BioH4

Divya

2022-11-21

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

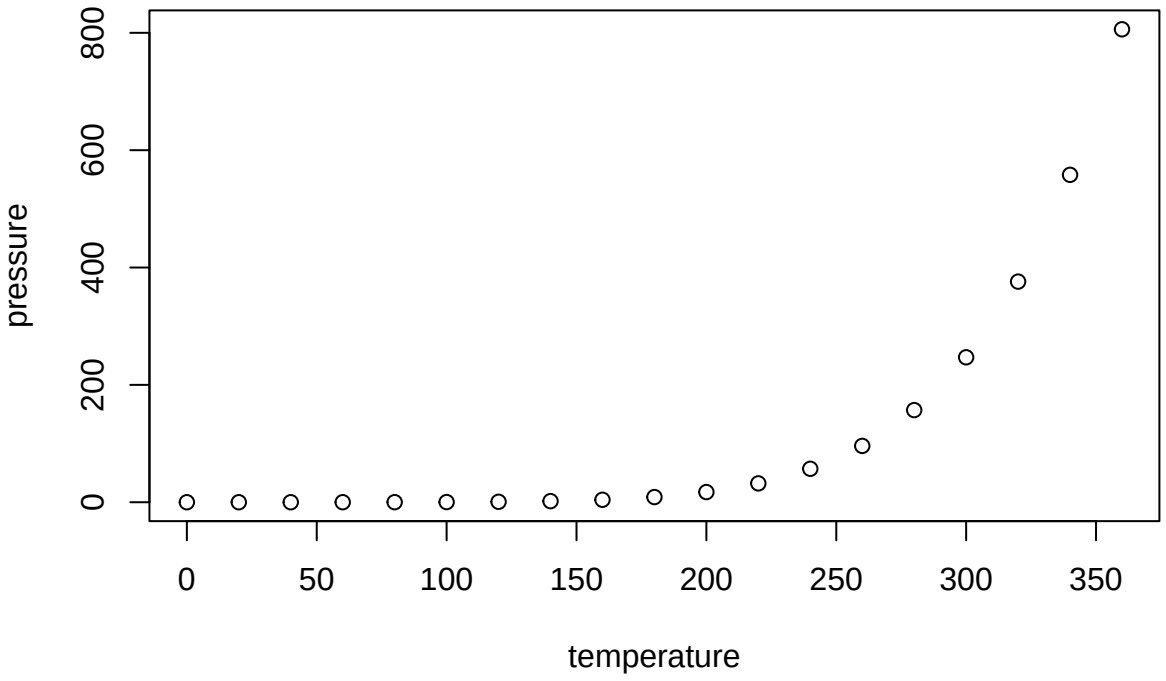
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

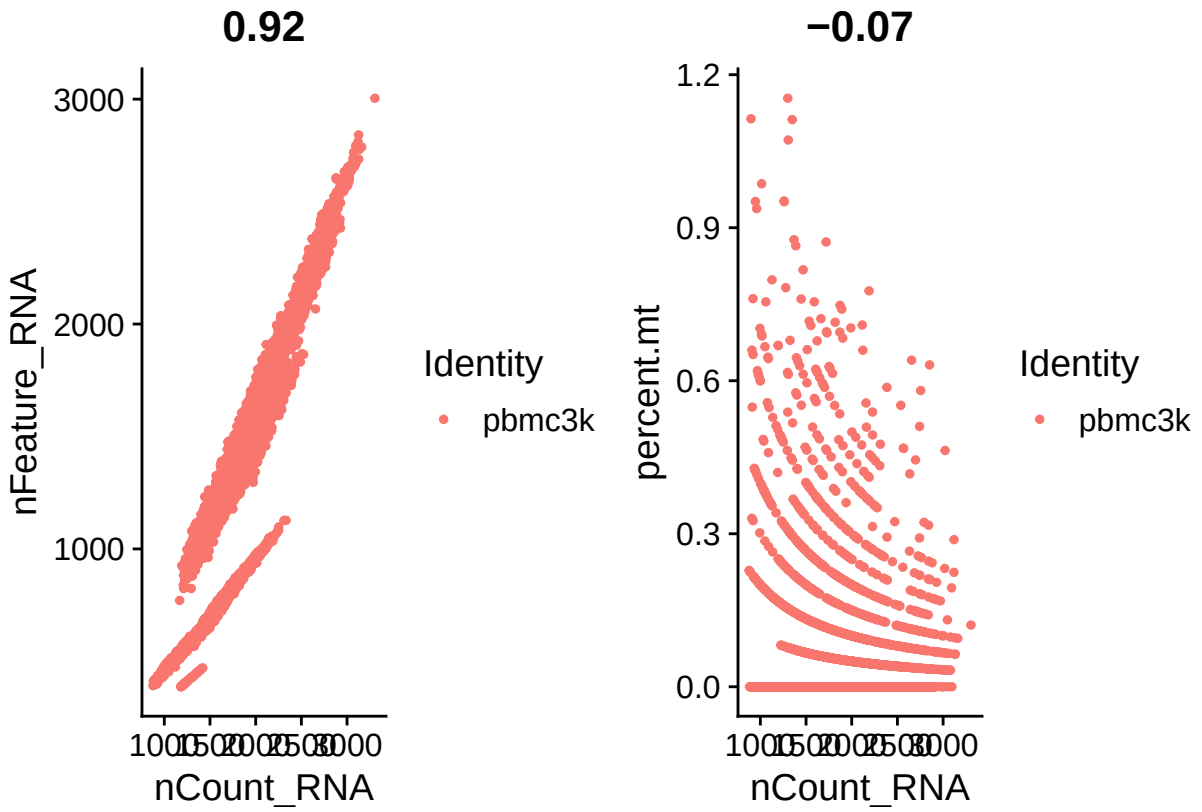
```
library(Seurat)
```

```
## Attaching SeuratObject
```

```
library(patchwork)
library(ggplot2)

pbmc.data <- Read10X(data.dir = "/Users/divya/Desktop/BioInfo/HW4/Vascular/")
pbmc <- CreateSeuratObject(counts = pbmc.data, project="pbmc3k", min.cells=3, min.features=200)
```

