# Integrate Hive Samples

Joan Abinet

2024-12-13 10:14:37 +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

Create your own data folder with the count Matrix from GEO (GSE276100)

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

list_sample <- list()
for (dir in all_dirs) {

    files <- list.files(dir)
    Count_file <- files[grep("TCM.tsv.gz$", files)]
    Countdata <- read.table(paste0(dir, "/", Count_file),
        sep = "\t", header = T, row.names = 1)

    Lavage_cellsHive <- CreateSeuratObject(counts = Countdata,
        project = str_sub(dir, -5, -1), min.features = 100)

    Lavage_cellsHive[["percent.mt"]] <- PercentageFeatureSet(Lavage_cellsHive,
        pattern = "^MT-")  # MT : human cells
    Lavage_cellsHive <- subset(Lavage_cellsHive,
        subset = nFeature_RNA > 400 & nCount_RNA >
```

```
            800 & nFeature_RNA < 8000 & percent.mt <
            20)

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

```
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)

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

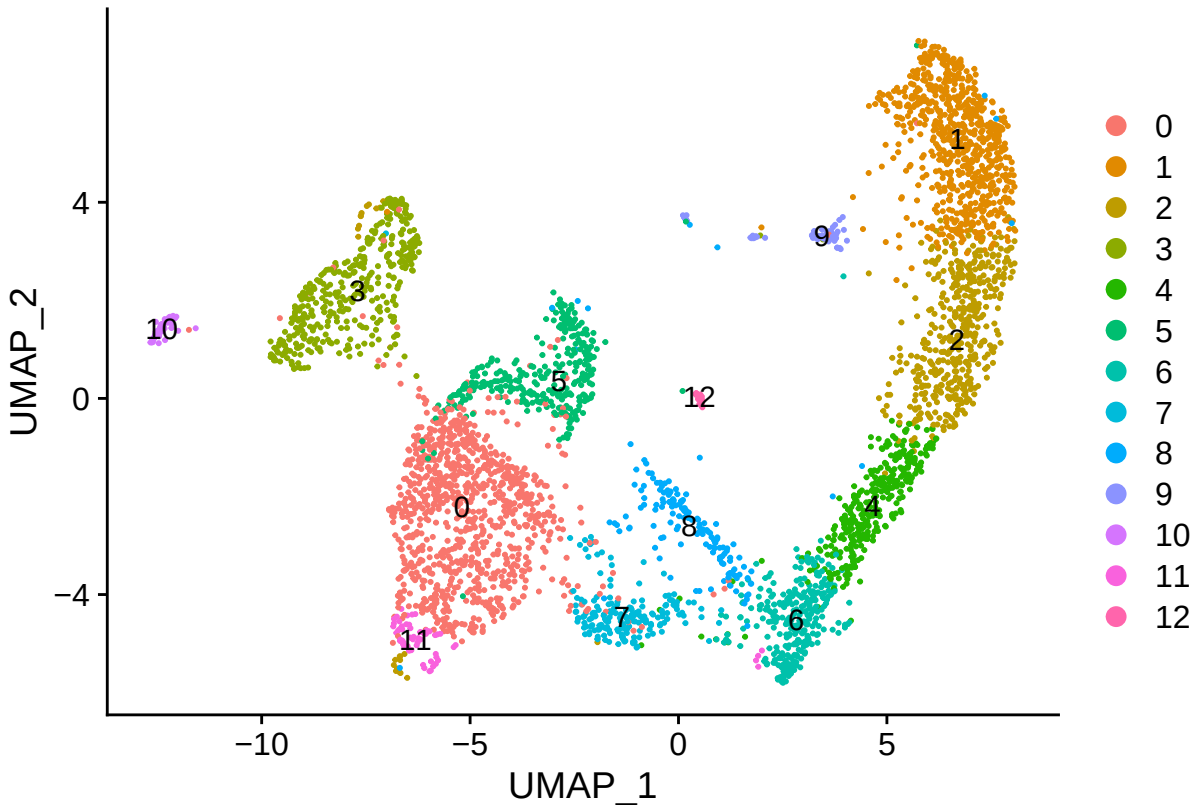
```
BAL.anchors <- FindIntegrationAnchors(object.list = list_sample,
    anchor.features = features, reduction = "rpca",
    verbose = F)

