sc-RNA-Seq

Lia Qiu

2023-02-13

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

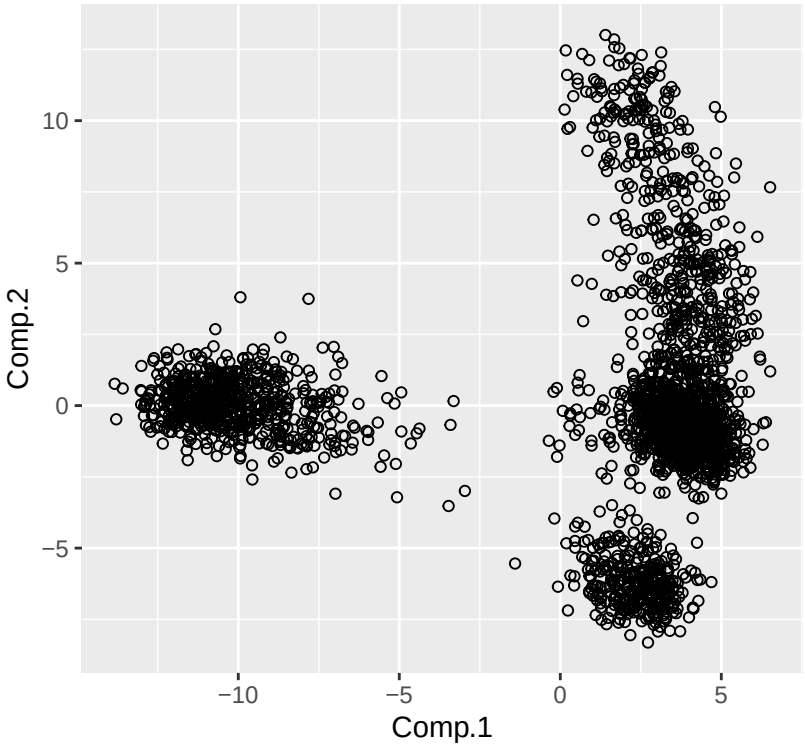
```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data = read.csv("https://raw.githubusercontent.com/gval38/PBMC_dataset/main/pbmcs.txt", sep = "\t", row
data = log(data+1)
```

PCA

```
pca=princomp(data)
p = as.data.frame(pca$scores[,1:2])
ggplot(p, aes(x=Comp.1, y=Comp.2)) + geom_point(shape=1)
```