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

```
#QC and filtering
pbmc[["percent.mt"]] <- PercentageFeatureSet(pbmc, pattern = "^MT-")

plot1 <- FeatureScatter(pbmc, feature1 = "nCount_RNA", feature2 = "nFeature_RNA")
plot2 <- FeatureScatter(pbmc, feature1 = "nCount_RNA", feature2 = "percent.mt")

plot1 + plot2
```




```
pbmc <- subset(pbmc, subset = nFeature_RNA > 200 & nFeature_RNA > 2500 & percent.mt < 5)
```

```
#normalize data
pbmc <- NormalizeData(pbmc, normalization.method = "LogNormalize", scale.factor = 10000)
pbmc <- NormalizeData(pbmc)
```

```
#Finding variable features
pbmc <- FindVariableFeatures(pbmc, selection.method = "vst", nFeatures=2000)
```

```
## Warning: The following arguments are not used: nFeatures
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at -1.9478
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.4904
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1.2527e-14
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 0.090619
```

```
top10 <- head(VariableFeatures(pbmc),10)
plot1 <- VariableFeaturePlot(pbmc)
plot2 <- LabelPoints(plot=plot1, points=top10, repel=TRUE)
```

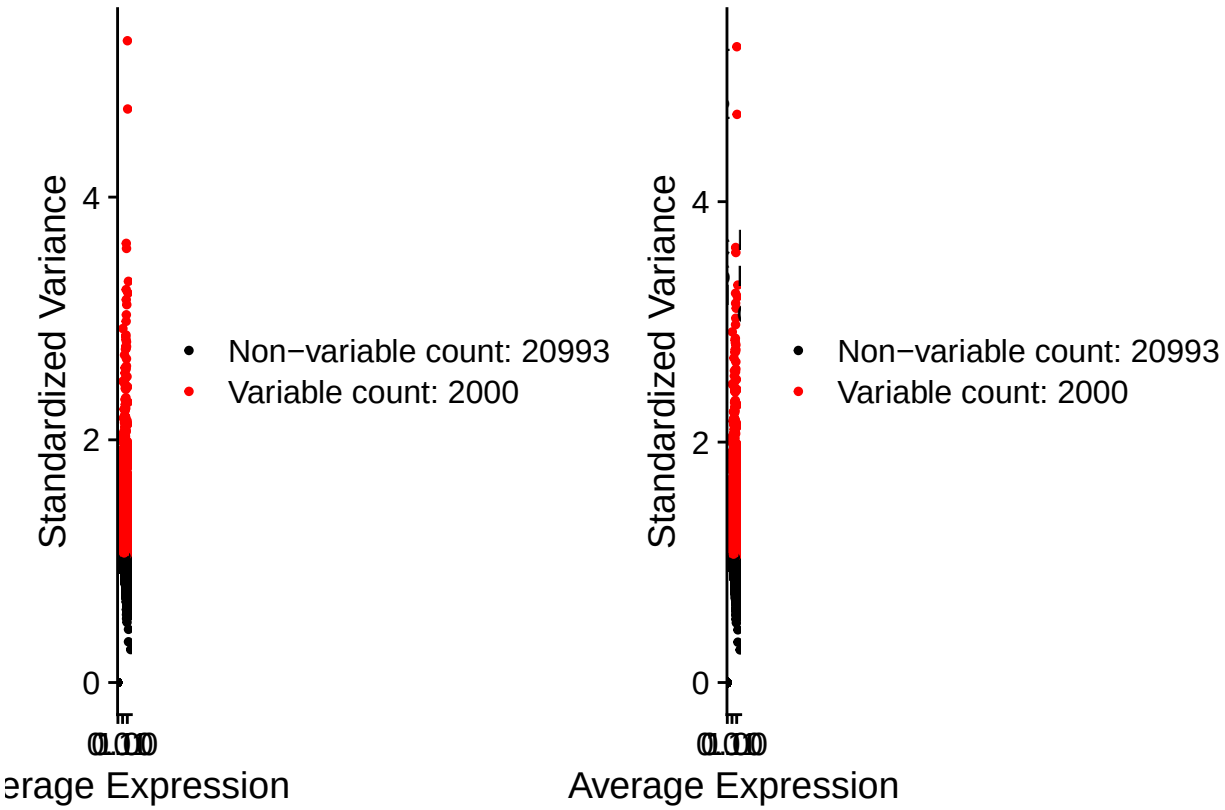
```
## When using repel, set xnudge and ynudge to 0 for optimal results
```

```
plot1+plot2
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```




```
#Scale the data
all.genes <- rownames(pbmc)
pre_scaling <- pbmc
pbmc <- ScaleData(pbmc,features = all.genes)
```

```
## Centering and scaling data matrix
```

```
#run linear dimentionally reduction
pbmc <- RunPCA(pbmc, features = VariableFeatures(object=pbmc))
```