BAL_Hive.integrated <- IntegrateData(anchorset = BAL.anchors,
    verbose = F)
```

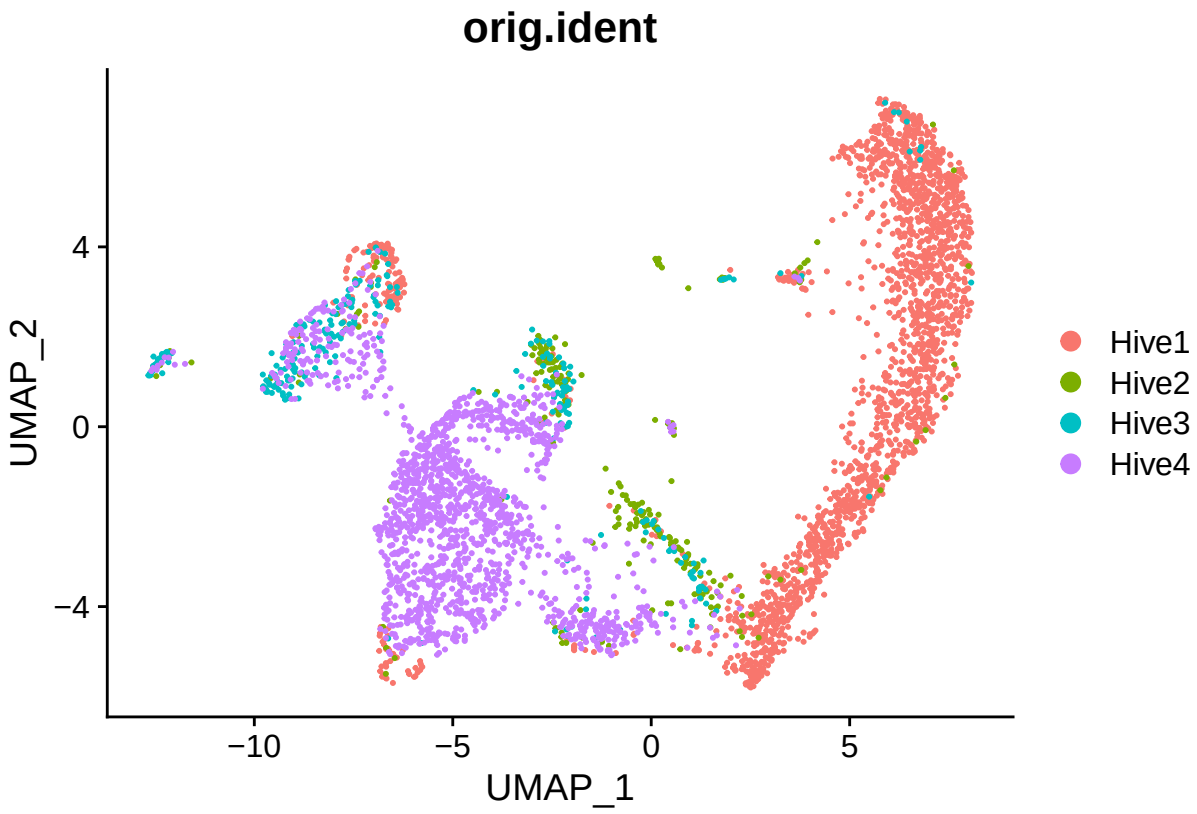
```
DefaultAssay(BAL_Hive.integrated) <- "integrated"

# Run the standard workflow for
# visualization and clustering
BAL_Hive.integrated <- ScaleData(BAL_Hive.integrated,
    verbose = FALSE)
BAL_Hive.integrated <- RunPCA(BAL_Hive.integrated,
    npcs = 30, verbose = FALSE)
BAL_Hive.integrated <- RunUMAP(BAL_Hive.integrated,
    reduction = "pca", dims = 1:15)
BAL_Hive.integrated <- FindNeighbors(BAL_Hive.integrated,
    reduction = "pca", dims = 1:15)
BAL_Hive.integrated <- FindClusters(BAL_Hive.integrated,
    resolution = 0.5)
```

