

Stage 2 Clustering

Sarthak Sharma

May 7, 2018

Load the data

Load the data from Stage 1 of clustering

```
cb.stage1 <- readRDS("~/cb_filtered.rds")
```

Subset the data

Since we were interested in the neural-specific cells, we used a subset of the original data for further analysis. Specifically, we used clusters 3,5,8 and 9 from the stage 1 of clustering.

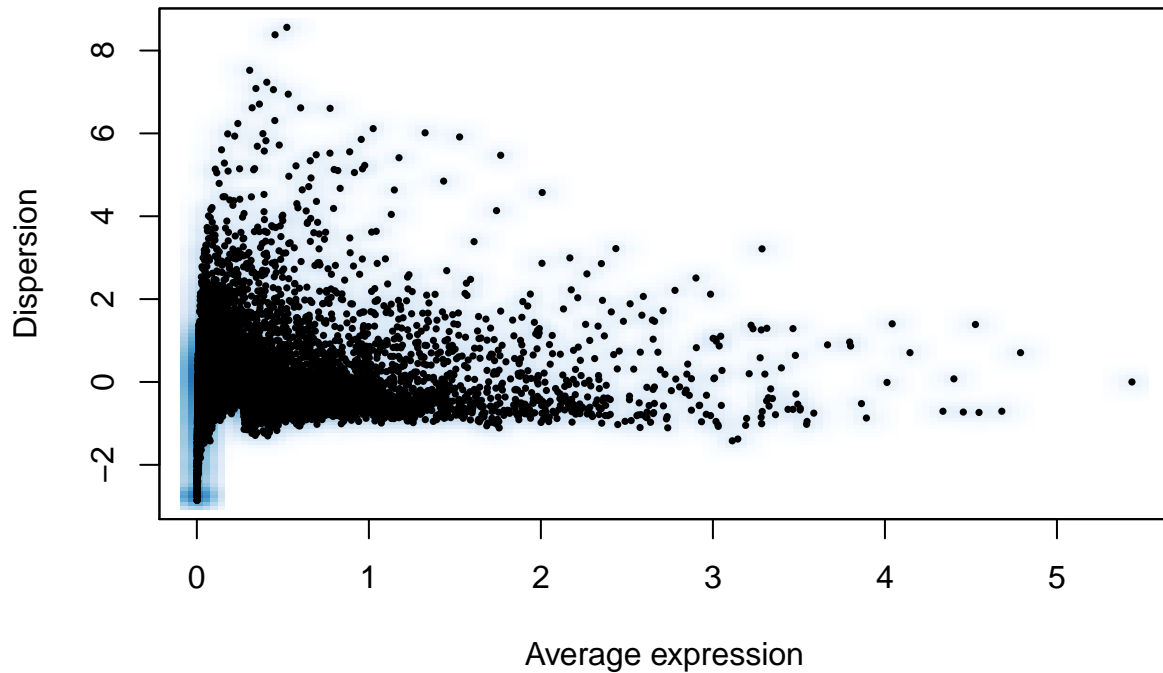
```
cb.neural <- SubsetData(cb.stage1, cells.use = WhichCells(cb.stage1, ident = c(3,5,8,9)))
```

Remove previous identity information

```
cb.neural <- SetAllIdent(cb.neural, id = NULL)
```

Identification of Variable Genes

```
cb.neural <- FindVariableGenes(object = cb.neural, x.low.cutoff = 0.0125,  
                              x.high.cutoff = 3, y.cutoff = 0.5, do.plot = F)
```



Dimensional Reduction Analysis

```
cb.neural <- ScaleData(object = cb.neural, vars.to.regress = c("nUMI"))

cb.neural <- RunPCA(cb.neural, pc.genes = cb.neural@var.genes, do.print = T, pcs.print = 1:5, genes.print = 1:5)

## [1] "PC1"
## [1] "KH2013:KH.C8.42" "KH2013:KH.C12.16" "KH2013:KH.C4.437"
## [4] "KH2013:KH.C9.478" "KH2013:KH.L164.21"
## [1] ""
## [1] "KH2013:KH.C8.121" "KH2013:KH.C9.543" "KH2013:KH.C1.212"
## [4] "KH2013:KH.C1.133" "KH2013:KH.L73.2"
## [1] ""
## [1] ""
## [1] "PC2"
## [1] "KH2013:KH.C8.332" "KH2013:KH.S643.1" "KH2013:KH.C7.461"
## [4] "KH2013:KH.S483.1" "KH2013:KH.S423.1"
## [1] ""
## [1] "KH2013:KH.C1.175" "KH2013:KH.C1.188" "KH2013:KH.C3.781"
## [4] "KH2013:KH.C2.251" "KH2013:KH.L132.16"
## [1] ""
## [1] ""
## [1] "PC3"
## [1] "KH2013:KH.C3.97" "KH2013:KH.C7.64" "KH2013:KH.C13.49"
## [4] "KH2013:KH.L116.39" "KH2013:KH.C10.200"
```

