Preprocessing And clustering

Abinet Joan

2023-06-01 09:24:26 +0200

Loading packages

```
suppressMessages(library(dplyr))
suppressMessages(library(Seurat))
suppressMessages(library(patchwork))
suppressMessages(library(ggplot2))
suppressMessages(library(dittoSeq))
suppressMessages(library(formatR))
```

Loading data

```
# Importing myeloid cells
path <- "../1-Preprocessing_workflow"
myeloid_cells <- readRDS(paste0(path, "/Myeloid_cells_Part1.rds"))

# Importing IM_cells
IMcells.data <- Read10X(data.dir = paste("barcode_j0", sep = ""))
IMcells <- CreateSeuratObject(counts = IMcells.data, project = "IM_cells", min.cells = 3,
    min.features = 200)

# QC on IM_cells
IMcells[["percent.mt"]] <- PercentageFeatureSet(IMcells, pattern = "^mt-")
IMcells <- subset(IMcells, subset = nFeature_RNA > 1000 & nFeature_RNA < 4000 &
nCount_RNA <
    20000 & percent.mt < 10)
```

Integretion workflow

```
seurat_list <- list(myeloid_cells, IMcells)

seurat_list <- lapply(X = seurat_list, FUN = function(x) {
    x <- NormalizeData(x)
    x <- FindVariableFeatures(x, selection.method = "vst", nfeatures = 2000)
})

features <- SelectIntegrationFeatures(object.list = seurat_list)

cells.anchors <- FindIntegrationAnchors(object.list = seurat_list, anchor.features =
features)
```

```
cells.combined <- IntegrateData(anchorset = cells.anchors)

cells.combined <- ScaleData(cells.combined, verbose = FALSE)
cells.combined <- RunPCA(cells.combined, npcs = 30, verbose = FALSE)

cells.combined <- RunUMAP(cells.combined, reduction = "pca", dims = 1:16)
```

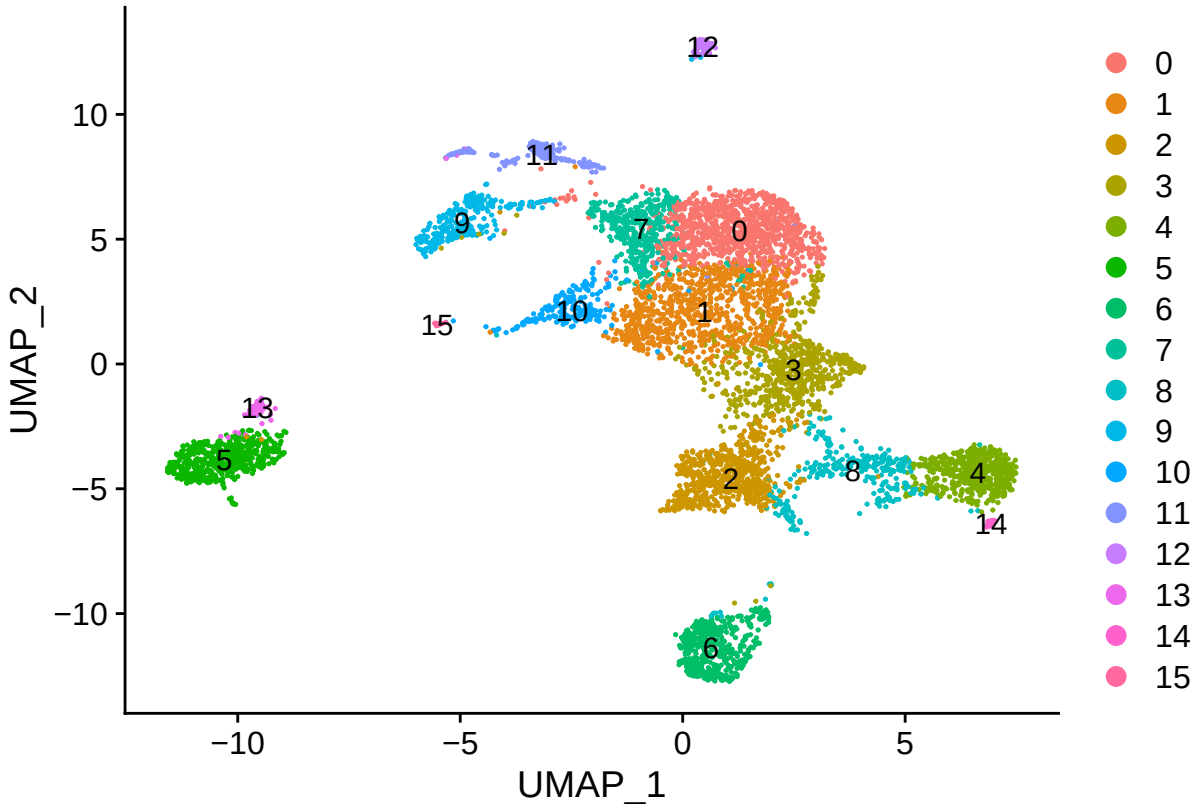
```
## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
## This message will be shown once per session
```

```
cells.combined <- FindNeighbors(cells.combined, reduction = "pca", dims = 1:16)
cells.combined <- FindClusters(cells.combined, resolution = 0.7)
```

```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 6311
## Number of edges: 226798
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.8713
## Number of communities: 16
## Elapsed time: 0 seconds
```

