

# Quality Control for scRNA and scATAC seq data

Load and Merge the data

Ximing Ran

2025-05-06

## Contents

1. Read the Raw Data	2
2. QC for data	3

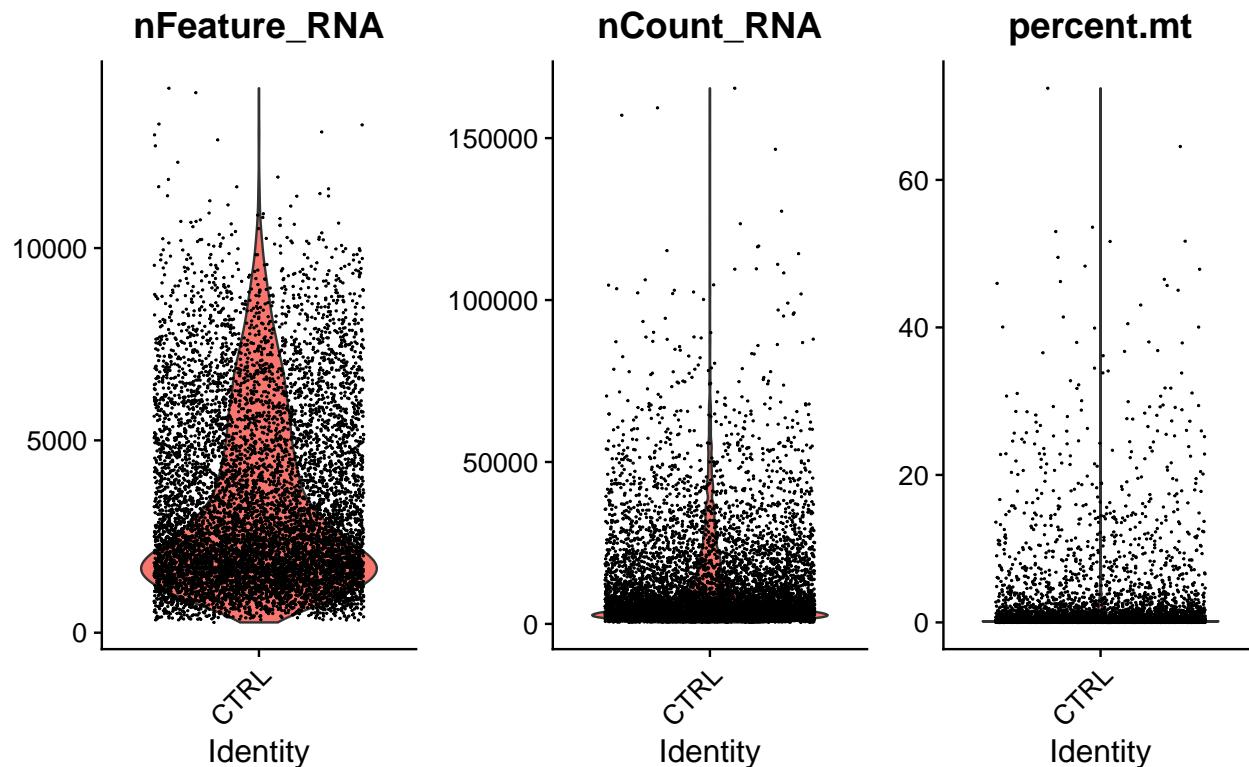
## 1. Read the Raw Data

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.

