Ovary_CellRanger-v3_SingleR

Clear workspace and load libraries

```
library(Seurat)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(SingleR)
library(pheatmap)
library(scales)
library(SingleCellExperiment)
library(scran)
library(ggpubr)
```

Load the processed Seurat object

```
#We followed the published script for processing;
#https://github.com/wagmag/SingleCellOvary/blob/master/Wagner20_Script_Unsorted.pdf
unsorted.V3<-readRDS("200920_Ovar_run2.integrated_MtFiltered_CRv3.rds")
unsorted.V3
```

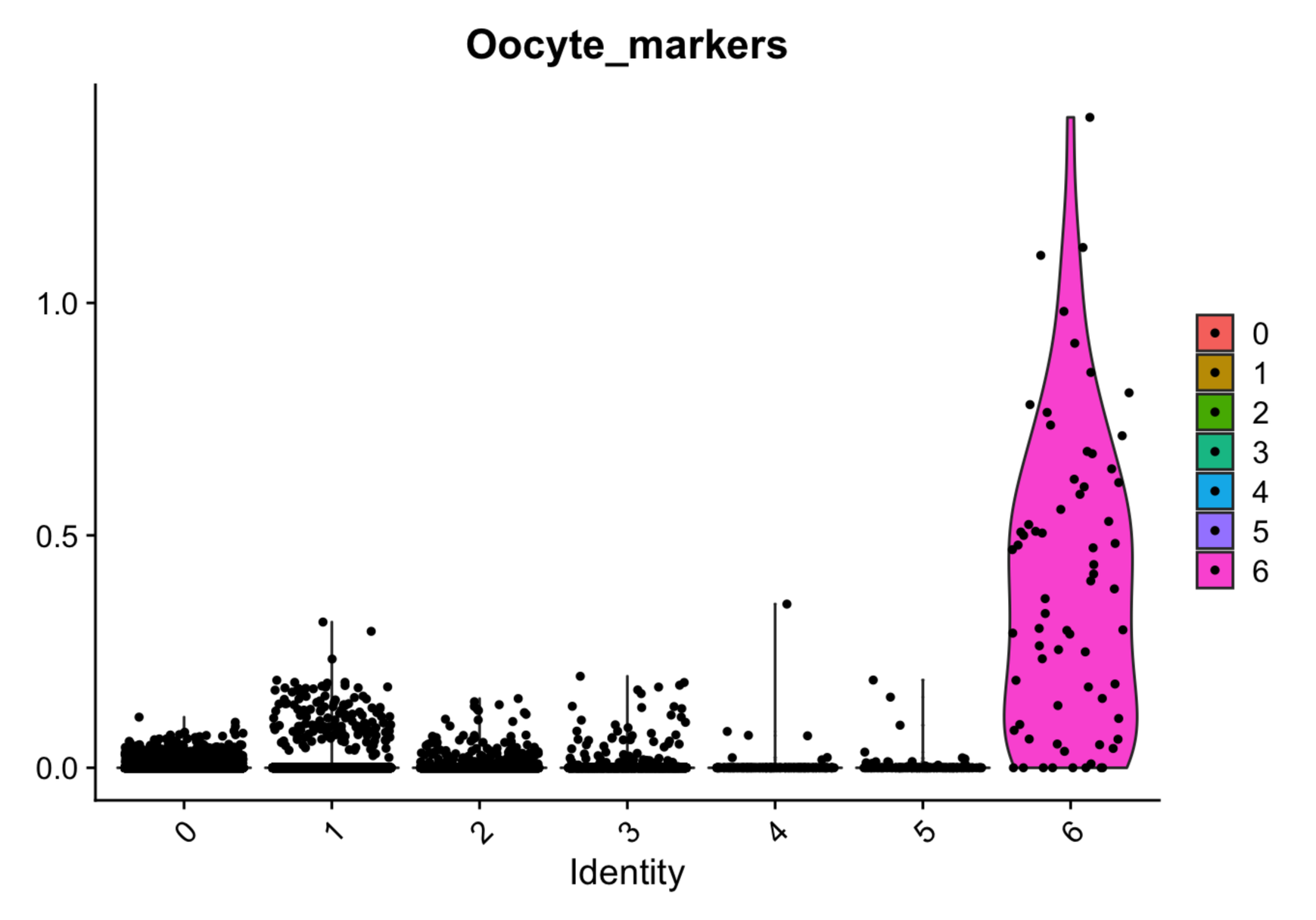
```
## An object of class Seurat
## 18392 features across 25789 samples within 2 assays
## Active assay: RNA (16392 features, 0 variable features)
##  1 other assay present: integrated
##  2 dimensional reductions calculated: pca, umap
```