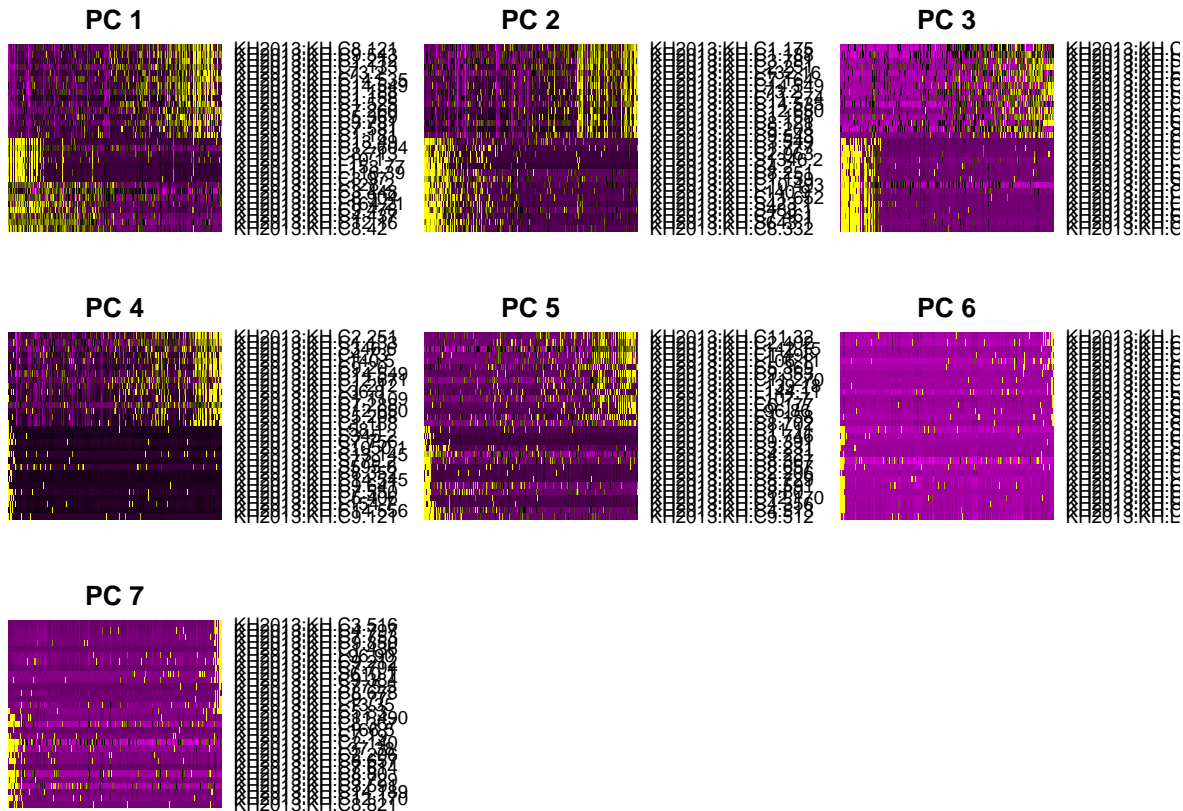
```
## [1] ""
## [1] "KH2013:KH.C8.717" "KH2013:KH.L142.15" "KH2013:KH.S618.2"
## [4] "KH2013:KH.L28.8" "KH2013:KH.L71.18"
## [1] ""
## [1] ""
## [1] "PC4"
## [1] "KH2013:KH.C9.121" "KH2013:KH.C14.556" "KH2013:KH.L154.7"
## [4] "KH2013:KH.C5.406" "KH2013:KH.C7.550"
## [1] ""
## [1] "KH2013:KH.C2.251" "KH2013:KH.C1.133" "KH2013:KH.S346.6"
## [4] "KH2013:KH.C4.73" "KH2013:KH.L140.5"
## [1] ""
## [1] ""
## [1] "PC5"
## [1] "KH2013:KH.C9.512" "KH2013:KH.C4.513" "KH2013:KH.C4.356"
## [4] "KH2013:KH.C12.170" "KH2013:KH.C8.80"
## [1] ""
## [1] "KH2013:KH.C11.32" "KH2013:KH.C2.402" "KH2013:KH.L142.15"
## [4] "KH2013:KH.C1.498" "KH2013:KH.L106.11"
## [1] ""
## [1] ""
```

```
cb.neural <- ProjectPCA(cb.neural, do.print = F)
```

