NK cell identification from GBM data

Code

The GBM data set is downloaded from GEO (GSE84465), including 3587 cells from four primary GBM patients. Darmanis S, Sloan SA, Croote D, Mignardi M et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. Cell Rep 2017 Oct 31;21(5):1399-1410. PMID: 29091775

Hide

```
GBM=read.csv('GSE84465_GBM_All_data.csv',sep="")
```

Hide

```
dim(GBM)
```

```
[1] 23465  3589
```

Run Seurat to identify cluster enriched NK cell

Hide

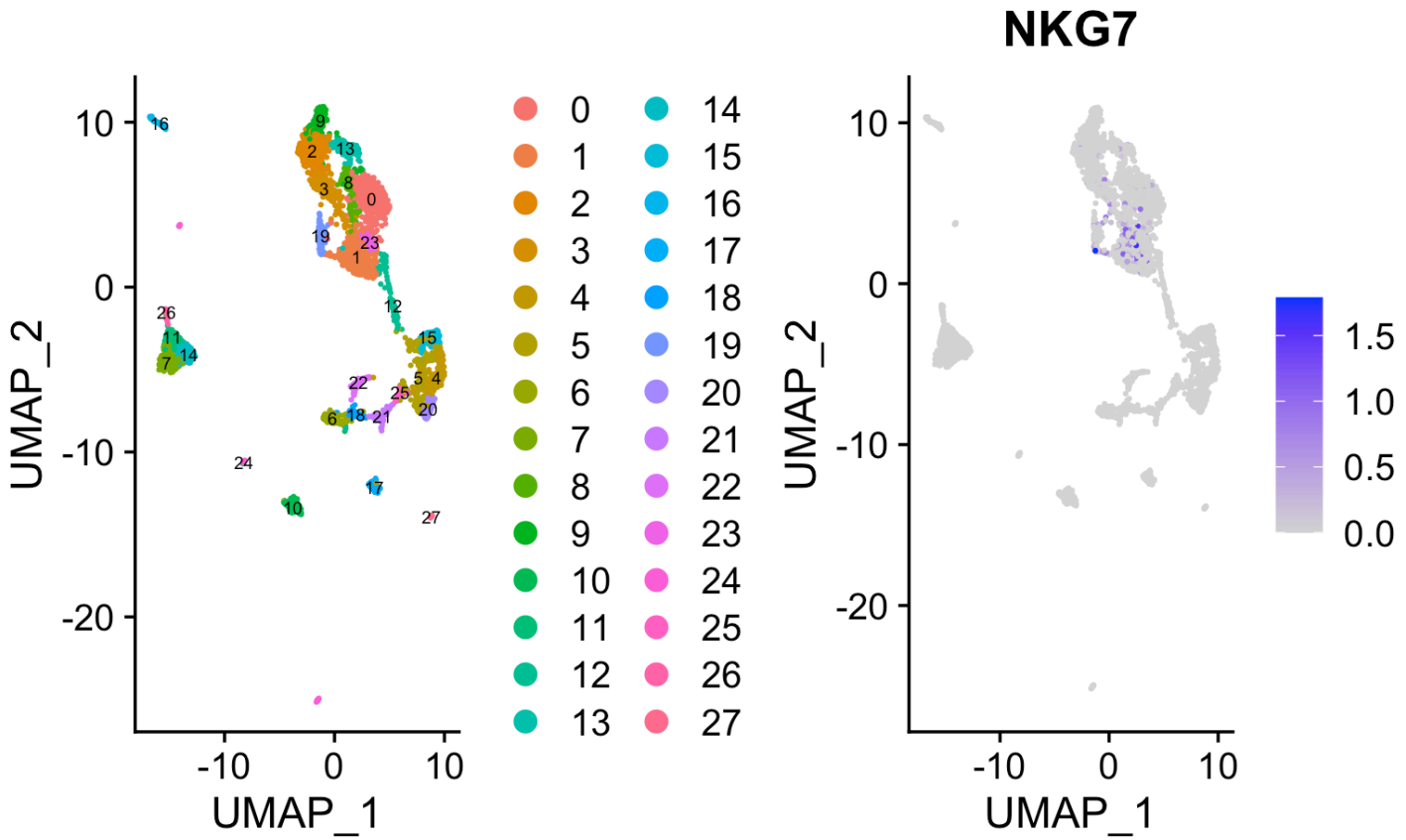
```
library(dplyr)
library(Seurat)
library(cowplot)
library(ggplot2)
```

