Integrate 10x Samples

Joan Abinet

2024-12-12 14:28:36 +0100

Contents

Load Packages

```
suppressMessages(library(dplyr))
suppressMessages(library(Seurat))
suppressMessages(library(patchwork))
suppressMessages(library(ggplot2))
suppressMessages(library(stringr))
suppressMessages(library(SingleCellExperiment))
suppressMessages(library(scDblFinder))
```

Loading data

```
all_dirs <- dir(path = "Data", full.names = T)

list_sample <- list()
for (i in 1:length(all_dirs)) {

    Seq_raw_file <- Read10X(data.dir = paste0(all_dirs[i]))
    Seurat_file <- CreateSeuratObject(counts = Seq_raw_file,
        project = str_sub(all_dirs[i], -8, -1), min.cells = 3,
        min.features = 150)

    Seurat_file[["percent.mt"]] <- PercentageFeatureSet(Seurat_file,
        pattern = "^MT-")
    Seurat_file <- subset(Seurat_file, subset = nFeature_RNA >
        400 & nCount_RNA > 800 & nFeature_RNA < 8000 & percent.mt <
        20)  # or 5

    list_sample <- append(list_sample, Seurat_file)
}
```

```
list_sample <- lapply(list_sample, function(x) {
    x <- NormalizeData(x, verbose = F)
    x <- FindVariableFeatures(x, selection.method = "vst", nfeatures = 2000,
        verbose = F)
})
```

```
features <- SelectIntegrationFeatures(list_sample, verbose = F)

list_sample <- lapply(list_sample, function(x) {
    x <- ScaleData(x, features = features, verbose = F)
    x <- RunPCA(x, features = features, verbose = F)
})
```

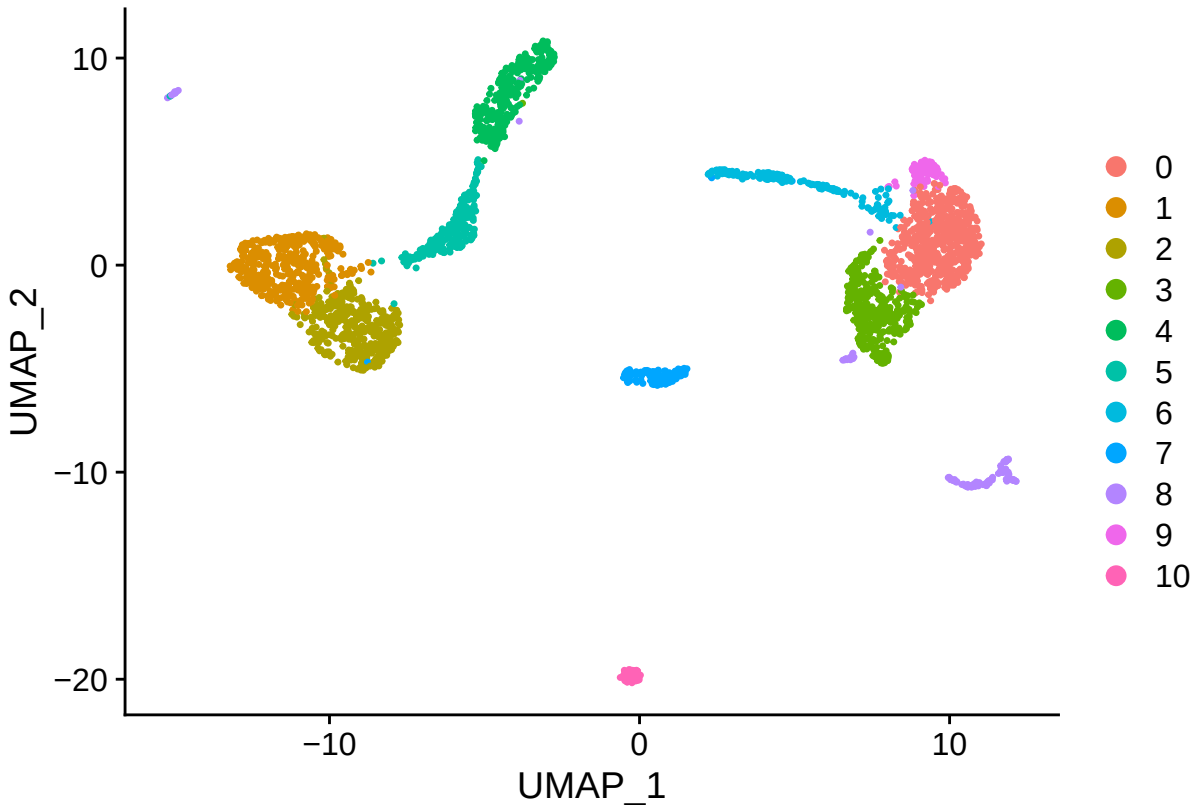
```
BAL.anchors <- FindIntegrationAnchors(object.list = list_sample,
    anchor.features = features, reduction = "rpca")

BAL_10x.integrated <- IntegrateData(anchorset = BAL.anchors)
```