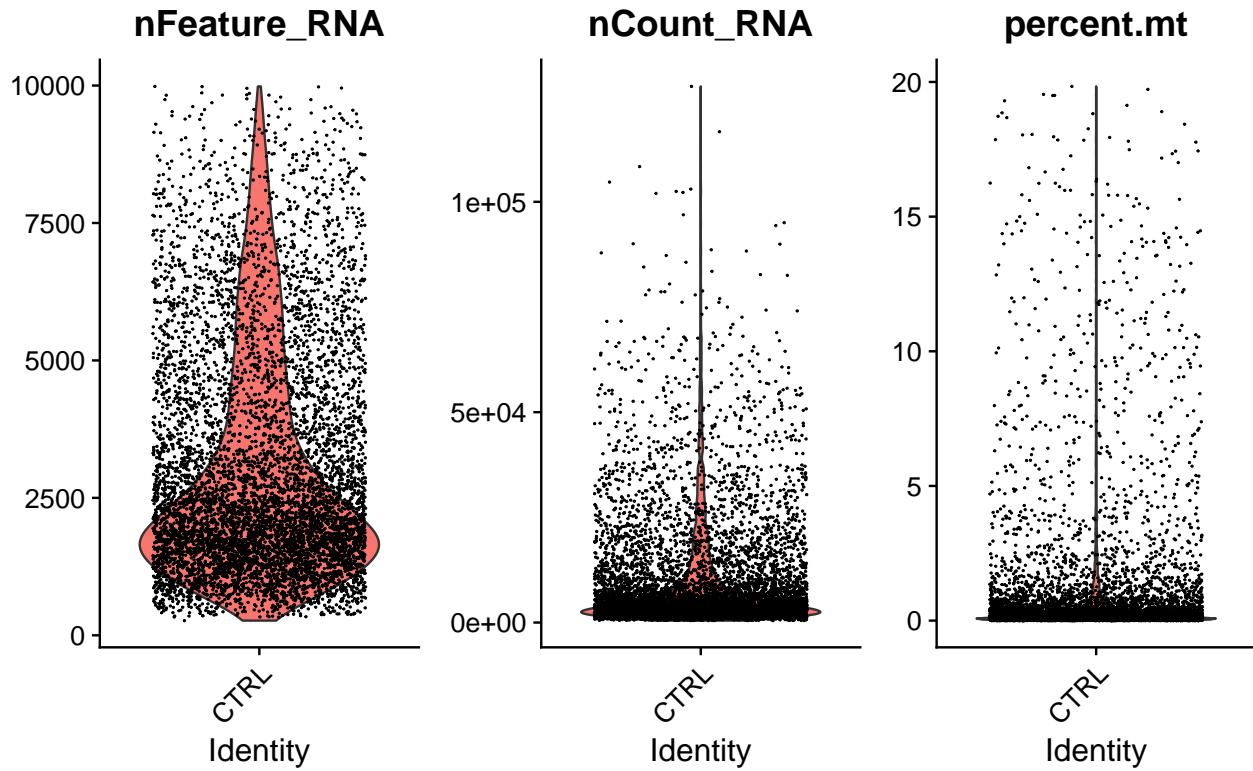
```
## [1] 1
## [1] ".../data/raw-data/cellrange-arc-out/CTRL/filtered_feature_bc_matrix.h5"
## [1] 2
## [1] ".../data/raw-data/cellrange-arc-out/TSC_edge/filtered_feature_bc_matrix.h5"
## [1] 3
## [1] ".../data/raw-data/cellrange-arc-out/TSC_tuber/filtered_feature_bc_matrix.h5"
## [1] 4
## [1] ".../data/raw-data/cellrange-arc-out/TSC_outside/filtered_feature_bc_matrix.h5"
```