Hide

```
GBMobj <- CreateSeuratObject(counts = GBM, project = "GBM")
GBMobj <- NormalizeData(object = GBMobj, verbose = FALSE)
GBMobj <- FindVariableFeatures(object = GBMobj, selection.method = "vst", nfeatures =
2000)
GBMobj <- ScaleData(GBMobj)
GBMobj <- RunPCA(GBMobj)
GBMobj <- FindNeighbors(GBMobj, dims = 1:25)
GBMobj <- FindClusters(GBMobj, resolution = 1.7)
GBMobj <- RunUMAP(GBMobj, dims = 1:25)
```

Hide

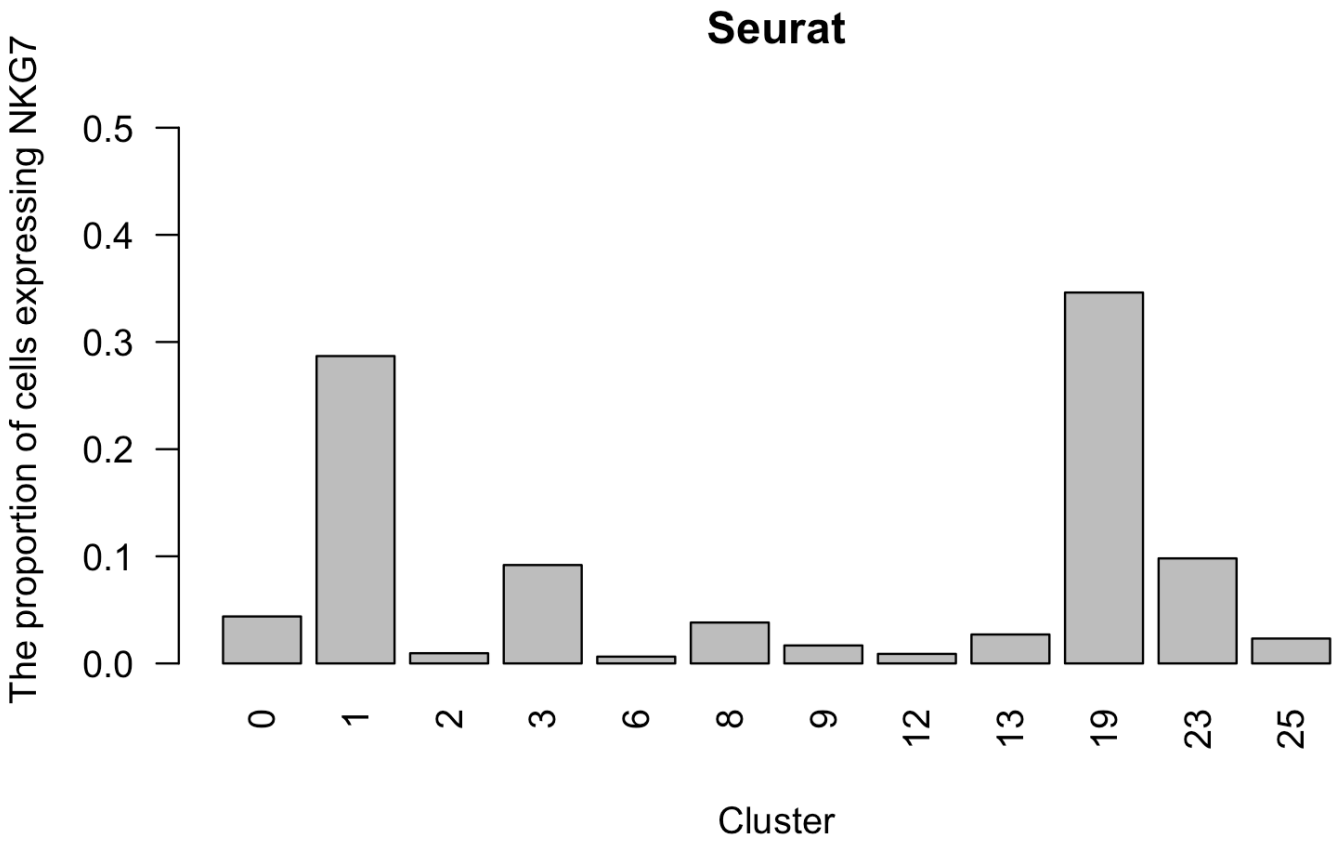
```
p1 <- DimPlot(GBMobj,label = TRUE,pt.size = 0.2,label.size=2)
p2 <- FeaturePlot(GBMobj, features = c("NKG7"),pt.size=0.2)
p1+p2
```



