SEP 12, 2023

OPEN ACCESS

scRNA Delivery

Tyler Stahl[1]

[1]Genomics Research Center

Tyler Stahl

ABSTRACT

Overview of GRC scRNA delivery links and folders

Protocol Citation: Tyler Stahl 2023. scRNA Delivery. **protocols.io** https://protocols.io/view/scrna-delivery-cy5txy6n

Protocol status: In development

We will continue to add additional resources and FAQs to this protocol

Created: Aug 24, 2023

Last Modified: Sep 12, 2023

PROTOCOL integer ID: 86931

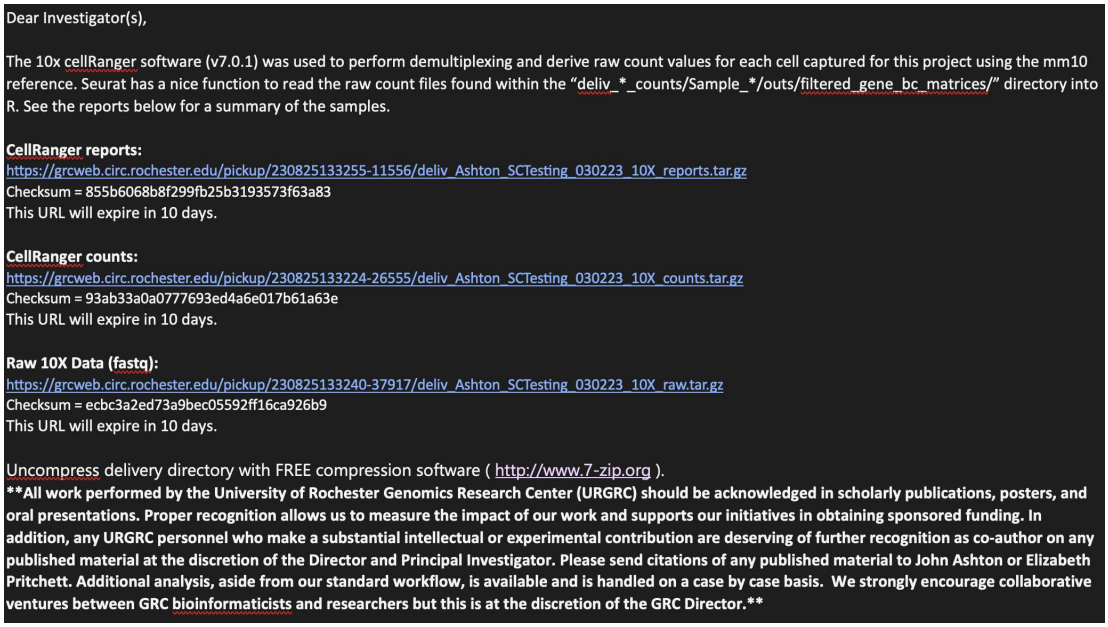
Overview of the Analysis

1 To analyze scRNA data, we use the cellRanger software provided by 10X. In brief, after sequencing, we demultiplex samples using cellRanger mkfastq. For each sample, we then run cellRanger counts. CellRanger counts performs cell counting, read alignment and gene expression quantification.

cellranger_methods.txt README.txt

Delivery Structure

2 In your GRC delivery email, there will be 3 delivery links: **1) Raw (fastq) data 2) counts data 3) reports**



Dear Investigator(s),

The 10x cellRanger software (v7.0.1) was used to perform demultiplexing and derive raw count values for each cell captured for this project using the mm10 reference. Seurat has a nice function to read the raw count files found within the "deliv_*_counts/Sample_*/outs/filtered_gene_bc_matrices/" directory into R. See the reports below for a summary of the samples.

CellRanger reports:
https://grcweb.circ.rochester.edu/pickup/230825133255-11556/deliv_Ashton_SCTesting_030223_10X_reports.tar.gz
Checksum = 855b6068b8f299fb25b3193573f63a83
This URL will expire in 10 days.

CellRanger counts:
https://grcweb.circ.rochester.edu/pickup/230825133224-26555/deliv_Ashton_SCTesting_030223_10X_counts.tar.gz
Checksum = 93ab33a0a0777693ed4a6e017b61a63e
This URL will expire in 10 days.

Raw 10X Data (fastq):
https://grcweb.circ.rochester.edu/pickup/230825133240-37917/deliv_Ashton_SCTesting_030223_10X_raw.tar.gz
Checksum = ecbc3a2ed73a9bec05592ff16ca926b9
This URL will expire in 10 days.

Uncompress delivery directory with FREE compression software (http://www.7-zip.org).
****All work performed by the University of Rochester Genomics Research Center (URGRC) should be acknowledged in scholarly publications, posters, and oral presentations. Proper recognition allows us to measure the impact of our work and supports our initiatives in obtaining sponsored funding. In addition, any URGRC personnel who make a substantial intellectual or experimental contribution are deserving of further recognition as co-author on any published material at the discretion of the Director and Principal Investigator. Please send citations of any published material to John Ashton or Elizabeth Pritchett. Additional analysis, aside from our standard workflow, is available and is handled on a case by case basis. We strongly encourage collaborative ventures between GRC bioinformaticists and researchers but this is at the discretion of the GRC Director.****

1) .fastq files contain nucleotide and quality information generated from the Illumina Sequencer. There is an unlikely reason that you would ever need to open these files directly; however, per the NIH data management policy fastq files need to be deposited online at the time of publication or end of the performance period and stored for 3 years after the grant. For more information on the NIH data management policy, please visit the following website: https://sharing.nih.gov/

2) Within the count's folder you can find the output from cell ranger counts for each sample. There are a lot of output files/folders from the pipeline; we define some of the important files/folders in the "Counts Output" section below

3) The cellranger web summary for each sample can be found in deliv_*_reports. Web summaries, give an overview of sample data including cell counting, mapping, and quality information.