Visualizing all clusters

```
DimPlot(cells.combined, reduction = "umap", label = T)
```



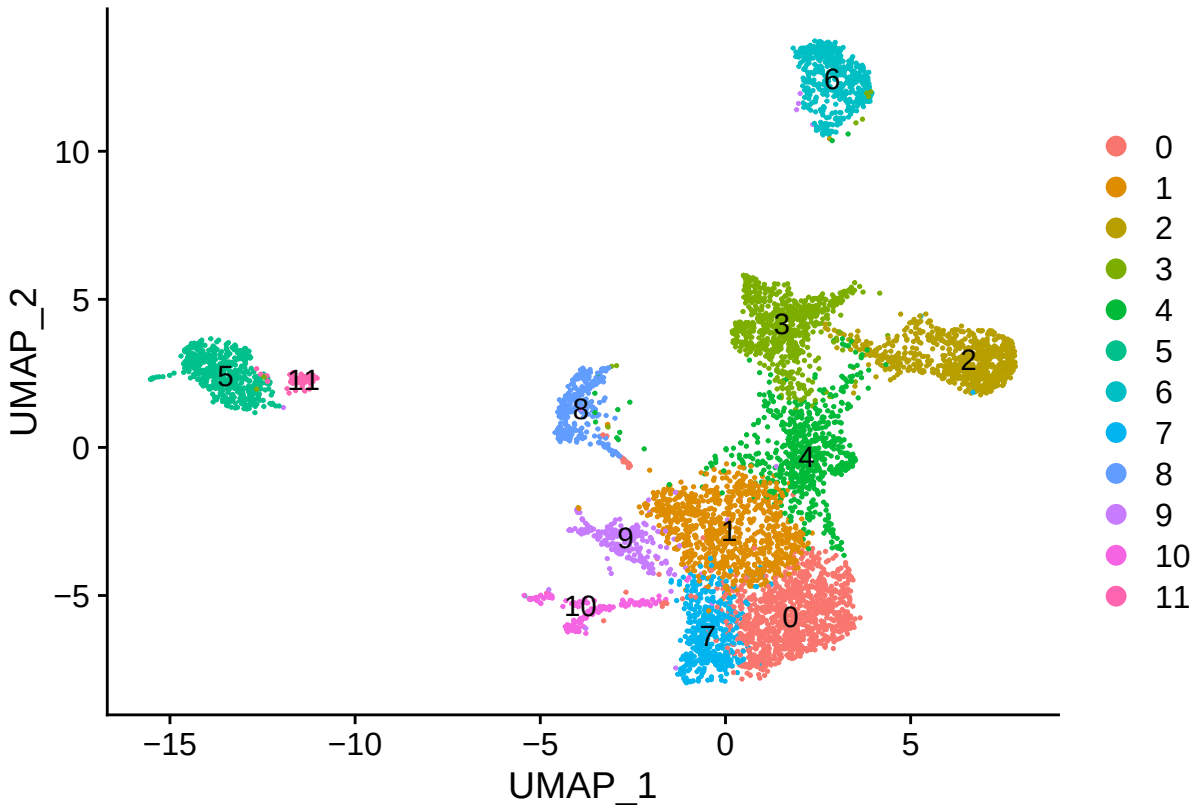
The clusters 12, 14, 15 were identified as contamination and had to be removed.

```
# Selecting the clusters we want to keep
myeloid_cells_filtered <- subset(cells.combined, seurat_clusters %in% c(0, 1, 2,
    3, 4, 5, 6, 7, 8, 9, 10, 11, 13))

# Re-running the linear reduction step
myeloid_cells_filtered <- FindVariableFeatures(myeloid_cells_filtered, selection.method =
"vst",
    nfeatures = 2000)
all.genes <- rownames(myeloid_cells_filtered)
myeloid_cells_filtered <- ScaleData(myeloid_cells_filtered, features = all.genes)
myeloid_cells_filtered <- RunPCA(myeloid_cells_filtered, features =
VariableFeatures(object = myeloid_cells_filtered))
plot1 <- DimPlot(myeloid_cells_filtered, reduction = "pca")


myeloid_cells_filtered <- JackStraw(myeloid_cells_filtered, num.replicate = 100)
myeloid_cells_filtered <- ScoreJackStraw(myeloid_cells_filtered, dims = 1:20)
plot2 <- JackStrawPlot(myeloid_cells_filtered, dims = 1:20)
plot3 <- ElbowPlot(myeloid_cells_filtered)

# Re running the clustering step
myeloid_cells_filtered <- RunUMAP(myeloid_cells_filtered, dims = 1:15)
myeloid_cells_filtered <- FindNeighbors(myeloid_cells_filtered, dims = 1:15)
myeloid_cells_filtered <- FindClusters(myeloid_cells_filtered, resolution = 0.7)  # the
resolution has been chosen based on trial and error, and 0.7 appear to be the most
relevant
```

Reversing umap coordiante for consistency

```
myeloid_cells_filtered@reductions$umap@cell.embeddings[, 2] <-
-myeloid_cells_filtered@reductions$umap@cell.embeddings[,
    2]
```

```
DimPlot(myeloid_cells_filtered, reduction = "umap", label = T)
```



Saving the results for further analysis

```
saveRDS(myeloid_cells_filtered, "Myeloid_cells_Part2.rds")
```

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.6 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/liblapack.so.3
##
## 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
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] formatR_1.14       dittoSeq_1.2.6     ggplot2_3.4.0      patchwork_1.1.2
```