Add oocyte and germline marker scores

```
oomarker <- c("GDF9", "ZP3", "FIGLA", "OOSP2")
oomarker_percent <- colSums(unsorted.V3@assays$RNA@counts[oomarker,])*100/colSums(unsorted.V3@assays$RNA@counts)
unsorted.V3 <- AddMetaData(object = unsorted.V3, metadata = oomarker_percent, col.name = "Oocyte_markers")

germ <- c("DAZL","DPPA3","PRDM1")
germ_percent <- colSums(unsorted.V3@assays$RNA@counts[germ,])*100/colSums(unsorted.V3@assays$RNA@counts)
unsorted.V3 <- AddMetaData(object = unsorted.V3, metadata = germ_percent, col.name = "Germ_cell_markers")
```

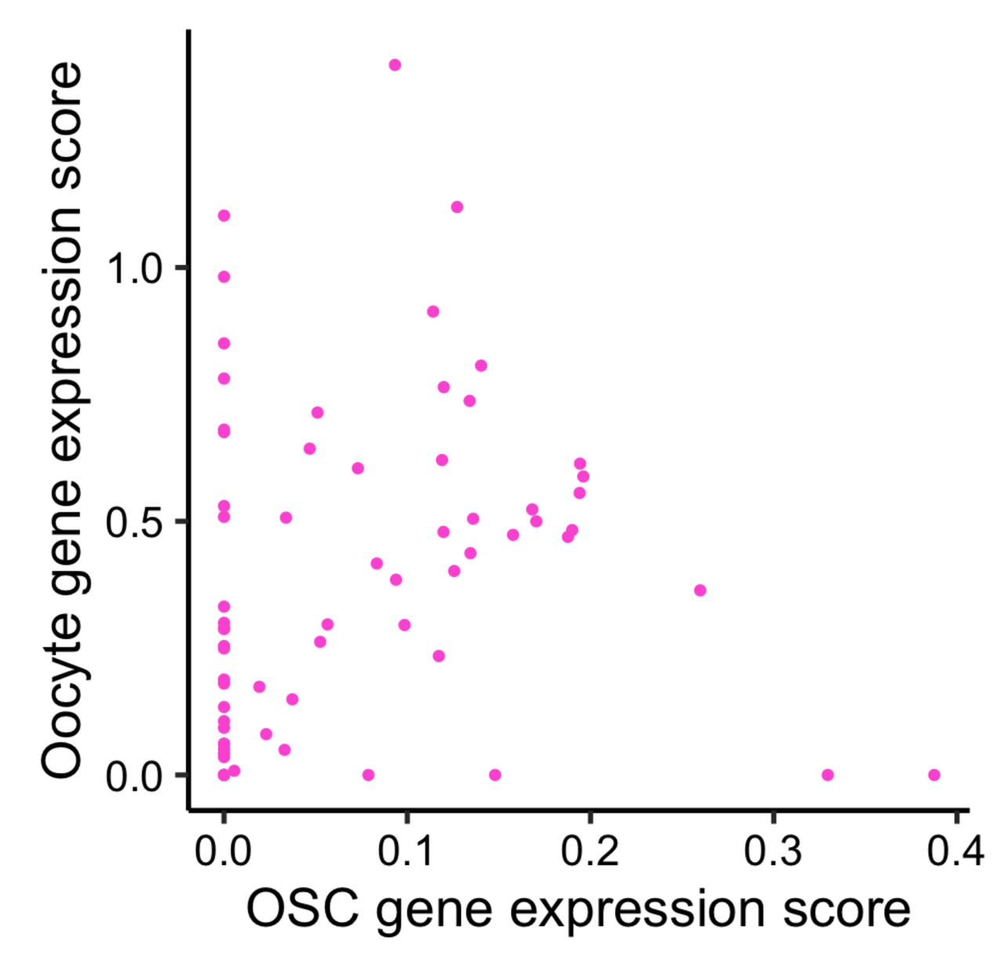
```
VlnPlot(unsorted.V3, features = "Oocyte_markers")
```




```
#Cluster 6 should be the oocyte cluster.

oocytes <- subset(unsorted.V3, idents = 6)
```

```
#Reproduce Fig.1c of Fleischmann et al.
ggplot(oocytes@meta.data, aes(x=Germ_cell_markers, y=Oocyte_markers))+geom_point(color=hue_pal()(7)[7],)+theme_classic(base_size=20)+ylab("Oocyte gene expression score")+xlab("OSC gene expression score")+theme(axis.text = element_text(color="black"), aspect.ratio = 1)
```