```
## 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:  RBM47, SYK, ABCA1, SLCO2B1, DOCK8, CTSB, SLC1A3, PTPRC, ACER3, FMN1
##     DOCK4, RIN3, DMXL2, PIK3R5, RNF130, CHST11, CD74, ITGAX, PPARG, MCTP1
##     FYB, MSR1, RNF13, BMP2K, TBXAS1, FAM49B, DAPK1, TANC2, MITF, DOCK2
## Negative:  PLA2R1, MSRB3, CACNA1C, HSPG2, DLG2, COL5A2, COL21A1, ZBTB16, CDH13, RBMS3
##     ITGA1, COL8A1, COL14A1, SGCD, SGIP1, SUGCT, LIMCH1, ANTXR1, MAP1B, RCAN2
##     PRKG1, FBXL7, LTBP1, ITGA8, ELN, RBPMS, PARD3B, NR2F2-AS1, CACNA2D1, EDIL3
## PC_ 2
## Positive:  ANPEP, CLEC5A, POLQ, SLC39A8, KIF18B, RP11-519G16.3, SLC16A10, SPP1, ASPM, FN1
##     RP11-556E13.1, S100A10, ALCAM, PTPRM, ITGAM, RAI14, SLC16A3, KCNQ3, UHRF1, NCAPG
##     PVT1, RAD54L, CENPK, NCAPH, IQGAP3, IL7R, FIRRE, FANCA, KCNK13, HELLS
## Negative:  SAMD4A, SLC18B1, PLA2G7, MS4A4A, A2M, HS3ST2, PLPP3, MGLL, HTRA4, RP11-106M3.2
##     PDK4, CPM, AXL, RP11-807H22.7, FUCA1, ABCC5, SOAT1, LINC01500, DHRS3, CCL18
```

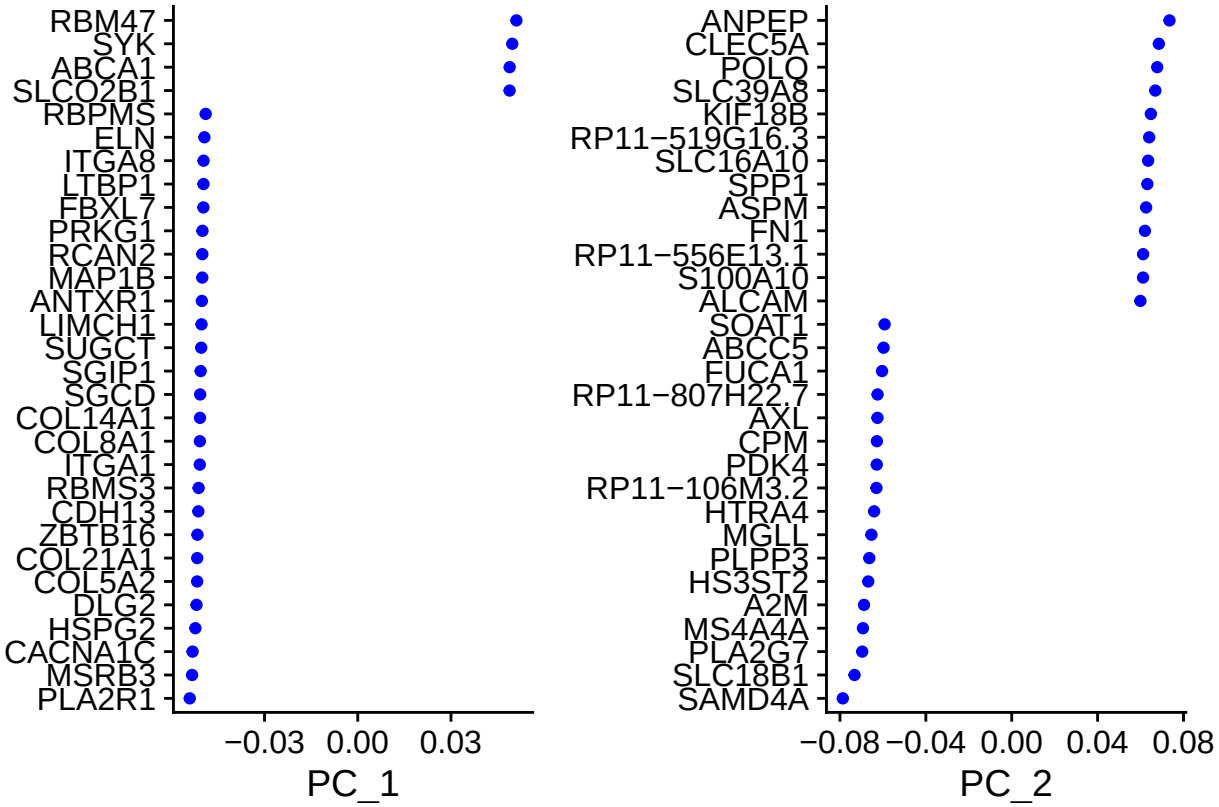
```
##     NCOA4, CR1, KLHL6, LRRC16A, FAM20A, TNFAIP2, TMEM255A, BASP1, SH3PXD2B, SAT1
## PC_ 3
## Positive:  COL6A3, CEMIP, ABI3BP, TNXB, OSMR-AS1, TBX15, MAPK10, FHL2, SLIT2, COL24A1
##     OSMR, FRMD6, ADAMTS17, MECOM, LDB2, CFH, PRICKLE2, DCLK1, MLIP, SFRP4
##     RUNX1T1, FGF7, PTGIS, CCDC39, ZNF385B, DDR2, CACNB4, WISP1, PLEKHS1, COL15A1
## Negative:  RGS5, LDB3, TC2N, MYBL1, NTN1, EFHD1, CATSPERB, GRIA2, KCTD16, C8orf34
##     FILIP1, UNC13C, TTLL7, NT5DC3, AC067959.1, LMOD1, SORBS1, SMOC2, RYR2, LINC01411
##     DTHD1, PTGER3, ST6GAL2, NRXN3, INCENP, CNTN4, SLC22A3, MYH11, COL27A1, PDE8B
## PC_ 4
## Positive:  CCDC68, RNF152, ID1, GUCY1A2, SLC12A2, FBXO16, HTR4, ADAMTS2, PROCR, PALM2-AKAP2
##     ANGPT2, NRG3, TNS2, CNTNAP3, LINC00607, MEG3, ADGRF5, ARHGAP44, LINC01091, FANK1
##     RARB, BRCA2, IMPG2, GALNT13, PCOLCE2, JAG1, LAMB1, RP1-90G24.10, RP11-96H19.1, PRDM5
## Negative:  TIMP3, CSRNP3, CTD-2269F5.1, PLEKHS1, OSMR-AS1, ROBO2, THSD4, FMO2, FHOD3, NOX4
##     MLIP, SFRP4, MFGE8, ADAMTSL3, ROR1, WISP1, DCN, FYN, GULP1, PPFIA2
##     CACNB4, ZFP36L2, RUNX1T1, ZNF385B, COL12A1, ZNF365, ITFG1, FRMD6, CLMP, GPR176
## PC_ 5
## Positive:  LDLRAD4, F13A1, CLEC10A, SH3KBP1, ADAM28, FCGR2A, POU2F2, COLEC12, CIITA, RTN1
##     SPRED1, KIAA1147, PALD1, ARHGAP18, TCF4, TGFBR2, LRRK2, MRC1, MS4A6A, CD4
##     GTDC1, RP1-69D17.4, MAN2A1, RP11-96H19.1, RP11-231C18.3, USP6NL, ARHGAP24, MAX, MNDA, FCGBP
## Negative:  BHLHE40, SEPT9, DIAPH3, ASPM, TFRC, KIF18B, ZBTB7C, KIF15, TULP4, POLQ
##     TDRD3, CCNF, ATP6V1A, GTSE1, DTYMK, NCAPD3, CHIT1, FAM134B, IQGAP3, RAI14
##     LGALS3, PAPLN, SH3BP5, NCAPH, HTRA4, C2orf48, BRIP1, CENPF, LPL, RP11-317J19.1
```

