

How to load 10x Genomics data

The **Load 10x data** Template for Qlucore Omics Explorer allows users to load 10x Genomics data sets generated with Cell Ranger 3.0 or later pipelines into Qlucore Omics Explorer with ease.

10x data is processed through the Cell Ranger pipelines, which outputs numerous files. The pipelines to take note of are the “count” and “aggr” pipelines. These pipelines output matrix files that contain sparse representations of the data in Matrix Market Exchange file format, as well as the necessary annotation tables for columns (samples/cases) and rows (genes/variables). These files are available in archives and summarized below. This document describes steps to use this template with your own 10x data. We will also show you where to find example files, downloadable from 10x Genomics. 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.

These are the file formats which are contained within the tar archives:

- barcodes.tsv.gz – One column, no header, unix line endings; Each row contains a barcode representing one cell in one sample.
- features.tsv.gz – Three columns, no header, unix line endings, tab separated; Each row contains a unique feature ID (col-1), a feature name (col-2), and a feature type (col-3).
- matrix.mtx.gz – Follows the Market Exchange Format (<https://math.nist.gov/MatrixMarket/formats.html>) for describing matrices, including sparse matrices; header lines are prefixed with a “%” character; the first non-header line indicates the size of the matrix and the number of entries to expect; all subsequent lines are in coordinate format (not array) and separated into 3 columns where the first 2 are matrix coordinates and the 3rd is the value.

STEP 1 - DOWNLOAD 10XG DATA OR USE YOU OWN

- 10x Genomics stores several data sets for test and evaluation. Any dataset containing results from their Cell Ranger **count** or **aggr** pipelines work. You can download the datasets from <https://www.10xgenomics.com/resources/datasets/>.
- The rest of this guide continues with the “5k Peripheral blood mononuclear cells (PBMCs) from a healthy donor with cell surface proteins (v3 chemistry)” data, found at https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.1.0/5k_pbmc_protein_v3.

- At the bottom of the page you will see several download links, we will continue with the “Feature / cell matrix (filtered)” data, it is approximately 35 MB large.
- 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 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.csv has been replaced by features.csv.gz 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.***

STEP 2 – LAUNCH THE 10XG ASSISTANT TEMPLATE

- Start Qlucore Omics Explorer
- Run the Template browser from the File menu or by clicking the “Templ.” button.
- Select the template named “**Load 10x data**” and click on the “Execute” button.
- This will bring up a dialog.
- Paste the full path to the data directory in the text box labeled, “Path”, and click the “OK” button.

NOTE – RUN TIME: The template will load and then convert the data. The loading and conversion can take minutes. One example on one normal desktop computer is that a 100 MB matrix file, takes 2 minutes.

STEP 3 – FILTER THE DATA

- Single cell data has many zeros and it is recommended to remove samples/cells that contain too little data. To do this...
 - Setup variable prefiltering in the “Data” tab.
 - Set in this example: “Remove variables less than” to 0.0001 and “if true in x% samples” to 25%.

STEP 5 – SUBSAMPLE (OPTIONAL)

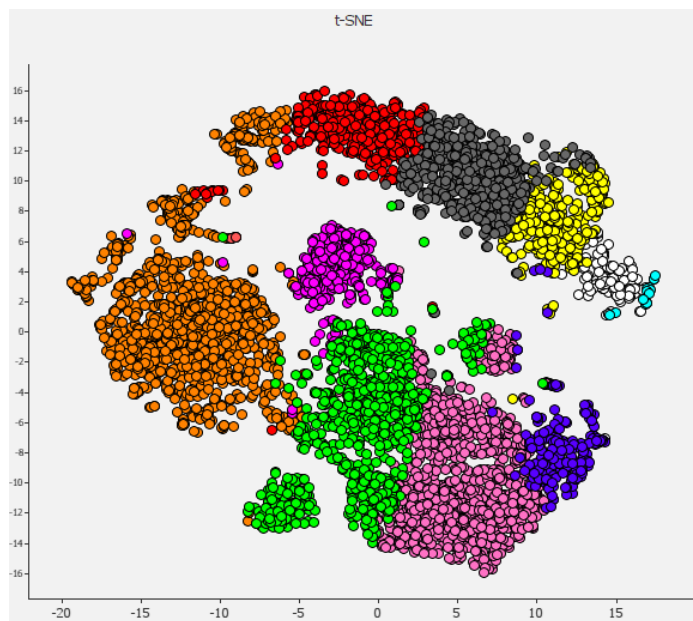
- Subsampling can help explore the data faster. Very large data sets will impact QOE performance for some actions, such as displaying a PCA plot. For a normal desktop computer, you can expect to see slower updates when the active samples and variables exceed 20000 and 3000, respectively.
- You can find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

STEP 6 – 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 “Stat.” button.

STEP 7 - t-SNE & Clustering

- Use for instance t-SNE with high perplexity (here equal to 60) and k-means clustering with k=11 to view structure in single cell expression profiles. The plot is colored according to the annotation generated with k-means clustering
- All other plots and functionality in QOE such as PCA are also available for single cells data.



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.