The proportion of cells expressing NK cell marker NKG7 in each cluster

Hide

```
subGBM=subset(GBMobj,NKG7>0)
index=match(names(table(Idents(subGBM))),names(table(Idents(GBMobj))))
seurat=table(Idents(subGBM))/table(Idents(GBMobj))[index]
barplot(seurat,ylim = c(0,0.55),las=2,main='Seurat',xlab="Cluster",ylab="The proporti
on of cells expressing NKG7")
```



Integration of Seurat and SCMarker

Hide

```
library(SCMarker)
GBM1=log(GBM+1)
scMarker.res=ModalFilter(data=GBM1,geneK=10,cellK=10,width=2)# default width = 1 for
UMI data, width =2 for TPM data.
scMarker.res=GeneFilter(obj=scMarker.res)
scMarker.res=getMarker(obj=scMarker.res,k=300,n=30)
```

Extract markers identified by SCMarker

Hide

```
scmarker=scMarker.res$marker
length(scmarker)
```

```
[1] 840
```

Hide

```
head(scmarker)
```

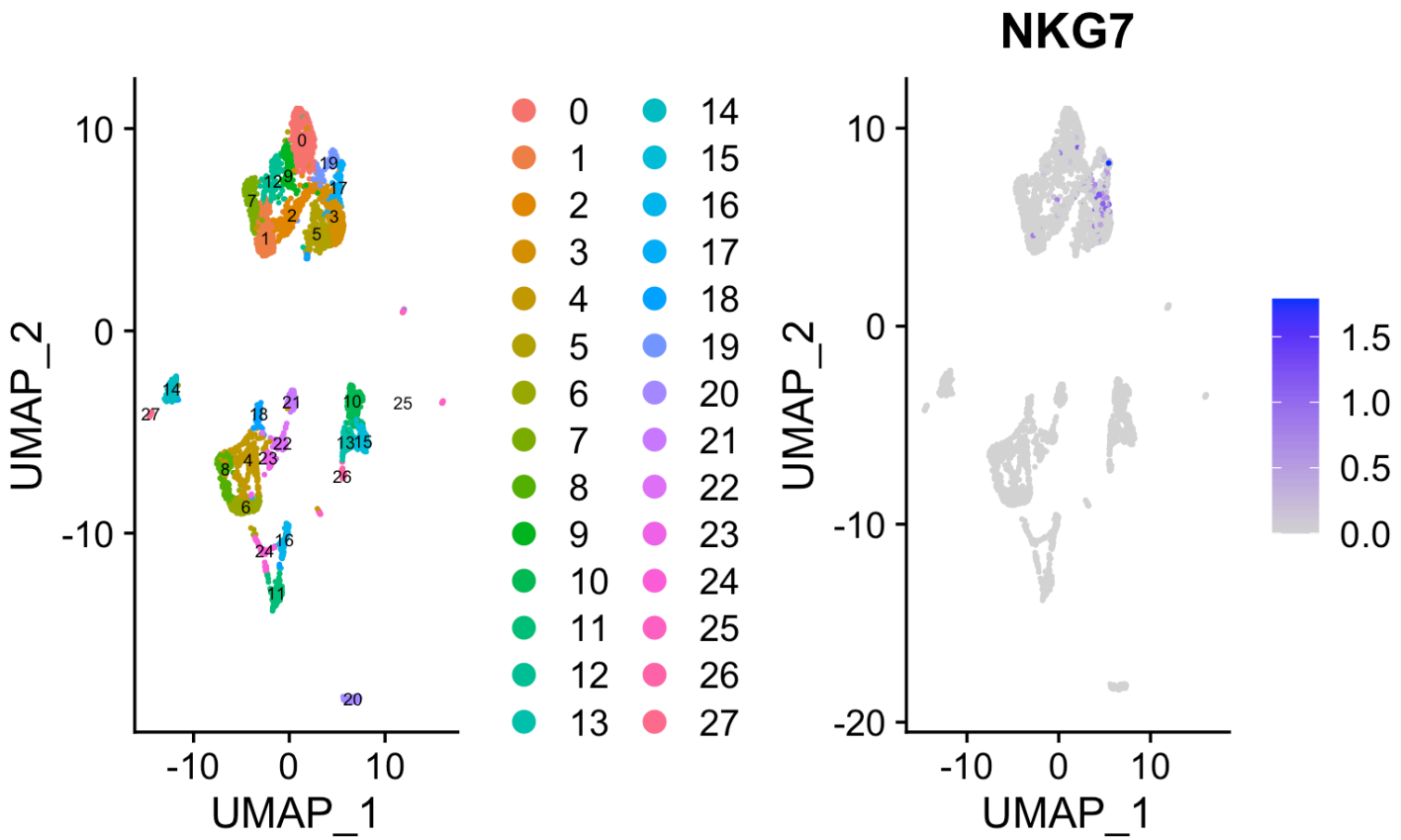
```
[1] "ABL2"   "ACSL1"  "ADORA3" "AIF1"   "AKR1B1" "ALOX5"
```