The fastq and counts folders are typically large (> 100GB) and will likely take a few hours to

download.

If you are on a PC, you will need to have compression software downloaded such as 7zip to unzip the folders. Macs have built-in zip software.

10X Web Summaries

3 The web_summary typically provides a user-friendly, interactive overview of the key metrics and results obtained during the scRNA-seq analysis. It's designed to help researchers quickly assess the quality of their scRNA-seq data and the success of the analysis. Here are some common components you might find in a 10x web_summary:

1. **Summary Metrics**: This section often includes basic information about the dataset, such as the number of cells, reads per cell, and the percentage of reads that were successfully mapped to the reference genome.
2. **Cell Barcode Filtering**: Information about how many cells passed quality control filters and were retained in the analysis.
3. **Library Complexity Metrics**: Metrics related to library complexity, such as the number of unique molecular identifiers (UMIs) per cell, which can provide insights into the quality of the library preparation.
4. **Gene Expression Metrics**: Summary statistics about gene expression, such as the number of detected genes per cell and the distribution of gene expression levels.
5. **t-SNE or UMAP Plots**: Dimensionality reduction plots, such as t-SNE (t-distributed stochastic neighbor embedding) or UMAP (Uniform Manifold Approximation and Projection), to visualize the clustering of cells based on their gene expression profiles.
6. **Gene Expression Heatmaps**: Heatmaps displaying the expression patterns of specific marker genes or genes of interest across different cell clusters or groups.
7. **Quality Control Metrics**: Information on the quality control measures applied during the analysis, including metrics related to cell doublets, mitochondrial gene content, and more.

Counts Output

4 After you download the deliv_*counts folder, you will find folders for each sample. There are lot of files that are created by the software that can be otherwise ignored. We will outline some important files within the sample "outs" folder:

1. **cloupe.cloupe**: A cloupe file (.cloupe) is a binary file format used to store and visualize single-cell data, typically produced after running cellRanger. It contains information about cell barcodes, feature (gene) counts, and allows for interactive exploration of scRNA-seq data using 10x Genomics' Loupe Cell Browser.
2. **filtered_feature_bc_matrix**: This directory contains the filtered and normalized feature-barcode matrix of your scRNA-seq data. It's the processed expression matrix where rows represent genes (features) and columns represent individual cells. This matrix has low-quality cells and ambient RNA removed.
3. **filtered_feature_bc_matrix.h5**: This is the HDF5 file containing the filtered and normalized feature-barcode matrix. It's a binary file format suitable for efficient storage and retrieval of

large datasets.

4. **metrics_summary.csv**: A CSV file that contains summary metrics and statistics from the cellRanger analysis. It includes information about the number of cells, sequencing metrics, and other quality control metrics.
5. **molecule_info.h5**: This HDF5 file stores detailed information about each sequenced molecule, including cell barcode, UMI (Unique Molecular Identifier), and other metadata. It can be used for downstream analyses and quality control.
6. **possorted_genome_bam.bam**: This is a BAM file containing aligned and sorted sequencing reads mapped to the reference genome. It represents the raw sequencing data for scRNA-seq and can be used for more in-depth analysis or visualization.
7. **possorted_genome_bam.bam.bai**: This is the index file for the BAM file mentioned above. It helps in efficient querying and retrieval of specific regions from the BAM file.
8. **raw_feature_bc_matrix**: This directory contains the raw feature-barcode matrix, which is the initial unprocessed count matrix from scRNA-seq data. It includes all cells, including low-quality ones.
9. **raw_feature_bc_matrix.h5**: Similar to the filtered version, this is the HDF5 file containing the raw feature-barcode matrix.
10. **web_summary.html**: An HTML file that provides a web-based summary report of the cellRanger analysis results. It often includes visualizations, quality control metrics, and an overview of the scRNA-seq data.

Further Analysis with Seurat

5 For users that know or are looking to learn R, Seurat is a great user friendly package to analyze scRNA-seq data. The following seraut tutorials provides information on Quality Control, Normalization, Clustering, and cell typing. https://satijalab.org/seurat/articles/pbmc3k_tutorial

A full list of Seurat tutorials can be found here: https://satijalab.org/seurat/articles/get_started

Here is an example on how to read in your 10X data from a grc_delivery. In this case, we point to the filtered_feature_bc_matrix

```
sample.data <- Read10X(data.dir =
"../deliv_Ashton_SCTesting_030223_10X_counts/Sample_WT_BM_1/outs/fil
tered_feature_bc_matrix")
```

Further Analysis with Loupe Browser

6 For users unfamilar with R, loupe browser is a no-code analysis software provided by 10X. While there may be less custom functionality compared to using R, loupe browser provides a great out of the box solution without the barrier of learning R. Within deliv_*_counts folder the cloupe.cloupe file can be read into loupe browser.

To download loupe browser see the following link:
https://www.10xgenomics.com/products/loupe-browser#single-cell-analysis

A full list of tutorials can be found here:
https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/tutorial

FAQ

7