

How to load 10x Genomics data

INTRODUCTION

Import of 10x Genomics Cell Ranger data into Qlucore Omics Explorer is easily done using Qlucore Templates. There are two templates for data import with support for different file formats. The templates are available through the Template Browser in Qlucore Omics:

10X H5 data import. This template loads input data in the H5 (also called HDF5) format.

10X MEX data import. This template loads input data in gz compressed tsv /mtx file format.

10X H5 DATA IMPORT TEMPLATE

The 10X H5 data import template loads 10x Genomics data sets from Cell Ranger pipelines (aggr, count or multi) saved as H5 files into Qlucore Omics Explorer. The template includes the following features:

- Support for multiplexing chemistry
- Support for aggregated datasets
- Support for importing multiple files, including datasets of mixed type
- Automatic annotation of samples for multiplexed and aggregated datasets
- Automatic filtering of data
- Data normalization

AUTOMATIC FILTERING OF DATA

The template will during import filter the data. The purpose is to enable removal of cells and genes that have too little information. The filter cutoffs can be adjusted. The automatic filtering uses the following criteria:

Count value filter. The count value is set to 0 if count value is less than the specified count input value (default input value 10)

Gene measurement filter. A gene is removed if only measured in less or equal than the specified percentage input value of the cells (default input value 1% of the cells)

Cell measurement filter. A cell is removed if it has measurements for less or equal than specified percentage input value of the genes (default input value 10% of the genes)

DATA NORMALIZATION

The template will detect if the data is normalized or not by checking if data is integers or not. If integer values are detected, then data is assumed to be raw, otherwise the template will assume that data has already been normalized / transformed in some way.

Raw data will be filtered and normalized using LogRPM. It is the log₂ relative abundance of a gene in each cell expressed in millions, i.e., $\log_2((\text{gene_count} * 1E6) / \text{total_gene_count} + 1)$ for each cell.

LAUNCH THE 10X H5 DATA IMPORT TEMPLATE

- Start Qlucore Omics Explorer
- Run the Template browser from the File menu or by clicking the “Template” icon in the Launch section
- Select the template named “10X H5 Data Import”.
- Paste the full path to the data directory in the text box labeled “Path”
- Select filter threshold values (default values are shown)
- Click the “Execute” button

NOTE: The template will load and then convert the data. The loading and conversion can take minutes. As an example (using a normal desktop computer) a 100 MB file takes 2 minutes.

DOWNLOAD 10X GENOMICS DATA OR USE YOUR OWN

- 10x Genomics stores several data sets for test and evaluation. Any dataset containing results from their Cell Ranger count, aggr or multi pipelines work. Example datasets in H5 format are available at <https://www.10xgenomics.com/resources/datasets/>.
- Record the full path to the directory where the data is stored. The directory must contain a H5 file to work.

ANALYSIS

STEP 1 - FILTER THE DATA: The template has inbuilt support for filtering, see “AUTOMATIC FILTERING OF DATA”. The automatic filtering removes the low-quality data, and the filtering of data in the “Data” tab may not be needed.

STEP 2 - SUBSAMPLE: Subsampling can help explore the data faster. Very large data sets will impact QOE performance for some actions, such as displaying a PCA plot. For a normal desktop computer, you can expect to see slower updates when the active samples and variables exceed 20000 and 3000, respectively. You can find the subsampling functionality in the File menu.

STEP 3 - VARIANCE FILTERING AND STATISTICS: Just as it is pointless keeping data containing too many zeros, it is not optimal to keep variables with too little information. Use the variance filtering functionality in the statistics dialogue by clicking on the “Statistics” icon in the Launch section.

STEP 4 - UMAP, t-SNE & Clustering: Use for instance UMAP or t-SNE to view structure in single cell expression profiles. All other plots and functionality in QOE such as PCA are also available for single cells data. You can also select to do kmeans++ clustering.

10X MEX DATA IMPORT TEMPLATE

The 10X MEX data import template for Qlucore Omics Explorer allows users to easily load 10X Genomics data sets from Cell Ranger pipelines in MEX data format (aggr or count saved as a set of three files; a barcode, a features and a matrix file) into Qlucore Omics Explorer with ease.

These pipelines output matrix files that contain sparse representations of the data in Market Exchange (MEX) format, as well as the necessary annotation tables for columns (samples/cases) and rows (genes/variables). If you have your own data, you may continue to import custom annotations; for example, perhaps you have reclassified the cells by flow cytometry or there are case-control groups that require statistical comparison.

DOWNLOAD 10X DATA OR USE YOUR OWN

- 10x Genomics stores several data sets for test and evaluation. Any dataset containing results from their Cell Ranger count or aggr pipelines works. Example datasets in MEX format are available at <https://www.10xgenomics.com/resources/datasets/>
- Once downloaded, make sure to unpack the archive into its own directory. Your own data will likely not be archived, and this step will not be necessary.
- Record the full path to the directory where the data is stored. The directory must contain the following 3 files to make the template work:
 - barcodes.tsv.gz
 - features.tsv.gz
 - matrix.mtx.gz

NOTE 1 - NORMALIZATION: The Cell Ranger pipelines output data that may or may not be normalized. The 10X MEX data import template is for output data that has already been normalized by the pipeline (no normalization will be done in the template). Qlucore Omics Explorer will scale and center the data so that the variance is 1 and the mean is 0. This template keeps this setting on and we recommend leaving it as is for most datasets.

NOTE 2 - VERSIONS: Prior to Cell Ranger 3.0 the output matrix file format was different. In particular, the file genes.tsv has been replaced by features.tsv to account for Feature Barcoding technology, and the matrix and barcode files are now gzipped. This template supports data from version 3.0 and onwards.

LAUNCH THE 10X MEX DATA IMPORT TEMPLATE

- Start Qlucore Omics Explorer
- Run the Template browser from the File menu or by clicking the “Template” icon in the Launch section
- Select the template named “10X MEX data import”.
- Paste the full path to the data directory in the text box labeled “Path”
- Click the “Execute” button.

How To Load 10X data rev B



NOTE: The template will load and then convert the data. The loading and conversion can take minutes. As an example (using a normal desktop computer) a 100 MB file takes 2 minutes.

ANALYSIS

See the analysis chapter after the H5 above.

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.

TRADEMARK LIST

10X, 10X GENOMICS and CELL RANGER are trademarks of 10X Genomics, Inc.