

Data exploratory analysis

Antonia Chroni for SJCRH DNB_BINF_Core

Contents

1	Information about this notebook	3
2	Set up	3
3	Directories and paths to file Inputs/Outputs	3
4	Read raw data file	3
5	Read processed data file	3
6	Update processed data file by Jackie Norrie on 2025-01-03	3
7	10x matched scRNA-seq and scATAC-seq cohort	3
7.1	Type of sequencing assay and unit	3
7.2	Number of samples per experiment	3
7.3	Number of samples per experiment, seq_unit, assay, and condition	4
7.4	Number of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and sample	5
8	10x Genomics Multiome cohort	6
8.1	Type of sequencing assay and seq_unit	6
8.2	Number of samples per experiment	6
8.3	Number of samples per experiment, seq_unit, assay, and condition	7
8.4	Number of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and sample	8
9	Notes	9
10	References	9
11	Session Info	10
## The following object is masked _by_ .GlobalEnv:		
##		
##	root_dir	

PI: NA

Project: test-dataset

Task: NA

Project Lead(s): NA

Department: Developmental Neurobiology

DNB Bioinformatics Core Analysis Team:

Lead Analyst(s): Antonia Chroni, PhD

Group Lead: Cody A. Ramirez, PhD

Contact E-mail: antonia.chroni@stjude.org¹

DNB Bioinformatics Core Pipeline: Standard sc-/sn-ATAC-Seq Analysis in 10X Genomics data

Date started: Dec-16-2024

Date completed: ONGOING

Report generated: 11:49:20 CST 01/07/2025

Reviewed by: _____ Date: _____

¹<mailto:antonia.chroni@stjude.org>

1 Information about this notebook

This is an exploratory analysis of the data available for the testing phase for the **sc-atac-seq** pipeline(s). We are investigating number of samples overall per condition and variables as defined in the **params**.

We are looking for cohorts that fit the following criteria as described in the `./analyses/README.md`².

- Control vs condition (min. 3+3 samples)
- Number of cells/sample - size of datasets might determined packages/pipelines to be used
- Single cell/Single-nucleus ATAC
- Integration of scRNA-seq and scATAC-seq data from the same biological system (multiple modalities)
- Pipeline for same samples but not same cells with scRNA-seq and scATAC-seq
- Annotate scATAC-seq cells via label transfer by using scRNA data: cell type annotation
- Available bulk ATAC-seq data for the same samples (matched) - this could be used for cell type annotation

2 Set up

3 Directories and paths to file Inputs/Outputs

4 Read raw data file

5 Read processed data file

6 Update processed data file by Jackie Norrie on 2025-01-03

This is the `cohorts_10x_rna_atac_testing_phase_2024-12-17.tsv` file updated by Jackie Norrie on 2025-01-03 to fix empty cells and/or inconsistencies.

7 10x matched scRNA-seq and scATAC-seq cohort

7.1 Type of sequencing assay and unit

We should investigate if there are samples from different sequencing technologies and unit.

Single cell sequencing was done by using cell, nucleus and there are 10Xv3_RNA, 10Xv2_ATAC sequencing technologies and assays in the database. Cohort is formed and processed accordingly.

7.2 Number of samples per experiment

Table 1: Summary of samples per experiment

experiment	n
Cerebellum	9
More Retina	7
Stressed Retina	3
Victoria Knockout	8

²<https://github.com/stjude-dnb-binfcore/sc-atac-seq/tree/main/analyses>

7.3 Number of samples per experiment, seq_unit, assay, and condition

Table 2: Summary of samples per experiment, seq_unit, assay, and condition

experiment	seq_unit	assay	condition	n
Cerebellum	cell	RNA	age	2
Cerebellum	cell	RNA	sorted	2
Cerebellum	nucleus	ATAC	age	3
Cerebellum	nucleus	ATAC	sorted	2
More Retina	cell	ATAC	age	2
More Retina	cell	ATAC	none	2
More Retina	cell	ATAC	sorted	1
More Retina	cell	RNA	age	1
More Retina	cell	RNA	sorted	1
Stressed Retina	cell	RNA	LPS Injection	1
Stressed Retina	cell	RNA	PBS Injection	1
Stressed Retina	nucleus	ATAC	LPS Injection	1
Victoria Knockout	cell	RNA	knock-out	2
Victoria Knockout	cell	RNA	wt	2
Victoria Knockout	nucleus	ATAC	knock-out	2
Victoria Knockout	nucleus	ATAC	wt	2

7.4 Number of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and sample

Table 3: Summary of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and Sample