```
## [5] SeuratObject_4.1.3 Seurat_4.3.0       dplyr_1.0.10
##
## loaded via a namespace (and not attached):
##   [1] plyr_1.8.8                  igraph_1.4.1
##   [3] lazyeval_0.2.2              sp_1.6-0
##   [5] splines_4.0.3               listenv_0.9.0
##   [7] scattermore_0.8             GenomeInfoDb_1.26.7
##   [9] digest_0.6.31               htmltools_0.5.4
##  [11] fansi_1.0.4                 magrittr_2.0.3
##  [13] tensor_1.5                  cluster_2.1.0
##  [15] ROCR_1.0-11                 limma_3.46.0
##  [17] globals_0.16.2              matrixStats_0.63.0
##  [19] R.utils_2.12.2              spatstat.sparse_3.0-0
##  [21] colorspace_2.1-0            ggrepel_0.9.2
##  [23] xfun_0.37                   crayon_1.5.2
##  [25] RCurl_1.98-1.10             jsonlite_1.8.4
##  [27] progressr_0.13.0            spatstat.data_3.0-0
##  [29] survival_3.2-7              zoo_1.8-11
##  [31] glue_1.6.2                  polyclip_1.10-4
##  [33] gtable_0.3.1                zlibbioc_1.36.0
##  [35] XVector_0.30.0              leiden_0.4.3
##  [37] DelayedArray_0.16.3         future.apply_1.10.0
##  [39] SingleCellExperiment_1.12.0 BiocGenerics_0.36.1
##  [41] abind_1.4-5                 scales_1.2.1
##  [43] pheatmap_1.0.12             DBI_1.1.3
##  [45] edgeR_3.32.1                spatstat.random_3.1-3
##  [47] miniUI_0.1.1.1              Rcpp_1.0.10
##  [49] viridisLite_0.4.1           xtable_1.8-4
##  [51] reticulate_1.27             stats4_4.0.3
##  [53] htmlwidgets_1.6.1           httr_1.4.5
##  [55] RColorBrewer_1.1-3          ellipsis_0.3.2
##  [57] ica_1.0-3                   farver_2.1.1
##  [59] pkgconfig_2.0.3             R.methodsS3_1.8.2
##  [61] uwot_0.1.14                 deldir_1.0-6
##  [63] locfit_1.5-9.4              utf8_1.2.3
##  [65] labeling_0.4.2              tidyselect_1.2.0
##  [67] rlang_1.0.6                 reshape2_1.4.4
##  [69] later_1.3.0                 munsell_0.5.0
##  [71] tools_4.0.3                 cli_3.6.0
##  [73] generics_0.1.3              ggridges_0.5.4
##  [75] evaluate_0.20               stringr_1.5.0
##  [77] fastmap_1.1.1               yaml_2.3.7
##  [79] goftest_1.2-3               knitr_1.42
##  [81] fitdistrplus_1.1-8          purrr_1.0.1
##  [83] RANN_2.6.1                  pbapply_1.7-0
##  [85] future_1.32.0               nlme_3.1-162
##  [87] mime_0.12                   R.oo_1.25.0
##  [89] compiler_4.0.3              rstudioapi_0.14
##  [91] plotly_4.10.1               png_0.1-8
##  [93] spatstat.utils_3.0-1        tibble_3.1.8
##  [95] stringi_1.7.12              highr_0.10
##  [97] lattice_0.20-41             Matrix_1.5-3
##  [99] vctrs_0.5.2                 pillar_1.8.1
## [101] lifecycle_1.0.3             spatstat.geom_3.0-6
```

```
## [103] lmtest_0.9-40              RcppAnnoy_0.0.20
## [105] data.table_1.14.8          cowplot_1.1.1
## [107] bitops_1.0-7               irlba_2.3.5.1
## [109] httpuv_1.6.9               GenomicRanges_1.42.0
## [111] R6_2.5.1                   promises_1.2.0.1
## [113] KernSmooth_2.23-20         gridExtra_2.3
## [115] IRanges_2.24.1             parallelly_1.34.0
## [117] codetools_0.2-19           MASS_7.3-53
## [119] assertthat_0.2.1           SummarizedExperiment_1.20.0
## [121] withr_2.5.0                sctransform_0.3.5
## [123] S4Vectors_0.28.1           GenomeInfoDbData_1.2.4
## [125] parallel_4.0.3             grid_4.0.3
## [127] tidyr_1.2.1                rmarkdown_2.19
## [129] MatrixGenerics_1.2.1       Rtsne_0.16
## [131] spatstat.explore_3.0-5     Biobase_2.50.0
## [133] shiny_1.7.4
```