

# Quality Control for scRNA seq data

Load and Merge the data

Ximing Ran

2025-05-15

## Contents

<b>1. Read the Raw Data</b>	<b>2</b>
<b>2. QC for data</b>	<b>3</b>

## 1. Read the Raw Data

In this section, we will read the raw data from the cellranger-arc output files. We have 4 samples: PSZ-6, TSC-tube, TSC-edge, and CTRL. We will rename the samples as CTRL, TSC-tuber, TSC-edge, and TSC-outside, respectively.

```
## [1] 1
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s001_C1/outs/filtered_feature_bc_mat.
## [1] 2
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s002_C2/outs/filtered_feature_bc_mat.
## [1] 3
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s003_C3/outs/filtered_feature_bc_mat.
## [1] 4
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s004_Q1/outs/filtered_feature_bc_mat.
## [1] 5
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s005_Q2/outs/filtered_feature_bc_mat.
## [1] 6
## [1] "../../../data/synaptosomes_scRNA/raw/p24020_Lyra_NovaSeqS4-s006_Q3/outs/filtered_feature_bc_mat.
```

## 2. QC for data

```
## [1] 1
## [1] "C1"
## [1] "Number of cells before QC: 9979"
## [1] "Number of cells after QC: 8642"

## [1] 2
## [1] "C2"
## [1] "Number of cells before QC: 9557"
## [1] "Number of cells after QC: 8524"

## [1] 3
## [1] "C3"
## [1] "Number of cells before QC: 10223"
## [1] "Number of cells after QC: 8929"

## [1] 4
## [1] "Q1"
## [1] "Number of cells before QC: 11206"
## [1] "Number of cells after QC: 9863"

## [1] 5
## [1] "Q2"
## [1] "Number of cells before QC: 8626"
## [1] "Number of cells after QC: 7693"

## [1] 6
## [1] "Q3"
## [1] "Number of cells before QC: 10951"
## [1] "Number of cells after QC: 9560"
```

```

## R version 4.4.0 (2024-04-24)
## Platform: aarch64-apple-darwin20
## Running under: macOS 15.4
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib; LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats4      stats      graphics  grDevices  utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] scDbfFinder_1.18.0      SingleCellExperiment_1.26.0
## [3] SummarizedExperiment_1.34.0 MatrixGenerics_1.16.0
## [5] matrixStats_1.5.0      data.table_1.17.2
## [7] AnnotationHub_3.12.0    BiocFileCache_2.12.0
## [9] dbplyr_2.5.0            simspec_0.0.0.9000
## [11] cowplot_1.1.3           EnsDb.Hsapiens.v86_2.99.0
## [13] ensemblDb_2.28.1        AnnotationFilter_1.28.0
## [15] GenomicFeatures_1.56.0  AnnotationDbi_1.66.0
## [17] Biobase_2.64.0          Signac_1.14.0
## [19] Seurat_5.3.0            SeuratObject_5.1.0
## [21] sp_2.2-0                rtracklayer_1.64.0
## [23] GenomicRanges_1.56.2    GenomeInfoDb_1.40.1
## [25] IRanges_2.38.1          S4Vectors_0.42.1
## [27] BiocGenerics_0.50.0     knitr_1.50
## [29] lubridate_1.9.4         forcats_1.0.0
## [31] stringr_1.5.1           dplyr_1.1.4
## [33] purrr_1.0.4             readr_2.1.5
## [35] tidyr_1.3.1             tibble_3.2.1
## [37] ggplot2_3.5.2           tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] ProtGenerics_1.36.0      spatstat.sparse_3.1-0
## [3] bitops_1.0-9            httr_1.4.7
## [5] RColorBrewer_1.1-3       tools_4.4.0
## [7] sctransform_0.4.2        R6_2.6.1
## [9] lazyeval_0.2.2          uwot_0.2.3
## [11] withr_3.0.2             gridExtra_2.3
## [13] progressr_0.15.1        textshaping_1.0.1
## [15] cli_3.6.5               spatstat.explore_3.4-2
## [17] fastDummies_1.7.5       labeling_0.4.3
## [19] spatstat.data_3.1-6     ggribes_0.5.6
## [21] pbapply_1.7-2           systemfonts_1.2.3
## [23] Rsamtools_2.20.0        dichromat_2.0-0.1
## [25] scater_1.32.1           parallelly_1.44.0
## [27] limma_3.60.6            rstudioapi_0.17.1

