

mesh_to_seurat.R

lubianat

2019-09-18

```
library(data.table)
library(Seurat)

## Registered S3 method overwritten by 'R.oo':
##   method      from
##   throw.default R.methodsS3

mesh <- as.data.frame(t(fread("../data/mesh_hca.txt")))
mesh <- mesh[-1,]
pbmc <- CreateSeuratObject(counts = mesh, project = "pbmc3k", min.cells = 1, min.features = 0)

## Warning: Feature names cannot have underscores ('_'), replacing with dashes
## ('-')

pbmc

## An object of class Seurat
## 633 features across 82 samples within 1 assay
## Active assay: RNA (633 features)

pbmc <- NormalizeData(pbmc)

pbmc <- FindVariableFeatures(pbmc, selection.method = "vst", nfeatures = 2000)

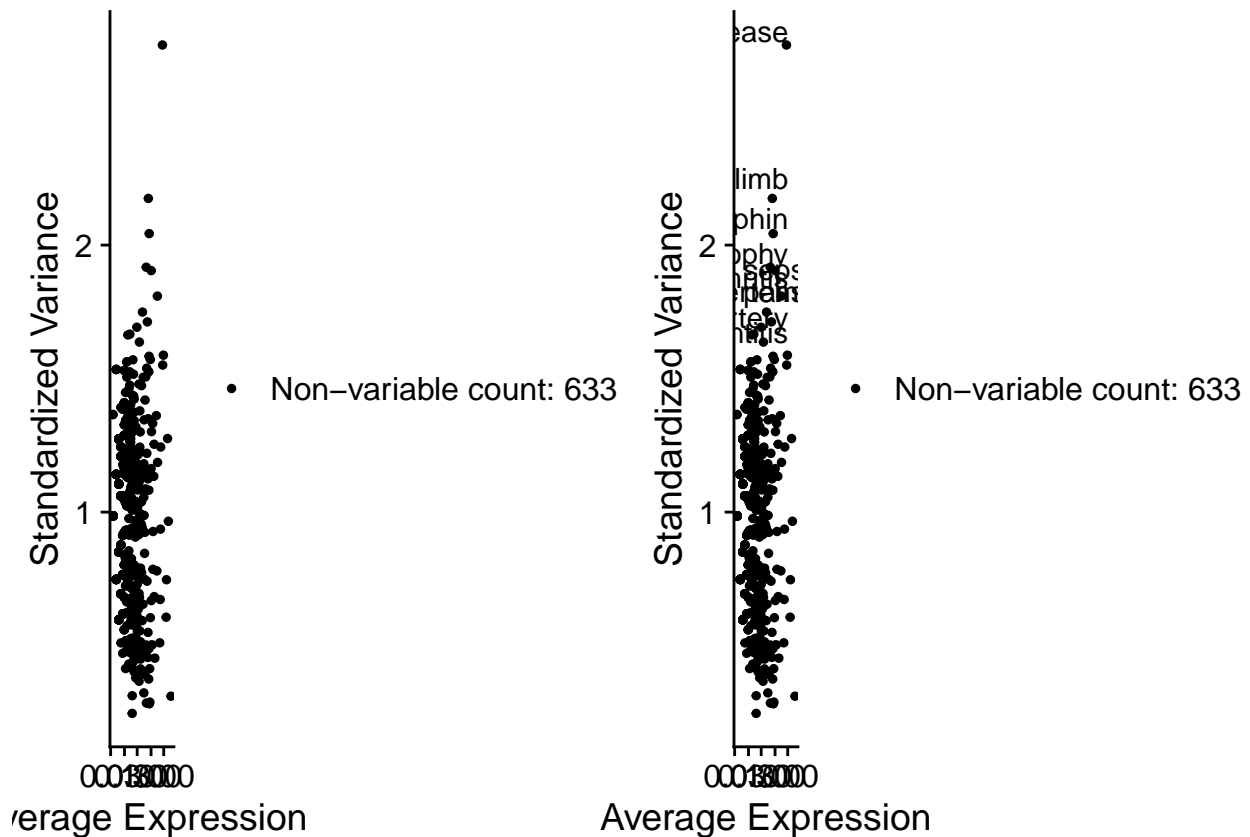
# Identify the 10 most highly variable genes
top10 <- head(VariableFeatures(pbmc), 10)

# plot variable features with and without labels
plot1 <- VariableFeaturePlot(pbmc)
plot2 <- LabelPoints(plot = plot1, points = top10, repel = TRUE)