```
print(pbmc[["pca"]],dims=1:5,nFeatures=5)
```

```
## Warning: The following arguments are not used: nFeatures
```

```
## PC_ 1
## Positive:  RBM47, SYK, ABCA1, SLCO2B1, DOCK8, CTSB, SLC1A3, PTPRC, ACER3, FMN1
##     DOCK4, RIN3, DMXL2, PIK3R5, RNF130, CHST11, CD74, ITGAX, PPARG, MCTP1
## Negative:  PLA2R1, MSRB3, CACNA1C, HSPG2, DLG2, COL5A2, COL21A1, ZBTB16, CDH13, RBMS3
##     ITGA1, COL8A1, COL14A1, SGCD, SGIP1, SUGCT, LIMCH1, ANTXR1, MAP1B, RCAN2
## PC_ 2
## Positive:  ANPEP, CLEC5A, POLQ, SLC39A8, KIF18B, RP11-519G16.3, SLC16A10, SPP1, ASPM, FN1
##     RP11-556E13.1, S100A10, ALCAM, PTPRM, ITGAM, RAI14, SLC16A3, KCNQ3, UHRF1, NCAPG
## Negative:  SAMD4A, SLC18B1, PLA2G7, MS4A4A, A2M, HS3ST2, PLPP3, MGLL, HTRA4, RP11-106M3.2
##     PDK4, CPM, AXL, RP11-807H22.7, FUCA1, ABCC5, SOAT1, LINC01500, DHRS3, CCL18
## PC_ 3
## Positive:  COL6A3, CEMIP, ABI3BP, TNXB, OSMR-AS1, TBX15, MAPK10, FHL2, SLIT2, COL24A1
##     OSMR, FRMD6, ADAMTS17, MECOM, LDB2, CFH, PRICKLE2, DCLK1, MLIP, SFRP4
## Negative:  RGS5, LDB3, TC2N, MYBL1, NTN1, EFHD1, CATSPERB, GRIA2, KCTD16, C8orf34
##     FILIP1, UNC13C, TTLL7, NT5DC3, AC067959.1, LMOD1, SORBS1, SMOC2, RYR2, LINC01411
## PC_ 4
## Positive:  CCDC68, RNF152, ID1, GUCY1A2, SLC12A2, FBXO16, HTR4, ADAMTS2, PROCR, PALM2-AKAP2
##     ANGPT2, NRG3, TNS2, CNTNAP3, LINC00607, MEG3, ADGRF5, ARHGAP44, LINC01091, FANK1
## Negative:  TIMP3, CSRNP3, CTD-2269F5.1, PLEKHS1, OSMR-AS1, ROBO2, THSD4, FMO2, FHOD3, NOX4
##     MLIP, SFRP4, MFGE8, ADAMTSL3, ROR1, WISP1, DCN, FYN, GULP1, PPFIA2
## PC_ 5
## Positive:  LDLRAD4, F13A1, CLEC10A, SH3KBP1, ADAM28, FCGR2A, POU2F2, COLEC12, CIITA, RTN1
##     SPRED1, KIAA1147, PALD1, ARHGAP18, TCF4, TGFBR2, LRRK2, MRC1, MS4A6A, CD4
## Negative:  BHLHE40, SEPT9, DIAPH3, ASPM, TFRC, KIF18B, ZBTB7C, KIF15, TULP4, POLQ
##     TDRD3, CCNF, ATP6V1A, GTSE1, DTYMK, NCAPD3, CHIT1, FAM134B, IQGAP3, RAI14
```

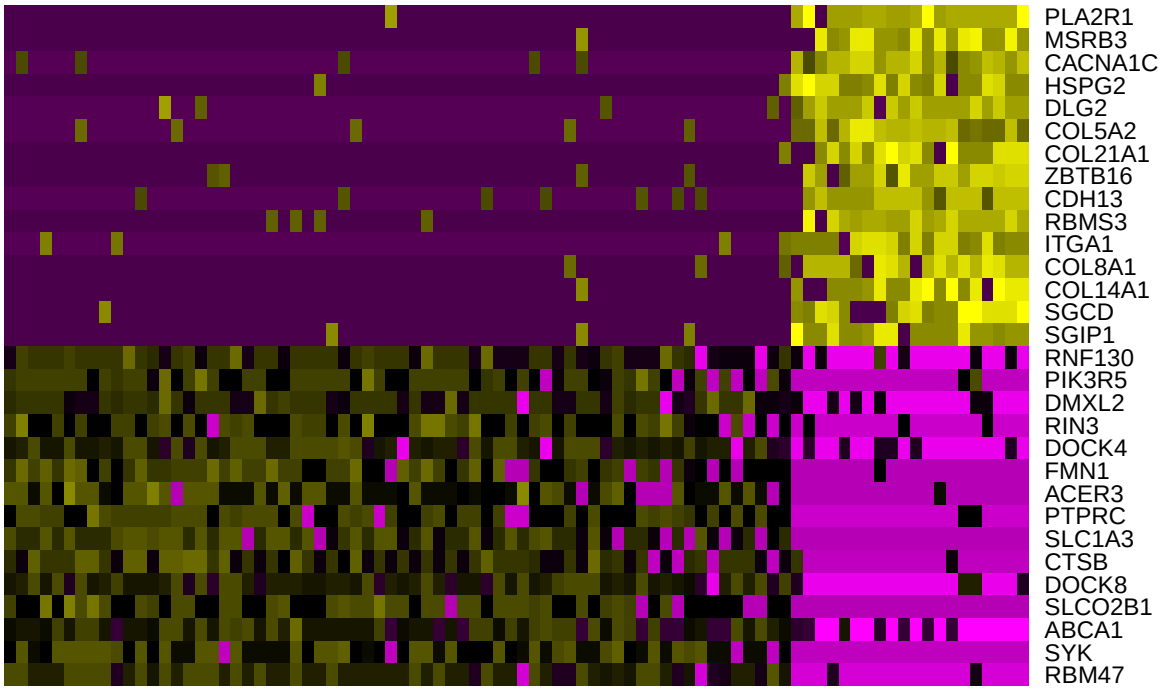
```
VizDimLoadings(pbmc,dims=1:2,reduction="pca")
```




