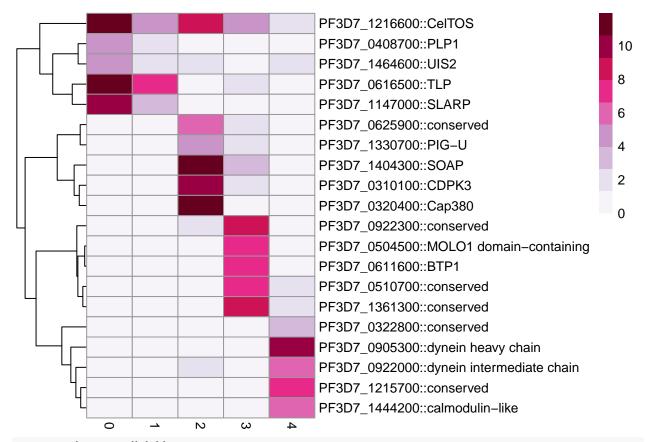
looking at cons markers and DE between species

```
setwd("/Users/vh3/Documents/PfMCA/ANALYSIS 2/pbpf")
library(scater, quietly = TRUE)
library(scmap)
library(Seurat)
library(scater)
library(scran)
library(cowplot)
library(RColorBrewer)
library(pheatmap)
library(scales)
p.combined <- readRDS("pbpfcombo_SeuratInt_20200702.rds")</pre>
p.sce <- as.SingleCellExperiment(p.combined)</pre>
mf <- read.csv("male_female_specific.csv", header = TRUE)</pre>
mf <- mf[, 1:2]
male.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/male_markers.csv",
    header = TRUE, row.names = 1)
male.markers$stage <- rep("male", length(male.markers$Pbe_p_val))</pre>
female.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/female_markers.csv",
    header = TRUE, row.names = 1)
female.markers$stage <- rep("female", length(female.markers$Pbe_p_val))</pre>
ook.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/ook_markers.csv",</pre>
    header = TRUE, row.names = 1)
ook.markers$stage <- rep("ook", length(ook.markers$Pbe_p_val))</pre>
spz.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/spz1_markers.csv",</pre>
    header = TRUE, row.names = 1)
spz.markers$stage <- rep("spz", length(spz.markers$Pbe_p_val))</pre>
m5 <- male.markers[1:5, ]</pre>
f5 <- female.markers[1:5, ]</pre>
o5 <- ook.markers[1:5, ]
s5 <- spz.markers[1:5, ]</pre>
all <- rbind(m5, f5)
all <- rbind(all, o5)
all <- rbind(all, s5)
p.sce.m <- p.sce[rownames(p.sce) %in% rownames(all), ]</pre>
lcpm_mat <- as.data.frame(assays(p.sce.m)[["logcounts"]])</pre>
tmat <- t(lcpm_mat)</pre>
tmat <- as.data.frame(tmat)</pre>
```

```
tmat$seurat_cluster <- as.factor(p.sce$seurat_clusters)</pre>
tmat <- as.data.frame(tmat)</pre>
# lcpm_mat_sub <- lcpm_mat[!is.na(lcpm_mat$Cluster_k15), ]</pre>
clustmean <- aggregate(tmat[, 1:20], by = list(as.factor(as.character(tmat$seurat_cluster))),</pre>
rownames(clustmean) <- clustmean$Group.1</pre>
clustmean2 <- clustmean[, 2:21]</pre>
tcm2 <- as.data.frame(t(clustmean2))</pre>
pheatmap(tcm2, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE,
    color = brewer.pal(9, "PuRd"))
                                                                          ORTH-300
                                                                          ORTH-3075
                                                                                          10
                                                                          ORTH-1259
                                                                                          8
                                                                          ORTH-3513
                                                                          ORTH-612
                                                                                          6
                                                                          ORTH-3594
                                                                          ORTH-1420
                                                                          ORTH-3352
                                                                                          2
                                                                          ORTH-3864
                                                                          ORTH-1886
                                                                          ORTH-2903
                                                                          ORTH-3401
                                                                          ORTH-1699
                                                                          ORTH-3461
                                                                          ORTH-3715
                                                                          ORTH-1862
                                                                          ORTH-3921
                                                                          ORTH-2900
                                                                          ORTH-291
                                                                          ORTH-1074