## Warning: Using `as.character()` on a quosure is deprecated as of rlang 0.3.0.
## Please use `as_label()` or `as_name()` instead.
## This warning is displayed once per session.

## When using repel, set xnudge and ynudge to 0 for optimal results

CombinePlots(plots = list(plot1, plot2))
```



```
all.genes <- rownames(pbmcc)
pbmcc <- ScaleData(pbmcc, features = all.genes)

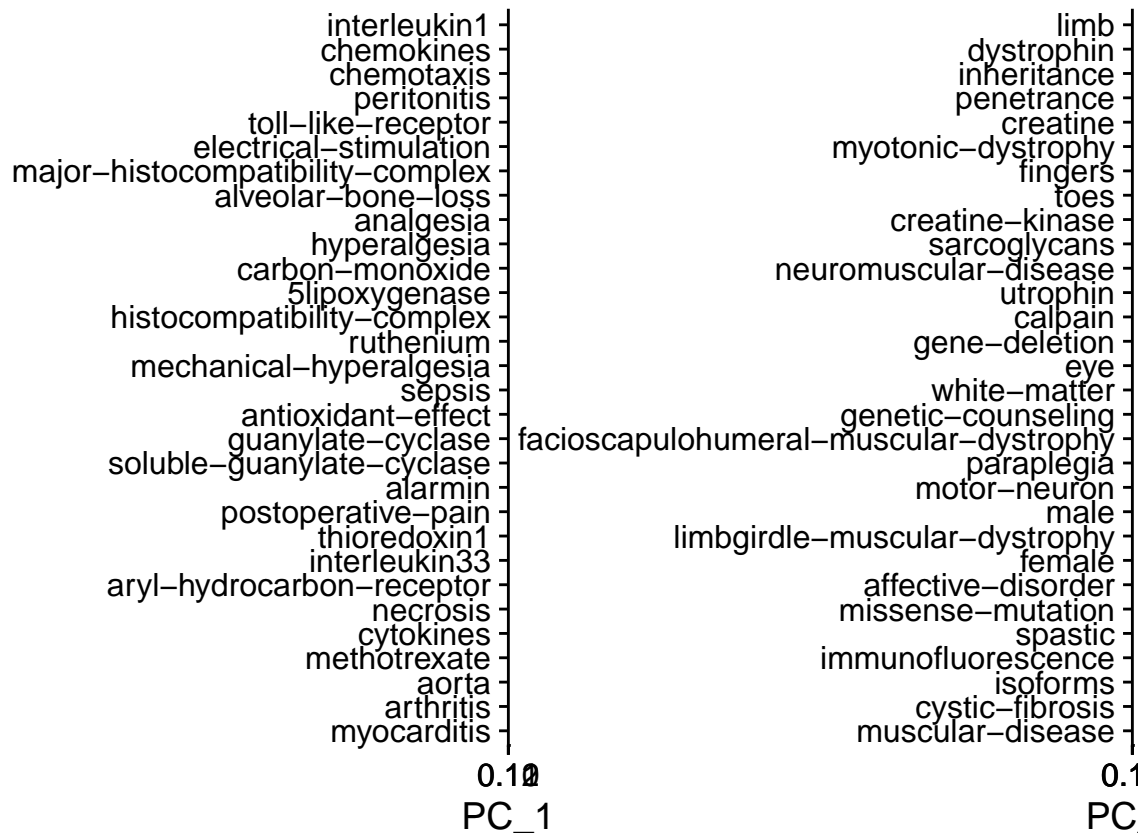
## Centering and scaling data matrix
pbmcc <- RunPCA(pbmcc, features = VariableFeatures(object = pbmcc))

## Warning in irlba(A = t(x = object), nv = npcs, ...): You're computing too
## large a percentage of total singular values, use a standard svd instead.

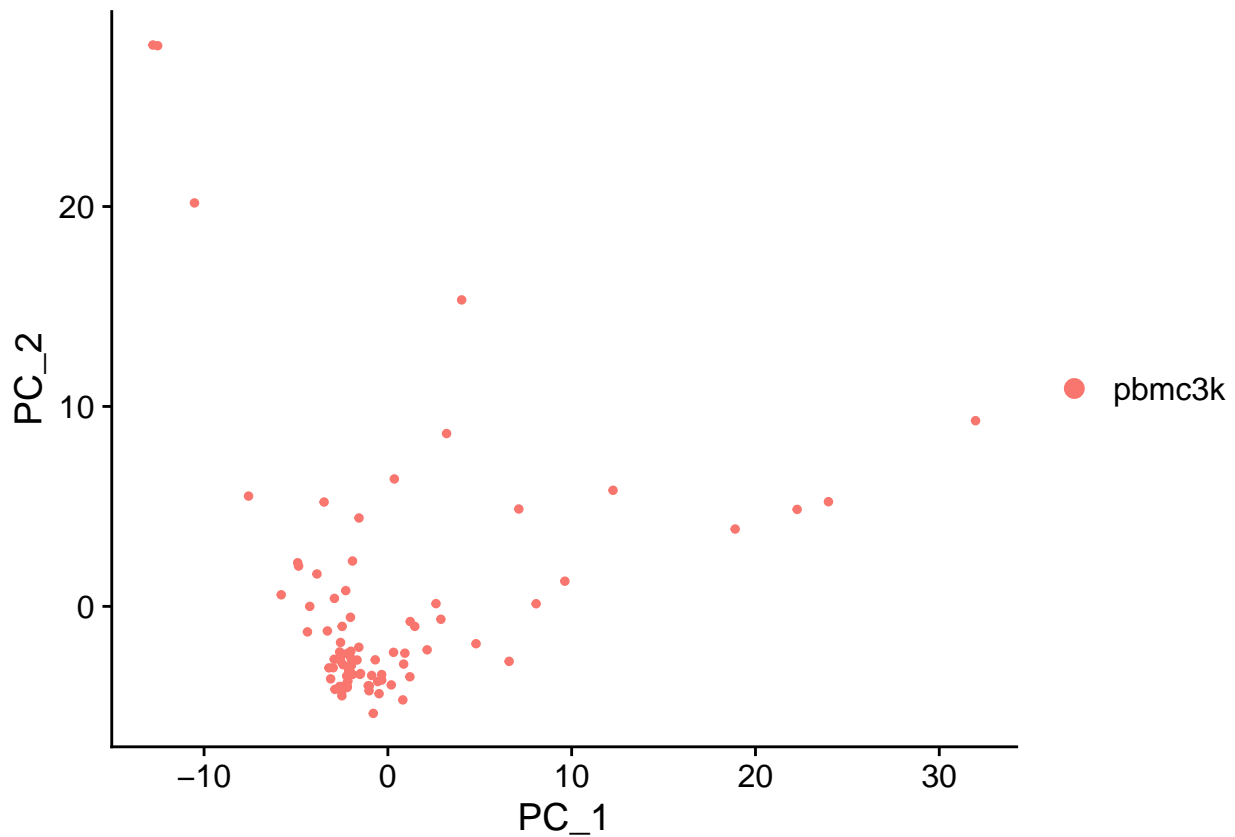
## PC_ 1
## Positive: interleukin1, chemokines, chemotaxis, peritonitis, toll-like-receptor, electrical-stimula
## carbon-monoxide, 5lipoxygenase, histocompatibility-complex, ruthenium, mechanical-hyperalgesia, s
## postoperative-pain, thioredoxin1, interleukin33, aryl-hydrocarbon-receptor, necrosis, cytokines,
## Negative: mental-retardation, eye, muscular-disease, inheritance, genetic-counseling, myotonic-dyst