```
DimPlot(BAL_Hive.integrated, reduction = "umap",
    label = T)
```




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



## Identifying Doublets

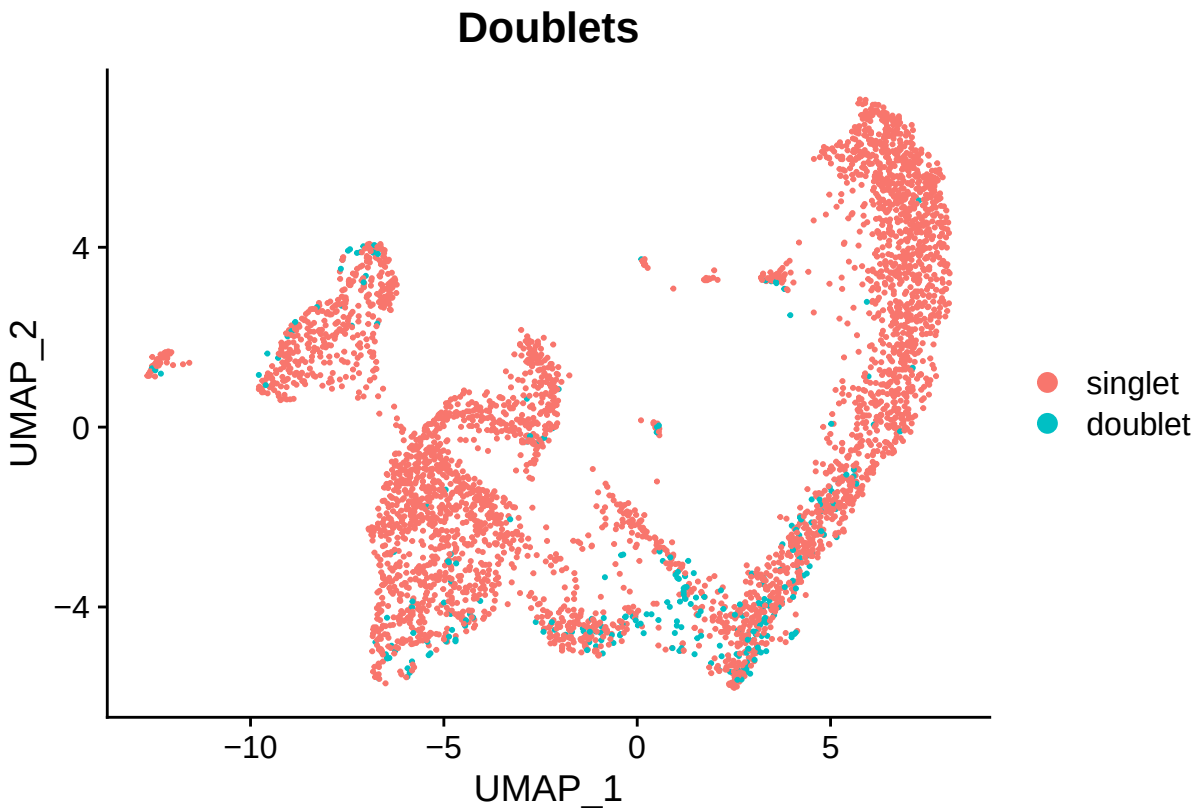
293 (7.3%) doublets called

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

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

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

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




## removing doublets

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

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

```r
saveRDS(BAL_Hive.integrated, "Hive_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              tibble_3.2.1
##   [7] polyclip_1.10-4           XML_3.99-0.14
##   [9] lifecycle_1.0.3           edgeR_3.42.4
##  [11] globals_0.16.2            lattice_0.22-5
##  [13] MASS_7.3-60.0.1           magrittr_2.0.3
##  [15] limma_3.56.2              plotly_4.10.2
##  [17] rmarkdown_2.23            yaml_2.3.7
##  [19] metapod_1.8.0             httpuv_1.6.11
##  [21] sctransform_0.3.5         spam_2.9-1
##  [23] sp_2.0-0                  spatstat.sparse_3.0-2
##  [25] reticulate_1.30           cowplot_1.1.1
##  [27] pbapply_1.7-2             RColorBrewer_1.1-3
##  [29] abind_1.4-5               zlibbioc_1.46.0
```

