Quality Control for scRNA and scATAC seq data

Load and Merge the data

Ximing Ran

2025-03-15

Contents


```
library(tibble)
library(tidyr)
library(dplyr)
library(rtracklayer)
library(dplyr)
library(Seurat)
library(Signac)
library(EnsDb.Hsapiens.v86)
library(ggplot2)
library(cowplot)
library(simspec)
library(cowplot)
library(AnnotationHub)
```

```
# load function from local files
```

```
# load function from local files
load(here::here("data","reference","annotations.rdata")) # load the annotations
```

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 PSZ, TSC-tube, TSC-edge, and TSC-outside, respectively.

```
options(Seurat.object.assay.version = "v3")

counts.psz <- Read10X_h5(here::here("data","raw-data", "cellrange-arc-out",
                                    "PSZ-6","filtered_feature_bc_matrix.h5"))
seurat.psz <- CreateSeuratObject(counts = counts.psz$`Gene Expression`,
                                 assay = "RNA",
                                 project = "PSZ-6")
fragments_path <- here::here("data","raw-data","cellrange-arc-out",
                             "PSZ-6","atac_fragments.tsv.gz")
seurat.psz[['ATAC']] <- CreateChromatinAssay(counts = counts.psz$`Peaks`,
                                             annotation = annotations,
                                             fragments = fragments_path,
                                             sep = c(":", "-"),
                                             genome = 'hg38')
```

```
counts.tube <- Read10X_h5(here::here("data","raw-data", "cellrange-arc-out",
                                     "TSC-tube","filtered_feature_bc_matrix.h5"))
seurat.tube <- CreateSeuratObject(counts = counts.tube$`Gene Expression`,
                                  assay = "RNA",
                                  project = "TSC-tube")
fragments_path <- here::here("data","raw-data","cellrange-arc-out",
                             "TSC-tube","atac_fragments.tsv.gz")
seurat.tube[['ATAC']] <- CreateChromatinAssay(counts = counts.tube$`Peaks`,
                                              annotation = annotations,
                                              fragments = fragments_path,
                                              sep = c(":", "-"),
                                              genome = 'hg38')
```

```
counts.edge <- Read10X_h5(here::here("data","raw-data", "cellrange-arc-out",
                                     "TSC-edge","filtered_feature_bc_matrix.h5"))
seurat.edge <- CreateSeuratObject(counts = counts.edge$`Gene Expression`,
                                  assay = "RNA",
                                  project = "TSC-edge")
fragments_path <- here::here("data","raw-data","cellrange-arc-out",
                             "TSC-edge","atac_fragments.tsv.gz")
seurat.edge[['ATAC']] <- CreateChromatinAssay(counts = counts.edge$`Peaks`,
                                              annotation = annotations,
                                              fragments = fragments_path,
                                              sep = c(":", "-"),
                                              genome = 'hg38')
```

```
counts.outside <- Read10X_h5(here::here("data","raw-data","cellrange-arc-out",
                                        "CTRL","filtered_feature_bc_matrix.h5"))
seurat.outside <- CreateSeuratObject(counts = counts.outside$`Gene Expression`,
                                     assay = "RNA",
                                     project = "TSC-outside")
fragments_path <- here::here("data","raw-data","cellrange-arc-out",
                             "CTRL","atac_fragments.tsv.gz")
seurat.outside[['ATAC']] <- CreateChromatinAssay(counts = counts.outside$`Peaks`,
```

```
                                                annotation = annotations,
                                                fragments = fragments_path,
                                                sep = c(":", "-"),
                                                genome = 'hg38')
```