## creatine-kinase, chromosome, neuromuscular-disease, sarcoglycans, family, utrophin, limb-girdle-m
## limb, diagnosis, paraplegia, twins, immunofluorescence, maps, gene-deletion, fluorescence, in-si
## PC_ 2
## Positive: limb, dystrophin, inheritance, penetrance, creatine, myotonic-dystrophy, fingers, toes, c
## neuromuscular-disease, utrophin, calpain, gene-deletion, eye, white-matter, genetic-counseling,
## male, limb-girdle-muscular-dystrophy, female, affective-disorder, missense-mutation, spastic, immu
## Negative: malaria, cell-cycle, vaccination, parasite, merozoite, flagellin, prostate, merozoite-sur
## surface-antigen, sporozoite, melatonin, binding-protein, epithelial-cell, cancer, infection, try
## transcription-factor, antigen, papillary-thyroid-cancer, epitopes, erythrocytes, surface-protein
## PC_ 3
## Positive: mechanical-hyperalgesia, antioxidant-effect, alarmin, postoperative-pain, thioredoxin1, i
## phages-t, pain, dystrophin, neuropathic-pain, central-nervous-system, nervous-system, chemotaxis
## sarcoglycans, creatine-kinase, toes, fingers, myotonic-dystrophy, inheritance, neuromuscular-dis
## Negative: cohort-study, hypercholesterolemia, diabetes-mellitus, heart-disease, infarction, myocard
```