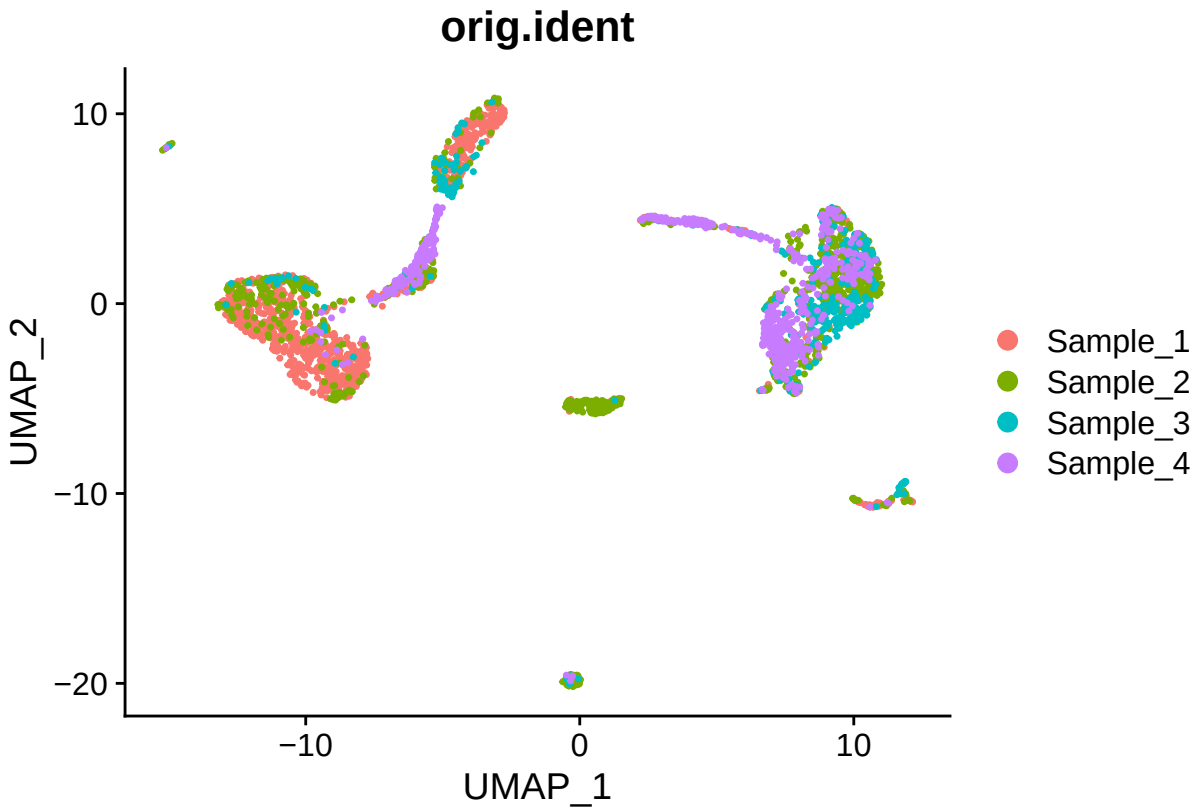
```
DefaultAssay(BAL_10x.integrated) <- "integrated"

# Run the standard workflow for visualization and
# clustering
BAL_10x.integrated <- ScaleData(BAL_10x.integrated, verbose = FALSE)
BAL_10x.integrated <- RunPCA(BAL_10x.integrated, npcs = 30, verbose = FALSE)
BAL_10x.integrated <- RunUMAP(BAL_10x.integrated, reduction = "pca",
    dims = 1:15)
BAL_10x.integrated <- FindNeighbors(BAL_10x.integrated, reduction = "pca",
    dims = 1:15)
BAL_10x.integrated <- FindClusters(BAL_10x.integrated, resolution = 0.35)
```

```
DimPlot(BAL_10x.integrated, reduction = "umap")
```




```
DimPlot(BAL_10x.integrated, reduction = "umap", group.by = "orig.ident")
```