2. Merge the samples

```
# Add unique identifiers to the cell names for each Seurat object
seurat.psz <- RenameCells(seurat.psz,
                          new.names = paste("psz", Cells(seurat.psz), sep = "_"))
seurat.tube <- RenameCells(seurat.tube,
                           new.names = paste("tube", Cells(seurat.tube), sep = "_"))
seurat.edge <- RenameCells(seurat.edge,
                           new.names = paste("edge", Cells(seurat.edge), sep = "_"))
seurat.outside <- RenameCells(seurat.outside,
                              new.names = paste("outside",
                                                Cells(seurat.outside), sep = "_"))

# Now merge the Seurat objects
seurat <- merge(seurat.psz, y = list(seurat.tube, seurat.edge, seurat.outside))

# remake the peaks link for the
peaks <- reduce(unlist(as(c(seurat.psz@assays$ATAC@ranges,
                            seurat.tube@assays$ATAC@ranges,
                            seurat.edge@assays$ATAC@ranges,
                            seurat.outside@assays$ATAC@ranges),
                            "GRangesList")))
peakwidths <- width(peaks)
peaks <- peaks[peakwidths < 10000 & peakwidths > 20]

counts_atac_merged <- FeatureMatrix(seurat@assays$ATAC@fragments,
                                    features = peaks,
                                    cells = colnames(seurat))
seurat[['ATAC']] <- CreateChromatinAssay(counts_atac_merged,
                                         fragments = seurat@assays$ATAC@fragments,
                                         annotation = seurat@assays$ATAC@annotation,
                                         sep = c(":","-"),
                                         genome = "hg38")
save(seurat, file = here::here("data","processed-data","seurat-merged_2025-03-15.rdata"))
# save(seurat, file = here::here("data","processed-data","seurat-merged_2025-03-14.rdata"))
```

```r
sessionInfo()
```

```
## R version 4.4.0 (2024-04-24)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sonoma 14.3.1
##
## 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] AnnotationHub_3.12.0      BiocFileCache_2.12.0
##  [3] dbplyr_2.5.0              simspec_0.0.0.9000
##  [5] cowplot_1.1.3             EnsDb.Hsapiens.v86_2.99.0
##  [7] ensembldb_2.28.1          AnnotationFilter_1.28.0
##  [9] GenomicFeatures_1.56.0    AnnotationDbi_1.66.0
## [11] Biobase_2.64.0            Signac_1.14.0
## [13] Seurat_5.2.1              SeuratObject_5.0.2
## [15] sp_2.2-0                  rtracklayer_1.64.0
## [17] GenomicRanges_1.56.2      GenomeInfoDb_1.40.1
## [19] IRanges_2.38.1            S4Vectors_0.42.1
## [21] BiocGenerics_0.50.0       knitr_1.49
## [23] lubridate_1.9.4           forcats_1.0.0
## [25] stringr_1.5.1             dplyr_1.1.4
## [27] purrr_1.0.4               readr_2.1.5
## [29] tidyr_1.3.1               tibble_3.2.1
## [31] ggplot2_3.5.1             tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##   [1] RcppAnnoy_0.0.22           splines_4.4.0
##   [3] later_1.4.1                BiocIO_1.14.0
##   [5] filelock_1.0.3             bitops_1.0-9
##   [7] polyclip_1.10-7            XML_3.99-0.18
##   [9] fastDummies_1.7.5          lifecycle_1.0.4
##  [11] rprojroot_2.0.4            hdf5r_1.3.12
##  [13] globals_0.16.3             lattice_0.22-6
##  [15] MASS_7.3-64                magrittr_2.0.3
##  [17] plotly_4.10.4              rmarkdown_2.29
##  [19] yaml_2.3.10                httpuv_1.6.15
##  [21] sctransform_0.4.1          spam_2.11-1
##  [23] spatstat.sparse_3.1-0      reticulate_1.40.0
##  [25] pbapply_1.7-2              DBI_1.2.3
##  [27] RColorBrewer_1.1-3         abind_1.4-8
##  [29] zlibbioc_1.50.0            Rtsne_0.17
```

```
##  [31] RCurl_1.98-1.16             rappdirs_0.3.3
##  [33] GenomeInfoDbData_1.2.12     ggrepel_0.9.6
##  [35] irlba_2.3.5.1               listenv_0.9.1
##  [37] spatstat.utils_3.1-2        goftest_1.2-3
##  [39] RSpectra_0.16-2             spatstat.random_3.3-2
##  [41] fitdistrplus_1.2-2          parallelly_1.42.0
##  [43] codetools_0.2-20            DelayedArray_0.30.1
##  [45] RcppRoll_0.3.1              tidyselect_1.2.1
##  [47] UCSC.utils_1.0.0            farver_2.1.2
##  [49] matrixStats_1.5.0           spatstat.explore_3.3-4
##  [51] GenomicAlignments_1.40.0    jsonlite_1.9.0
##  [53] progressr_0.15.1            ggridges_0.5.6
##  [55] survival_3.8-3              tools_4.4.0
##  [57] ica_1.0-3                   Rcpp_1.0.14
##  [59] glue_1.8.0                  gridExtra_2.3
##  [61] SparseArray_1.4.8           here_1.0.1
##  [63] xfun_0.51                   MatrixGenerics_1.16.0
##  [65] withr_3.0.2                 BiocManager_1.30.25
##  [67] fastmap_1.2.0               digest_0.6.37
##  [69] timechange_0.3.0            R6_2.6.1
##  [71] mime_0.12                   colorspace_2.1-1
##  [73] scattermore_1.2             tensor_1.5
##  [75] spatstat.data_3.1-4         RSQLite_2.3.9
##  [77] generics_0.1.3              data.table_1.16.4
##  [79] httr_1.4.7                  htmlwidgets_1.6.4
##  [81] S4Arrays_1.4.1              uwot_0.2.2
##  [83] pkgconfig_2.0.3             gtable_0.3.6
##  [85] blob_1.2.4                  lmtest_0.9-40
##  [87] XVector_0.44.0              htmltools_0.5.8.1
##  [89] dotCall64_1.2               ProtGenerics_1.36.0
##  [91] scales_1.3.0                png_0.1-8
##  [93] spatstat.univar_3.1-1       rstudioapi_0.17.1
##  [95] tzdb_0.4.0                  reshape2_1.4.4
##  [97] rjson_0.2.23                nlme_3.1-167
##  [99] curl_6.2.1                  cachem_1.1.0
## [101] zoo_1.8-12                  BiocVersion_3.19.1
## [103] KernSmooth_2.23-26          parallel_4.4.0
## [105] miniUI_0.1.1.1              restfulr_0.0.15
## [107] pillar_1.10.1               grid_4.4.0
## [109] vctrs_0.6.5                 RANN_2.6.2
## [111] promises_1.3.2              xtable_1.8-4
## [113] cluster_2.1.8               evaluate_1.0.3
## [115] cli_3.6.4                   compiler_4.4.0
## [117] Rsamtools_2.20.0            rlang_1.1.5
## [119] crayon_1.5.3                future.apply_1.11.3
## [121] plyr_1.8.9                  stringi_1.8.4
## [123] viridisLite_0.4.2           deldir_2.0-4
## [125] BiocParallel_1.38.0         munsell_0.5.1
## [127] Biostrings_2.72.1           lazyeval_0.2.2
## [129] spatstat.geom_3.3-5         Matrix_1.7-2
## [131] RcppHNSW_0.6.0              hms_1.1.3
## [133] patchwork_1.3.0             bit64_4.6.0-1
## [135] future_1.34.0               KEGGREST_1.44.1
## [137] shiny_1.10.0                SummarizedExperiment_1.34.0
```

```
## [139] ROCR_1.0-11                memoise_2.0.1
## [141] igraph_2.1.4               fastmatch_1.1-6
## [143] bit_4.5.0.1
```