```
## gene-pol, familial-hypercholesterolemia, cardiomyopathy, atherosclerosis, reninangiotensin-system
## obesity, safety, lipids, artery, sleep, restriction-fragment-length-polymorphism, vascular-disease
## PC_ 4
## Positive: artery, hypertension, reninangiotensin-system, obesity, glucose, blood-pressure, coronary-
## relaxation, lipoprotein, endothelin, rat, sleep, atherosclerosis, familial-hypercholesterolemia,
## restriction-fragment-length-polymorphism, endothelial-nitric-oxide-synthase, atorvastatin, insulin
## Negative: granulocytemacrophage-colonystimulating-factor, vaccine-immunogenicity, macrophage-colony-
## serogroup, wasp, epitope-mapping, ventricular-dysfunction, colonystimulating-factor, cellular-imm
## immune-system, allergy, binding-protein, pancreatic-islet, myosin, hamster, nasal-polyp, cutaneous
## PC_ 5
## Positive: smooth-muscle, coinfection, risk-factor, intensive-care, parvovirus, myocardial-infarction
## herpes-simplex-virus, influenza, reninangiotensin-system, hypertrophy, epidemiology, blood-pressure
## epidemic, stroke, le-cell, erythrovirus, flow-cytometry, enzyme-immunoassay, endothelial-nitric-
## Negative: adenocarcinoma, breast, head-and-neck-cancer, ductal-carcinoma, neck-cancer, expressed-seq
## alcohol-consumption, genetic-marker, breast-cancer, energy-metabolism, biopsy, transcriptome, na
## cell-line, alternative-splicing, metastasis, tumor-suppressor-gene, gene-expression, exon, micro
```

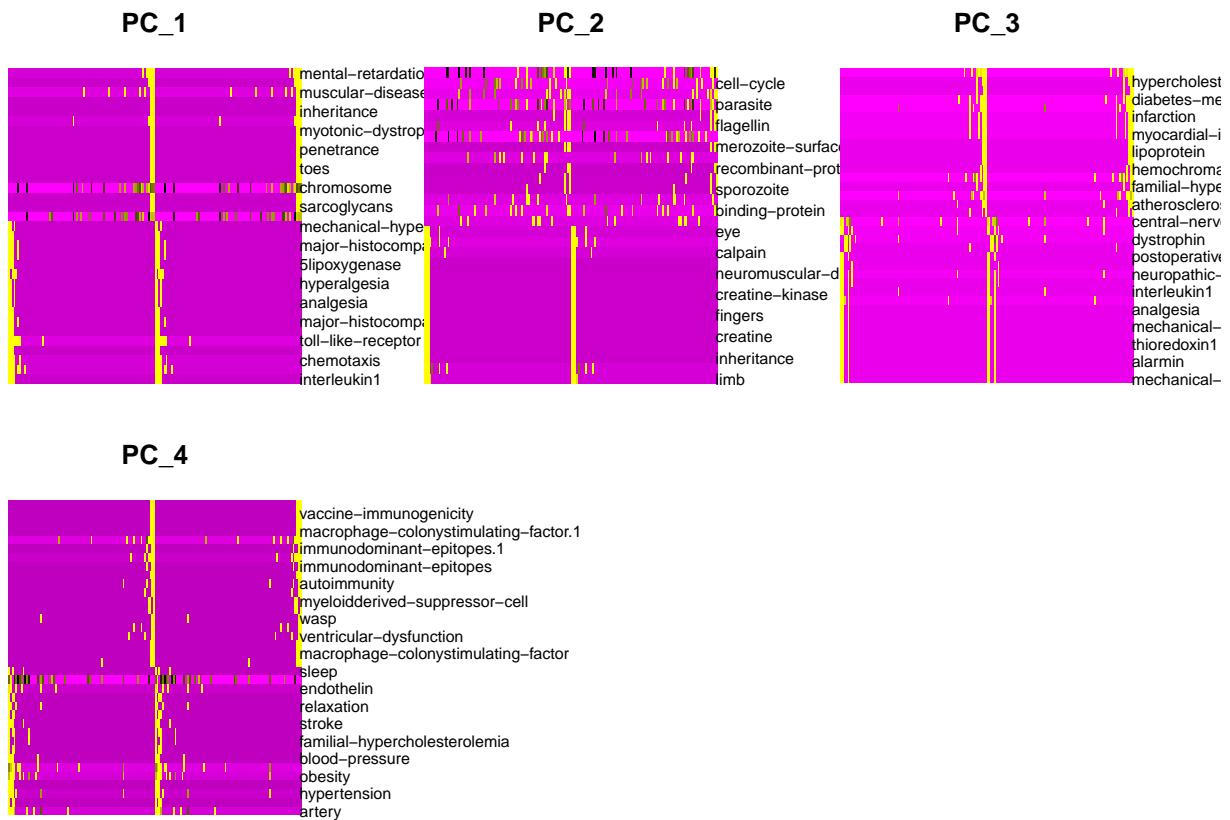
```
VizDimLoadings(pbmc, dims = 1:2, reduction = "pca")
```



```
DimPlot(pbmc, reduction = "pca")
```

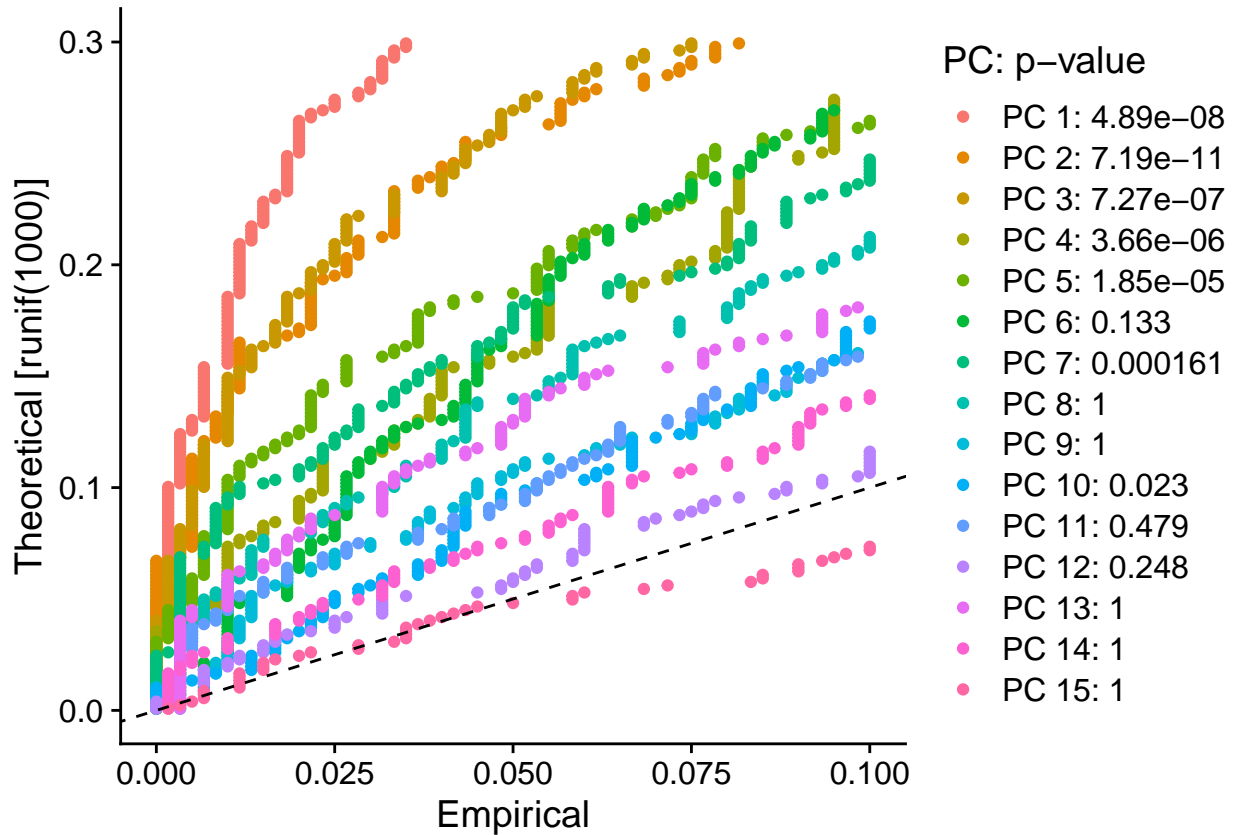


```
DimHeatmap(pbmc, dims = 1:4, cells = 500, balanced = TRUE)
```

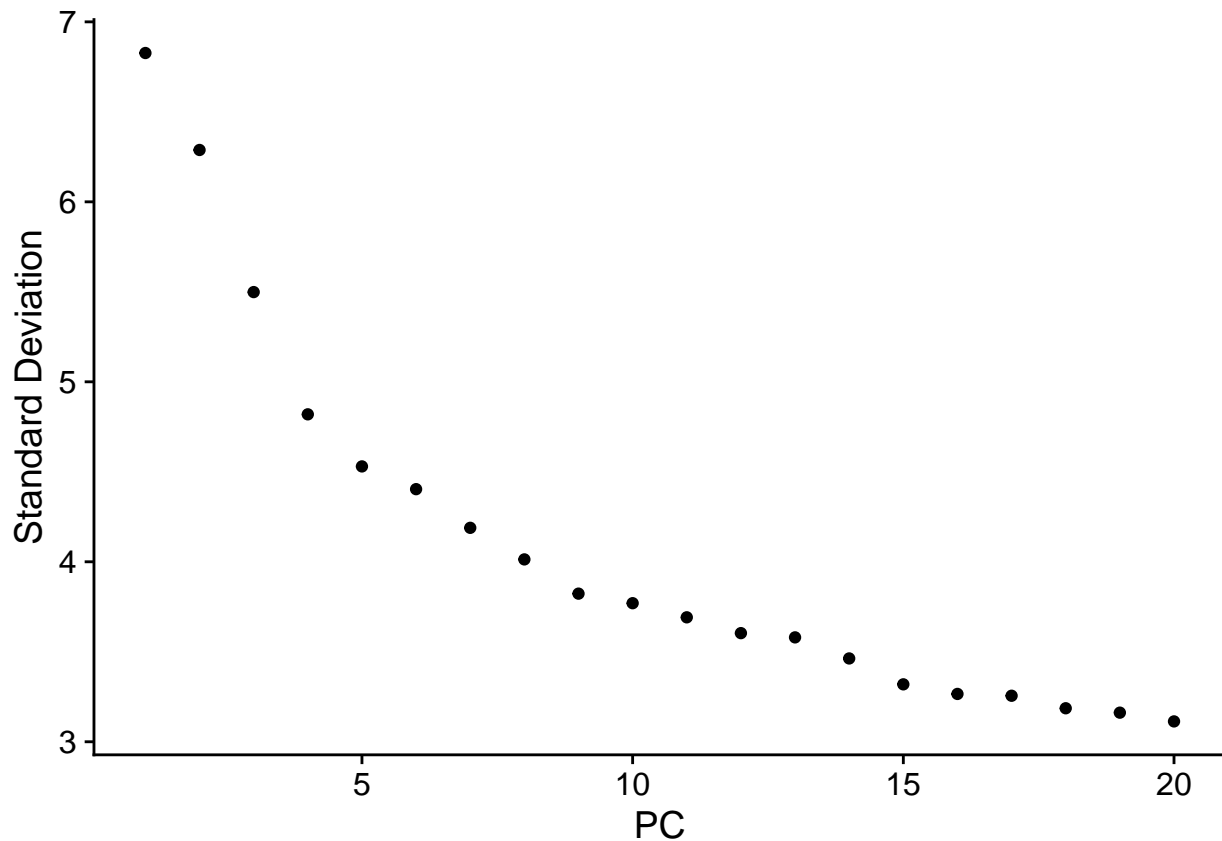


```
pbmc <- JackStraw(pbmc, num.replicate = 100)
pbmc <- ScoreJackStraw(pbmc, dims = 1:20)
JackStrawPlot(pbmc, dims = 1:15)
```

Warning: Removed 7480 rows containing missing values (geom_point).



```
# True dimensionality ~ 7
ElbowPlot(pbmc)
```



```
pbmc <- FindNeighbors(pbmc, dims = 1:7)
```

```
## Computing nearest neighbor graph
```

```
##Computing SNN
```

```
pbmc <- FindClusters(pbmc, resolution = 1.2)
```

```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
```

```
##
```

```
## Number of nodes: 82
```

```
## Number of edges: 2754
```

```
##
```