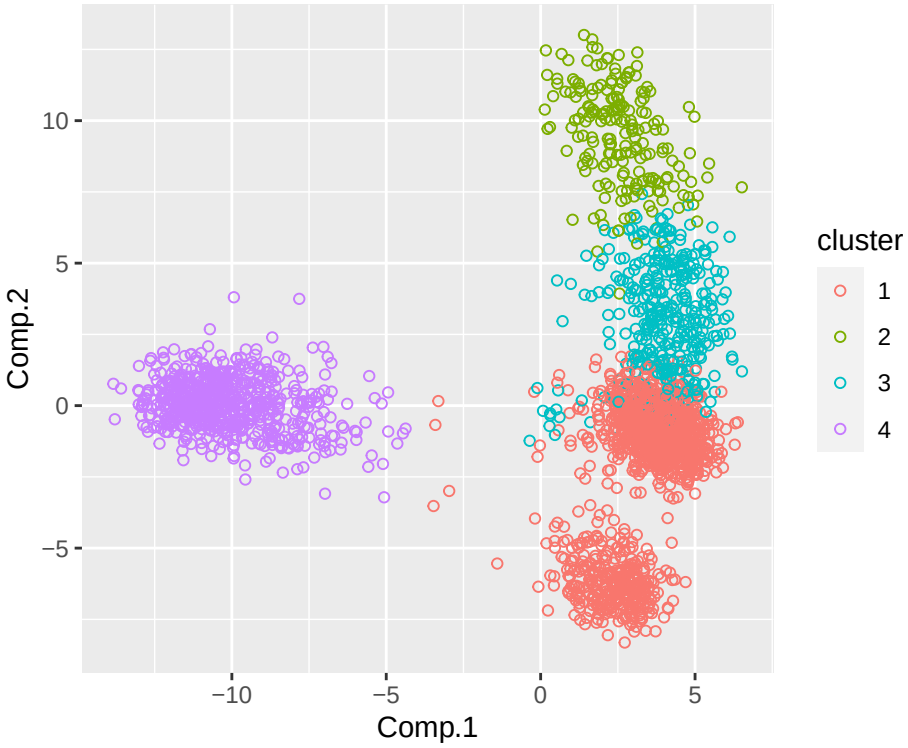


colored by kmeans

```
p$cell = rownames(p)
p = as_tibble(p)

km4 = kmeans(pca$scores, 4)
p = p %>% mutate(cluster=as.factor(km4$cluster[cell]))

p %>% ggplot(aes(x=Comp.1, y=Comp.2, color = cluster)) +
  geom_point(shape=1)
```



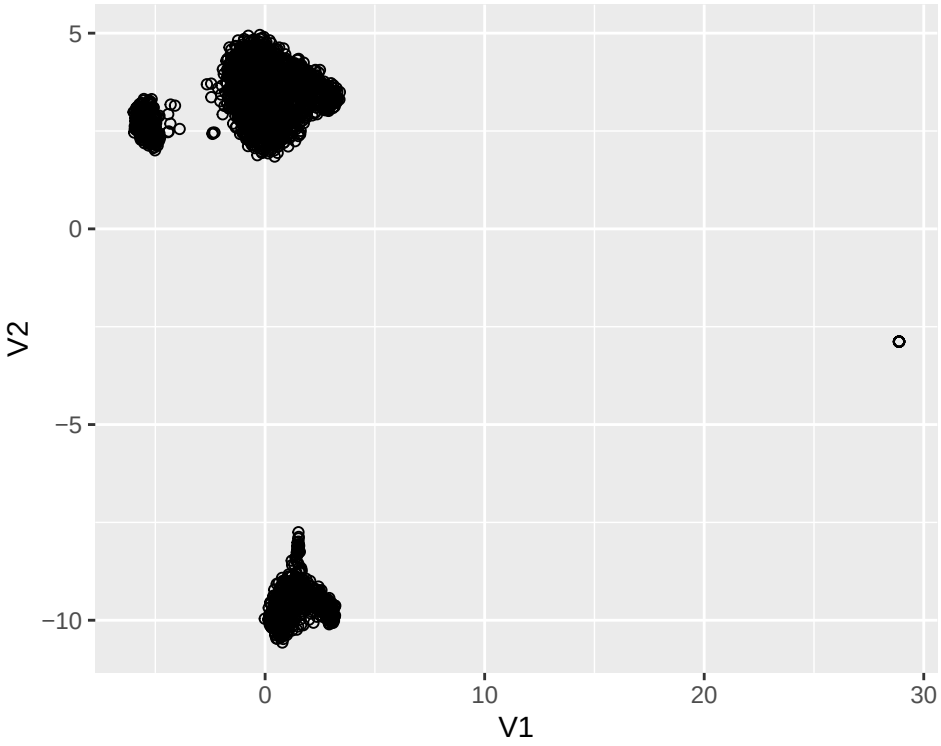
UMAP

```
library(umap)
```

```
## Warning: package 'umap' was built under R version 4.1.3
```

```
ump <- umap(data)

u = as.data.frame(ump$layout[,1:2])
ggplot(u, aes(x=V1, y=V2)) + geom_point(shape=1)
```