```
DimHeatmap(pbmc,dims=1,cells=500)
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

PC_1




```
DimHeatmap(pbmc,dims=1:15,cells=500)
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

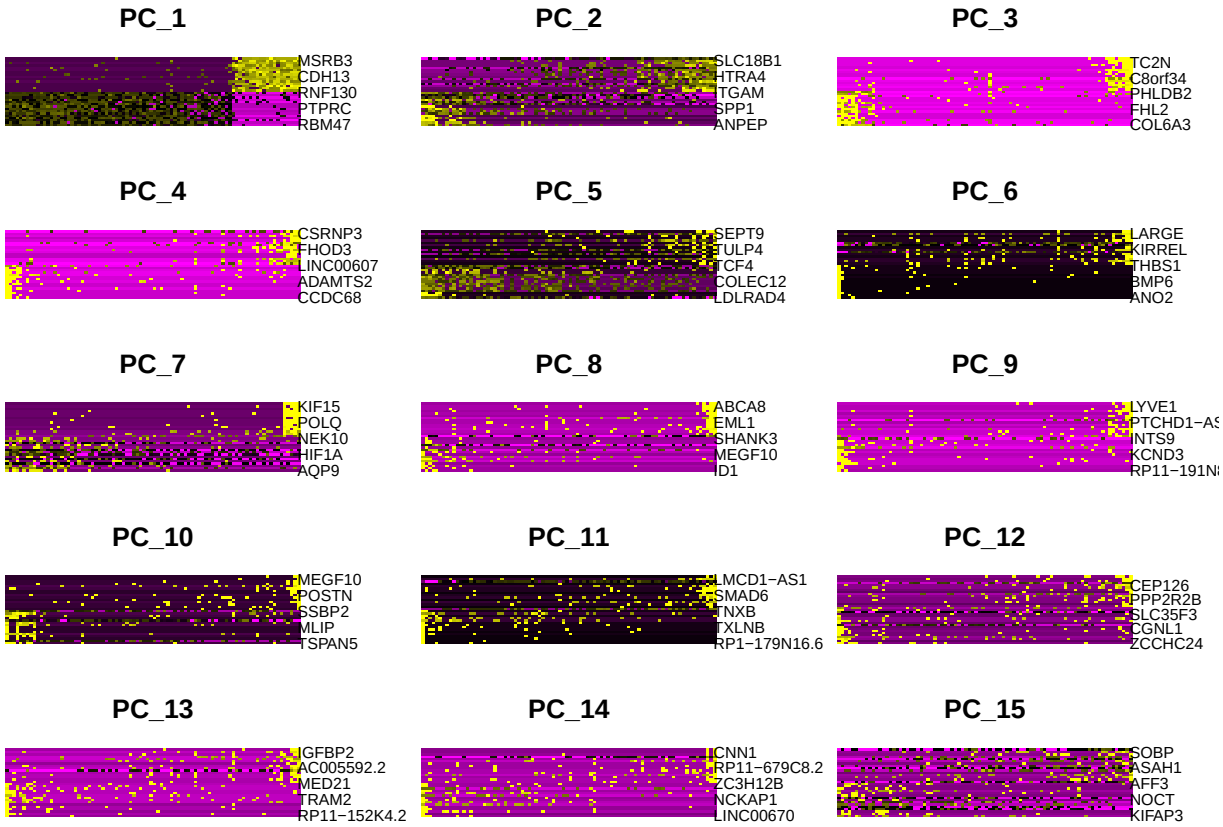
```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

```
## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```

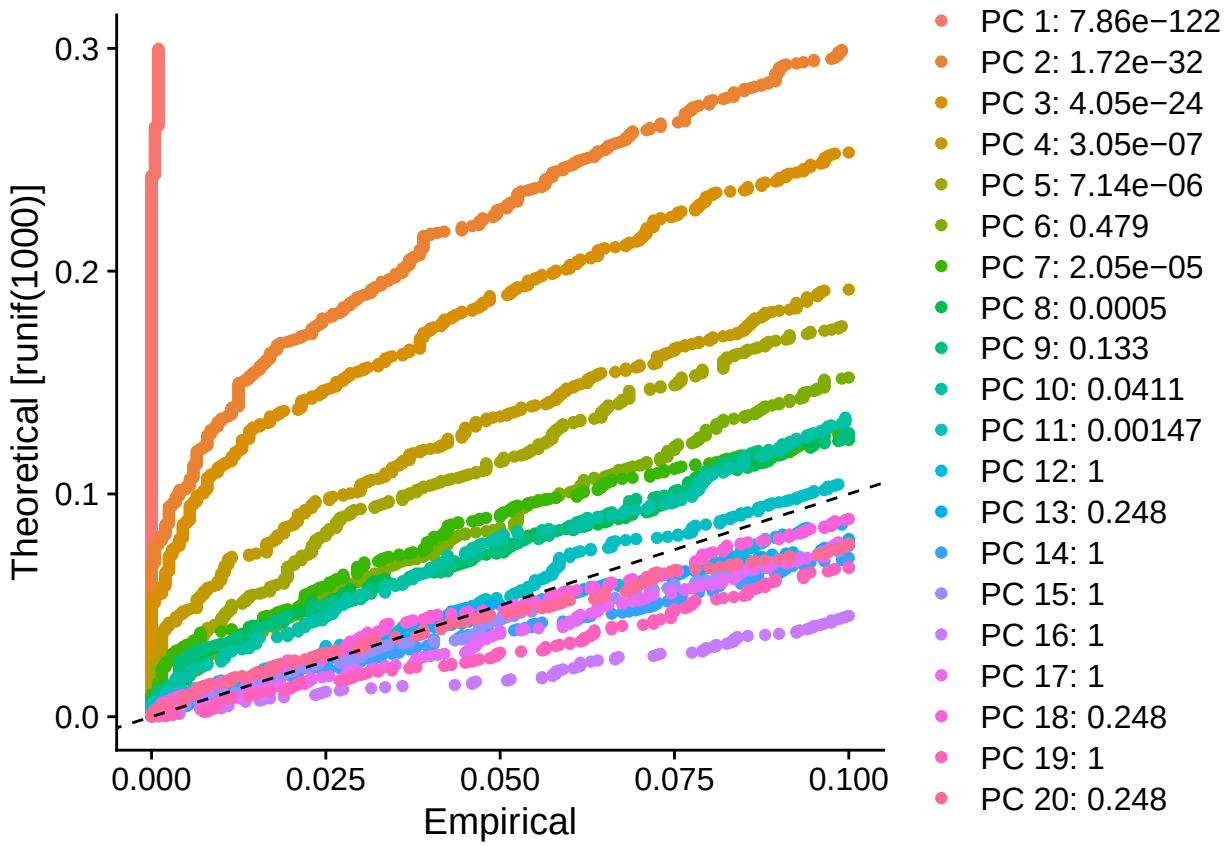
```
## Warning: Requested number is larger than the number of available items (86).
```