Identifying Doublets

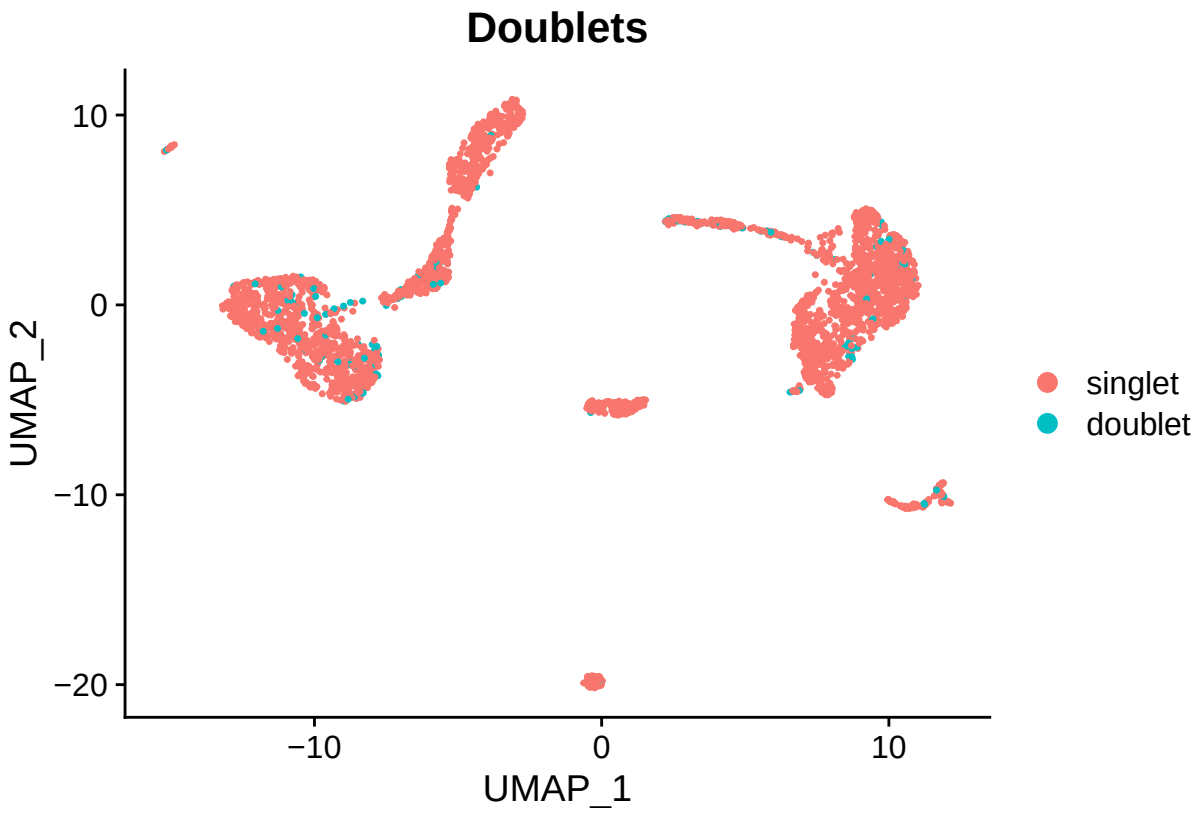
173 (6.1%) doublets called

```
DefaultAssay(BAL_10x.integrated) <- "RNA"
sce <- as.SingleCellExperiment(BAL_10x.integrated)
sce <- scDblFinder(sce, clusters = "seurat_clusters")

setequal(colnames(sce), colnames(BAL_10x.integrated))

BAL_10x.integrated$Doublets <- sce$scDblFinder.class
```

```
DimPlot(BAL_10x.integrated, group.by = "Doublets")
```