                0
                                         2
# stage <- as.data.frame(colData(mca.qc.sub)['stage']) cluster <-</pre>
# as.data.frame(colData(mca.qc.sub)['seurat_clusters'])
# pheatmap(clustmean2, annotation_col = stage, show_colnames = FALSE,
# show_rownames = TRUE, color = brewer.pal(9, 'PuRd'))
all$pbpfgene <- paste(all$pfgene, all$pbgene, sep = "
all$orthname <- rownames(all)</pre>
p.sce.m <- p.sce[rownames(p.sce) %in% rownames(all), ]</pre>
```

```
rowData(p.sce.m)$pbpfgene <- all[match(rownames(rowData(p.sce.m)), all[, 17]), 16]</pre>
rowData(p.sce.m)$pfgene <- all[match(rownames(rowData(p.sce.m)), all[, 17]), 14]</pre>
geneanno <- read.csv("/Users/vh3/Documents/allpf_geneanno.csv", header = TRUE)</pre>
rowData(p.sce.m) $gene_name <- geneanno[match(rowData(p.sce.m) $pfgene, geneanno[, 1]), 5]
rowData(p.sce.m)$product_desc <- geneanno[match(rowData(p.sce.m)$pfgene, geneanno[, 1]), 3]</pre>
rowData(p.sce.m) $genesanno <- paste(rowData(p.sce.m) $pbpfgene, rowData(p.sce.m) $gene_name, sep="
rowData(p.sce.m) name_desc <- paste(rowData(p.sce.m) pfgene, rowData(p.sce.m) gene_name, rowData(p.sce.m)
rd <- read.csv("pscemrowdata_mananno.csv", header=TRUE, row.names = 1)</pre>
rowData(p.sce.m) <- rd</pre>
lcpm_mat <- as.data.frame(assays(p.sce.m)[["logcounts"]])</pre>
rownames(lcpm_mat) <- rowData(p.sce.m)$man_anno</pre>
tmat <- t(lcpm_mat)</pre>
tmat <- as.data.frame(tmat)</pre>
tmat$seurat_cluster <- as.factor(p.sce.m$seurat_clusters)</pre>
tmat <- as.data.frame(tmat)</pre>
#lcpm_mat_sub <- lcpm_mat[!is.na(lcpm_mat$Cluster_k15), ]</pre>
#colnames(tmat) <- rowData(p.sce.m)$pbpfgene</pre>
clustmean <- aggregate(tmat[, 1:20], by = list(as.factor(as.character(tmat$seurat_cluster))),</pre>
rownames(clustmean) <- clustmean$Group.1</pre>
clustmean2 <- clustmean[, 2:21]</pre>
tcm2 <- as.data.frame(t(clustmean2))</pre>
pheatmap(tcm2, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE,
   color = brewer.pal(9, "PuRd"))
```



show_col(hue_pal()(5))



```
cols <- c("spz"="#F8766D", "ook"="#00BF7D", "female"="#00B0F6", "male"="#E76BF3")

rowData(p.sce.m)$stage <- all[match(rownames(rowData(p.sce.m)), all[, 17]), 15]

#rowData(p.sce.m)$orthname <- rownames(rowData(p.sce.m))
rownames(rowData(p.sce.m)) <- rowData(p.sce.m)$man_anno

rd <- as.data.frame(rowData(p.sce.m))
rownames(rd) <- rd$man_anno</pre>
```

```
stage <- rd["stage"]</pre>
ann_c <- list(stage = cols)</pre>
pheatmap(tcm2, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE,
    color = brewer.pal(9, "PuRd"), annotation_row=stage, fontsize = 8, annotation_colors = ann_c)
                                                    PF3D7_1216600::CelTOS
                                                                                                   spz
                                                    PF3D7 0408700::PLP1
                                                                                              10
                                                                                                   ook
                                                    PF3D7 1464600::UIS2
                                                                                                    female
                                                    PF3D7_0616500::TLP
                                                                                                   male
                                                    PF3D7_1147000::SLARP
                                                    PF3D7_0625900::conserved
                                                    PF3D7_1330700::PIG-U
                                                                                              4
                                                    PF3D7_1404300::SOAP
                                                    PF3D7_0310100::CDPK3
                                                                                              2
                                                    PF3D7_0320400::Cap380
                                                    PF3D7_0922300::conserved
                                                    PF3D7_0504500::MOLO1 domain-containing
