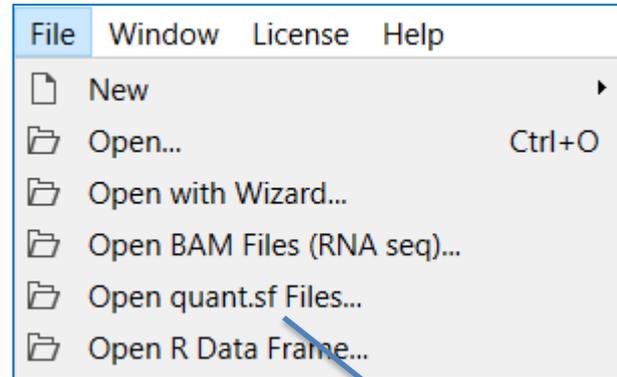
 Data Import Wizard

Select the feature (variable) annotation that contains normalization weights

	1	2	3	4	5	6	7	8	9	10	11
1					SampleID	5303	5308	5302	5309	5312	5305
2					Age	20	26	28	30	40	40
3					Gender	Female	Female	Male	Male	Male	Male
4					Treatm...	Drug 2	Placebo	Drug 2	Drug 1	Drug 1	Placebo
5					Gender ...	P3	P1	P3	P4	P5	P2
6	Variabl...	Symbol	Name	Feature...							
7	ID_01	ENC1	Variable ...	1000		17681	16698	15745	15736	14753	16772
8	ID_02	CDK8	Variable ...	763		67677	10977	28151	4508	21552	35296
9	ID_03	PEX7	Variable ...	900		20543	21216	11456	16594	27958	7340
10	ID_04	VPS13D	Variable ...	860		31843	91052	36436	213600	240177	108948

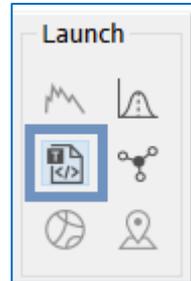
Salmon pipeline data (quant.sf files)

- quant.sf files can come from the Salmon program library and is also output format used by Illumina owned platforms
- The top folder is selected, then the quant.sf files in the tree are automatically identified for import



TCGA data

- Launch the Template Browser
- Use the predefined Qlucore template for import of data from TCGA
- Note that the TCGA Template can download both normalized data and un-normalized data
- Note: Qlucore Templates are Python scripts



Template Browser

The following templates are available in your template folders. To interact with a template, select it below.

10X Genomics Assistant	Load single cell data from 10X Genomics
FTP download, load, and preprocess	Download RNA abundance data from EN
Heatmap sample clustering	Plots a heatmap and orders samples by p
Multigroup anova	Performs an anova on a grouping annota
t-test	Performs a t-test and displays a heatmap
t-test & fold-change	Filter by fold-change and p-value before
t-test (paired)	Performs a paired t-test and displays a he
TCGA RSEM	Download RNA expression data from TCG
Three plots	Displays a sample PCA plot, a heat map a

This particular template will download TCGA mRNA gene expression data from GDAC (Broad Institute) and downloading constitutes agreement to the TCGA data usage policy. Only datasets with mRNA expression levels (abundances) processed are available through this template, i.e. 27 of 38 cohorts.

This is an advanced template designed to both provide easy access to data but also provide an example for developers/bioinformaticians to write custom scripts for processing on-line information and loading it into Qlucore Omics Explorer.

 The following TCGA cohorts are available for download:
 Adrenocortical carcinoma (ACC)



This is a template for download and processing of TCGA data. Using it constitutes agreement to the TCGA data usage policy.

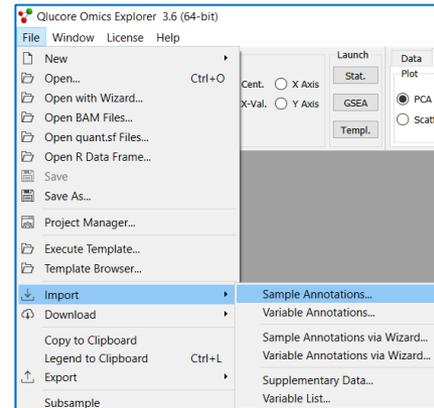
--- Input fields ---
 'Cohort': Input only one cohort in this field and it must be the abbreviated name of that cohort.
 See below for a list of cohorts.
 'Normalize': Write "yes" to download normalized gene expression data. Write "no" to download raw counts. If you intend to use the limma methodology we recommend using the raw data.
 --- User notes ---
 It may take a few minutes to download and process the data, please be patient.
 Choose a cohort and click OK to execute when you are ready.