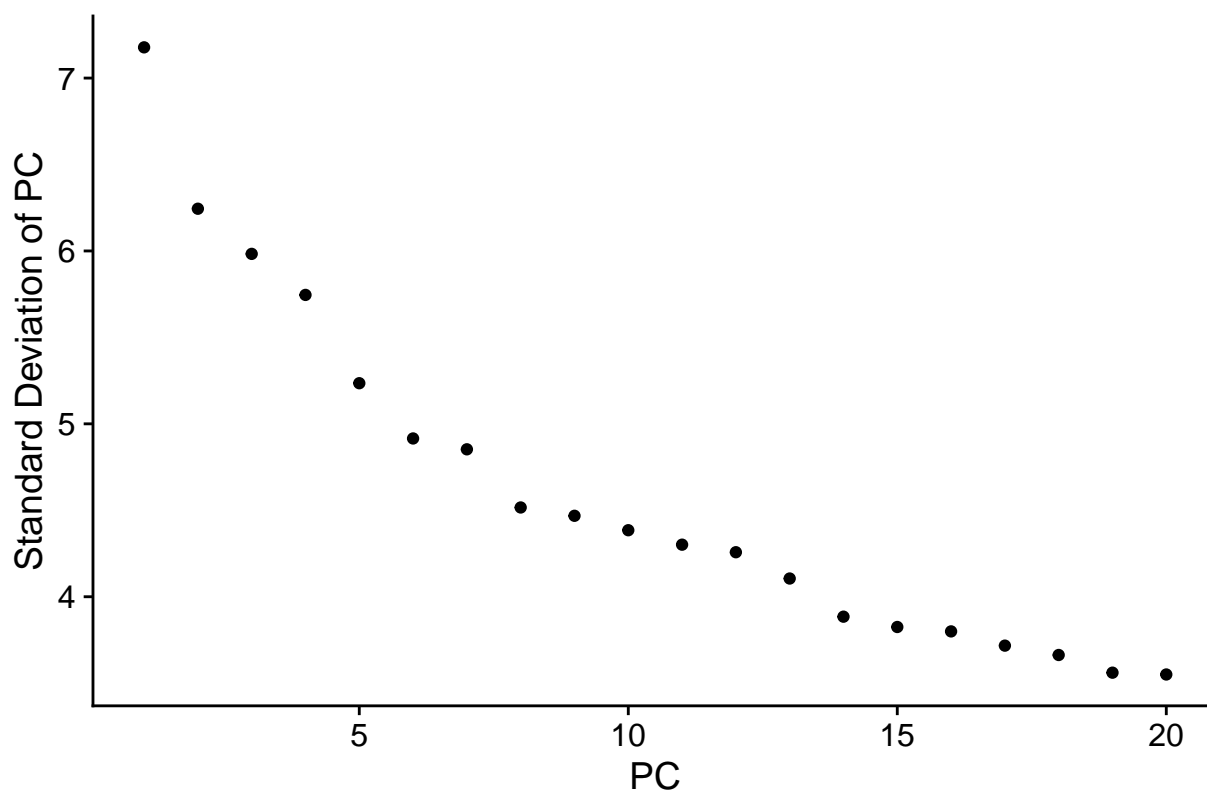
Determination of statistically significant PCs