Hide

```
all.genes <- rownames(GBMobj)
GBMscmarker <- ScaleData(GBMobj,features = all.genes)
###Run PCA based on union of markers from SCMarker and Seurat
GBMscmarker <- RunPCA(GBMscmarker, features = union(scmarker,VariableFeatures(GBMobj)
))
GBMscmarker <- FindNeighbors(GBMscmarker, dims = 1:25)
GBMscmarker <- FindClusters(GBMscmarker, resolution = 1.7)
GBMscmarker <- RunUMAP(GBMscmarker, dims = 1:25)
```

Hide

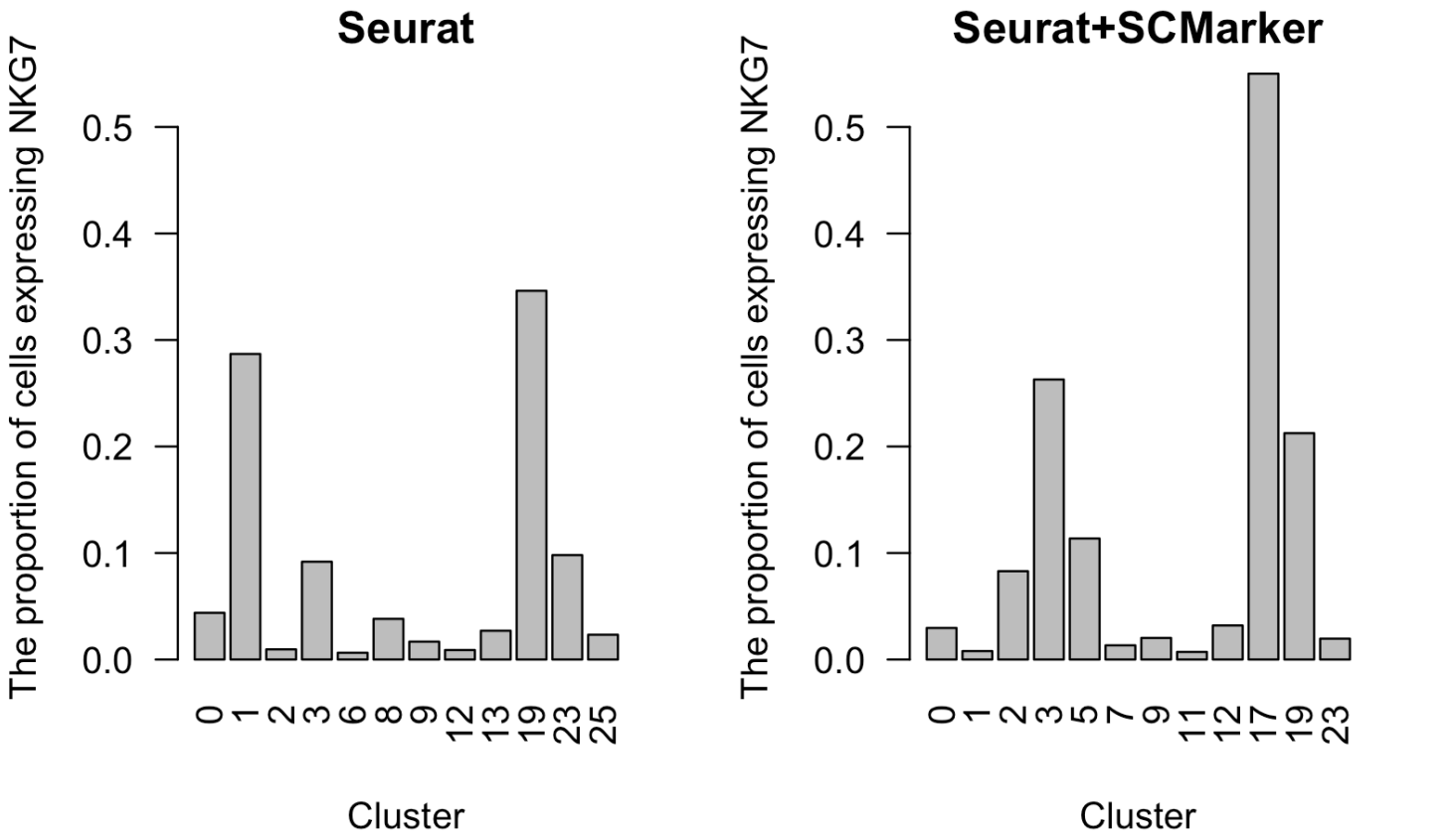
```
p1 <- DimPlot(GBMscmarker,label = TRUE,pt.size = 0.2,label.size=2)
p2 <- FeaturePlot(GBMscmarker, features = c("NKG7"),pt.size=0.2)
p1+p2
```



Hide

Hide

```
subGBMmarker=subset(GBMscmarker,NKG7>0)
index=match(names(table(Idents(subGBMmarker))),names(table(Idents(GBMscmarker))))
scmarkercluster=table(Idents(subGBMmarker))/table(Idents(GBMscmarker))[index]
par(mfrow=c(1,2))
barplot(seurat,ylim = c(0,0.55),las=2,main='Seurat',xlab="Cluster",ylab="The proporti
on of cells expressing NKG7")
barplot(scmarkercluster,ylim = c(0,0.55),las=2,main='Seurat+SCMarker',xlab="Cluster",
ylab="The proportion of cells expressing NKG7")
```



FindMarker for each cluster

Hide

```
cluster=unique(Idents(GBMscmarker))
cluster=as.numeric(as.character(cluster[order(cluster)]))
gene=c()
for (i in cluster){
  cluster1.markers <- FindMarkers(GBMscmarker, only.pos = TRUE,ident.1 = i, min.pct =
0.25)
  if (length(grep("ERCC-",row.names(cluster1.markers)))!=0){
    cluster1.markers=cluster1.markers[-grep("ERCC-",row.names(cluster1.markers)),]
  }
  cluster1.markers=cluster1.markers[order(cluster1.markers$p_val_adj),]
  gene=union(gene[!is.na(gene)],row.names(cluster1.markers)[1:6])
}
```

