Clustering

The codes below will detail the gene clustering process

Find differentially expressed genes

Identifies the positive and negative markers of a single cluster. min.pct requires a feature to be detected at a minimum percentage in either of the two groups. thresh.test requires a feature to be differentially expressed by some amount between the two groups.

```
cluster2.markers <-FindMarkers(eb, ident.1 = 2, min.pct = 0.25)
head(cluster2.markers, n = 5)
```

```
##               p_val avg_log2FC pct.1 pct.2     p_val_adj
## LECT1 6.067760e-176  1.1217750 0.872 0.360 1.073326e-171
## TERF1 1.890121e-164  1.1285013 0.996 0.745 3.343435e-160
## MFGE8 1.233187e-157  1.0874545 0.928 0.576 2.181385e-153
## TPI1  7.102153e-153  0.6041106 1.000 0.997 1.256300e-148
## COX7C 2.465760e-139  0.5417917 0.997 0.994 4.361683e-135
```

Find all markers differently expressed in cluster 5 from clusters 0 and 3

```
cluster5.markers <- FindMarkers(eb, ident.1 = 5, ident.2 = c(0,3),min.pct = 0.25)
head(cluster5.markers, n=5)
```

```
##                  p_val avg_log2FC pct.1 pct.2     p_val_adj
## TMSB4X   3.377654e-114  1.2064893 1.000 0.993 5.974733e-110
## EOMES    7.434328e-106  0.7632151 0.307 0.001 1.315058e-101
## TUBB2A    9.331031e-98  1.8845941 0.899 0.554  1.650566e-93
## SERPINB9  7.320498e-88  0.8033117 0.341 0.021  1.294923e-83
## FST       1.827158e-74  1.4416593 0.856 0.369  3.232060e-70
```

Find markers for every cluster compared to all the other cells, return only the positive ones (estimate 80 seconds)

```
eb.markers <- FindAllMarkers(eb, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25
)
eb.markers %>%
  group_by(cluster) %>%
  top_n(n=2, wt= avg_log2FC)
```