Execute Close

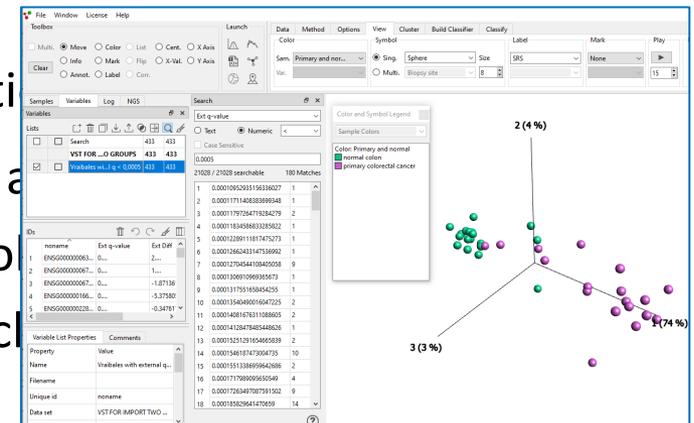
Add annotations

- The unique sample/variable id in the first column must match the id in the imported dataset

SampleID	Age	Gender	Treatment	Rank	Censor
5303	20	Female	Drug 2	Very low	1
5308	26	Female	Placebo	Very low	1
5302	28	Male	Drug 2	Low	1
5309	30	Male	Drug 1	Low	1
5312	40	Male	Drug 1	Medium	1
5305	40	Male	Placebo	Medium	0
5311	43	Female	Drug 1	Medium	1
5307	48	Male	Placebo	High	1
5301	54	Male	Drug 2	High	1
5310	56	Female	Drug 1	Very high	0
5306	56	Female	Placebo	Very high	1
5304	63	Female	Drug 2	Very high	1

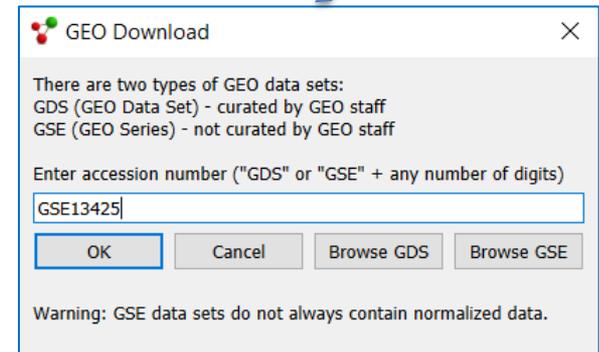
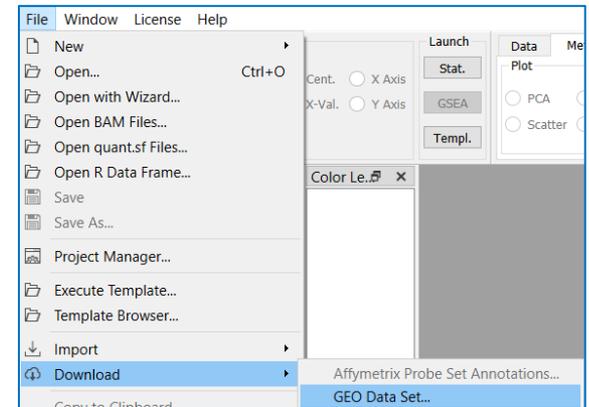
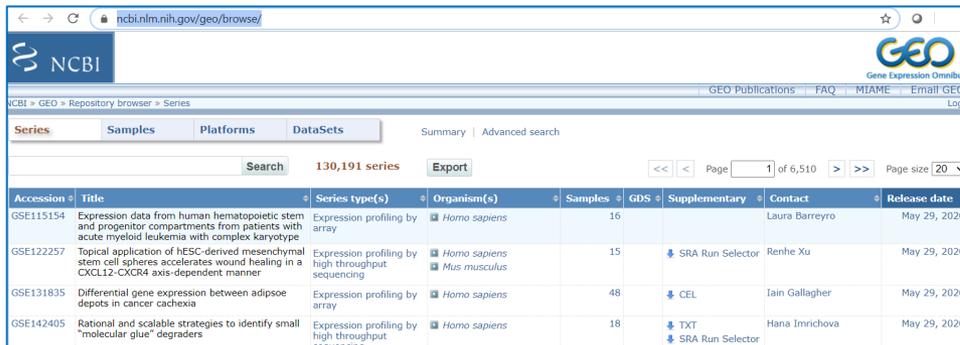