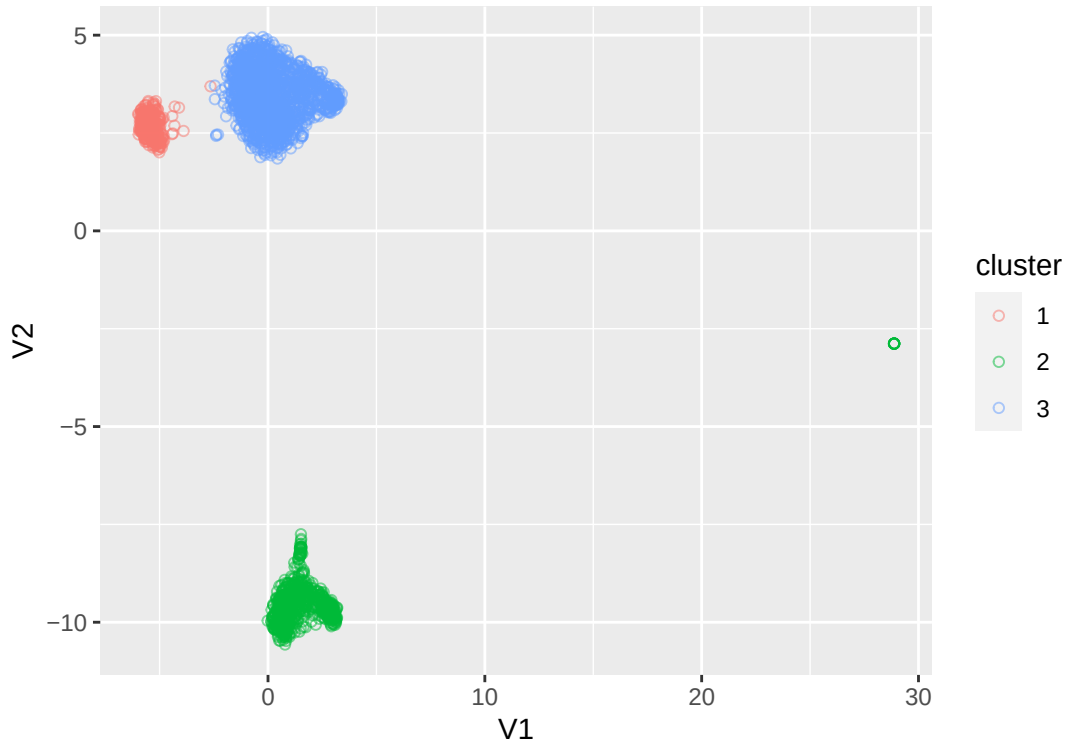


colored by kmeans

```
u$cell = rownames(u)
u = as_tibble(u)

km3 = kmeans(ump$layout, 3)
u = u %>% mutate(cluster=as.factor(km3$cluster[cell]))

ggplot(u, aes(x=V1, y=V2, color=cluster)) +
  geom_point(shape=1, alpha=0.5)
```