Similar to Stage 1 of clustering

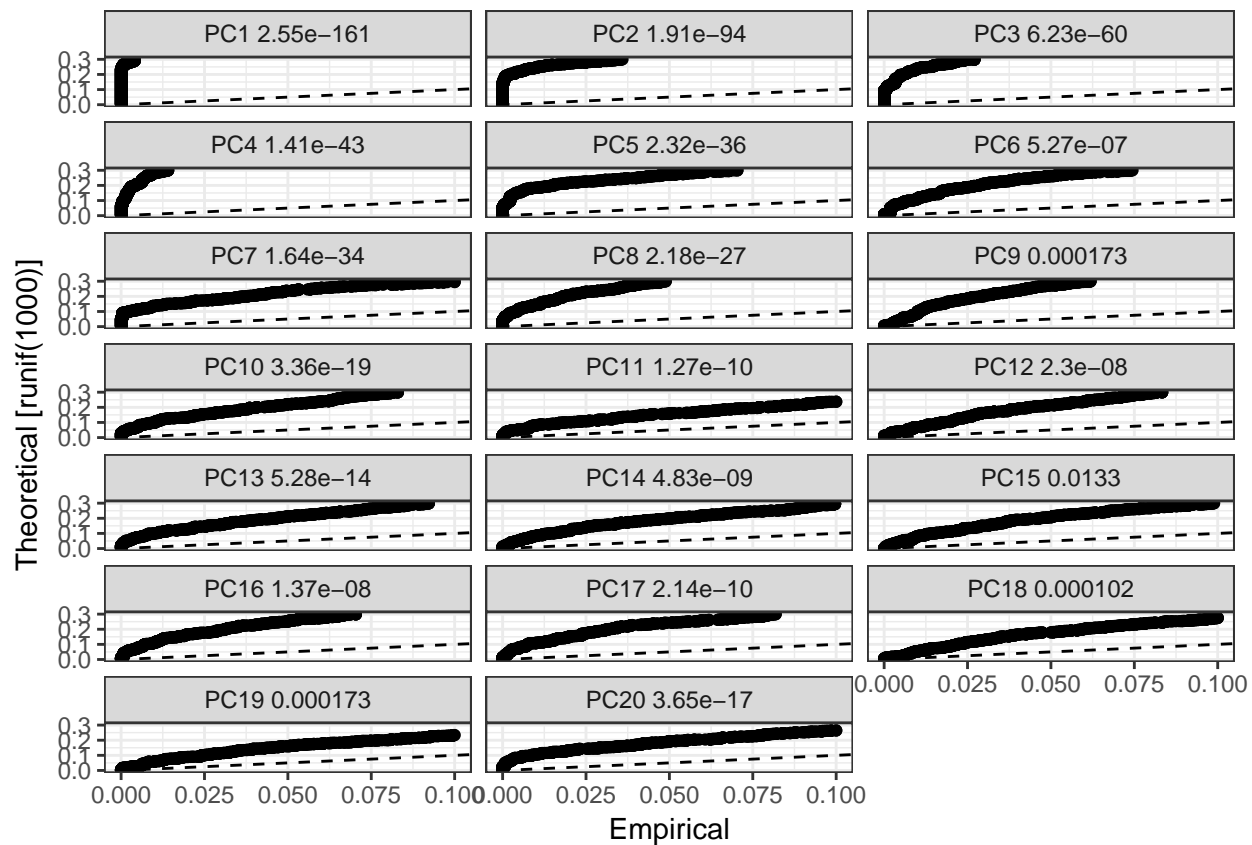
```
PCHearmap(cb.neural,pc.use = 1:7,cells.use = 500,do.balanced = T, label.columns = F, use.full = F)
```



```
PCElbowPlot(object=cb.neural)
```