```
## Running Louvain algorithm...
```

```
## Maximum modularity in 10 random starts: 0.1017
```

```
## Number of communities: 5
```

```
## Elapsed time: 0 seconds
```

```
pbmc <- RunUMAP(pbmc, dims = 1:7)
```

```
## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R
```

```
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
```

```
## This message will be shown once per session
```

```
## 17:46:33 Read 82 rows and found 7 numeric columns
```

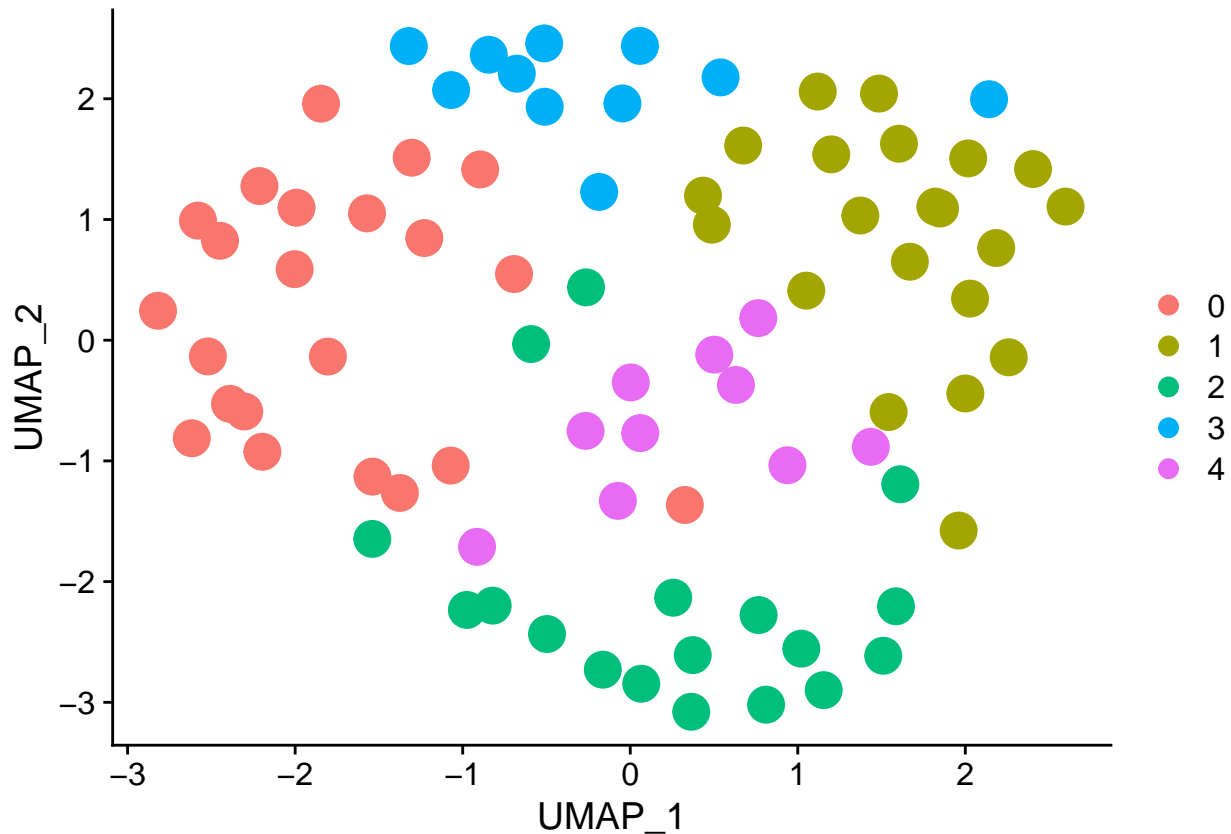
```
## 17:46:33 Using Annoy for neighbor search, n_neighbors = 30
```

```
## 17:46:33 Building Annoy index with metric = cosine, n_trees = 50
```

```
## 0% 10 20 30 40 50 60 70 80 90 100%
```

```
## [----|----|----|----|----|----|----|----|----|
## *****|
## 17:46:33 Writing NN index file to temp file /tmp/RtmpRKEkoV/file6aafeba33e4
## 17:46:33 Searching Annoy index using 1 thread, search_k = 3000
## 17:46:33 Annoy recall = 100%
## 17:46:33 Commencing smooth kNN distance calibration using 1 thread
## 17:46:33 Initializing from normalized Laplacian + noise
## 17:46:33 Commencing optimization for 500 epochs, with 2994 positive edges
## 17:46:34 Optimization finished
```