heatmap

```
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.1.3
```

```
library(gplots) #colorpanel
```

```
## Warning: package 'gplots' was built under R version 4.1.3
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

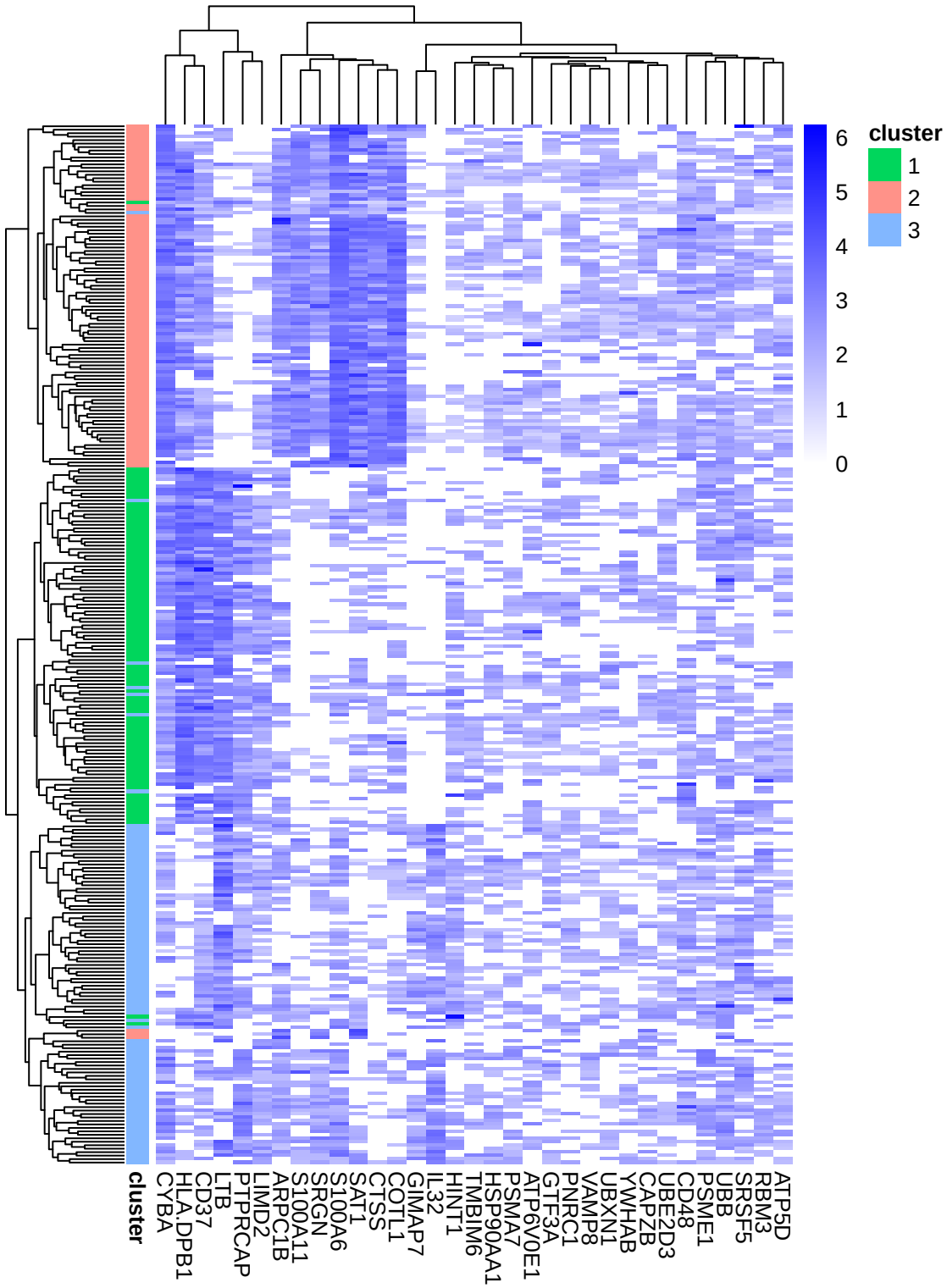
```
# screen genes by variability
mad.gene <- apply(data, 2, mad)
#hist(mad.gene[which(mad.gene>0)], breaks = 10)
sigGenes <- names(mad.gene[which(mad.gene>0)])

# pick 100 random cells from each group
cells.g1 <- u %>% filter(cluster == 1) %>% pull(cell) %>% sample(100)
cells.g2 <- u %>% filter(cluster == 2) %>% pull(cell) %>% sample(100)
cells.g3 <- u %>% filter(cluster == 3) %>% pull(cell) %>% sample(100)

var.data <- data[c(cells.g1, cells.g2, cells.g3),sigGenes]

cell.grp = as.data.frame(km3$cluster[rownames(var.data)])
colnames(cell.grp)[1] = "cluster"
```