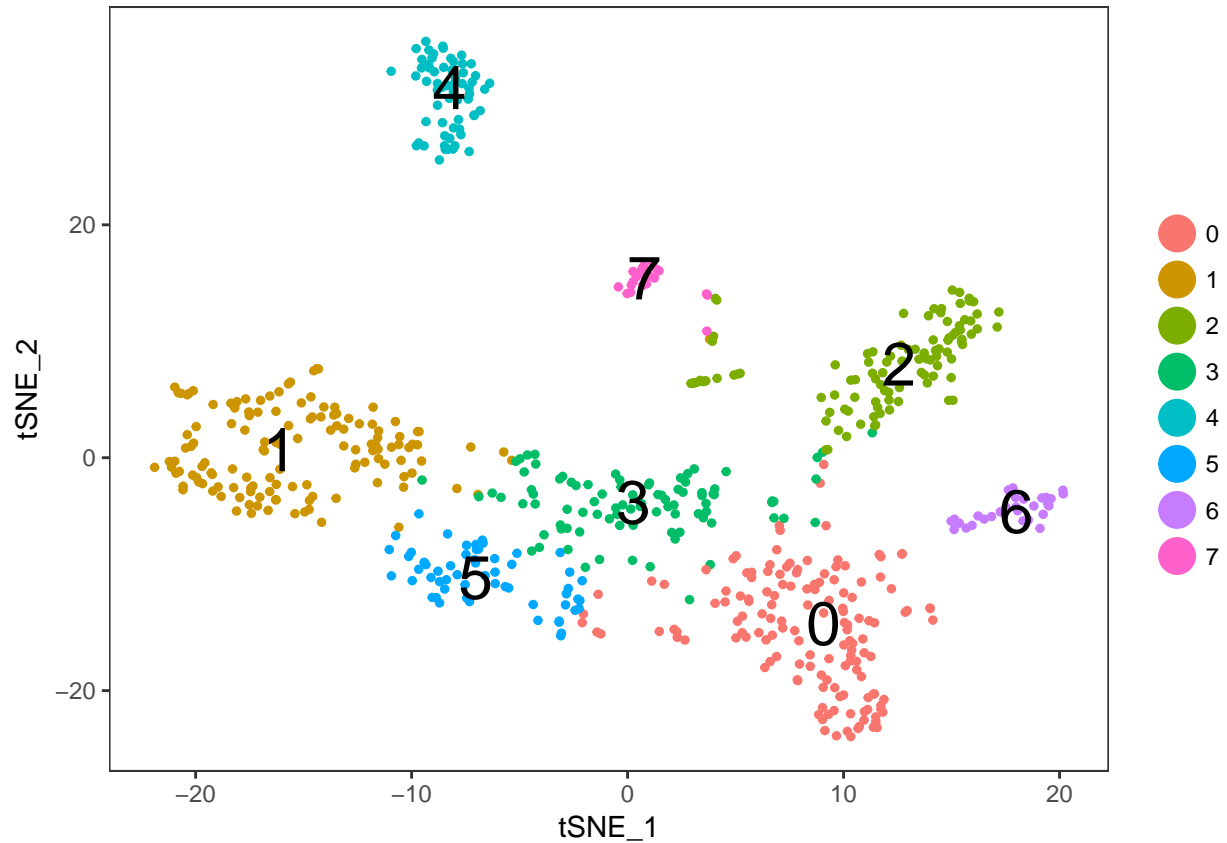
```
cb.neural <- JackStraw(cb.neural, num.replicate = 100, do.print = F)
```



Single Cell Clustering and differential gene expression

Find and Plot Clusters

```
cb.neural <- FindClusters(cb.neural, reduction.type = "pca",
                        dims.use = 1:7, resolution = 0.8,
                        print.output = 0, save.SNN = F, force.recalc = T)
cb.neural<- RunTSNE(cb.neural, dims.use = 1:7, do.fast = T)
TSNEPlot(cb.neural, do.label = T, label.size = 8,pt.size = 1)
```



Find Differentially Expressed Gene Markers

```
cb.neural.allMarkers <- FindAllMarkers(cb.neural)
```