- Note: You can also import variable annotations
- NOTE: You can import external test results and variable annotations (like p/q-values, log fold change/fold change) and filter using "Search"



Download data from GEO

- GEO – Gene Expression Omnibus – is an online repository for data.
- In Qlucore you can directly download datasets from GEO in the SOFT file format
- In case where data is stored as raw data or in the SRA (Sequence Read Archive), then you need to pre-processed data before import into Omics Explorer.

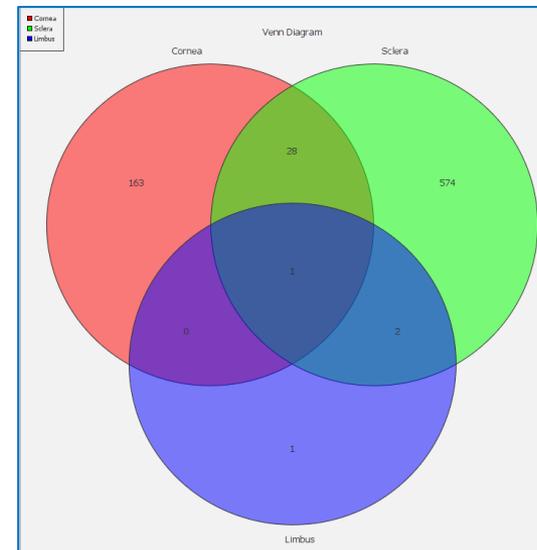
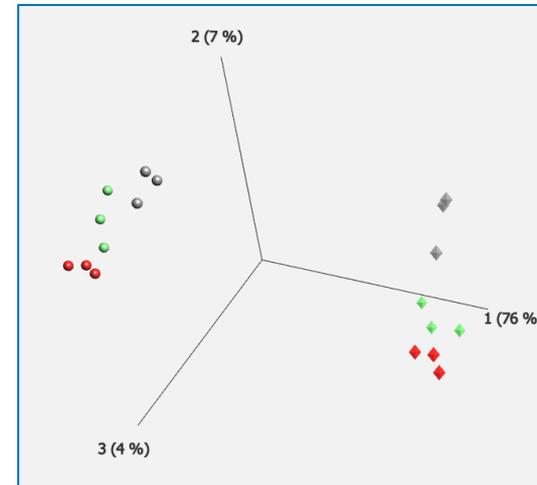



The screenshot shows the NCBI GEO website interface. The 'DataSets' tab is selected, and a search for '130,191 series' is displayed. The following table shows the first four rows of the dataset list:

Accession	Title	Series type(s)	Organism(s)	Samples	GDS	Supplementary	Contact	Release date
GSE115154	Expression data from human hematopoietic stem and progenitor compartments from patients with acute myeloid leukemia with complex karyotype	Expression profiling by array	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	16			Laura Barreyro	May 29, 2020
GSE122257	Topical application of hESC-derived mesenchymal stem cell spheres accelerates wound healing in a CXCL12-CXCR4 axis-dependent manner	Expression profiling by high throughput sequencing	<input checked="" type="checkbox"/> <i>Homo sapiens</i> <input checked="" type="checkbox"/> <i>Mus musculus</i>	15		<input checked="" type="checkbox"/> SRA Run Selector	Renhe Xu	May 29, 2020
GSE131835	Differential gene expression between adipose depots in cancer cachexia	Expression profiling by array	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	48		<input checked="" type="checkbox"/> CEL	Iain Gallagher	May 29, 2020
GSE142405	Rational and scalable strategies to identify small "molecular glue" degraders	Expression profiling by high throughput sequencing	<input checked="" type="checkbox"/> <i>Homo sapiens</i>	18		<input checked="" type="checkbox"/> TXT <input checked="" type="checkbox"/> SRA Run Selector	Hana Imrichova	May 29, 2020