```
##  [31] Rtsne_0.16                purrr_1.0.1
##  [33] RCurl_1.98-1.12           GenomeInfoDbData_1.2.10
##  [35] ggrepel_0.9.3             irlba_2.3.5.1
##  [37] listenv_0.9.0             spatstat.utils_3.0-3
##  [39] goftest_1.2-3             dqrng_0.3.0
##  [41] spatstat.random_3.1-5     fitdistrplus_1.1-11
##  [43] parallelly_1.36.0         DelayedMatrixStats_1.22.1
##  [45] leiden_0.4.3              codetools_0.2-19
##  [47] DelayedArray_0.26.3       scuttle_1.10.1
##  [49] tidyselect_1.2.0          farver_2.1.1
##  [51] viridis_0.6.3             ScaledMatrix_1.8.1
##  [53] spatstat.explore_3.2-1    GenomicAlignments_1.36.0
##  [55] jsonlite_1.8.7            BiocNeighbors_1.18.0
##  [57] ellipsis_0.3.2            progressr_0.13.0
##  [59] ggridges_0.5.4            survival_3.5-8
##  [61] scater_1.28.0             tools_4.3.3
##  [63] ica_1.0-3                 Rcpp_1.0.11
##  [65] glue_1.6.2                gridExtra_2.3
##  [67] xfun_0.39                 withr_2.5.0
##  [69] formatR_1.14              fastmap_1.1.1
##  [71] bluster_1.10.0            fansi_1.0.4
##  [73] digest_0.6.33             rsvd_1.0.5
##  [75] R6_2.5.1                  mime_0.12
##  [77] colorspace_2.1-0          scattermore_1.2
##  [79] tensor_1.5                spatstat.data_3.0-1
##  [81] utf8_1.2.3                tidyr_1.3.0
##  [83] generics_0.1.3            data.table_1.14.8
##  [85] rtracklayer_1.60.0        httr_1.4.6
##  [87] htmlwidgets_1.6.2         S4Arrays_1.2.1
##  [89] uwot_0.1.16               pkgconfig_2.0.3
##  [91] gtable_0.3.3              lmtest_0.9-40
##  [93] XVector_0.40.0            htmltools_0.5.5
##  [95] dotCall64_1.0-2           scales_1.2.1
##  [97] png_0.1-8                 scran_1.28.2
##  [99] knitr_1.43                rstudioapi_0.14
## [101] reshape2_1.4.4            rjson_0.2.21
## [103] nlme_3.1-164              zoo_1.8-12
## [105] KernSmooth_2.23-22        vipor_0.4.5
## [107] parallel_4.3.3            miniUI_0.1.1.1
## [109] restfulr_0.0.15           pillar_1.9.0
## [111] grid_4.3.3                vctrs_0.6.3
## [113] RANN_2.6.1                promises_1.2.0.1
## [115] BiocSingular_1.16.0       beachmat_2.16.0
## [117] xtable_1.8-4              cluster_2.1.6
## [119] beeswarm_0.4.0            evaluate_0.21
## [121] locfit_1.5-9.8            cli_3.6.1
## [123] compiler_4.3.3            Rsamtools_2.16.0
## [125] rlang_1.1.1               crayon_1.5.2
## [127] future.apply_1.11.0       labeling_0.4.2
## [129] ggbeeswarm_0.7.2          plyr_1.8.8
## [131] stringi_1.7.12            viridisLite_0.4.2
## [133] deldir_1.0-9              BiocParallel_1.34.2
## [135] munsell_0.5.0             Biostrings_2.68.1
## [137] lazyeval_0.2.2            spatstat.geom_3.2-4
```

```
## [139] Matrix_1.6-1              sparseMatrixStats_1.12.0
## [141] future_1.33.0             statmod_1.5.0
## [143] shiny_1.7.4.1             highr_0.10
## [145] ROCR_1.0-11               igraph_1.5.0.1
## [147] xgboost_1.7.8.1
```