```
DimPlot(pbmc, reduction = "umap", pt.size = 6)
```



```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:data.table':
##
##   between, first, last
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
pbmc.markers <- FindAllMarkers(pbmc, only.pos = TRUE, min.pct = 0.15, logfc.threshold = 0.25)

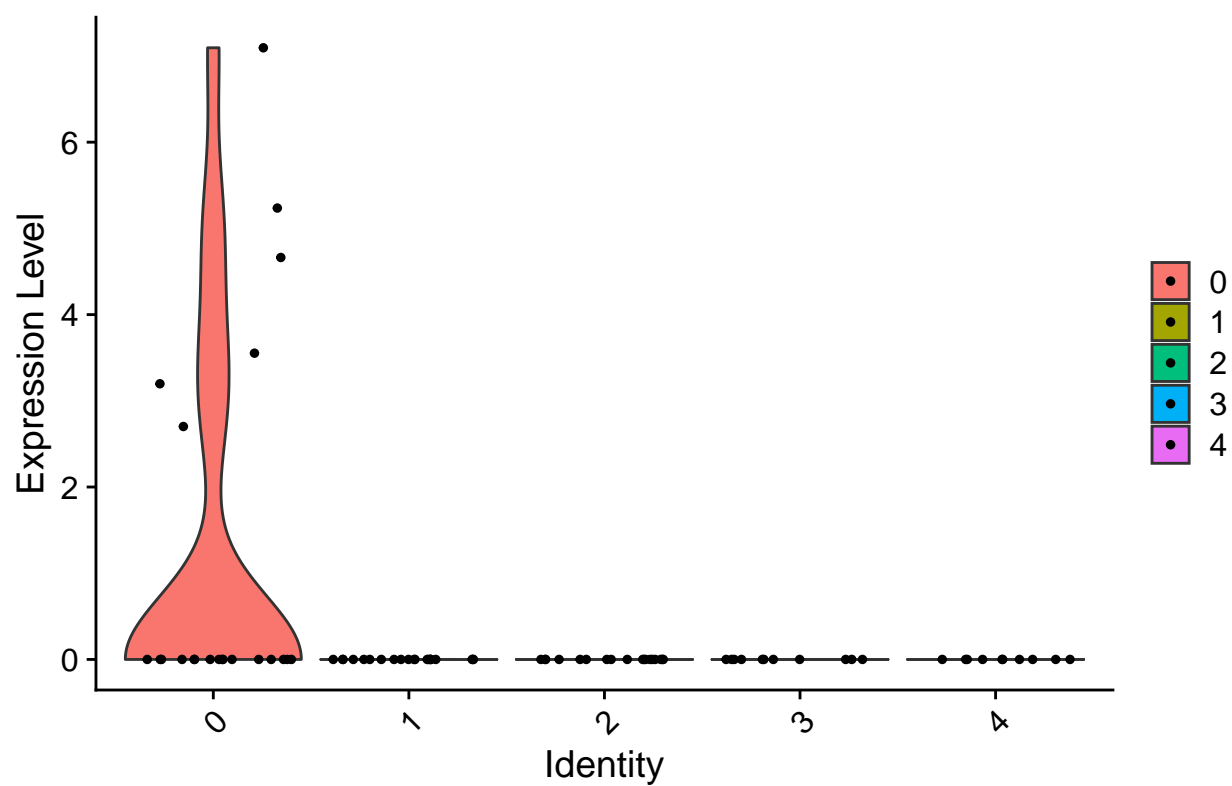
## Calculating cluster 0
## Calculating cluster 1
## Calculating cluster 2
## Calculating cluster 3
## Calculating cluster 4

pbmc.markers %>% group_by(cluster) %>% top_n(n = 4, wt = avg_logFC)

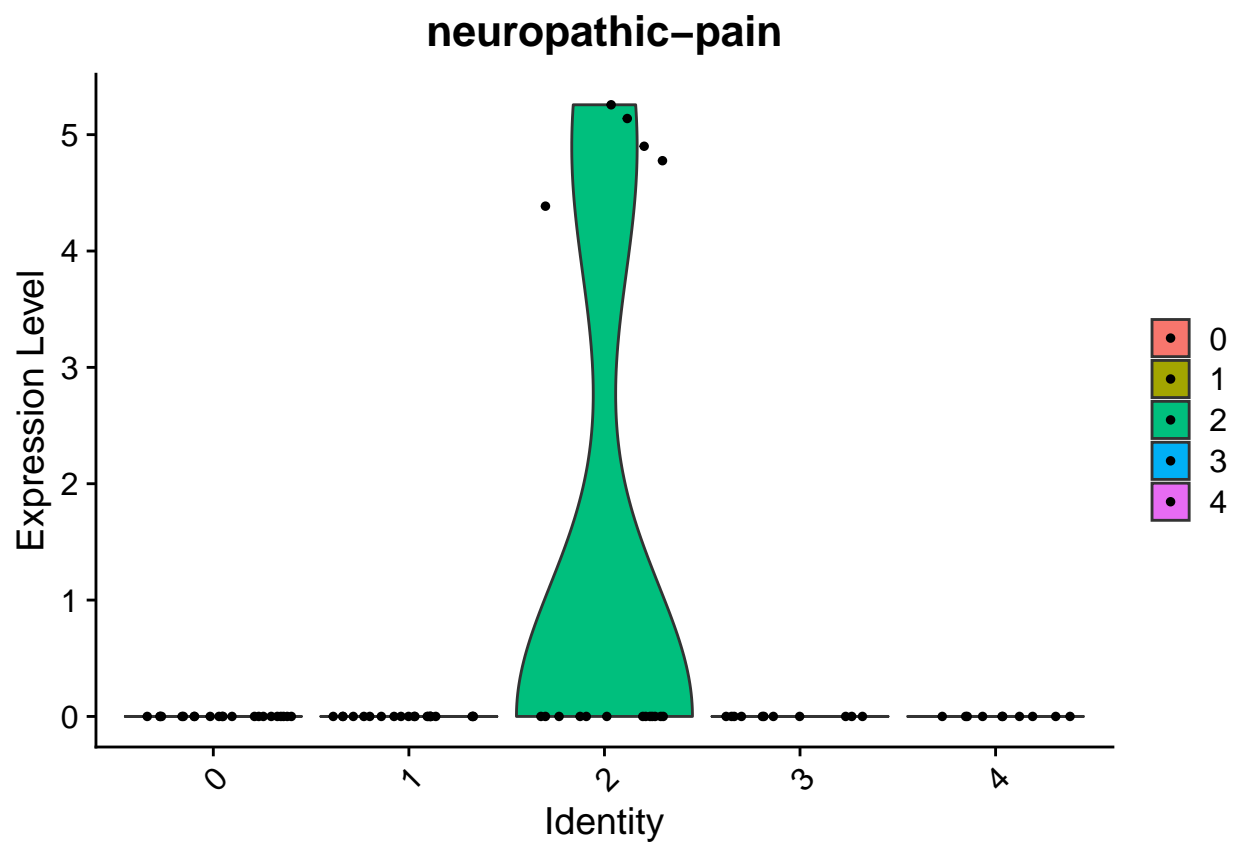
## # A tibble: 17 x 7
## # Groups:   cluster [5]
##       p_val avg_logFC pct.1 pct.2 p_val_adj cluster gene
##       <dbl>   <dbl> <dbl> <dbl>   <dbl>   <fct>   <chr>
## 1  3.17e-5     4.28 0.273 0     0.0201    0   hearing-loss
## 2  1.62e-4     4.14 0.227 0     0.102     0   limb-girdle-muscular~
## 3  8.07e-4     4.13 0.182 0     0.511     0   lip
## 4  8.07e-4     4.01 0.182 0     0.511     0   limb
## 5  