Quality control analysis for 10X snRNA-seq V.2

Daniel Jacobsen[1], Dinh H Diep[1]

[1]University of California, San Diego

Version 2

Jul 19, 2022

1 Works for me

Share

dx.doi.org/10.17504/protocols.io.261genbqjg47/v2

Dinh Diep
University of California, San Diego

ABSTRACT

Here we describe a computational protocol for performing quality control analysis on shallow sequencing data obtained from 10X snRNA-seq experiments. The workflow starts with raw MiSeq run folders and uses cellranger to generate count matrices. The raw count matrices are analyzed and sequencing saturation plots are generated. The saturation plots are then compared against plots from a reference set of libraries with varying qualities (bad, fair, good, great), thus allowing for the determination of sequencing requirements as well as an assessment of the overall quality of each 10X snRNA experiment.

ATTACHMENTS

10X_snRNA_preseq_analysis_v1.0.tar.gz

DOI

dx.doi.org/10.17504/protocols.io.261genbqjg47/v2

PROTOCOL CITATION

Daniel Jacobsen, Dinh H Diep 2022. Quality control analysis for 10X snRNA-seq . **protocols.io** https://dx.doi.org/10.17504/protocols.io.261genbqjg47/v2 Version created by Dinh Diep

CREATED

Jul 19, 2022

LAST MODIFIED

Jul 19, 2022

PROTOCOL INTEGER ID

67087

MATERIALS TEXT

cellranger software
bcl2fastq software

1 Install cellranger using instructions from https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation

Make sure that cellranger is in the environment's path, otherwise modify commands to include the full path to cellranger.

2 Download the tar.gz file from this protocol.

3 Extract the tar.gz file from this protocol.
<FILENAME> is the name of the downloaded tar.gz file.

Extract file

tar -xzf <FILENAME>

This command extracts a tar.gz file
linux

4 Install anaconda or miniconda Python distributions following given instructions.
Get anaconda from here: https://www.anaconda.com/products/distribution , OR
get miniconda from here: https://docs.conda.io/en/latest/miniconda.html

5 Install preseq using given instructions from http://smithlabresearch.org/software/preseq/.

Citation: Daniel Jacobsen, Dinh H Diep Quality control analysis for 10X snRNA-seqÃÂ
https://dx.doi.org/10.17504/protocols.io.261genbqjg47/v2

preseq must be in the environment's path.

6 Preseq requires the GSL libraries. Install GSL using the instructions from https://www.gnu.org/software/gsl/.

7 Create a symbolic link so that preseq can find the required gsl library.

Create a symbolic link to GSL library

sudo ln -s /usr/local/lib/libgsl.so /usr/lib/libgsl.so.0

This command creates a symbolic link to a gsl library for preseq
linux

8 Install samtools using given instructions from http://www.htslib.org/download/.

9 Use conda to install bcl2fastq with the following terminal command:

install bcl2fastq

conda install -c dranew bcl2fastq

This command installs bcl2fastq for demultiplexing Illumina sequencing runs.
linux

10 Use conda to install required python packages with the following terminal command:

Citation: Daniel Jacobsen, Dinh H Diep Quality control analysis for 10X snRNA-seqÃÂ
https://dx.doi.org/10.17504/protocols.io.261genbqjg47/v2

Install python packages for 10X_snRNA_preseq_analysis package

conda install -c conda-forge numpy seaborn matplotlib pandas

This command installs required python packages for generating preseq plots for 10X_snRNA_preseq_analysis package
linux

11 Run cellranger mkfastq to generate fastq files. Make sure that the following placeholders are set to the correct paths and desired names.
<FASTQ_OUT> is the name of the output folder
<RUN> is the path to the MiSeq run folder
<CSV> is the path to the sample-sheet.csv file

Generate fastq files from raw run folders

cellranger mkfastq --id=<FASTQ_OUT> --run=<RUN> --sample-sheet=<CSV>

This command generates cellranger mkfastq results.
linux

12 Run cellranger count. Make sure that the following placeholders are set to the correct paths and desired names.
<ID> is the name of the output folder for the sample
<SAMPLE> is the sample name used for the sample in the sample-sheet.csv file
<REF> is the path to the cellranger reference data folder
<NUM> is the number of expected cells from the experiment

Citation: Daniel Jacobsen, Dinh H Diep Quality control analysis for 10X snRNA-seqÃÂ
https://dx.doi.org/10.17504/protocols.io.261genbqjg47/v2

Generate the count matrix

cellranger count --id <ID> --fastqs <FASTQ_OUT> --sample <SAMPLE> --transcriptome <REF> --include-introns --expect-cells <NUM>

This command generates cellranger count results.
linux

13 Run the preseq script in the folder downloaded from this protocol. Make sure that the following placeholders are sset to the correct paths and names.

<PATH_TO_FOLDER> is the path to the folder that was extracted from the tar.gz file.
<ID> is the name of the output folder for the sample generated with cellranger.

Generate preseq plots from 10X snRNA output folder

<PATH_TO_FOLDER>/scripts/loop.preseq.r.sh <ID>

This command generates the preseq results for 10X snRNA experiments.
linux

14 View outputs.

<ID>.lc_extrap_log.txt contains preseq statistics
<ID>.lc_extrap_output.png to view the sequencing saturation plots
<ID>/outs/web_summary.html to view the cellranger analyses
<ID>/outs/summary.csv to view quality statistics generated by cellranger

Citation: Daniel Jacobsen, Dinh H Diep Quality control analysis for 10X snRNA-seqÃÂ
https://dx.doi.org/10.17504/protocols.io.261genbqjg47/v2