Looking into sc3 run on scran normalized data

library(scater)

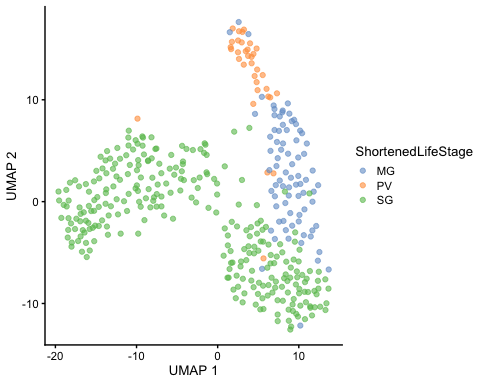
## Warning: package 'BiocGenerics' was built under R version 4.0.5

## Warning: package 'GenomeInfoDb' was built under R version 4.0.5

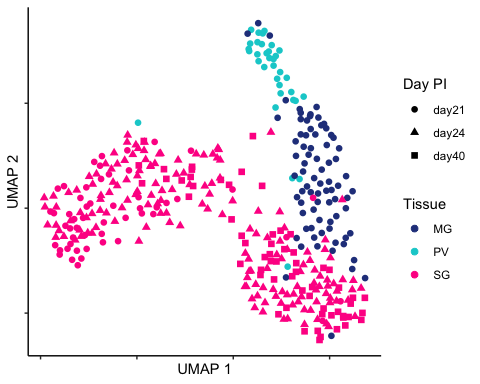
library(SC3)  
library(gridExtra)  
library(viridis)  
library(gplots)  
library(dplyr)  
library(pheatmap)  
library(RColorBrewer)  
library(colorspace)  
library(devtools)  
setwd("/Users/virginiahowick/Documents/Tryps/")

tca.qc.scran <- readRDS("/Users/virginiahowick/Documents/Tryps/for\_github/sc\_data/howick\_tryps\_sce.rds")

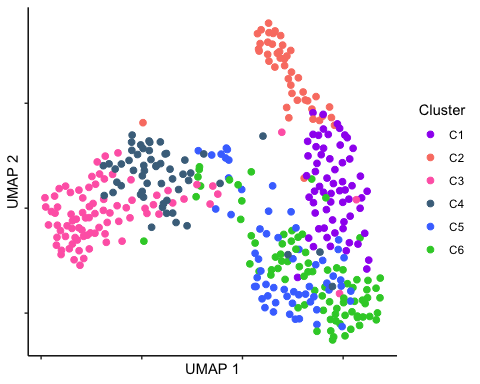
set.seed(7)  
tca.qc.scran <- runUMAP(tca.qc.scran, n\_neighbors = 5, min\_dist = 1, spread = 3,   
 ntop = 200)  
plotUMAP(tca.qc.scran, colour\_by = "ShortenedLifeStage")



um <- plotUMAP(tca.qc.scran, colour\_by = "ShortenedLifeStage", shape\_by = "time")  
umdat <- um$data  
  
  
colors <- c(MG = "royalblue4", PV = "darkturquoise", SG = "deeppink")  
  
ggplot(umdat, aes(X, Y)) + geom\_point(aes(colour = colour\_by, shape = shape\_by),   
 size = 2) + labs(x = "UMAP 1", y = "UMAP 2", color = "Tissue", shape = "Day PI") +   
 scale\_color\_manual(values = colors) + theme\_classic() + theme(axis.text = element\_blank())



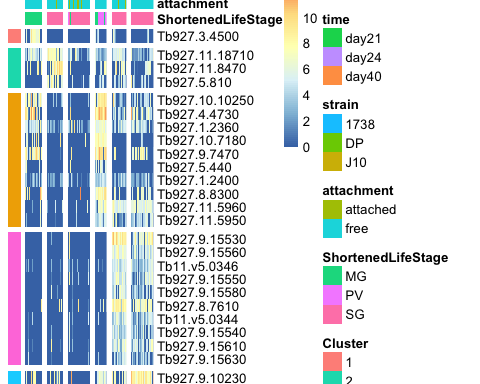
um <- plotUMAP(tca.qc.scran, colour\_by = "sc3\_6\_clusters", shape\_by = "ShortenedLifeStage")  
umdat <- um$data  
  
clustcol <- c(`1` = "purple", `2` = "royalblue1", `3` = "limegreen", `4` = "salmon",   
 `5` = "skyblue4", `6` = "hotpink")  
  
  
ggplot(umdat, aes(X, Y)) + geom\_point(aes(colour = colour\_by), size = 2) + labs(x = "UMAP 1",   
 y = "UMAP 2", color = "Cluster") + scale\_color\_manual(values = clustcol, breaks = c("1",   
 "4", "6", "5", "2", "3"), labels = c("C1", "C2", "C3", "C4", "C5", "C6")) + theme\_classic() +   
 theme(axis.text = element\_blank())



col2hex(c("purple", "royalblue1", "limegreen", "salmon", "skyblue4", "hotpink"))