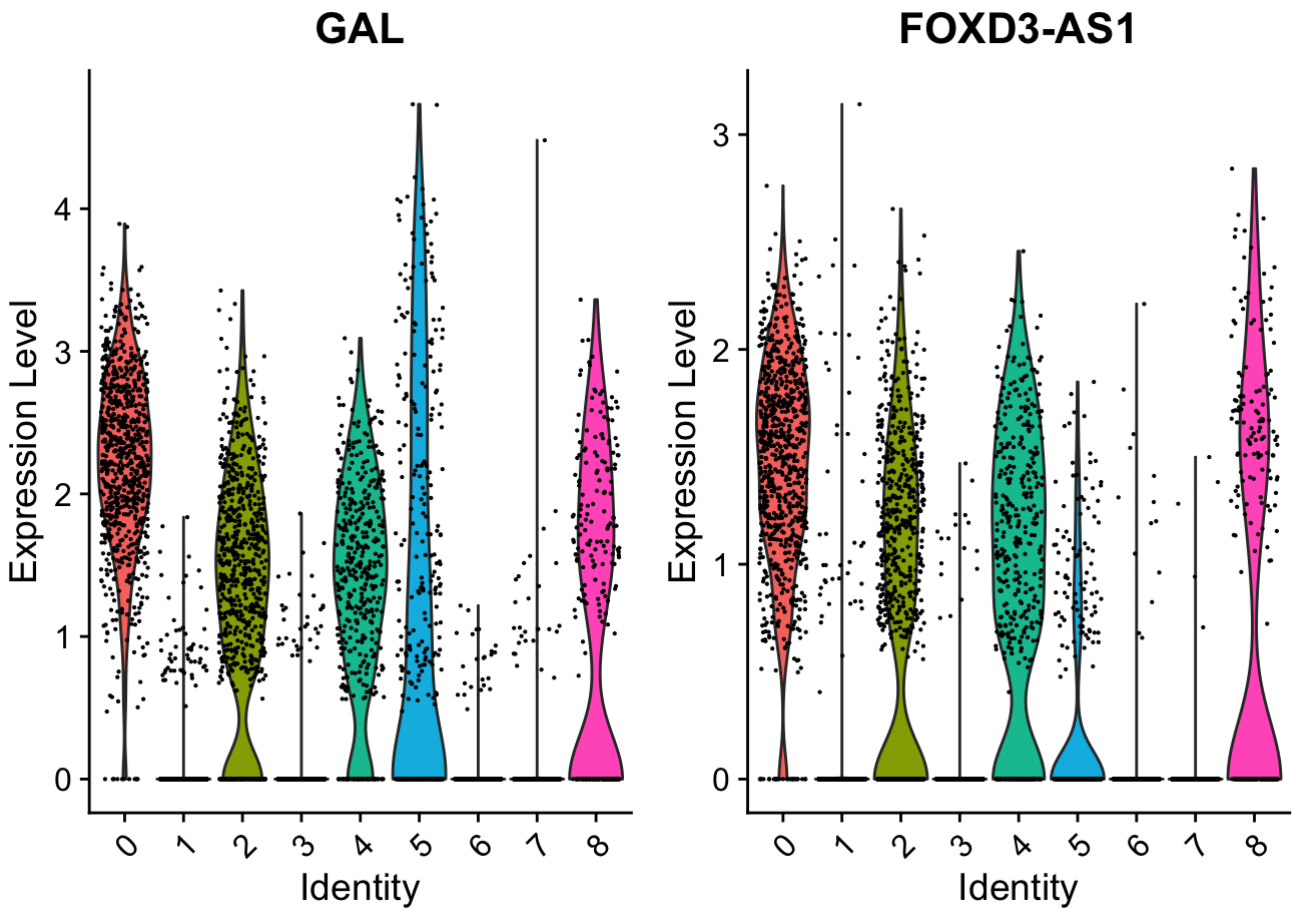
```
## # A tibble: 18 x 7
## # Groups:   cluster [9]
##        p_val avg_log2FC pct.1 pct.2 p_val_adj cluster gene
##        <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>   <chr>
##  1 0              1.72  0.986 0.387 0         0       GAL
##  2 1.31e-151      1.46  0.443 0.087 2.32e-147 0       LINC00458
##  3 2.12e-173      1.10  0.723 0.232 3.75e-169 1       FRZB
##  4 1.12e-164      1.19  0.975 0.664 1.98e-160 1       CRABP2
##  5 6.07e-176      1.12  0.872 0.36  1.07e-171 2       LECT1
##  6 1.89e-164      1.13  0.996 0.745 3.34e-160 2       TERF1
##  7 0              2.63  0.934 0.232 0         3       DLK1
##  8 1.41e-268      1.87  0.917 0.361 2.50e-264 3       TPBG
##  9 5.51e-102      1.07  0.976 0.469 9.75e- 98 4       TDGF1
## 10 2.50e- 97      0.982 0.91  0.484 4.42e- 93 4       UBE2S
## 11 6.71e- 97      1.61  0.899 0.559 1.19e- 92 5       TUBB2A
## 12 4.38e- 41      1.65  0.428 0.171 7.75e- 37 5       NTS
## 13 5.53e-138      1.97  0.99  0.637 9.78e-134 6       UBE2C
## 14 1.41e-123      1.57  0.99  0.72  2.49e-119 6       ARL6IP1
## 15 0              3.14  0.934 0.09  0         7       HAPLN1
## 16 4.96e-167      3.11  0.965 0.492 8.78e-163 7       S100A11
## 17 2.67e- 44      1.13  0.855 0.729 4.72e- 40 8       UGP2
## 18 3.36e- 36      0.865 0.934 0.907 5.94e- 32 8       PGK1
```

```
head(eb.markers, n = 6)
```