## 2. QC for data

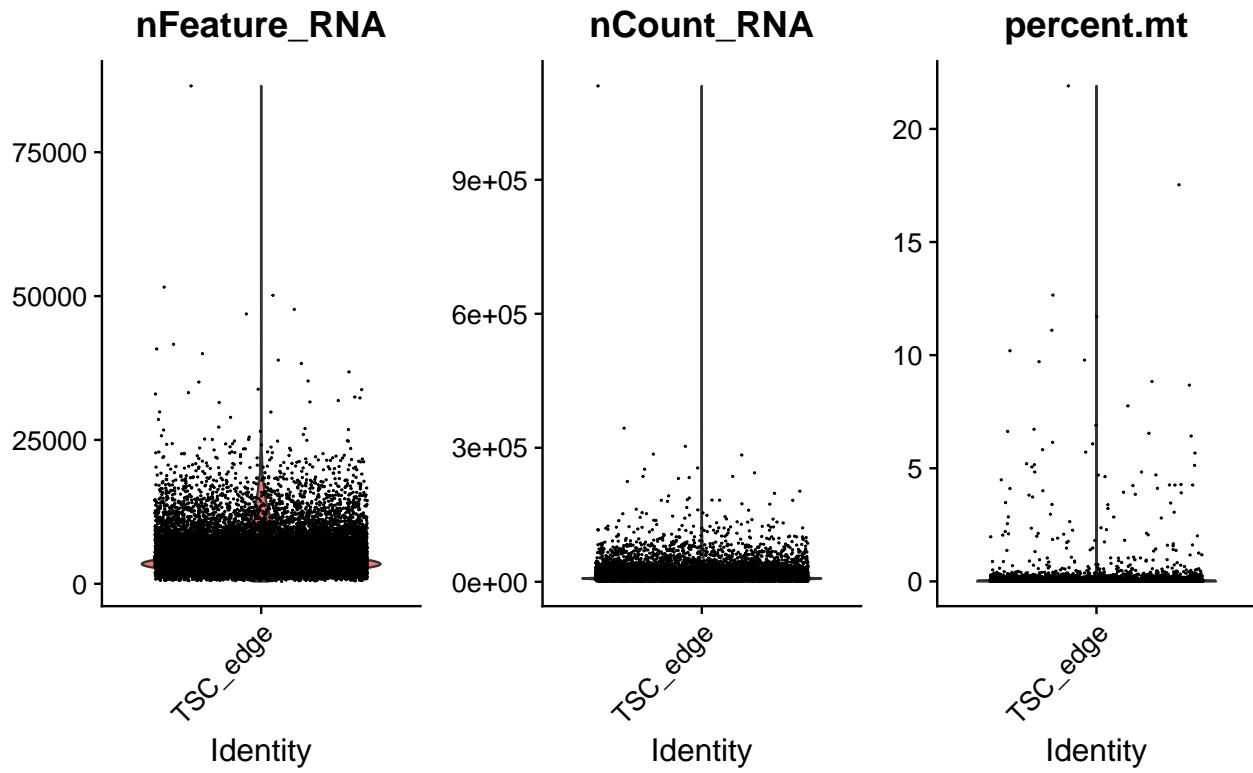
```
## [1] 1  
## [1] "CTRL"  
## [1] "Number of cells before QC: 8654"
```



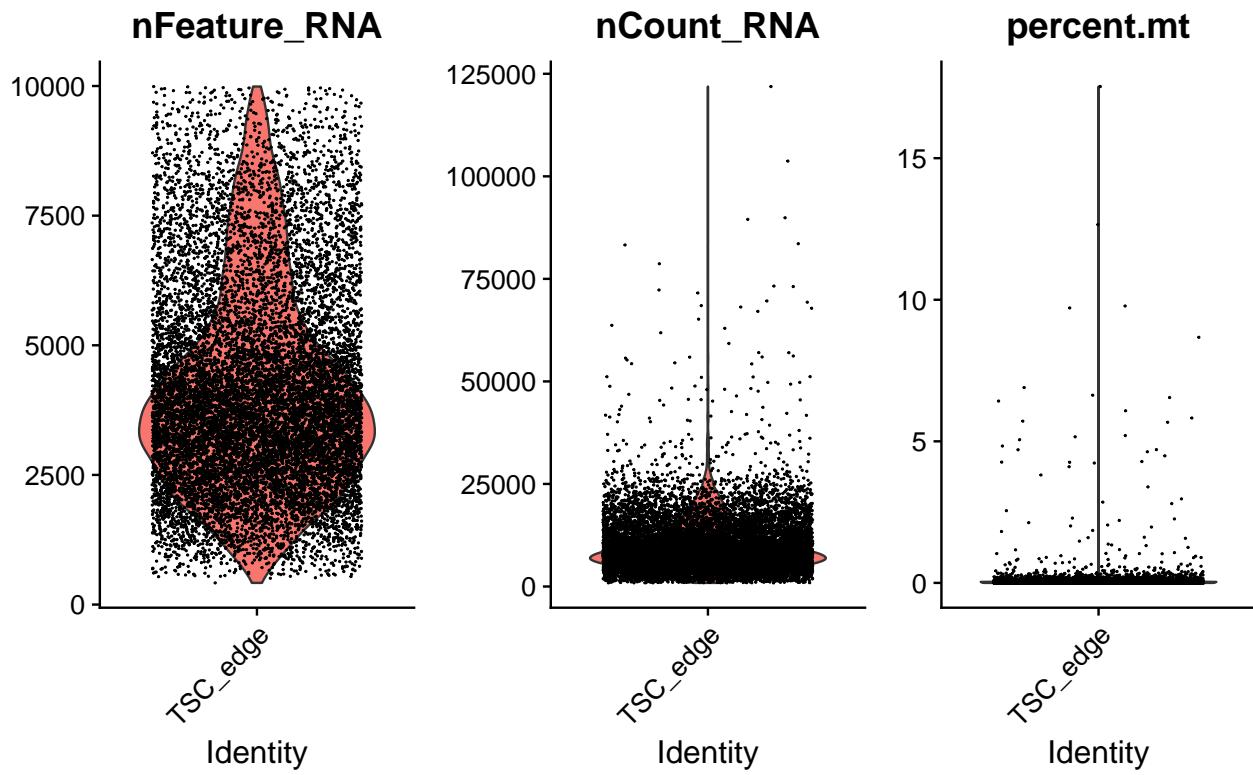
```
## [1] "Number of cells after QC: 7329"
```