## [1] "#A020F0" "#4876FF" "#32CD32" "#FA8072" "#4A708B" "#FF69B4"

hm <- sc3\_plot\_markers(tca.qc.scran, k = 6, auroc = 0.75, show\_pdata = c("ShortenedLifeStage",   
 "strain", "attachment", "time", "status"))



unable to recreate SC3 markers heatmap so using their functions from github

get\_processed\_dataset <- function(object) {  
 dataset <- logcounts(object)  
 if (!is.null(rowData(object)$sc3\_gene\_filter)) {  
 dataset <- dataset[rowData(object)$sc3\_gene\_filter, ]  
 }  
 return(dataset)  
}  
  
  
make\_col\_ann\_for\_heatmaps <- function(object, show\_pdata) {  
 if (any(!show\_pdata %in% colnames(colData(object)))) {  
 show\_pdata\_excl <- show\_pdata[!show\_pdata %in% colnames(colData(object))]  
 show\_pdata <- show\_pdata[show\_pdata %in% colnames(colData(object))]  
 message(paste0("Provided columns '", paste(show\_pdata\_excl, collapse = "', '"),   
 "' do not exist in the phenoData table!"))  
 if (length(show\_pdata) == 0) {  
 return(NULL)  
 }  
 }  
 ann <- NULL  
 if (is.null(metadata(object)$sc3$svm\_train\_inds)) {  
 ann <- colData(object)[, colnames(colData(object)) %in% show\_pdata]  
 } else {  
 ann <- colData(object)[metadata(object)$sc3$svm\_train\_inds, colnames(colData(object)) %in%   
 show\_pdata]  
 }  
 # remove columns with 1 value only  
 if (length(show\_pdata) > 1) {  
 keep <- unlist(lapply(ann, function(x) {  
 length(unique(x))  
 })) > 1  
 if (!all(keep)) {  
 message(paste0("Columns '", paste(names(keep)[!keep], collapse = "', '"),   
 "' were excluded from annotation since they contained only a single value."))  
 }  
 ann <- ann[, names(keep)[keep]]  
 if (ncol(ann) == 0) {  
 ann <- NULL  
 } else {  
 ann <- as.data.frame(lapply(ann, function(x) {  
 if (nlevels(as.factor(x)) > 9)   
 x else as.factor(x)  
 }))  
 # convert outlier scores back to numeric  
 for (i in grep("\_log2\_outlier\_score", colnames(ann))) {  
 if (class(ann[, i]) == "factor") {  
 ann[, i] <- as.numeric(levels(ann[, i]))[ann[, i]]  
 }  
 }  
 }  
 } else {  
 if (length(unique(ann)) > 1) {  
 ann <- as.data.frame(ann)  
 colnames(ann) <- show\_pdata  
 if (!grepl("\_log2\_outlier\_score", show\_pdata)) {  
 ann <- as.data.frame(lapply(ann, function(x) {  
 if (nlevels(as.factor(x)) > 9)   
 return(x) else return(as.factor(x))  
 }))  
 }  
 } else {  
 message(paste0("Column '", show\_pdata, "' was excluded from annotation since they contained only a single value."))  
 ann <- NULL  
 }  
 }  
 return(ann)  
}  
  
  
organise\_marker\_genes <- function(object, k, p\_val, auroc) {  
 dat <- rowData(object)[, c(paste0("sc3\_", k, "\_markers\_clusts"), paste0("sc3\_",   
 k, "\_markers\_auroc"), paste0("sc3\_", k, "\_markers\_padj"), "feature\_symbol")]  
 dat <- dat[dat[, paste0("sc3\_", k, "\_markers\_padj")] < p\_val & !is.na(dat[, paste0("sc3\_",   
 k, "\_markers\_padj")]), ]  
 dat <- dat[dat[, paste0("sc3\_", k, "\_markers\_auroc")] > auroc, ]  
   