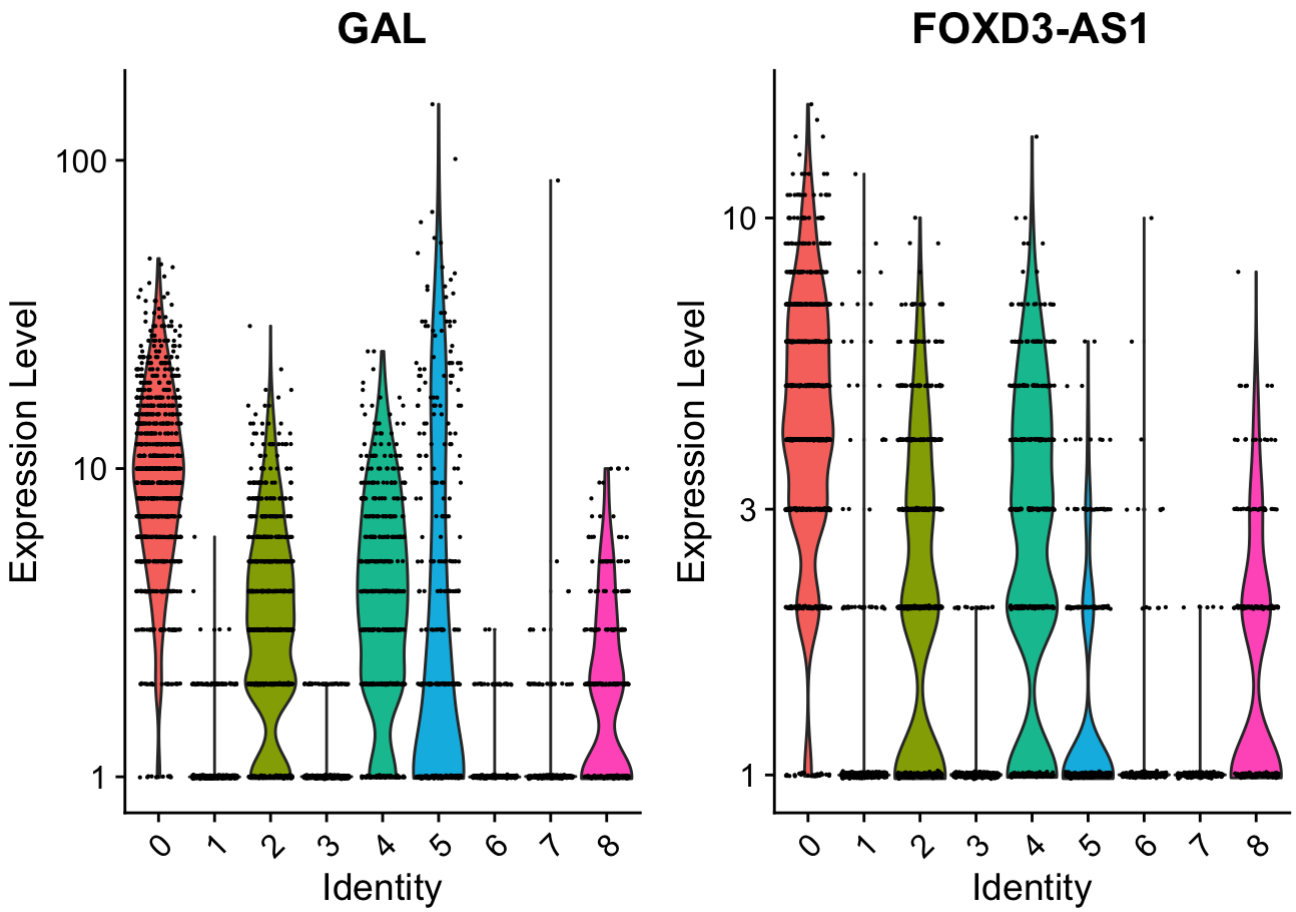
```
##                      p_val avg_log2FC pct.1 pct.2     p_val_adj cluster
## GAL           0.000000e+00  1.7222944 0.986 0.387  0.000000e+00       0
## FOXD3-AS1     0.000000e+00  1.3519286 0.966 0.288  0.000000e+00       0
## D21S2088E    3.075692e-268  0.8361288 0.620 0.089 5.440591e-264       0
## THY1         1.689104e-234  1.0609575 0.831 0.251 2.987857e-230       0
## RPS26        3.065307e-219  0.7175786 0.999 0.986 5.422221e-215       0
## RP11-132A1.3 6.543218e-216  0.8542234 0.679 0.151 1.157430e-211       0
##                        gene
## GAL                     GAL
## FOXD3-AS1         FOXD3-AS1
## D21S2088E         D21S2088E
## THY1                   THY1
## RPS26                 RPS26
## RP11-132A1.3 RP11-132A1.3
```

##Visualize marker expression Violin plots

```
VlnPlot(eb, features = c("GAL","FOXD3-AS1"))
```