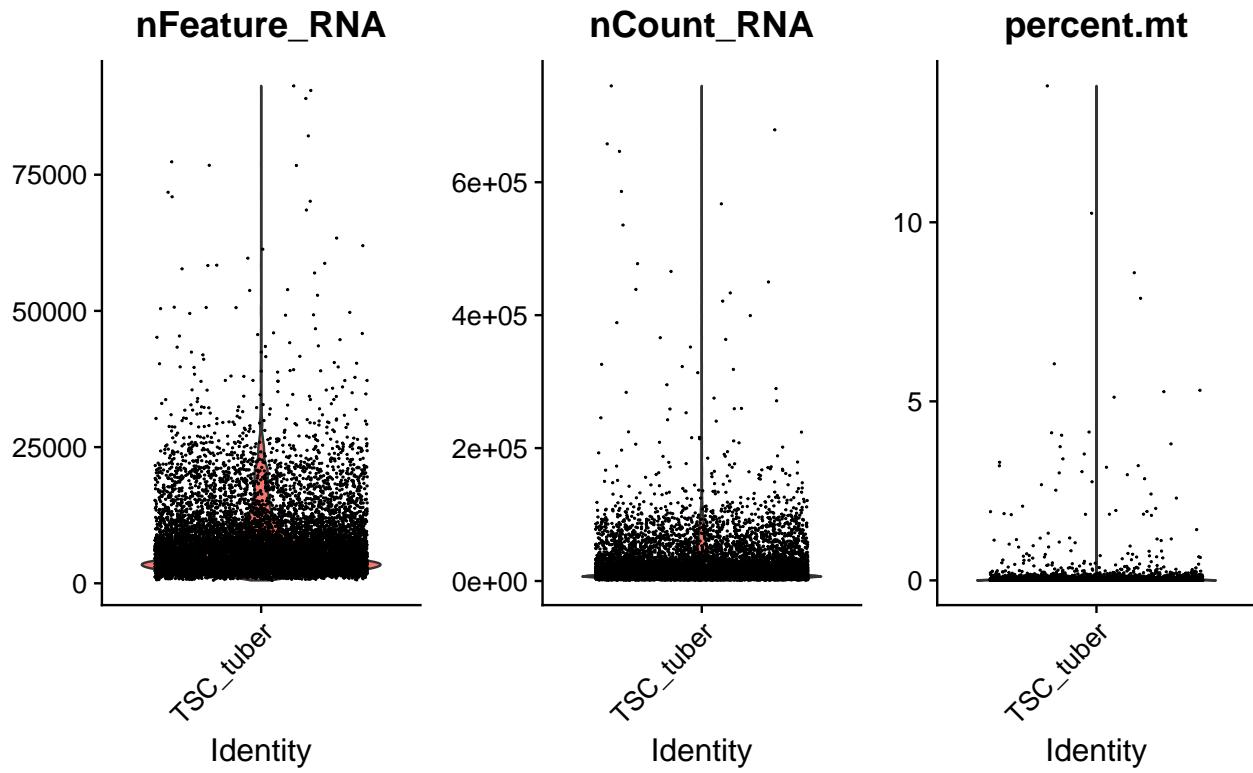
```
## [1] 2
## [1] "TSC_edge"
## [1] "Number of cells before QC: 16701"
```