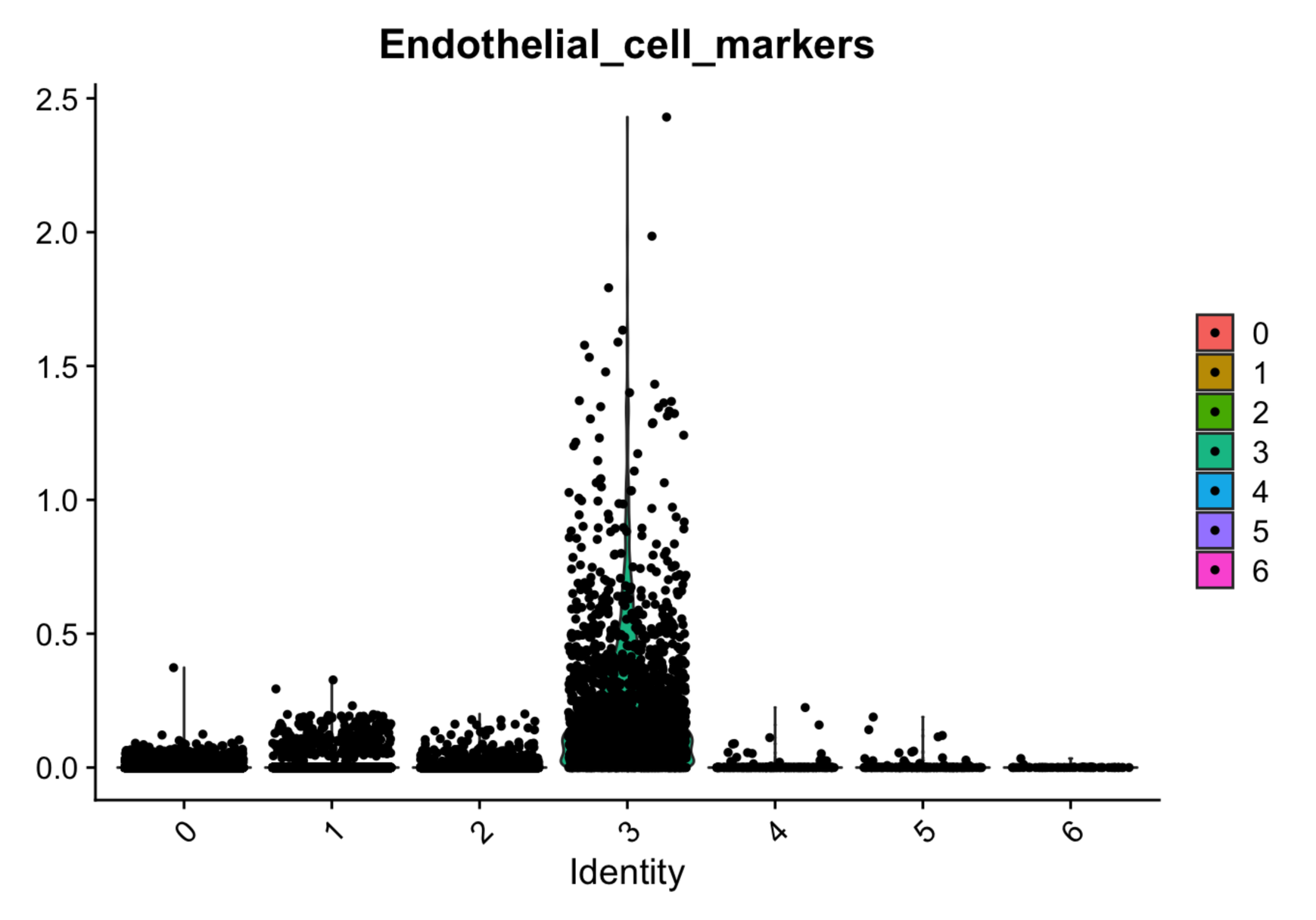

```
#'Two presumptive OSCs' should be:
oocytes@meta.data %>% dplyr::filter(Germ_cell_markers > 0.3)
```

```
##                             orig.ident nCount_RNA nFeature_RNA percent.mt
## CCTAGCTCAAGGGTCA-1_1 10XProject-062        516          425  0.7751938
## TGAGAGGAGAATCTCC-1_2 10XProject-063        607          474  2.6359143
##                      integrated_snn_res.0.1 seurat_clusters Oocyte_markers
## CCTAGCTCAAGGGTCA-1_1                      6               6              0
## TGAGAGGAGAATCTCC-1_2                      6               6              0
##                      Germ_cell_markers
## CCTAGCTCAAGGGTCA-1_1         0.3875969
## TGAGAGGAGAATCTCC-1_2         0.3294893
```

```
#Both of their UMI count are very low, which could make the 'OSC gene expression score' higher.
```