```
## Setting to 86.

## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.

## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.

## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.

## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.

## Warning: Requested number is larger than the number of available items (86).
## Setting to 86.
```




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

```
## Warning: Removed 34673 rows containing missing values (`geom_point()`).
```




```
#cluster
pbmc
```

```
## An object of class Seurat
## 22993 features across 86 samples within 1 assay
## Active assay: RNA (22993 features, 2000 variable features)
##  1 dimensional reduction calculated: pca
```

```
pbmc <- FindNeighbors(pbmc, dims=1:20)
```

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

```
## Computing SNN
```

```
pbmc <- FindClusters(pbmc, resolution = 0.5)
```

```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 86
## Number of edges: 2138
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.6463
## Number of communities: 2
## Elapsed time: 0 seconds
```

```
head(Idents(pbmc),5)
```

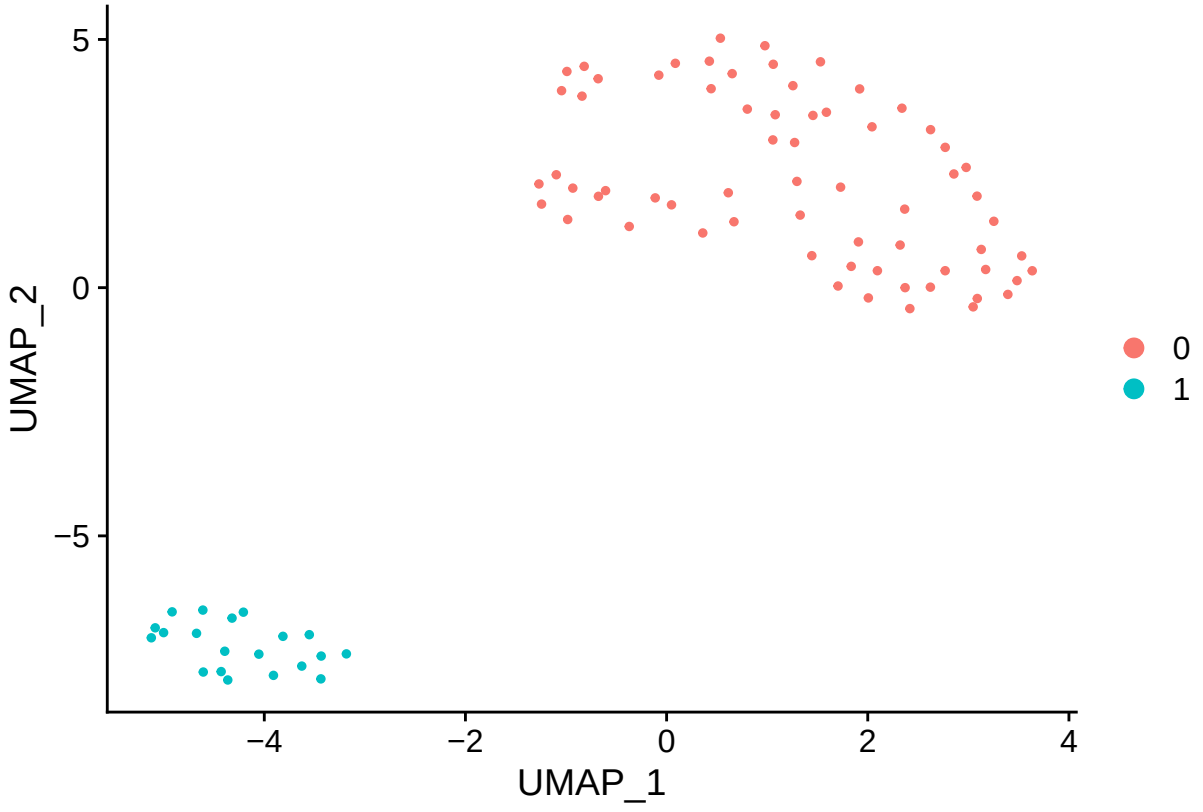
```
## ACCAACAAGCACTTTG-1-0 CCACCATAGTTGCCCG-1-0 CGTTAGAAGGGAGGTG-1-0
##                    1                    1                    1
```