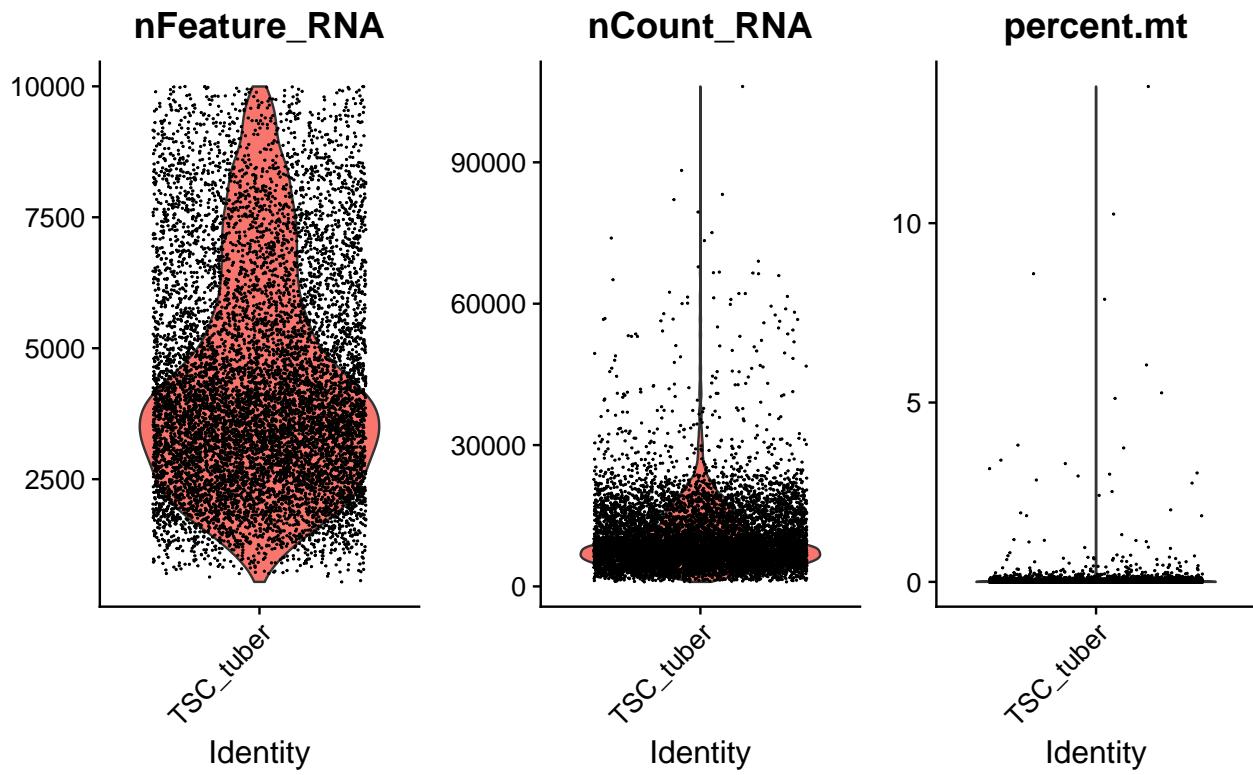
```
## [1] "Number of cells after QC: 12877"
```