 d <- NULL  
   
 for (i in sort(unique(dat[, paste0("sc3\_", k, "\_markers\_clusts")]))) {  
 tmp <- dat[dat[, paste0("sc3\_", k, "\_markers\_clusts")] == i, ]  
 tmp <- tmp[order(tmp[, paste0("sc3\_", k, "\_markers\_auroc")], decreasing = TRUE),   
 ]  
 d <- rbind(d, tmp)  
 }  
   
 if (nrow(dat) > 0) {  
 return(d)  
 } else {  
 return(NULL)  
 }  
}  
  
markers\_for\_heatmap <- function(markers) {  
 res <- NULL  
 for (i in unique(markers[, 1])) {  
 tmp <- markers[markers[, 1] == i, ]  
 if (nrow(tmp) > 10) {  
 res <- rbind(res, tmp[1:10, ])  
 } else {  
 res <- rbind(res, tmp)  
 }  
 }  
   
 return(res)  
}  
  
  
  
  
sc3\_plot\_markers.ginny <- function(object, k, auroc, p.val, show\_pdata) {  
 if (is.null(metadata(object)$sc3$consensus)) {  
 warning(paste0("Please run sc3\_consensus() first!"))  
 return(object)  
 }  
 hc <- metadata(object)$sc3$consensus[[as.character(k)]]$hc  
 dataset <- get\_processed\_dataset(object)  
 if (!is.null(metadata(object)$sc3$svm\_train\_inds)) {  
 dataset <- dataset[, metadata(object)$sc3$svm\_train\_inds]  
 }  
   
 add\_ann\_col <- FALSE  
 ann <- NULL  
 if (!is.null(show\_pdata)) {  
 ann <- make\_col\_ann\_for\_heatmaps(object, show\_pdata)  
 if (!is.null(ann)) {  
 add\_ann\_col <- TRUE  
 # make same names for the annotation table  
 rownames(ann) <- colnames(dataset)  
 }  
 }  
   
 # get all marker genes  
 markers <- organise\_marker\_genes(object, k, p.val, auroc)  
   
 if (!is.null(markers)) {  
 # get top 10 marker genes of each cluster  
 markers <- markers\_for\_heatmap(markers)  
 # added the following line to add anno  
 markers$anno <- anno[match(markers$feature\_symbol, anno[, 1]), 10]  
 markers$label <- paste(markers$feature\_symbol, markers$anno, sep = "::")  
 row.ann <- data.frame(Cluster = factor(markers[, 1], levels = unique(markers[,   
 1])))  
 # changed below from feat symb to label  
 rownames(row.ann) <- markers$label  
 # changed from feat to label  
 do.call(pheatmap::pheatmap, c(list(dataset[markers$label, , drop = FALSE],   
 show\_colnames = FALSE, cluster\_rows = FALSE, cluster\_cols = hc, cutree\_cols = k,   
 annotation\_row = row.ann, annotation\_names\_row = FALSE, color = inferno(10),   
 annotation\_colors = ann\_c, gaps\_row = which(diff(markers[, 1]) != 0),   
 cellheight = 10), list(annotation\_col = ann)[add\_ann\_col]))  
 } else {  
 message("No markers have been found, try to lower significance thresholds!")  
 }  
}

### Set your color annotations here!!!  
  
time\_col <- c(day40 = "dodgerblue", day21 = "darkmagenta", day24 = "darkolivegreen1")  
col2hex(c("dodgerblue", "darkmagenta", "darkolivegreen1"))

## [1] "#1E90FF" "#8B008B" "#CAFF70"

clustcol <- c(`1` = "purple", `2` = "royalblue1", `4` = "salmon", `5` = "skyblue4",   
 `6` = "hotpink")  
  
sequential\_hcl(2, palette = "Dark Mint")

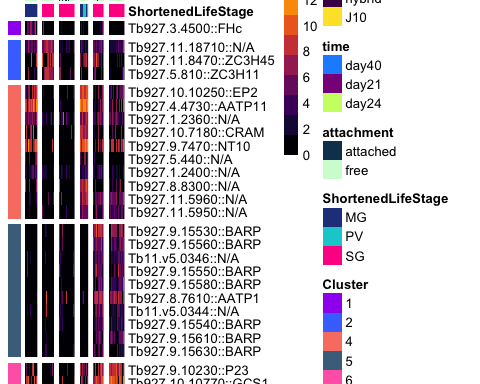
## [1] "#0E3F5C" "#D1FBD4"

attach\_col <- c(attached = "#0E3F5C", free = "#D1FBD4")  
sequential\_hcl(3, palette = "Viridis")