Save the R object

```
saveRDS(cb.neural, file = "~/cb_neural.rds")
```

Miscellaneous

We changed the identity of clusters to character labels to distinguish these from the clusters of stage 1

```
ident0.cells <- WhichCells(cb.neural, ident = 0)
ident1.cells <- WhichCells(cb.neural, ident = 1)
ident2.cells <- WhichCells(cb.neural, ident = 2)
ident3.cells <- WhichCells(cb.neural, ident = 3)
ident4.cells <- WhichCells(cb.neural, ident = 4)
ident5.cells <- WhichCells(cb.neural, ident = 5)
ident6.cells <- WhichCells(cb.neural, ident = 6)
ident7.cells <- WhichCells(cb.neural, ident = 7)

cb.neural <- SetIdent(object = cb.neural, cells.use = ident0.cells, ident.use = "A")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident1.cells, ident.use = "B")
```

```

cb.neural <- SetIdent(object = cb.neural, cells.use = ident2.cells, ident.use = "C")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident3.cells, ident.use = "D")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident4.cells, ident.use = "E")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident5.cells, ident.use = "F")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident6.cells, ident.use = "G")
cb.neural <- SetIdent(object = cb.neural, cells.use = ident7.cells, ident.use = "H")

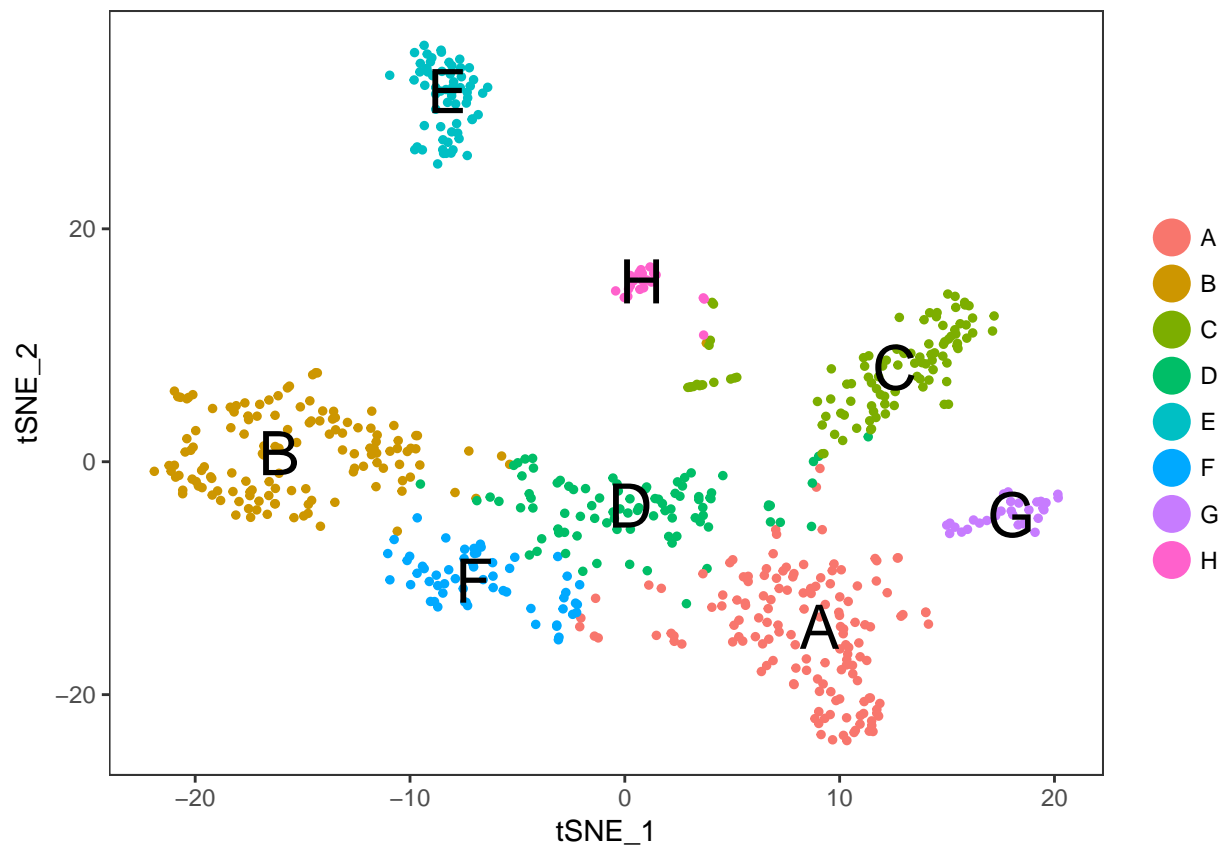
```

TSNE Plot with the new identities

```

TSNEPlot(cb.neural, do.label = T, label.size = 8, pt.size = 1)

```



Manual Splitting of Clusters

Based on our observations of gene markers, we decided to split certain clusters manually. Refer the main paper for the details. The process of manual splitting involves interaction with the plot. Details for the same can be found on 'Seurat' website (<http://satijalab.org/seurat/>).

```

select.cells.1 <- FeaturePlot(object = cb.neural, features.plot = "KH2013:KH.C2.442" ,
                             cells.use = WhichCells(cb.neural, ident = "B"), do.identify = T)
cb.neural <- SetIdent(object = cb.neural, cells.use = select.cells.1, ident.use = "I")

select.cells.2 <- FeaturePlot(object = cb.neural, features.plot = "KH2013:KH.S605.3",
                             do.identify = T)
cb.neural <- SetIdent(object = cb.neural, cells.use = select.cells.2, ident.use = "J")

```