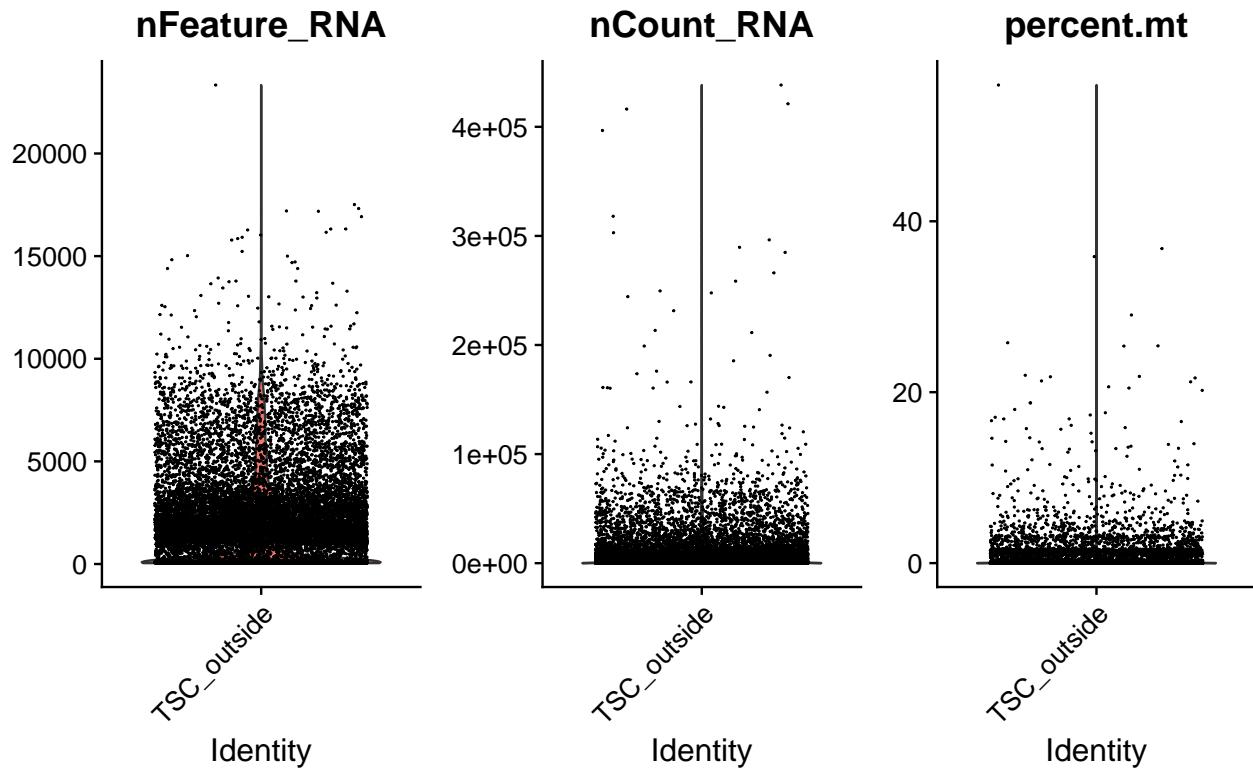
```
## [1] 3
## [1] "TSC_tuber"
## [1] "Number of cells before QC: 12764"
```