```
## CTGAGCGCAGGCATGA-1-0 CTTGATTGTTCACGAT-1-0
##                    1                    1
## Levels: 0 1
```

```
#run non linear dimentionality reduction
pbmc <- RunUMAP(pbmc, dims=1:10)
```

```
## 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
## 22:13:32 UMAP embedding parameters a = 0.9922 b = 1.112
## 22:13:32 Read 86 rows and found 10 numeric columns
## 22:13:32 Using Annoy for neighbor search, n_neighbors = 30
## 22:13:32 Building Annoy index with metric = cosine, n_trees = 50
## 0%   10   20   30   40   50   60   70   80   90   100%
## [----|----|----|----|----|----|----|----|----|----|
## **************************************************|
## 22:13:33 Writing NN index file to temp file /var/folders/3b/l3s7zj8x6pn97rzdz4qcs28c0000gn/T//RtmpCK
## 22:13:33 Searching Annoy index using 1 thread, search_k = 3000
## 22:13:33 Annoy recall = 100%
## 22:13:33 Commencing smooth kNN distance calibration using 1 thread with target n_neighbors = 30
## 22:13:33 Initializing from normalized Laplacian + noise (using irlba)
## 22:13:33 Commencing optimization for 500 epochs, with 2710 positive edges
## 22:13:33 Optimization finished
```

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



```
#assign biological meaning to these clusters
pbmc.markers <- FindAllMarkers(pbmc, only.pos = TRUE, min.pct=0.25, logFC.threshold = 0.25)
```

```
## Calculating cluster 0
```

```
## For a more efficient implementation of the Wilcoxon Rank Sum Test,
## (default method for FindMarkers) please install the limma package
## --------------------------------------------
## install.packages('BiocManager')
## BiocManager::install('limma')
## --------------------------------------------
## After installation of limma, Seurat will automatically use the more
## efficient implementation (no further action necessary).
## This message will be shown once per session
```

```
## Calculating cluster 1
```

```
pbmc.markers %>% group_by(cluster) %>% slice_max(n=2,order_by = avg_log2FC)
```

```
## # A tibble: 4 x 7
## # Groups:   cluster [2]
##      p_val avg_log2FC pct.1 pct.2 p_val_adj cluster gene
##      <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>   <chr>
## 1 9.28e-10       2.96 0.894 0      2.13e- 5 0       SLC1A3
## 2 1.33e-10       2.62 0.97  0.1    3.05e- 6 0       RBM47
## 3 3.77e-14       3.08 1     0.212  8.66e-10 1       PRKG1
## 4 7.29e-17       3.01 1     0.076  1.68e-12 1       CACNA1C
```

```
pbmc
```

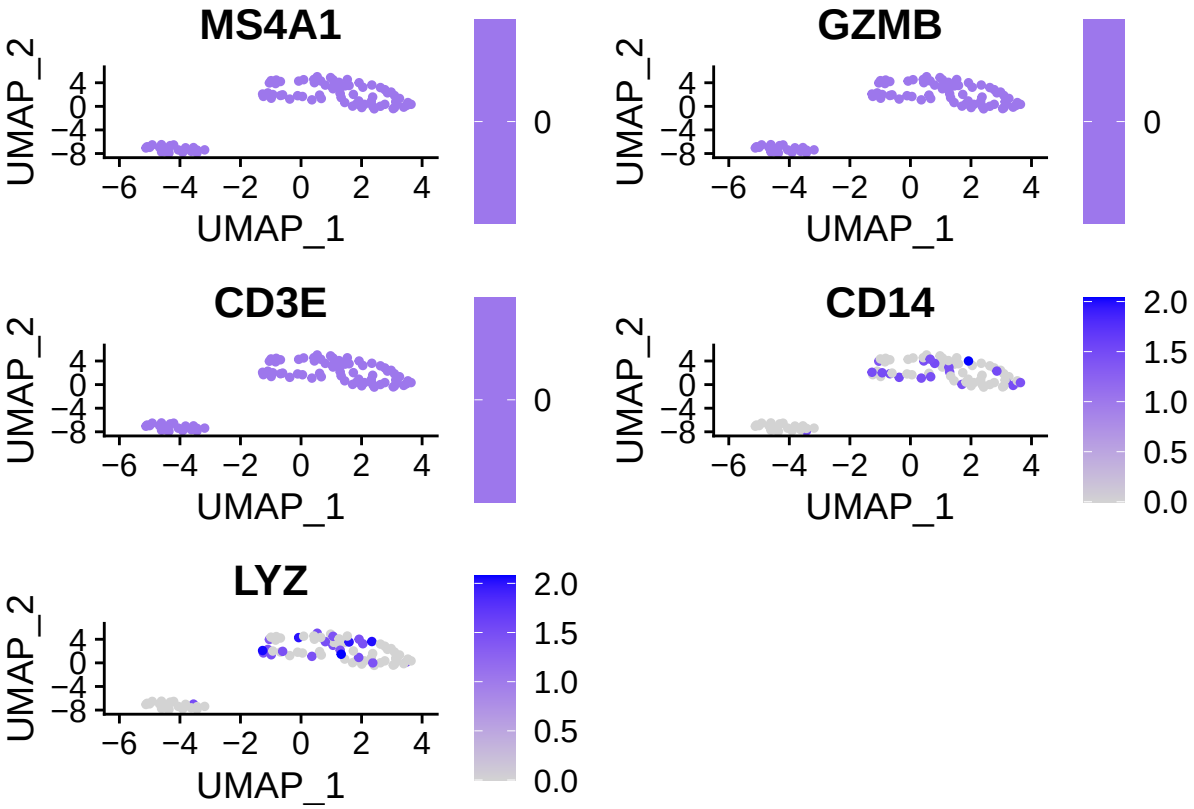
```
## An object of class Seurat
## 22993 features across 86 samples within 1 assay
## Active assay: RNA (22993 features, 2000 variable features)
##  2 dimensional reductions calculated: pca, umap
```