experiment	tissue	seq_unit	seq_technology_assay	condition	Sample	n
Cerebellum	Cerebellum	cell	10Xv3_RNA	age	6 week cerebellum	1
Cerebellum	Cerebellum	cell	10Xv3_RNA	age	P0 cerebellum	1
Cerebellum	Cerebellum	cell	10Xv3_RNA	sorted	ATOH HI	1
Cerebellum	Cerebellum	cell	10Xv3_RNA	sorted	ATOH Low	1
Cerebellum	Cerebellum	nucleus	10Xv2_ATAC	age	6 week cerebellum	2
Cerebellum	Cerebellum	nucleus	10Xv2_ATAC	age	P0 cerebellum	1
Cerebellum	Cerebellum	nucleus	10Xv2_ATAC	sorted	ATOH HI	1
Cerebellum	Cerebellum	nucleus	10Xv2_ATAC	sorted	ATOH Low	1
More Retina	Retina	cell	10Xv2_ATAC	age	E14.5 Retina	2
More Retina	Retina	cell	10Xv2_ATAC	none	Retina	2
More Retina	Retina	cell	10Xv2_ATAC	sorted	NRL Dep	1
More Retina	Retina	cell	10Xv3_RNA	age	E14.5 Retina	1
More Retina	Retina	cell	10Xv3_RNA	sorted	NRL Dep	1
Stressed Retina	Retina	cell	10Xv3_RNA	LPS	LPS	1
Stressed Retina	Retina	cell	10Xv3_RNA	Injection PBS	PBS	1
Stressed Retina	Retina	nucleus	10Xv2_ATAC	Injection LPS	LPS_ATAC	1
Victoria Knockout	Retina	cell	10Xv3_RNA	Injection knock-out	SEKO_21	1
Victoria Knockout	Retina	cell	10Xv3_RNA	knock-out	SEKO_25	1
Victoria Knockout	Retina	cell	10Xv3_RNA	wt	Wt1	1
Victoria Knockout	Retina	cell	10Xv3_RNA	wt	Wt2	1
Victoria Knockout	Retina	nucleus	10Xv2_ATAC	knock-out	SEKO_21	1
Victoria Knockout	Retina	nucleus	10Xv2_ATAC	knock-out	SEKO_25	1
Victoria Knockout	Retina	nucleus	10Xv2_ATAC	wt	Wt1	1
Victoria Knockout	Retina	nucleus	10Xv2_ATAC	wt	Wt2	1

8 10x Genomics Multiome cohort

8.1 Type of sequencing assay and seq_unit

We should investigate if there are samples from two different sequencing technologies and unit.

Single cell sequencing was done by using nucleus and there are 10X_RNA, 10X_ATAC sequencing technologies and assays in the database. Cohort is formed and processed accordingly.

8.2 Number of samples per experiment

Table 4: Summary of samples per experiment

experiment	n
Multiome Retina	6

8.3 Number of samples per experiment, seq_unit, assay, and condition

Table 5: Summary of samples per experiment, seq_unit, assay, and condition

experiment	seq_unit	assay	condition	n
Multiome Retina	nucleus	ATAC	age	3
Multiome Retina	nucleus	RNA	age	3

8.4 Number of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and sample

Table 6: Summary of samples per experiment, tissue, seq_unit, seq_technology_assay, condition, and Sample

experiment	tissue	seq_unit	seq_technology_assay	condition	Sample	n
Multiome Retina	Retina	nucleus	10X_ATAC	age	Adult Retina	1
Multiome Retina	Retina	nucleus	10X_ATAC	age	E14.5 Retina	1
Multiome Retina	Retina	nucleus	10X_ATAC	age	P0 Retina	1
Multiome Retina	Retina	nucleus	10X_RNA	age	Adult Retina	1
Multiome Retina	Retina	nucleus	10X_RNA	age	E14.5 Retina	1
Multiome Retina	Retina	nucleus	10X_RNA	age	P0 Retina	1

9 Notes

I have identified the following datasets that almost fit the criteria for the testing phase:

- **10x matched scRNA-seq and scATAC-seq:** Victoria Knockout experiment with 1 replicate/knock-out group (4 samples for 10x RNA + 4 samples 10x ATAC). Replicates could be potentially grouped together and have 2 replicates for WT and 2 replicates for knock-out.
- **10x Genomics Multiome:** Multiome Retina experiment with 1 replicate/age group (3 samples for 10x RNA (Multiome) + 3 samples for 10x ATAC (multiome)).

10 References

- More Retina experiment published by Norrie et al., 2025³.
- Victoria Knockout experiment published by Honnell et al., 2022⁴.

³<https://doi.org/10.1016/j.devcel.2024.12.014>

⁴<https://www.nature.com/articles/s41467-021-27924-y#Sec13>

11 Session Info

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Red Hat Enterprise Linux 8.8 (Ootpa)
##
## Matrix products: default
## BLAS: /usr/lib64/libblas.so.3.8.0
## LAPACK: /usr/lib64/liblapack.so.3.8.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/Chicago
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] readxl_1.4.3   knitr_1.48     lubridate_1.9.3 forcats_1.0.0
##  [5] stringr_1.5.1  dplyr_1.1.4    purrr_1.0.2     readr_2.1.5
##  [9] tidyr_1.3.1    tibble_3.2.1   ggplot2_3.5.1   tidyverse_2.0.0
## [13] yaml_2.3.10
##
## loaded via a namespace (and not attached):
##  [1] bit_4.0.5      gtable_0.3.5   jsonlite_1.8.8  crayon_1.5.3
##  [5] compiler_4.4.0 tidyselect_1.2.1 parallel_4.4.0  jquerylib_0.1.4
##  [9] scales_1.3.0   fastmap_1.2.0  mime_0.12       R6_2.5.1
## [13] generics_0.1.3 munsell_0.5.1  bslib_0.8.0     pillar_1.9.0
## [17] tzdb_0.4.0     rlang_1.1.4    utf8_1.2.4      stringi_1.8.4
## [21] cachem_1.1.0   xfun_0.47      sass_0.4.9      bit64_4.0.5
## [25] timechange_0.3.0 cli_3.6.3      withr_3.0.1     magrittr_2.0.3
## [29] digest_0.6.37  grid_4.4.0     vroom_1.6.5     hms_1.1.3
## [33] lifecycle_1.0.4 vctrs_0.6.5    evaluate_0.24.0 glue_1.7.0
## [37] cellranger_1.1.0 fansi_1.0.6     colorspace_2.1-1 rmarkdown_2.28
## [41] tools_4.4.0    pkgconfig_2.0.3 htmltools_0.5.8.1
```