                                                    PF3D7_0611600::BTP1
                                                    PF3D7_0510700::conserved
                                                    PF3D7_1361300::conserved
                                                    PF3D7_0322800::conserved
                                                    PF3D7_0905300::dynein heavy chain
                                                    PF3D7_0922000::dynein intermediate chain
                                                    PF3D7_1215700::conserved
                                                    PF3D7_1444200::calmodulin-like
                 0
                                N
```

```
male.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/male_pf_specific.csv", head
male.markers$stage <- rep("male", length(male.markers$p_val))
female.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/female_pf_specific.csv",
female.markers$stage <- rep("female", length(female.markers$p_val))

ook.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/ook_pf_specific.csv", heade
ook.markers$stage <- rep("ook", length(ook.markers$p_val))

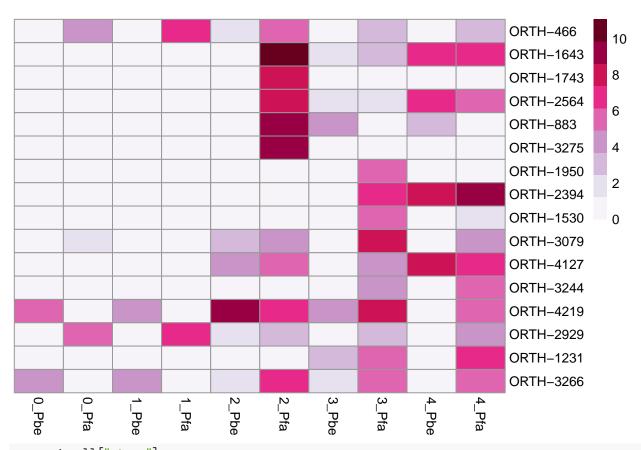
spz.markers <- read.csv("/Users/vh3/Documents/PfMCA/ANALYSIS_2/pbpf/markers/spz1_pf_specific.csv", head
spz.markers$stage <- rep("spz", length(spz.markers$p_val))

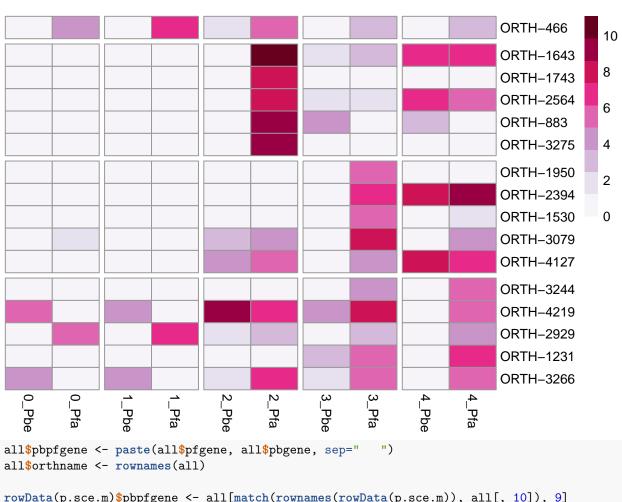
m5 <- male.markers[1:5, ]
f5 <- female.markers[1:5, ]
s5 <- spz.markers[1:1, ]

all <- rbind(s5, o5)
all <- rbind(all, f5)
all <- rbind(all, m5)</pre>
```

```
p.sce.m <- p.sce[rownames(p.sce) %in% rownames(all), ]</pre>
lcpm_mat <- as.data.frame(assays(p.sce.m)[["logcounts"]])</pre>
tmat <- t(lcpm_mat)</pre>
tmat <- as.data.frame(tmat)</pre>
tmat$seurat_cluster <- as.factor(p.sce$ident)</pre>
tmat <- as.data.frame(tmat)</pre>
#lcpm_mat_sub <- lcpm_mat[!is.na(lcpm_mat$Cluster_k15), ]</pre>
clustmean <- aggregate(tmat[, 1:16], by = list(as.factor(as.character(tmat$seurat_cluster))),</pre>
    mean)
rownames(clustmean) <- clustmean$Group.1</pre>
clustmean2 <- clustmean[, 2:17]</pre>
tcm2 <- as.data.frame(t(clustmean2))</pre>
pheatmap(tcm2, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE,
    color = brewer.pal(9, "PuRd"))
                                                                              ORTH-4219
                                                                                              10
                                                                              ORTH-2394
                                                                                             8
                                                                              ORTH-4127
                                                                              ORTH-1643
                                                                                              6
                                                                              ORTH-2564
                                                                              ORTH-883
                                                                                              4
                                                                              ORTH-1743
                                                                                              2
                                                                              ORTH-3275
                                                                              ORTH-2929
                                                                              ORTH-466
                                                                              ORTH-3079
                                                                              ORTH-3266
                                                                              ORTH-1950
                                                                              ORTH-1530
                                                                              ORTH-3244
                                                                              ORTH-1231
                                1_Pfa
                                       2_Pbe