```
cell.grp = cell.grp %>% mutate(cluster = as.character(cluster))
p <- pheatmap(var.data, col=colorpanel(360, "white","blue"),  annotation_row = cell.grp, show_rownames=F
```

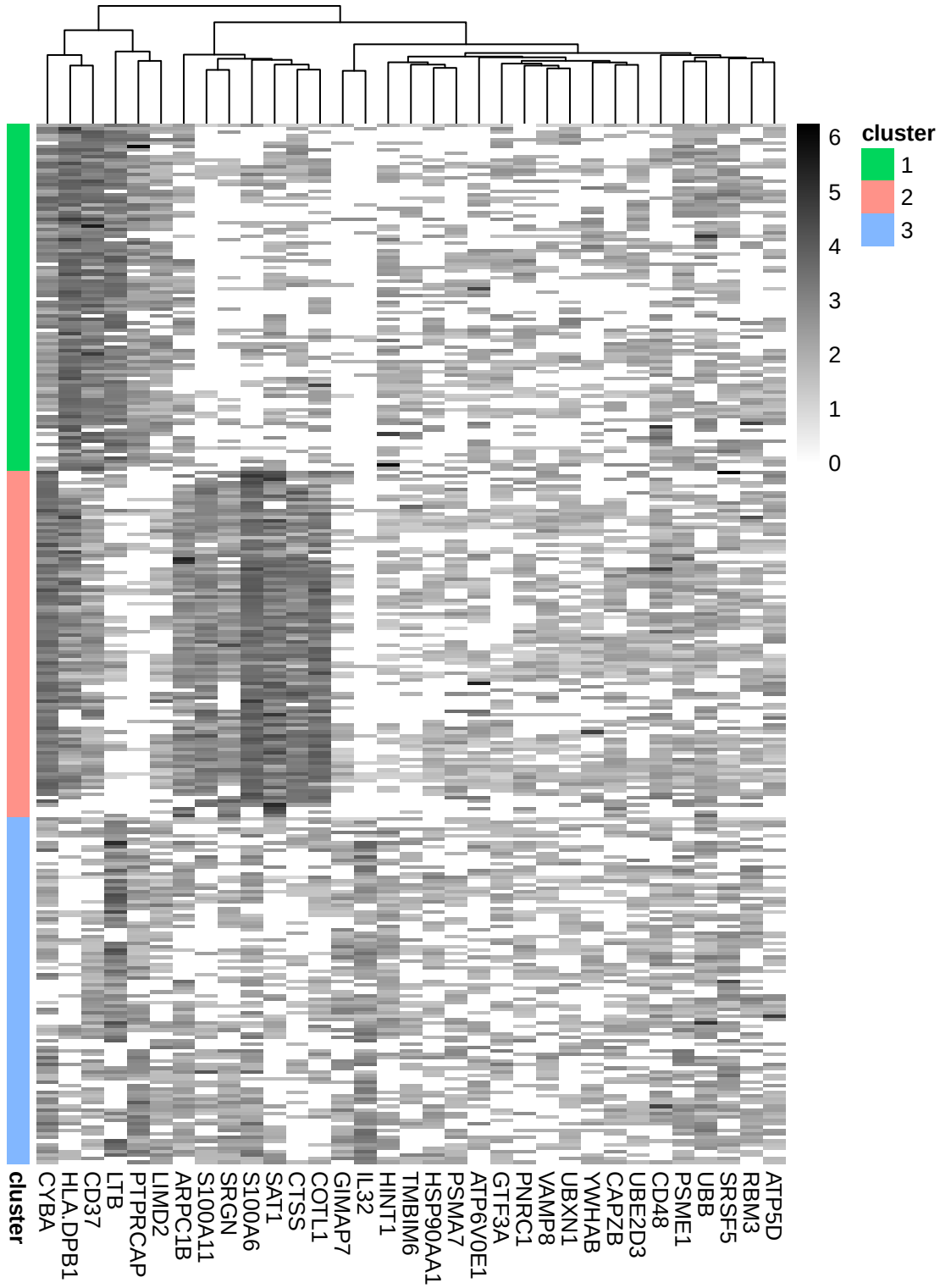


order cell by UMAP group

```
var.data.order = var.data[rownames(var.data)[p$tree_row[["order"]]],]
var.data.order$cluster = cell.grp[rownames(var.data.order),1]

var.data.order = var.data.order[order(var.data.order$cluster),]
var.data.order = var.data.order[,-ncol(var.data.order)]

p.order = pheatmap(var.data.order, col=colorpanel(360, "white","black"), cluster_rows=F, annotation_row
```



choose sig genes by tree

```
sigGene = colnames(var.data.order)[p.order$tree_col[["order"]]][0:15]
```

```
data.sig = data[,sigGene]
data.sig$cell = rownames(data.sig)
data.sig = as_tibble(data.sig)

x = u %>% left_join(data.sig, by="cell") %>% select(-cluster)
long = x %>% pivot_longer(4:(ncol(x)), names_to = "gene", values_to = "log.FPKM")

long %>% ggplot(aes(x=V1, y=V2, color=log.FPKM)) +
  geom_point(shape=1) +
  facet_wrap(~gene, nrow=3) +
  scale_colour_gradientn(colours=c("#eeeeee","black")) +
  theme_bw() +
  theme(legend.key.size = unit(0.5, "cm"))
```