Endothelial cell marker score

```
endo <- c("CD34", "VWF", "FLI1", "CDH5")
endo_percent <- colSums(unsorted.V3@assays$RNA@counts[endo,]*100)/colSums(unsorted.V3@assays$RNA@counts)
unsorted.V3 <- AddMetaData(object = unsorted.V3, metadata = endo_percent , col.name = "Endothelial_cell_markers")
VlnPlot(unsorted.V3, features = c("Endothelial_cell_markers"))
```




```
#Cluster 3 should be the endothelial cluster.
```

Load Li et al. data

```
#Original data were downloaded from: https://github.com/zorrodong/germcell
#Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. Cell Stem Cell 20, 858–873 (2017).

#Clustering (female only)
final<-read.delim("190529Li-Run3/Li-final-cluster.txt",header = 1,sep="\t",check.names = F)
dim(final)
```

```
## [1] 992   2
```

```
#992 female annotataed samples in Li et al.

#Expression (including male samples)
Li.tpm <- read.table("190529Li-Run3/FGC_tpm_expr.txt", row.names = 1, header = T) %>% as.matrix()

#Extract only cluster-annotated expression data
Li.tpm.F <- Li.tpm[,final$sample]
dim(Li.tpm.F)
```

```
## [1] 24153   992
```

```
#992 samples

#Change the cluster names
Li.cluster <- final %>% dplyr::mutate(cluster = case_when(cluster == "Female_FGC#1" ~ "Mitotic FGCs", cluster == "Female_FGC#2" ~ "RA responsive FGCs", cluster == "Female_FGC#3" ~ "Meiotic FGCs", cluster == "Female_FGC#4" ~ "Fetal oocytes", cluster == "Female_Soma#1" ~ "Endothelial", cluster == "Female_Soma#2" ~ "Granulosa (w07-10)", cluster == "Female_Soma#3" ~ "Granulosa (w10-20)", cluster == "Female_Soma#4" ~ "Granulosa (w20-26)"))
Li.cluster <- transform(Li.cluster, cluster= factor(cluster, levels = c("Mitotic FGCs", "RA responsive FGCs", "Meiotic FGCs", "Fetal oocytes", "Endothelial", "Granulosa (w07-10)", "Granulosa (w10-20)", "Granulosa (w20-26)")))
```

Prepare SingleCellExperiment for SingleR annotation reference

```
Li.sce <- SingleCellExperiment(assays = list(normcounts = Li.tpm.F), colData = DataFrame(cluster=Li.cluster$cluster))
logcounts(Li.sce) <- log2(normcounts(Li.sce) + 1)
Li.sce
```

```
## class: SingleCellExperiment
## dim: 24153 992
## metadata(0):
## assays(2): normcounts logcounts
## rownames(24153): A1BG A1BG-AS1 ... ZZEF1 ZZZ3
## rowData names(0):
## colnames(992): F_5W_embryo1_sc1 F_5W_embryo1_sc10 ...
##   F_26W_embryo1_sc83 F_26W_embryo1_sc86
## colData names(1): cluster
## reducedDimNames(0):
## altExpNames(0):
```