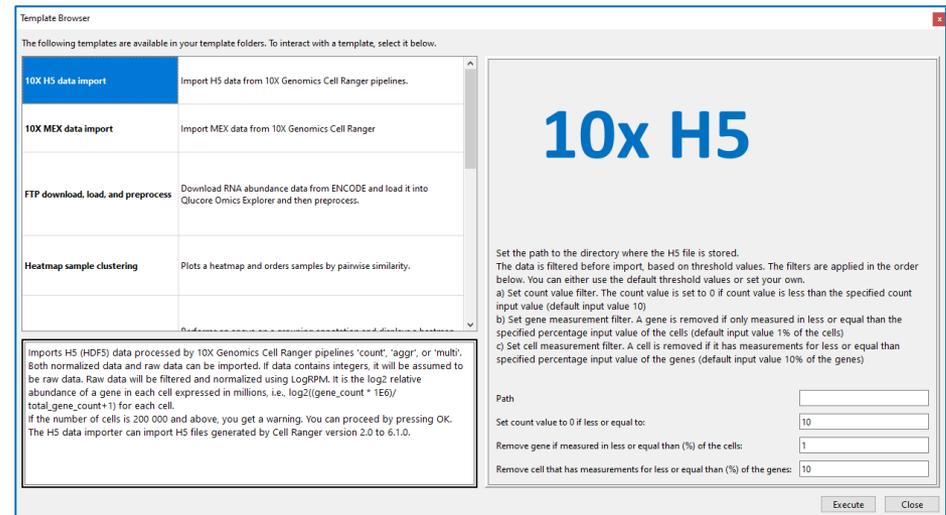
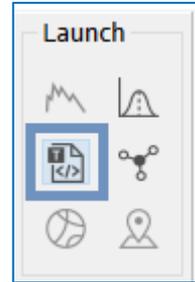
An example with RNAseq data

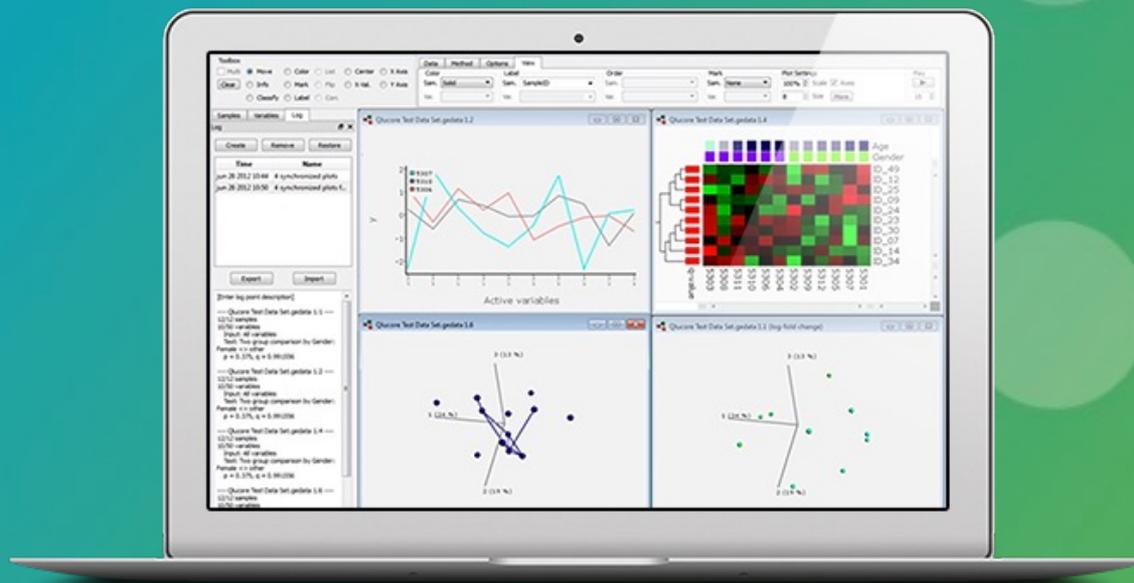
- SARS-CoV-2 from human tissue (eye)
- Raw count data in a matrix, GSE16407
- Using Wizard to import data, TPM normalization as in article
- Add sample annotations
- Set up t-tests for three types of tissue, infected against control
- Compare variable lists ($q = 0,05$)



10XGenomics Cell Ranger H5 data

- Launch the Template Browser
- Use the predefined Qlucore template for import of H5 data from 10x Genomics Cell Ranger pipeline
- Cell Ranger version 3.0 and above supported
- Cell Ranger can produce both normalized data and unnormalized data
- The template normalizes and log transforms raw count data, can apply filters on counts and on measurements





[Visit www.glucore.com](http://www.glucore.com)