2.91e-3     2.35 0.429 0.148 1     1   parasite
## 6  1.65e-5     3.67 0.278 0     0.0104    2   neuropathic-pain
## 7  1.65e-5     3.51 0.278 0     0.0104    2   wound
## 8  1.00e-3     4.75 0.167 0     0.636     2   pain
## 9  1.00e-3     3.56 0.167 0     0.636     2   hyperalgesia
## 10 5.77e-3     3.43 0.182 0.014 1     3   metabolic-pathway
## 11 5.77e-3     3.20 0.182 0.014 1     3   egg
## 12 5.77e-3     3.20 0.182 0.014 1     3   histone-methyltrans~
## 13 5.77e-3     3.00 0.182 0.014 1     3   pancreatic-cancer
## 14 5.10e-8     4.01 0.4   0     0.0000323 4   adenovirus
## 15 2.82e-6     4.38 0.3   0     0.00179   4   parvovirus
## 16 9.22e-5     3.93 0.3   0.014 0.0584   4   systemic-lupus-eryt~
## 17 1.15e-3     3.86 0.4   0.069 0.728    4   virus

VlnPlot(pbmc, features = c("hearing-loss"))
```


hearing-loss

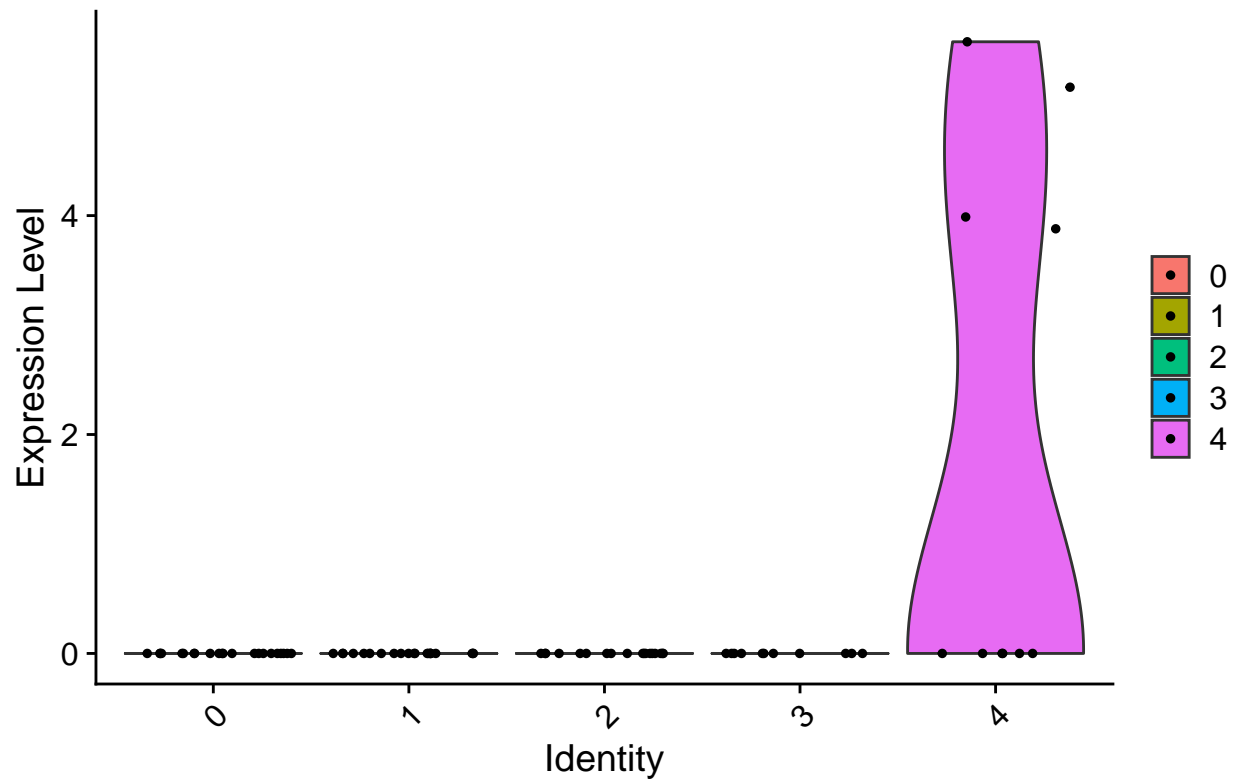


```
VlnPlot(pbmc, features = c("parasite"))
```

```
VlnPlot(pbmc, features = c("adenovirus"))
```

adenovirus



```
FeaturePlot(pbm, features = c("hearing-loss", "parasite", "neuropathic-pain", "adenovirus"))
```