```
FeaturePlot(pbmc, features = c("MS4A1","GZMB","CD3E","CD14","LYZ"))
```

```
## Warning in FeaturePlot(pbmc, features = c("MS4A1", "GZMB", "CD3E", "CD14", : All
## cells have the same value (0) of MS4A1.
```

```
## Warning in FeaturePlot(pbmc, features = c("MS4A1", "GZMB", "CD3E", "CD14", : All
## cells have the same value (0) of GZMB.
```

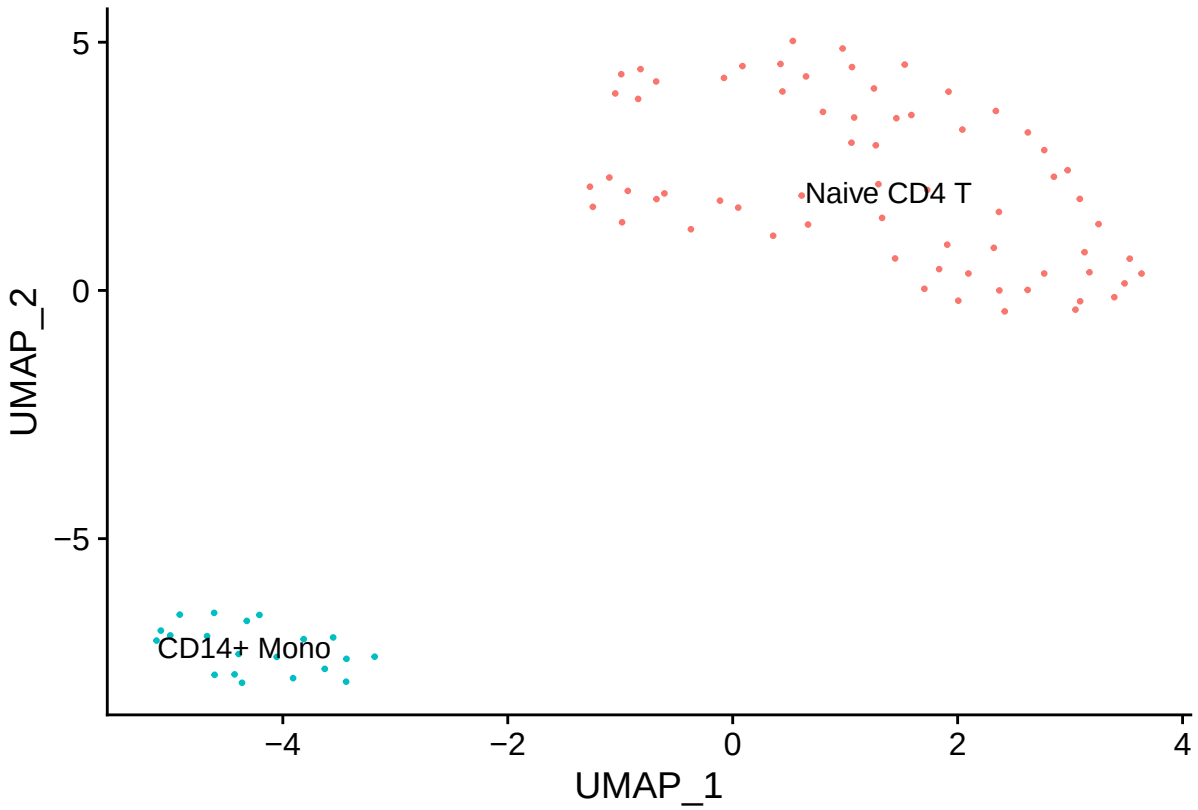
```
## Warning in FeaturePlot(pbmc, features = c("MS4A1", "GZMB", "CD3E", "CD14", : All
## cells have the same value (0) of CD3E.
```




```
#talk to a biologist

new.cluster.ids <- c("Naive CD4 T","CD14+ Mono", "Memory CD4 T","B","CD8 T","FCGR3A+ Mono","NK","DC","P
names(new.cluster.ids) <- levels(pbmc)
pbmc <- RenameIdents(pbmc,new.cluster.ids)
```

```
## Warning: Cannot find identity NA

## Warning: Cannot find identity NA

## Warning: Cannot find identity NA

## Warning: Cannot find identity NA

## Warning: Cannot find identity NA

## Warning: Cannot find identity NA

## Warning: Cannot find identity NA
```

```
DimPlot(pbmc, reduction="umap",label=TRUE,pt.size=0.5) + NoLegend()
```


Naive CD4 T
CD14+ Mono
UMAP_2
UMAP_1
5
0
−5
−4
−2
0
2
4