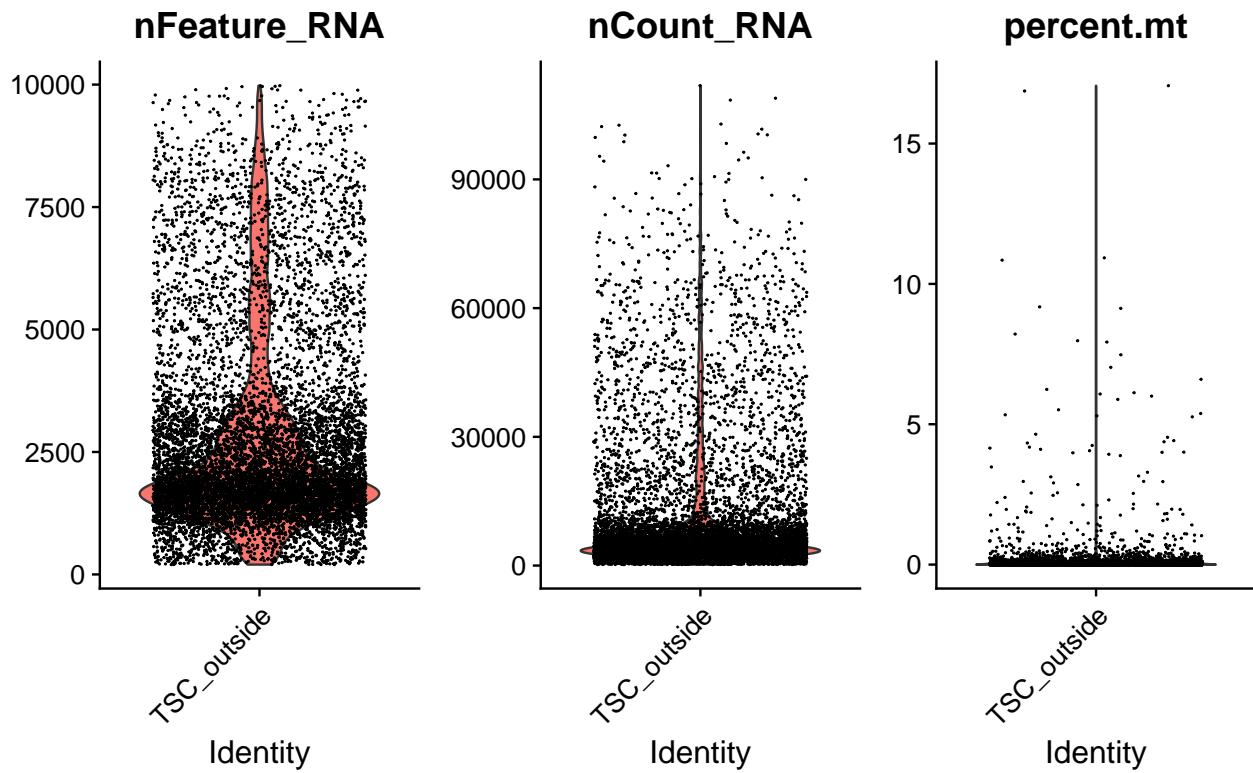
```
## [1] "Number of cells after QC: 9309"
```



```
## [1] 4
## [1] "TSC_outside"
## [1] "Number of cells before QC: 20000"
```



```
## [1] "Number of cells after QC: 9873"
```



```

## 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] scDbIFinder_1.18.0           SingleCellExperiment_1.26.0
## [3] SummarizedExperiment_1.34.0  MatrixGenerics_1.16.0
## [5] matrixStats_1.5.0            data.table_1.17.0
## [7] BSgenome.Hsapiens.UCSC.hg38_1.4.5 BSgenome_1.72.0
## [9] BiocIO_1.14.0                Biostrings_2.72.1
## [11] XVector_0.44.0              AnnotationHub_3.12.0
## [13] BiocFileCache_2.12.0         dbplyr_2.5.0
## [15] simspec_0.0.0.9000          cowplot_1.1.3
## [17] EnsDb.Hsapiens.v86_2.99.0   ensemblDb_2.28.1
## [19] AnnotationFilter_1.28.0     GenomicFeatures_1.56.0
## [21] AnnotationDbi_1.66.0         Biobase_2.64.0
## [23] Signac_1.14.0              Seurat_5.2.1
## [25] SeuratObject_5.0.2          sp_2.2-0
## [27] rtracklayer_1.64.0          GenomicRanges_1.56.2
## [29] GenomeInfoDb_1.40.1         IRanges_2.38.1
## [31] S4Vectors_0.42.1            BiocGenerics_0.50.0
## [33] knitr_1.50                  lubridate_1.9.4
## [35]forcats_1.0.0               stringr_1.5.1
## [37] dplyr_1.1.4                 purrr_1.0.4
## [39] readr_2.1.5                 tidyR_1.3.1
## [41] tibble_3.2.1                ggplot2_3.5.2
## [43] 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_0.4.1          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_1.0.0
## [15] cli_3.6.4                 spatstat.explore_3.4-2
## [17] fastDummies_1.7.5          labeling_0.4.3
## [19] spatstat.data_3.1-6        ggridges_0.5.6
## [21] pbapply_1.7-2              systemfonts_1.2.2