Removing doublets and reclusters

```
DefaultAssay(BAL_10x.integrated) <- "integrated"
BAL_10x.integrated <- subset(BAL_10x.integrated, Doublets ==
    "singlet")

BAL_10x.integrated <- RunUMAP(BAL_10x.integrated, reduction = "pca",
    dims = 1:15)
BAL_10x.integrated <- FindNeighbors(BAL_10x.integrated, reduction = "pca",
    dims = 1:15)
BAL_10x.integrated <- FindClusters(BAL_10x.integrated, resolution = 0.5)
```

```
saveRDS(BAL_10x.integrated, "BAL_10x.integrated_noDB.rds")
```

```r
sessionInfo()
```

```
## R version 4.3.3 (2024-02-29)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=fr_BE.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=fr_BE.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=fr_BE.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=fr_BE.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Brussels
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] scDblFinder_1.14.0          SingleCellExperiment_1.22.0
##  [3] SummarizedExperiment_1.30.2 Biobase_2.60.0
##  [5] GenomicRanges_1.52.0        GenomeInfoDb_1.36.0
##  [7] IRanges_2.34.0              S4Vectors_0.38.1
##  [9] BiocGenerics_0.46.0         MatrixGenerics_1.12.2
## [11] matrixStats_1.0.0           stringr_1.5.0
## [13] ggplot2_3.4.2               patchwork_1.1.2
## [15] SeuratObject_4.1.3          Seurat_4.3.0
## [17] dplyr_1.1.2
##
## loaded via a namespace (and not attached):
##   [1] RcppAnnoy_0.0.21          splines_4.3.3
##   [3] later_1.3.1               BiocIO_1.10.0
##   [5] bitops_1.0-7              R.oo_1.25.0
##   [7] tibble_3.2.1              polyclip_1.10-4
##   [9] XML_3.99-0.14             lifecycle_1.0.3
##  [11] edgeR_3.42.4              globals_0.16.2
##  [13] lattice_0.22-5            MASS_7.3-60.0.1
##  [15] magrittr_2.0.3            limma_3.56.2
##  [17] plotly_4.10.2             rmarkdown_2.23
##  [19] yaml_2.3.7                metapod_1.8.0
##  [21] httpuv_1.6.11             sctransform_0.3.5
##  [23] spam_2.9-1                sp_2.0-0
##  [25] spatstat.sparse_3.0-2     reticulate_1.30
##  [27] cowplot_1.1.1             pbapply_1.7-2
##  [29] RColorBrewer_1.1-3        abind_1.4-5
##  [31] zlibbioc_1.46.0           Rtsne_0.16
##  [33] R.utils_2.12.2            purrr_1.0.1
```

```