1st trial: Whether endothelial cluster cells are accurately annotated as endothelial cells

```
Li.cluster %>% group_by(cluster) %>% tally()
```

```
## # A tibble: 8 x 2
##   cluster                n
##   <fct>              <int>
## 1 Mitotic FGCs         453
## 2 RA responsive FGCs   102
## 3 Meiotic FGCs         110
## 4 Fetal oocytes         46
## 5 Endothelial            8
## 6 Granulosa (w07-10)    37
## 7 Granulosa (w10-20)    71
## 8 Granulosa (w20-26)   165
```

```
#only 8 samples are annotated as Endothelial cluster in Li et al.
#Here we first try whether our endothelial cluster cells are accurately annotated as endothelial cells.
```

```
#According to the SingleR vignette, Wilcoxon ranked sum test is suggested when using single-cell references.
#https://bioconductor.org/packages/release/bioc/vignettes/SingleR/inst/doc/SingleR.html
endo <- subset(unsorted.V3, idents = 3)
endo.sce <- as.SingleCellExperiment(endo)
pred.endo.wilcox <- SingleR(test=endo.sce, ref=Li.sce, labels=Li.sce$cluster, de.method="wilcox")
table(pred.endo.wilcox$labels)[which(names(table(pred.endo.wilcox$labels))=="Endothelial")]/nrow(endo@meta.data)*100
```

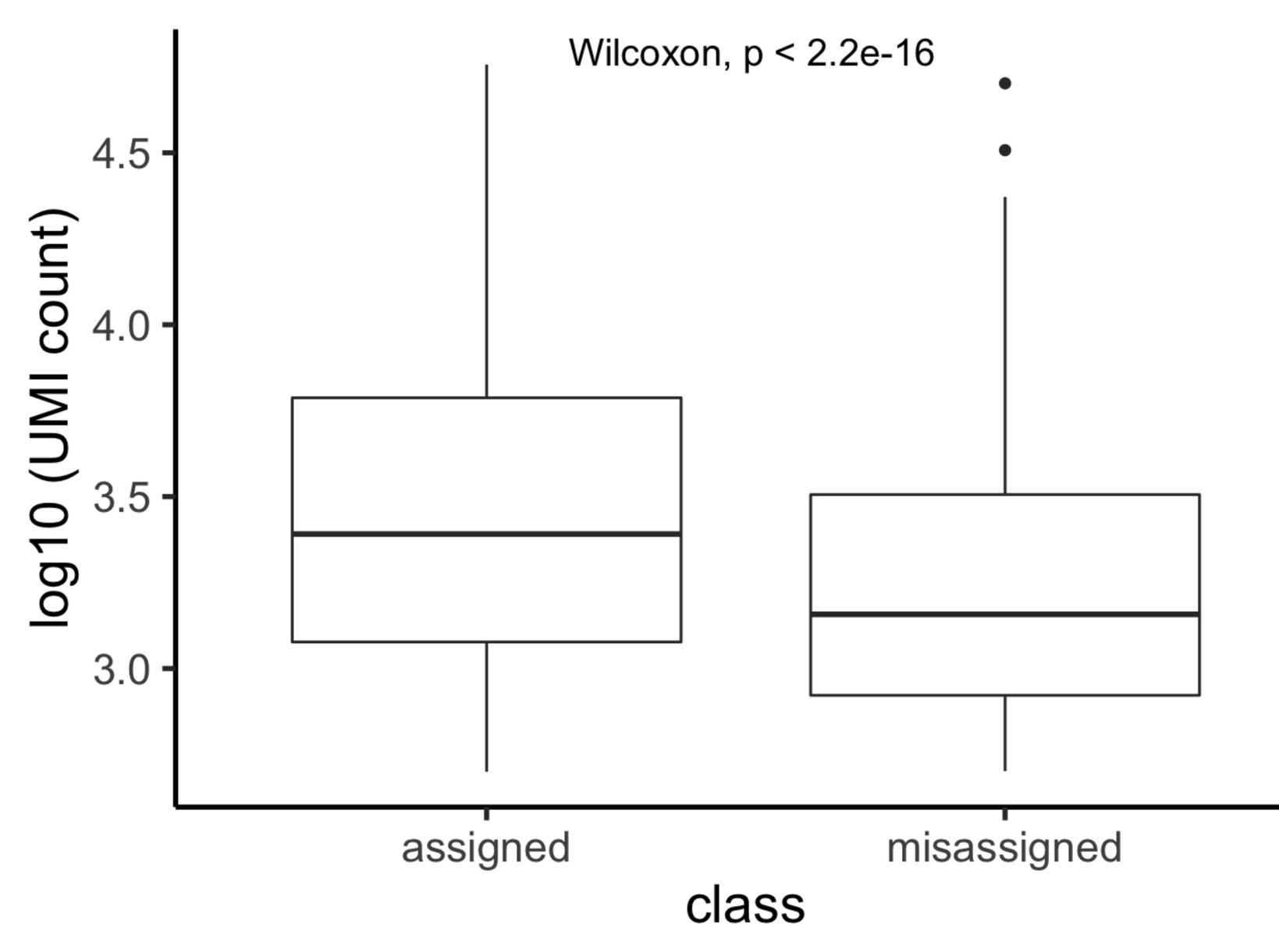
```
## Endothelial
##    56.99732
```

```
#However, only 57% cells were annotated as endothelial cells.
```