```

## [29]	RSQLite_2.3.11	generics_0.1.4
## [31]	BiocIO_1.14.0	ica_1.0-3
## [33]	spatstat.random_3.3-3	Matrix_1.7-3
## [35]	ggbeeswarm_0.7.2	abind_1.4-8
## [37]	lifecycle_1.0.4	edgeR_4.2.2
## [39]	yaml_2.3.10	SparseArray_1.4.8
## [41]	Rtsne_0.17	grid_4.4.0
## [43]	blob_1.2.4	promises_1.3.2
## [45]	dqrng_0.4.1	crayon_1.5.3
## [47]	miniUI_0.1.2	lattice_0.22-7
## [49]	beachmat_2.20.0	KEGGREST_1.44.1
## [51]	metapod_1.12.0	pillar_1.10.2
## [53]	rjson_0.2.23	xgboost_1.7.10.1
## [55]	future.apply_1.11.3	codetools_0.2-20
## [57]	fastmatch_1.1-6	glue_1.8.0
## [59]	spatstat.univar_3.1-3	vctrs_0.6.5
## [61]	png_0.1-8	spam_2.11-1
## [63]	gtable_0.3.6	assertthat_0.2.1
## [65]	cachem_1.1.0	xfun_0.52
## [67]	S4Arrays_1.4.1	mime_0.13
## [69]	survival_3.8-3	RcppRoll_0.3.1
## [71]	statmod_1.5.0	bluster_1.14.0
## [73]	fitdistrplus_1.2-2	ROCR_1.0-11
## [75]	nlme_3.1-168	bit64_4.6.0-1
## [77]	filelock_1.0.3	RcppAnnoy_0.0.22
## [79]	rprojroot_2.0.4	irlba_2.3.5.1
## [81]	vipor_0.4.7	KernSmooth_2.23-26
## [83]	colorspace_2.1-1	DBI_1.2.3
## [85]	ggrastr_1.0.2	tidyselect_1.2.1
## [87]	bit_4.6.0	compiler_4.4.0
## [89]	curl_6.2.2	BiocNeighbors_1.22.0
## [91]	hdf5r_1.3.12	DelayedArray_0.30.1
## [93]	plotly_4.10.4	scales_1.4.0
## [95]	lmtest_0.9-40	rappdirs_0.3.3
## [97]	digest_0.6.37	goftest_1.2-3
## [99]	spatstat.utils_3.1-3	rmarkdown_2.29
## [101]	XVector_0.44.0	htmltools_0.5.8.1
## [103]	pkgconfig_2.0.3	sparseMatrixStats_1.16.0
## [105]	fastmap_1.2.0	rlang_1.1.6
## [107]	htmlwidgets_1.6.4	UCSC.utils_1.0.0
## [109]	shiny_1.10.0	DelayedMatrixStats_1.26.0
## [111]	farver_2.1.2	zoo_1.8-14
## [113]	jsonlite_2.0.0	BiocParallel_1.38.0
## [115]	BiocSingular_1.20.0	RCurl_1.98-1.17
## [117]	magrittr_2.0.3	scuttle_1.14.0
## [119]	GenomeInfoDbData_1.2.12	dotCall64_1.2
## [121]	patchwork_1.3.0	Rcpp_1.0.14
## [123]	viridis_0.6.5	reticulate_1.42.0
## [125]	stringi_1.8.7	zlibbioc_1.50.0
## [127]	MASS_7.3-65	plyr_1.8.9
## [129]	parallel_4.4.0	listenv_0.9.1
## [131]	ggrepel_0.9.6	deldir_2.0-4
## [133]	Biostings_2.72.1	splines_4.4.0
## [135]	tensor_1.5	hms_1.1.3

## [137] locfit_1.5-9.12	igraph_2.1.4
## [139] spatstat.geom_3.3-6	RcppHNSW_0.6.0
## [141] reshape2_1.4.4	ScaledMatrix_1.12.0
## [143] BiocVersion_3.19.1	XML_3.99-0.18
## [145] evaluate_1.0.3	scran_1.32.0
## [147] BiocManager_1.30.25	tzdb_0.5.0
## [149] httpuv_1.6.16	RANN_2.6.2
## [151] polyclip_1.10-7	future_1.49.0
## [153] scattermore_1.2	rsvd_1.0.5
## [155] xtable_1.8-4	restfulr_0.0.15
## [157] RSpectra_0.16-2	later_1.4.2
## [159] ragg_1.4.0	viridisLite_0.4.2
## [161] memoise_2.0.1	beeswarm_0.4.0
## [163] GenomicAlignments_1.40.0	cluster_2.1.8.1
## [165] timechange_0.3.0	globals_0.18.0
## [167] here_1.0.1	