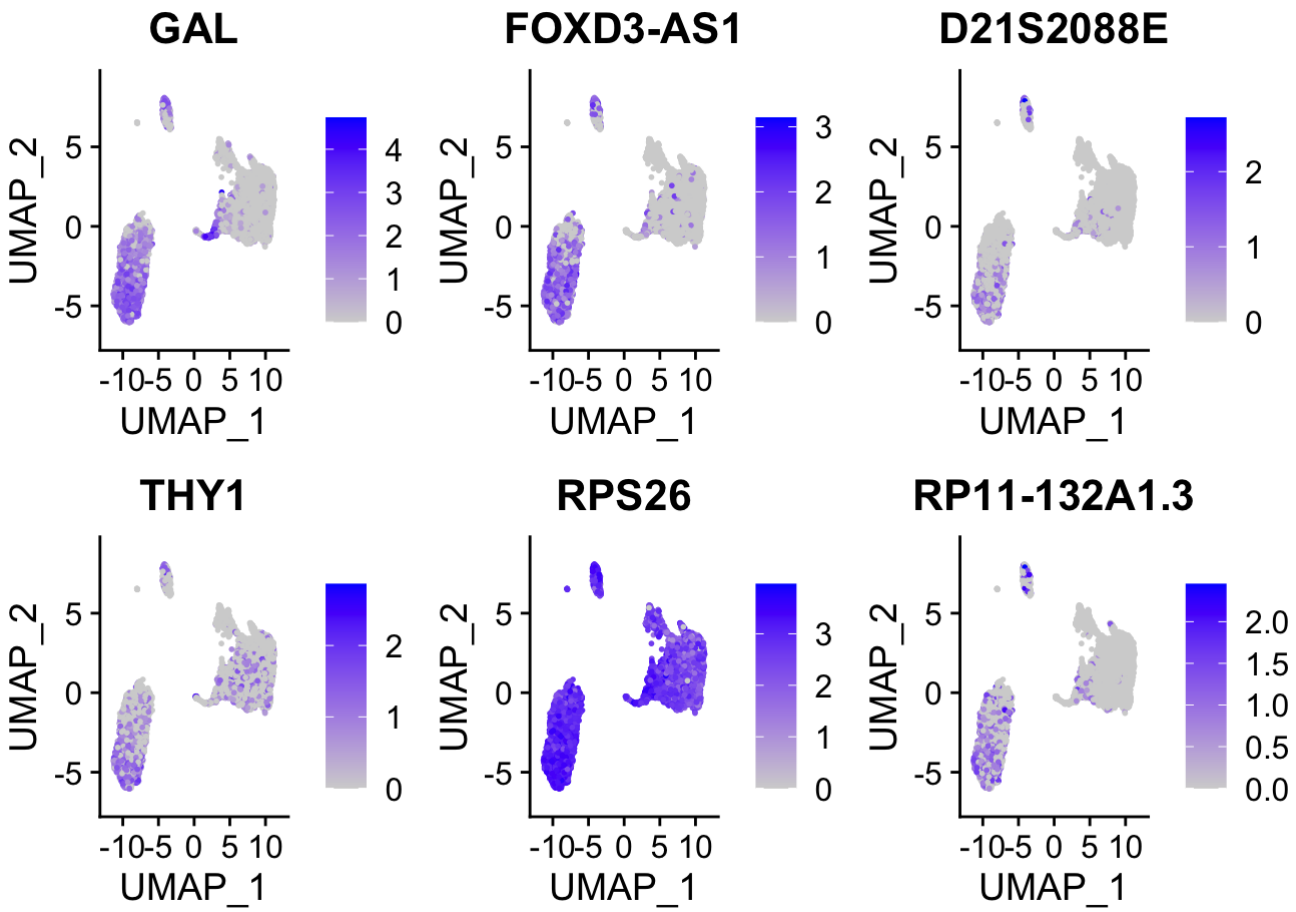



```
#plot raw counts
VlnPlot(eb, features = c("GAL","FOXD3-AS1"), slot = "counts", log = TRUE)
```



