Tutorial on how to use the single cell RNA seq Snap pipeline
2025-02-18

Contents

1 Introduction

1.1 Overview

This notebook offers guidelines and best practices for effectively using the single cell RNA seq Snap pipeline[1]. Please note that we will continue to maintain and update the pipeline with new analysis modules.

1.2 Prerequisites

We assume users have basic experience using Shell, git, and GitHub.

2 How to access and run the snap pipeline

We recommend that users fork the `sc-rna-seq-snap` repository. Team members should use the stjude-dnb-binfcore[2] account, while others can use their preferred GitHub account. We welcome collaborations, so please feel free to reach out if you're interested in being added to the `stjude-dnb-binfcore` account.

2.1 Naming your fork

- Team members: Retain the pipeline name and append the project name at the end, e.g., `sc-rna-seq-snap-Vsx2_SE`.
- External collaborators: Keep the pipeline name as-is, unless you plan to use the repository for multiple projects. In that case, follow the same naming convention as team members.

2.2 Managing the privacy of your fork

Please be aware that forked repositories are public by default and will contain the code from the main branch of the `sc-rna-seq-snap` repository, which has been reviewed and tested.

No results will be published unless the project is under review or has been officially published.

Furthermore, we adhere to strict guidelines to ensure privacy and protect sensitive data. Sensitive file paths and patient-related data must never be published on GitHub.

2.3 Synchronizing your fork

User needs to ensure that the main branch of the forked repository is always up to date with `stjude-dnb-binfcore/sc-rna-seq-snap:main`.

When syncing your forked repository with the main repository, please be cautious of any changes made to the following files, as they are typically modified and specified for project data analysis:

- `project_parameters.Config.yaml`
- `./analyses/cellranger-analysis/waiter.sh`

Before pulling the latest changes, stash any modifications you have made to these files. This ensures that you won't accidentally overwrite your changes when syncing with the main repository. For more details, see the README.md[3].

---

[1] https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap

[2] https://github.com/stjude-dnb-binfcore

[3] https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/blob/main/README.md#to-run-the-code-in-this-repository

2.4 Accessing the code

Next, open the terminal, navigate to your desired directory, and use `git clone` to clone the repository. For more details, see the README.md[4].

3 Data

3.1 Set up parameters for the project

Replace parameters in the `project_parameters.Config.yaml` with the correct file paths and parameters specific to your project. Ensure that you use the appropriate file paths and parameters for your data, such as assay type and condition. The default values are set for scRNA-seq data analysis without ambient RNA removal and doublet detection.

3.2 Genome references

Our team at the Bioinformatics core at DNB maintains the following genome references: `GRCh38`, `GRCm39`, `GRCh38ANDGRCm39`, and `GRCh38_GFP_tdTomato`, for human, mouse, and dual index genomes, respectively. Please submit an issue[5] to request the path to the reference genome of preference.

3.3 Cell type gene marker lists

Our team at the Bioinformatics core at DNB maintains the following cell type gene marker lists for cell type annotation: `mouse brain tissue`, `human adult retina tissue`, `human fetal retina tissue` and `mouse retina tissue`. Please submit an issue[6] to request the list of gene markers of preference.

4 Analysis

4.1 Analysis module

Please refer to the `<analysis_module>/README.md` files for instructions on how to run the specific analysis module. These files contain the required parameters and necessary files to successfully execute the module's pipeline.

4.2 Container Overview

We have generated a Docker image that contains all tools, packages, and dependencies necessary to run the code and analyses modules. The environment is specifically configured for `Rstudio/R v4.4.0` and `Seurat v4.4.0`. For more details, please refer to the README.md[7].

4.3 CPU and Memory Resources

While we have provided estimates for the computational resources required, users may need to adjust memory settings based on the size of the cohort and the specific demands of the analysis.

---

[4]https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/blob/main/README.md#to-access-the-code-in-this-repository

[5]https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/issues

[6]https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/issues

[7]https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/blob/main/run-container/README.md#load-specific-version-of-singularity

For reference, the estimated resources are for 8 samples with approximately 50,000 cells in total. For further details, please consult the README.md[8].

5 Contact

Contributions, issues, and feature requests are welcome! Please feel free to check issues[9].

5.1 Authors

Antonia Chroni, PhD (@AntoniaChroni[10])

These materials have been developed by the Bioinformatic core team at the St. Jude Children's Research Hospital[11]. These are open access materials distributed under the terms of the BSD 2-Clause License[12], which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

---

[8] https://github.com/stjude-dnb-binfcore/sc-rna-seq-snap/blob/main/README.md#to-request-memory-from-the-hpcf-cluster

[9] https://github.com/stjude-dnb-binfcore/trainings/issues

[10] https://github.com/AntoniaChroni

[11] https://www.stjude.org/

[12] https://opensource.org/license/bsd-2-clause