## [1] "#4B0055" "#009B95" "#FDE333"

strain\_col <- c(`1738` = "#009B95", hybrid = "#4B0055", J10 = "#FDE333")  
  
ann\_c <- list(ShortenedLifeStage = colors, time = time\_col, Cluster = clustcol, attachment = attach\_col,   
 spr\_strain = strain\_col)

anno <- read.csv("/Users/virginiahowick/Documents/sanger\_laptop\_tryps/ANALYSIS\_2/tbbtransanno.csv", header=TRUE)  
  
rowData(tca.qc.scran)$gene\_name <- anno[match(rowData(tca.qc.scran)$feature\_symbol, anno[, 1]), 10]  
rowData(tca.qc.scran)$anno <- paste(rowData(tca.qc.scran)$feature\_symbol, rowData(tca.qc.scran)$gene\_name, sep="::")  
  
tca.qc.scran2 <- tca.qc.scran  
rownames(tca.qc.scran2) <- rowData(tca.qc.scran)$anno  
rowData(tca.qc.scran)$feature\_symbol <- rownames(tca.qc.scran2)  
sc3\_plot\_markers.ginny(  
 tca.qc.scran2, k = 6, auroc = 0.75, p.val = 0.01,  
 show\_pdata = c(  
 "time",  
 "attachment",  
 "spr\_strain",  
 "ShortenedLifeStage"  
 )  
)