Why this happened?

```
endo.assigned <- pred.endo.wilcox %>% data.frame() %>% dplyr::filter(labels == "Endothelial")
endo.misassigned <- pred.endo.wilcox %>% data.frame() %>% dplyr::filter(labels != "Endothelial")
endo.SingleR.results <- endo@meta.data %>% dplyr::mutate(class = case_when(rownames(endo@meta.data) %in% rownames(endo.assigned) ~ "assigned", rownames(endo@meta.data) %in% rownames(endo.misassigned) ~ "misassigned"))

ggplot(endo.SingleR.results, aes(x=class, y=log10(nCount_RNA)))+geom_boxplot()+theme_classic(base_size = 20)+stat_compare_means(label.x = 1.3, size=5)+ylab("log10 (UMI count)")
```




```
#It is possible that low UMI count samples are not accurately assigned with Wilcoxon ranked sum test.
#'may fail for low-coverage data where the median is frequently zero' according to the SingleR vignette.
```

Here we use the default method “classic”.

```
pred.endo.default <- SingleR(test=endo.sce, ref=Li.sce, labels=Li.sce$cluster)
table(pred.endo.default$labels)[which(names(table(pred.endo.default$labels))=="Endothelial")]/nrow(endo@meta.data)*100
```

```
## Endothelial
##    88.09651
```

```
#Here 88% cells were annotated as endothelial cells. We decided to use the default method.
```

Oocyte cluster annotation

```
oocytes.sce <- as.SingleCellExperiment(oocytes)
pred.oocyte <- SingleR(test=oocytes.sce, ref=Li.sce, labels=Li.sce$cluster)
table(pred.oocyte$labels)
```

```
## 
##      Fetal oocytes Granulosa (w07-10) Granulosa (w10-20) Granulosa (w20-26)
##                 55                  2                  2                  8
```

Scores for ‘Two presumptive OSCs’

```
pred.oocyte["CCTAGCTCAAGGGTCA-1_1",]$scores
```

```
##      Endothelial Fetal oocytes Granulosa (w07-10) Granulosa (w10-20)
## [1,]  0.04326535     0.1313218         0.03561659         0.04361401
##      Granulosa (w20-26) Meiotic FGCs Mitotic FGCs RA responsive FGCs
## [1,]         0.03800417   0.07693175   0.06548468         0.09345143
```

```
pred.oocyte["TGAGAGGAGAATCTCC-1_2",]$scores
```

```
##      Endothelial Fetal oocytes Granulosa (w07-10) Granulosa (w10-20)
## [1,]  0.02779434      0.143503         0.03256882         0.03937629
##      Granulosa (w20-26) Meiotic FGCs Mitotic FGCs RA responsive FGCs
## [1,]         0.03008794    0.0583074   0.05814704         0.07184052
```

Heatmap

```
coldata <- data.frame(rownames(oocytes@meta.data), rep("Oocyte",67), log10(oocytes@meta.data$nCount_RNA))
colnames(coldata)<-c("cell", "Oocyte cluster", "log10 (UMI count)")
coldata <- coldata %>% dplyr::mutate(`Oocyte cluster` = case_when(cell == "CCTAGCTCAAGGGTCA-1_1" ~ "Presumptive OSCs", cell == "TGAGAGGAGAATCTCC-1_2" ~ "Presumptive OSCs", TRUE~"Other")) %>% column_to_rownames("cell")
plotScoreHeatmap(pred.oocyte, annotation_col = coldata)
```