                   0_Pfa
                                              2_Pfa
                                                    3_Pbe
                                                           3_Pfa
                                                                  4_Pbe
tcm3 <- tcm2[rownames(all), ]</pre>
pheatmap(tcm3, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE, cluster_rows = FALSE,
```

color = brewer.pal(9, "PuRd"))





```
all$pbfgene <- paste(all$pfgene, all$pbgene, sep=" ")
all$orthname <- rownames(all)

rowData(p.sce.m)$pbpfgene <- all[match(rownames(rowData(p.sce.m)), all[, 10]), 9]

rowData(p.sce.m)$pfgene <- all[match(rownames(rowData(p.sce.m)), all[, 10]), 7]

geneanno <- read.csv("/Users/vh3/Documents/allpf_geneanno.csv", header = TRUE)

rowData(p.sce.m)$gene_name <- geneanno[match(rowData(p.sce.m)$pfgene, geneanno[, 1]), 5]

rowData(p.sce.m)$genesanno <- paste(rowData(p.sce.m)$pfgene, rowData(p.sce.m)$gene_name, sep=" ")

rowData(p.sce.m)$genesanno <- paste(rowData(p.sce.m)$pfgene, rowData(p.sce.m)$gene_name, rowData(p.sce.m)

rowData(p.sce.m)$name_desc <- paste(rowData(p.sce.m)$pfgene, rowData(p.sce.m)$gene_name, rowData(p.sce.m)

rowData(p.sce.m) <- rd

rowData(p.sce.m) <- rd

rowData(p.sce.m)$stage <- all[match(rownames(rowData(p.sce.m)), all[, 10]), 8]

rowData(p.sce.m)$orthname <- rownames(rowData(p.sce.m))

rownames(rowData(p.sce.m)) <- rowData(p.sce.m))

#rownames(rowData(p.sce.m))

**rownames(rowData(p.sce.m))

***rownames(rowData(p.sce.m))

***rownames(rowData(p.sce.m))
```

```
ann_c <- list(stage = cols)</pre>
tcm4 <- tcm3
all$man_anno <- rd[match(rownames(all), rownames(rd)), 7]</pre>
rownames(tcm4) <- all[match(rownames(tcm4), all[, 10]), 11]</pre>
group <- all["stage"]</pre>
test <- which(group$stage != dplyr::lag(group$stage))</pre>
test2 <- test-1
pheatmap(tcm4, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE, cluster_rows = FALSE,
    color = brewer.pal(9, "PuRd"), gaps_row = test2, gaps_col = c(2, 4, 6, 8))
                                           PF3D7_1235500 ::mRNA methyltransferase
                                                                                            10
                                           PF3D7 0605600::nucleoside diphosphate kinase
                                                                                            8
                                           PF3D7_1004600::conserved
                                           PF3D7_1025500::conserved
                                                                                            6
                                           PF3D7_1116800::HSP101
                                           PF3D7_1413700::conserved
                                                                                            4
                                           PF3D7_0805400::N-acetyltransferase
                                                                                            2
                                           PF3D7_0209000::P230
                                                                                            0
                                           PF3D7_1344500::USB1
                                           PF3D7_0409200::SOF1
                                           PF3D7_0418000::conserved
                                           PF3D7 1417500::CBF5
                                           PF3D7_0103400::zinc-carboxypeptidase
                                           PF3D7 0925200::RRP8
                                           PF3D7_1461100::conserved
                                           PF3D7_1414600::Pgt1
                              3_Pfa
             1_Pfa
                  2_Pbe
                          3_Pbe
                                   4_Pbe
         1_Pbe
                      2_Pfa
ann_c <- list(stage = cols)</pre>
rownames(rd) <- rd$man_anno</pre>
stage <- rd["stage"]</pre>
pheatmap(tcm4, show_colnames = TRUE, show_rownames = TRUE, cluster_cols = FALSE, cluster_rows = FALSE,
    color = brewer.pal(9, "PuRd"), gaps_row = test2, gaps_col = c(2, 4, 6, 8), annotation_row=stage, for
```

