Feb 19, 2025

Single cell/nuclei RNAseq analysis

DOI
dx.doi.org/10.17504/protocols.io.n2bvj8bzbgk5/v1

Raquel Garza[1]

[1]Laboratory of Molecular Neurogenetics, Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, BMC A11, Lund University, 221 84 Lund, Sweden.

ASAP Collaborative Rese… Jakobsson

Raquel Garza
Lund University

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.n2bvj8bzbgk5/v1

Protocol Citation: Raquel Garza 2025. Single cell/nuclei RNAseq analysis. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.n2bvj8bzbgk5/v1**

Protocol status: Working
We use this protocol and it's working

Created: March 03, 2023

Last Modified: February 19, 2025

Protocol Integer ID: 78063

Keywords: ASAPCRN

Funders Acknowledgements:
Aligning Science Across Parkinson's through the Michael J. Fox Foundation for Parkinson's Research
Grant ID: ASAP-000520
Swedish Research Council
Grant ID: 2018-02694
Swedish Brain Foundation
Grant ID: FO2019-0098
Cancerfonden
Grant ID: 190326
Barncancerfonden
Grant ID: PR2017-0053
NIHR Cambridge Biomedical Research Centre
Grant ID: NIHR203312
Swedish Society for Medical Research
Grant ID: S19-0100
National Institutes of Health
Grant ID: HG002385
Swedish Research Council
Grant ID: 2021-03494
Swedish Research Council
Grant ID: 2020-01660

Abstract

This protocol describes the process for the single cell/nuclei RNA sequencing data of the manuscript "L1retrotransposons drive human neuronal transcriptome complexity and functional diversification " from fetal forebrain and adult prefrontal cortex tissue.

Preprocessing

1 The raw base calls were demultiplexed and converted to sample-specific fastq files using 10x Genomics Cell Ranger mkfastq (version 3.1.0; RRID:SCR_017344).

Gene expression

2 Cell Ranger count was run with default settings, using an mRNA reference for single-cell samples and a pre-mRNA reference (generated using 10x Genomics Cell Ranger 3.1.0 guidelines) for single nuclei samples.

3 To produce velocity plots, loom files were generated using velocyto (version 0.17.17; RRID:SCR_018167) run10x in default parameters, masking for TEs (same GTF file as input for TEtranscripts; see protocol Bulk RNA sequencing analysis (dx.doi.org/10.17504/protocols.io.yxmvm2m55g3p/v1) section TE subfamily quantification) and gencode annotation as guide for features.

4 Samples were analysed using Seurat (version 3.1.5; RRID:SCR_007322). Counts were normalized using the Centered Log Ratio (CLR) transformation (Seurat::NormalizeData) and clusters were found with a resolution 0.5 (Seurat::FindClusters).

Quality control

5 For each sample, cells were filtered out if the percentage of mitochondrial content was over 10% (perc_mitochondrial).

6 For adult samples, cells were discarded if the number of detected features (nFeature_RNA) was higher than 2 standard deviations over the mean in the sample (to avoid keeping doublets), or lower than a standard deviation below the mean in the sample (to avoid low quality cells). For fetal samples, cells were discarded if the number of detected features was higher than 2 standard deviations over the mean in the sample, or lower than 2,000 features detected.

TE quantification

7 Run trusTEr (version 0.1.1; doi:10.5281/zenodo.7589548).

7.1 All clustering, normalization and merging of samples were performed using the contained scripts of get_clusters.R (get_custers() from the Sample class) and merge_samples.R (merge_samples() from the Experiment class) of trusTEr (version 0.1.1; doi:10.5281/zenodo.7589548).

7.2 The function tsv_to_bam() backtraces cells barcodes to Cell Ranger's output BAM file. tsv_to_bam() runs using subset-bam from 10x Genomics version 1.0 (RRID:SCR_023216).

7.3 filter_UMIs() filters potential PCR duplicates in the BAM files; this step uses Pysam version 0.15.1 (RRID:SCR_021017).

7.4 Convert BAM to FastQ files using bamtofastq from 10x Genomics (version 1.2.0; RRID: SCR_023215)

7.5 Remapping for each cluster was performed using STAR aligner (version 2.7.8a; RRID:SCR_004463)

7.6 Quantification of TE subfamilies was done using TEcount (version 2.0.3; RRID:SCR_023208) and individual elements were quantified using featureCounts (Subread version 1.6.3; RRID:SCR_012919).

7.7 The normalization step of trusTEr (divide counts by number of cells in cluster) and the integration with Seurat and normalize TE subfamilies' expression, was performed using Seurat version 3.1.5 (RRID:SCR_007322).