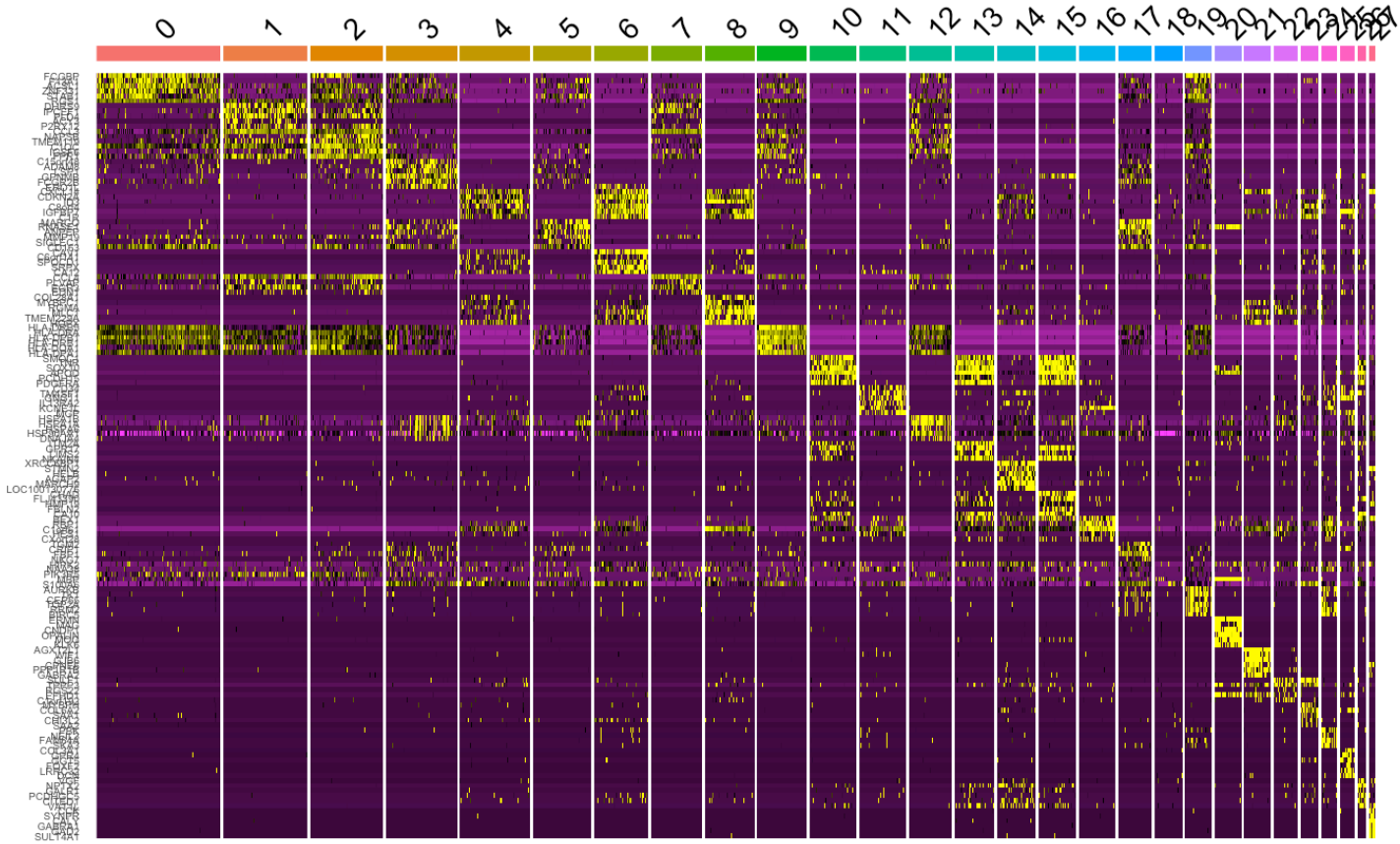
Hide

```
"NKG7" %in% gene
```

```
[1] TRUE
```

Hide

```
DoHeatmap(GBMscmarker, features = gene,size=3) + NoLegend()+ theme(axis.text.y = elem
ent_text(size = 3))
```


0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27