```

```

## [23] Rsamtools_2.20.0           scater_1.32.1
## [25] parallelly_1.43.0          limma_3.60.6
## [27] rstudioapi_0.17.1          RSQLite_2.3.9
## [29] generics_0.1.3              ica_1.0-3
## [31] spatstat.random_3.3-3       Matrix_1.7-3
## [33] ggbeeswarm_0.7.2           abind_1.4-8
## [35] lifecycle_1.0.4             edgeR_4.2.2
## [37] yaml_2.3.10                SparseArray_1.4.8
## [39] Rtsne_0.17                  grid_4.4.0
## [41] blob_1.2.4                 dqrng_0.4.1
## [43] promises_1.3.2              crayon_1.5.3
## [45] miniUI_0.1.2               lattice_0.22-7
## [47] beachmat_2.20.0             KEGGREST_1.44.1
## [49] metapod_1.12.0              pillar_1.10.2
## [51] rjson_0.2.23                xgboost_1.7.9.1
## [53] future.apply_1.11.3         codetools_0.2-20
## [55] fastmatch_1.1-6             glue_1.8.0
## [57] spatstat.univar_3.1-2       vctrs_0.6.5
## [59] png_0.1-8                  spam_2.11-1
## [61] gtable_0.3.6                assertthat_0.2.1
## [63] cachem_1.1.0                xfun_0.52
## [65] S4Arrays_1.4.1              mime_0.13
## [67] survival_3.8-3              RcppRoll_0.3.1
## [69] tinytex_0.57                statmod_1.5.0
## [71] bluster_1.14.0              fitdistrplus_1.2-2
## [73] ROCR_1.0-11                 nlme_3.1-168
## [75] bit64_4.6.0-1               filelock_1.0.3
## [77] RcppAnnoy_0.0.22            rprojroot_2.0.4
## [79] irlba_2.3.5.1               viper_0.4.7
## [81] KernSmooth_2.23-26           colorspace_2.1-1
## [83] DBI_1.2.3                   ggrastr_1.0.2
## [85] tidyselect_1.2.1              bit_4.6.0
## [87] compiler_4.4.0               curl_6.2.2
## [89] BiocNeighbors_1.22.0          hdf5r_1.3.12
## [91] DelayedArray_0.30.1          plotly_4.10.4
## [93] scales_1.3.0                 lmtest_0.9-40
## [95] rappdirs_0.3.3               digest_0.6.37
## [97] goftest_1.2-3                spatstat.utils_3.1-3
## [99] rmarkdown_2.29                htmltools_0.5.8.1
## [101] pkgconfig_2.0.3              sparseMatrixStats_1.16.0
## [103] fastmap_1.2.0               rlang_1.1.6
## [105] htmlwidgets_1.6.4            UCSC.utils_1.0.0
## [107] shiny_1.10.0                DelayedMatrixStats_1.26.0
## [109] farver_2.1.2                zoo_1.8-14
## [111] jsonlite_2.0.0               BiocParallel_1.38.0
## [113] BiocSingular_1.20.0           RCurl_1.98-1.17
## [115] magrittr_2.0.3               scuttle_1.14.0
## [117] GenomeInfoDbData_1.2.12      dotCall64_1.2
## [119] patchwork_1.3.0              munsell_0.5.1
## [121] Rcpp_1.0.14                  viridis_0.6.5
## [123] reticulate_1.42.0             stringi_1.8.7
## [125] zlibbioc_1.50.0              MASS_7.3-65
## [127] plyr_1.8.9                  parallel_4.4.0
## [129] listenv_0.9.1                ggrepel_0.9.6

```

```
## [131] deldir_2.0-4          splines_4.4.0
## [133] tensor_1.5              hms_1.1.3
## [135] locfit_1.5-9.12        igraph_2.1.4
## [137] spatstat.geom_3.3-6     RcppHNSW_0.6.0
## [139] reshape2_1.4.4           ScaledMatrix_1.12.0
## [141] BiocVersion_3.19.1      XML_3.99-0.18
## [143] evaluate_1.0.3           scran_1.32.0
## [145] BiocManager_1.30.25     tzdb_0.5.0
## [147] httpuv_1.6.16           RANN_2.6.2
## [149] polyclip_1.10-7         future_1.40.0
## [151] scattermore_1.2          rsvd_1.0.5
## [153] xtable_1.8-4            restfulr_0.0.15
## [155] RSpectra_0.16-2          later_1.4.2
## [157] ragg_1.4.0               viridisLite_0.4.2
## [159] beeswarm_0.4.0           memoise_2.0.1
## [161] GenomicAlignments_1.40.0 cluster_2.1.8.1
## [163] timechange_0.3.0          globals_0.17.0
## [165] here_1.0.1
```