session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.3 (2020-10-10)  
## os macOS Big Sur 10.16   
## system x86\_64, darwin17.0   
## ui X11   
## language (EN)   
## collate en\_GB.UTF-8   
## ctype en\_GB.UTF-8   
## tz Europe/London   
## date 2022-01-10   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] CRAN (R 4.0.2)  
## beachmat 2.6.3 2020-12-12 [1] Bioconductor   
## beeswarm 0.2.3 2016-04-25 [1] CRAN (R 4.0.2)  
## Biobase \* 2.50.0 2020-10-27 [1] Bioconductor   
## BiocGenerics \* 0.36.1 2021-04-16 [1] Bioconductor   
## BiocNeighbors 1.8.2 2020-12-07 [1] Bioconductor   
## BiocParallel 1.24.1 2020-11-06 [1] Bioconductor   
## BiocSingular 1.6.0 2020-10-27 [1] Bioconductor   
## bitops 1.0-7 2021-04-24 [1] CRAN (R 4.0.2)  
## callr 3.5.1 2020-10-13 [1] CRAN (R 4.0.2)  
## caTools 1.18.0 2020-01-17 [1] CRAN (R 4.0.2)  
## class 7.3-17 2020-04-26 [1] CRAN (R 4.0.3)  
## cli 3.0.1 2021-07-17 [1] CRAN (R 4.0.2)  
## cluster 2.1.0 2019-06-19 [1] CRAN (R 4.0.3)  
## codetools 0.2-18 2020-11-04 [1] CRAN (R 4.0.2)  
## colorspace \* 2.0-2 2021-06-24 [1] CRAN (R 4.0.2)  
## cowplot 1.1.0 2020-09-08 [1] CRAN (R 4.0.2)  
## crayon 1.4.1 2021-02-08 [1] CRAN (R 4.0.2)  
## DBI 1.1.0 2019-12-15 [1] CRAN (R 4.0.2)  
## DelayedArray 0.16.3 2021-03-24 [1] Bioconductor   
## DelayedMatrixStats 1.12.1 2020-11-24 [1] Bioconductor   
## DEoptimR 1.0-8 2016-11-19 [1] CRAN (R 4.0.2)  
## desc 1.2.0 2018-05-01 [1] CRAN (R 4.0.2)  
## devtools \* 2.3.2 2020-09-18 [1] CRAN (R 4.0.2)  
## digest 0.6.27 2020-10-24 [1] CRAN (R 4.0.2)  
## doParallel 1.0.16 2020-10-16 [1] CRAN (R 4.0.2)  
## doRNG 1.8.2 2020-01-27 [1] CRAN (R 4.0.2)  
## dplyr \* 1.0.7 2021-06-18 [1] CRAN (R 4.0.2)  
## e1071 1.7-8 2021-07-28 [1] CRAN (R 4.0.2)  
## ellipsis 0.3.2 2021-04-29 [1] CRAN (R 4.0.2)  
## evaluate 0.14 2019-05-28 [1] CRAN (R 4.0.1)  
## fansi 0.5.0 2021-05-25 [1] CRAN (R 4.0.2)  
## farver 2.1.0 2021-02-28 [1] CRAN (R 4.0.2)  
## fastmap 1.0.1 2019-10-08 [1] CRAN (R 4.0.2)  
## FNN 1.1.3 2019-02-15 [1] CRAN (R 4.0.2)  
## foreach 1.5.1 2020-10-15 [1] CRAN (R 4.0.2)  
## formatR 1.7 2019-06-11 [1] CRAN (R 4.0.2)  
## fs 1.5.0 2020-07-31 [1] CRAN (R 4.0.2)  
## generics 0.1.0 2020-10-31 [1] CRAN (R 4.0.2)  
## GenomeInfoDb \* 1.26.7 2021-04-08 [1] Bioconductor   
## GenomeInfoDbData 1.2.4 2020-12-11 [1] Bioconductor   
## GenomicRanges \* 1.42.0 2020-10-27 [1] Bioconductor   
## ggbeeswarm 0.6.0 2017-08-07 [1] CRAN (R 4.0.2)  
## ggplot2 \* 3.3.5 2021-06-25 [1] CRAN (R 4.0.2)  
## glue 1.4.2 2020-08-27 [1] CRAN (R 4.0.2)  
## gplots \* 3.1.1 2020-11-28 [1] CRAN (R 4.0.2)  
## gridExtra \* 2.3 2017-09-09 [1] CRAN (R 4.0.2)  
## gtable 0.3.0 2019-03-25 [1] CRAN (R 4.0.2)  
## gtools 3.8.2 2020-03-31 [1] CRAN (R 4.0.2)  
## htmltools 0.5.0 2020-06-16 [1] CRAN (R 4.0.2)  
## httpuv 1.5.4 2020-06-06 [1] CRAN (R 4.0.2)  
## IRanges \* 2.24.1 2020-12-12 [1] Bioconductor   
## irlba 2.3.3 2019-02-05 [1] CRAN (R 4.0.2)  
## iterators 1.0.13 2020-10-15 [1] CRAN (R 4.0.2)  
## KernSmooth 2.23-18 2020-10-29 [1] CRAN (R 4.0.2)  
## knitr 1.30 2020-09-22 [1] CRAN (R 4.0.2)  
## labeling 0.4.2 2020-10-20 [1] CRAN (R 4.0.2)  
## later 1.1.0.1 2020-06-05 [1] CRAN (R 4.0.2)  
## lattice 0.20-41 2020-04-02 [1] CRAN (R 4.0.3)  
## lifecycle 1.0.0 2021-02-15 [1] CRAN (R 