##  [35] RCurl_1.98-1.12           GenomeInfoDbData_1.2.10
##  [37] ggrepel_0.9.3             irlba_2.3.5.1
##  [39] listenv_0.9.0             spatstat.utils_3.0-3
##  [41] goftest_1.2-3             dqrng_0.3.0
##  [43] spatstat.random_3.1-5     fitdistrplus_1.1-11
##  [45] parallelly_1.36.0         DelayedMatrixStats_1.22.1
##  [47] leiden_0.4.3              codetools_0.2-19
##  [49] DelayedArray_0.26.3       scuttle_1.10.1
##  [51] tidyselect_1.2.0          farver_2.1.1
##  [53] viridis_0.6.3             ScaledMatrix_1.8.1
##  [55] spatstat.explore_3.2-1    GenomicAlignments_1.36.0
##  [57] jsonlite_1.8.7            BiocNeighbors_1.18.0
##  [59] ellipsis_0.3.2            progressr_0.13.0
##  [61] ggridges_0.5.4            survival_3.5-8
##  [63] scater_1.28.0             tools_4.3.3
##  [65] ica_1.0-3                 Rcpp_1.0.11
##  [67] glue_1.6.2                gridExtra_2.3
##  [69] xfun_0.39                 withr_2.5.0
##  [71] formatR_1.14              fastmap_1.1.1
##  [73] bluster_1.10.0            fansi_1.0.4
##  [75] digest_0.6.33             rsvd_1.0.5
##  [77] R6_2.5.1                  mime_0.12
##  [79] colorspace_2.1-0          scattermore_1.2
##  [81] tensor_1.5                spatstat.data_3.0-1
##  [83] R.methodsS3_1.8.2         utf8_1.2.3
##  [85] tidyr_1.3.0               generics_0.1.3
##  [87] data.table_1.14.8         rtracklayer_1.60.0
##  [89] httr_1.4.6                htmlwidgets_1.6.2
##  [91] S4Arrays_1.2.1            uwot_0.1.16
##  [93] pkgconfig_2.0.3           gtable_0.3.3
##  [95] lmtest_0.9-40             XVector_0.40.0
##  [97] htmltools_0.5.5           dotCall64_1.0-2
##  [99] scales_1.2.1              png_0.1-8
## [101] scran_1.28.2              knitr_1.43
## [103] rstudioapi_0.14           reshape2_1.4.4
## [105] rjson_0.2.21              nlme_3.1-164
## [107] zoo_1.8-12                KernSmooth_2.23-22
## [109] vipor_0.4.5               parallel_4.3.3
## [111] miniUI_0.1.1.1            restfulr_0.0.15
## [113] pillar_1.9.0              grid_4.3.3
## [115] vctrs_0.6.3               RANN_2.6.1
## [117] promises_1.2.0.1          BiocSingular_1.16.0
## [119] beachmat_2.16.0           xtable_1.8-4
## [121] cluster_2.1.6             beeswarm_0.4.0
## [123] evaluate_0.21             locfit_1.5-9.8
## [125] cli_3.6.1                 compiler_4.3.3
## [127] Rsamtools_2.16.0          rlang_1.1.1
## [129] crayon_1.5.2              future.apply_1.11.0
## [131] labeling_0.4.2            ggbeeswarm_0.7.2
## [133] plyr_1.8.8                stringi_1.7.12
## [135] viridisLite_0.4.2         deldir_1.0-9
## [137] BiocParallel_1.34.2       munsell_0.5.0
## [139] Biostrings_2.68.1         lazyeval_0.2.2
## [141] spatstat.geom_3.2-4       Matrix_1.6-1
```

```
## [143] sparseMatrixStats_1.12.0  future_1.33.0
## [145] statmod_1.5.0             shiny_1.7.4.1
## [147] highr_0.10                ROCR_1.0-11
## [149] igraph_1.5.0.1            xgboost_1.7.8.1
```