Feature plots

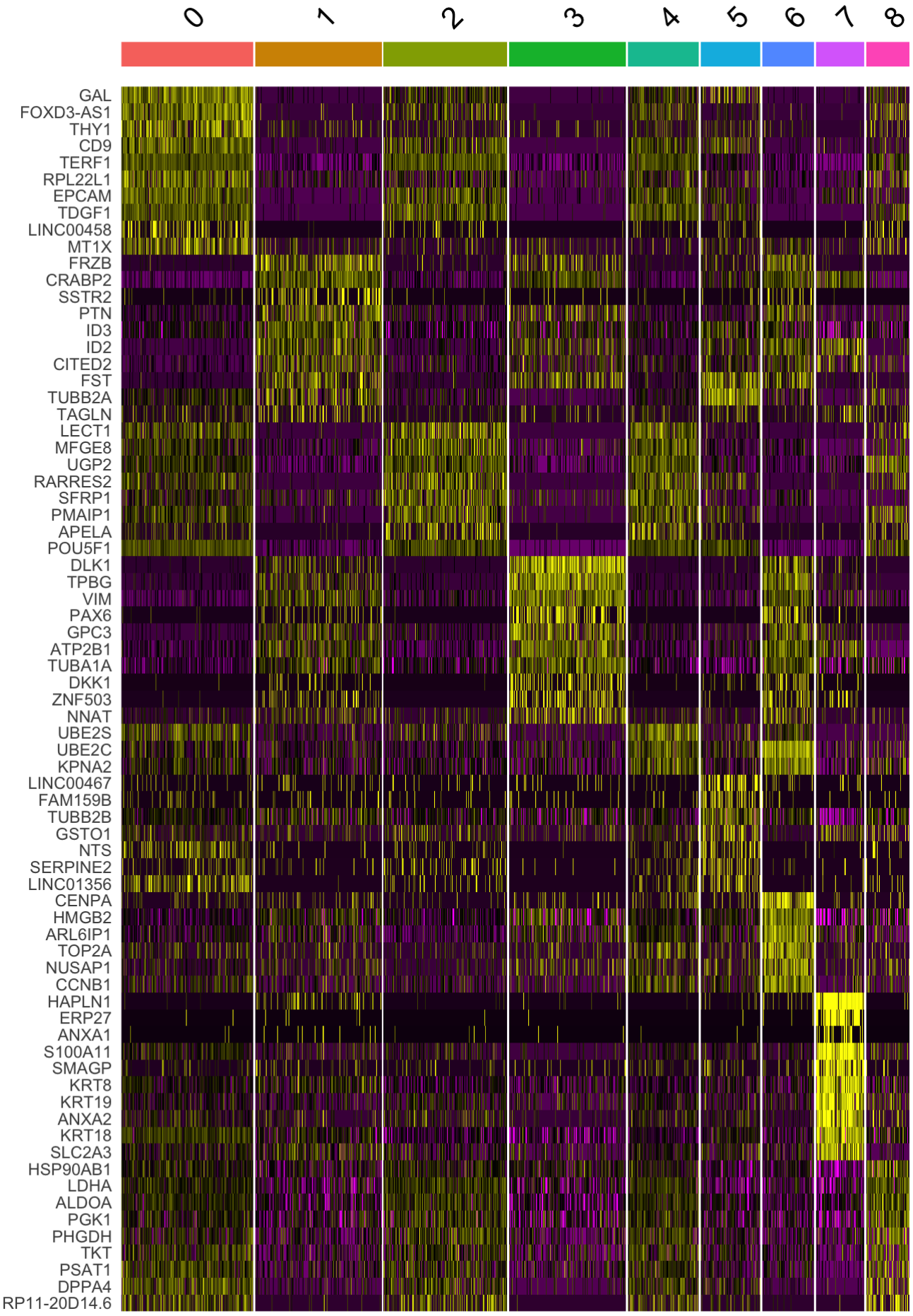
```
FeaturePlot(eb, features = c("GAL","FOXD3-AS1","D21S2088E","THY1","RPS26","RP11-132A1.3"
), ncol = 3)
```



Plot

top 20 markers on Heatmaps

```
eb.markers %>%
  group_by(cluster) %>%
  top_n(n = 10, wt = avg_log2FC) -> top10
DoHeatmap(eb, features = top10$gene) + NoLegend()
```


0
1
2
3
4
5
6
7
8
GAL
FOXD3-AS1
THY1
CD9
TERF1
RPL22L1
EPCAM
TDGF1
LINC00458
MT1X
FRZB
CRABP2
SSTR2
PTN
ID3
ID2
CITED2
FST
TUBB2A
TAGLN
LECT1
MFGE8
UGP2
RARRES2
SFRP1
PMAIP1
APELA
POU5F1
DLK1
TPBG
VIM
PAX6
GPC3
ATP2B1
TUBA1A
DKK1
ZNF503
NNAT
UBE2S
UBE2C
KPNA2
LINC00467
FAM159B
TUBB2B
GSTO1
NTS
SERPINE2
LINC01356
CENPA
HMGB2
ARL6IP1
TOP2A
NUSAP1
CCNB1
HAPLN1
ERP27
ANXA1
S100A11
SMAGP
KRT8
KRT19
ANXA2
KRT18
SLC2A3
HSP90AB1
LDHA
ALDOA
PGK1
PHGDH
TKT
PSAT1
DPPA4
RP11-20D14.6

Assign embryoid body types (currently under investigation)

```
# new.cluster.ids <- c("hemangioblast","cardiac","epicardial precursors","smooth muscle
 precursors","cardiac precursors","neuronal subtypes")
# names(new.cluster.ids) <- levels(eb)
# eb <- RenameIdents(eb, new.cluster.ids)
# DimPlot(eb, reduction = "umap", label = TRUE, pt.size = 0.5) + NoLegend()
```

Save the rds file