4.0.2)  
## magrittr 2.0.1 2020-11-17 [1] CRAN (R 4.0.2)  
## Matrix 1.2-18 2019-11-27 [1] CRAN (R 4.0.3)  
## MatrixGenerics \* 1.2.1 2021-01-30 [1] Bioconductor   
## matrixStats \* 0.60.0 2021-07-26 [1] CRAN (R 4.0.2)  
## memoise 1.1.0 2017-04-21 [1] CRAN (R 4.0.2)  
## mime 0.9 2020-02-04 [1] CRAN (R 4.0.2)  
## munsell 0.5.0 2018-06-12 [1] CRAN (R 4.0.2)  
## mvtnorm 1.1-1 2020-06-09 [1] CRAN (R 4.0.2)  
## pcaPP 1.9-73 2018-01-14 [1] CRAN (R 4.0.2)  
## pheatmap \* 1.0.12 2019-01-04 [1] CRAN (R 4.0.2)  
## pillar 1.6.2 2021-07-29 [1] CRAN (R 4.0.2)  
## pkgbuild 1.1.0 2020-07-13 [1] CRAN (R 4.0.2)  
## pkgconfig 2.0.3 2019-09-22 [1] CRAN (R 4.0.2)  
## pkgload 1.1.0 2020-05-29 [1] CRAN (R 4.0.2)  
## prettyunits 1.1.1 2020-01-24 [1] CRAN (R 4.0.2)  
## processx 3.4.5 2020-11-30 [1] CRAN (R 4.0.2)  
## promises 1.1.1 2020-06-09 [1] CRAN (R 4.0.2)  
## proxy 0.4-26 2021-06-07 [1] CRAN (R 4.0.2)  
## ps 1.5.0 2020-12-05 [1] CRAN (R 4.0.2)  
## purrr 0.3.4 2020-04-17 [1] CRAN (R 4.0.2)  
## R6 2.5.0 2020-10-28 [1] CRAN (R 4.0.2)  
## RColorBrewer \* 1.1-2 2014-12-07 [1] CRAN (R 4.0.2)  
## Rcpp 1.0.7 2021-07-07 [1] CRAN (R 4.0.2)  
## RCurl 1.98-1.3 2021-03-16 [1] CRAN (R 4.0.2)  
## remotes 2.2.0 2020-07-21 [1] CRAN (R 4.0.2)  
## rlang 0.4.11 2021-04-30 [1] CRAN (R 4.0.2)  
## rmarkdown 2.5 2020-10-21 [1] CRAN (R 4.0.3)  
## rngtools 1.5 2020-01-23 [1] CRAN (R 4.0.2)  
## robustbase 0.93-6 2020-03-23 [1] CRAN (R 4.0.2)  
## ROCR 1.0-11 2020-05-02 [1] CRAN (R 4.0.2)  
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.0.2)  
## rrcov 1.5-5 2020-08-03 [1] CRAN (R 4.0.2)  
## RSpectra 0.16-0 2019-12-01 [1] CRAN (R 4.0.2)  
## rstudioapi 0.13 2020-11-12 [1] CRAN (R 4.0.2)  
## rsvd 1.0.3 2020-02-17 [1] CRAN (R 4.0.2)  
## S4Vectors \* 0.28.1 2020-12-09 [1] Bioconductor   
## SC3 \* 1.18.0 2020-10-27 [1] Bioconductor   
## scales 1.1.1 2020-05-11 [1] CRAN (R 4.0.2)  
## scater \* 1.18.3 2020-11-08 [1] Bioconductor   
## scuttle 1.0.3 2020-11-23 [1] Bioconductor   
## sessioninfo 1.1.1 2018-11-05 [1] CRAN (R 4.0.2)  
## shiny 1.5.0 2020-06-23 [1] CRAN (R 4.0.2)  
## SingleCellExperiment \* 1.12.0 2020-10-27 [1] Bioconductor   
## sparseMatrixStats 1.2.0 2020-10-27 [1] Bioconductor   
## stringi 1.7.3 2021-07-16 [1] CRAN (R 4.0.2)  
## stringr 1.4.0 2019-02-10 [1] CRAN (R 4.0.2)  
## SummarizedExperiment \* 1.20.0 2020-10-27 [1] Bioconductor   
## testthat 3.0.0 2020-10-31 [1] CRAN (R 4.0.2)  
## tibble 3.1.3 2021-07-23 [1] CRAN (R 4.0.2)  
## tidyselect 1.1.1 2021-04-30 [1] CRAN (R 4.0.2)  
## usethis \* 2.0.0 2020-12-10 [1] CRAN (R 4.0.2)  
## utf8 1.2.2 2021-07-24 [1] CRAN (R 4.0.2)  
## uwot 0.1.9 2020-11-15 [1] CRAN (R 4.0.2)  
## vctrs 0.3.8 2021-04-29 [1] CRAN (R 4.0.2)  
## vipor 0.4.5 2017-03-22 [1] CRAN (R 4.0.2)  
## viridis \* 0.5.1 2018-03-29 [1] CRAN (R 4.0.2)  
## viridisLite \* 0.4.0 2021-04-13 [1] CRAN (R 4.0.2)  
## withr 2.4.2 2021-04-18 [1] CRAN (R 4.0.2)  
## WriteXLS 6.1.0 2020-11-23 [1] CRAN (R 4.0.2)  
## xfun 0.19 2020-10-30 [1] CRAN (R 4.0.2)  
## xtable 1.8-4 2019-04-21 [1] CRAN (R 4.0.2)  
## XVector 0.30.0 2020-10-28 [1] Bioconductor   
## yaml 2.2.1 2020-02-01 [1] CRAN (R 4.0.2)  
## zlibbioc 1.36.0 2020-10-28 [1] Bioconductor   
##   
## [1] /Library/Frameworks/R.framework/Versions/4.0/Resources/library