

Analysis 11 - Comparing clusters of tf2 large

Based off of lcmSOM_analysis10_122916_redo, with newly mapped genes.

Purpose

Comparing the clusters for two purposes

1. Quality of Analysis - To understand how consistent different clusteres are, therefore how conservative the results are.
2. Decrease the amount of interesting genes in clusters by comparing WT SOM with *tf2* superSOM.

Analysis Set-up

Required Libraries

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2
library(reshape)

## Warning: package 'reshape' was built under R version 3.3.2
library(kohonen)

## Warning: package 'kohonen' was built under R version 3.3.2
library(goseq)

## Loading required package: BiasedUrn
## Loading required package: geneLenDataBase
##
library(GO.db)

## Loading required package: AnnotationDbi
## Warning: package 'AnnotationDbi' was built under R version 3.3.1
## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':
## 
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
## 
##   IQR, mad, xtabs
```

```

## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, cbind, colnames,
##     do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, lengths, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff,
##     sort, table, tapply, union, unique, unsplit

## Loading required package: Biobase

## Welcome to Bioconductor
##
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: IRanges

## Loading required package: S4Vectors

## Warning: package 'S4Vectors' was built under R version 3.3.1

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:reshape':
##
##     expand, rename

## The following objects are masked from 'package:base':
##
##     colMeans, colSums, expand.grid, rowMeans, rowSums

library(knitr)

## Warning: package 'knitr' was built under R version 3.3.2
library(somtools)
library(cowplot)

## Warning: package 'cowplot' was built under R version 3.3.2

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':
##
##     ggsave

#source("./clusterFunctions.R")

```

Upload that dataset:

```

genes25 <- read.csv("../data/output/analysis4.top25_19Oct2017.csv")
colnames(genes25)

```

```

## [1] "X"          "gene"        "tf2ambr"     "tf2aother"   "tf2bmbr"
## [6] "tf2bother"  "tf2cmbr"     "tf2cother"   "wtambr"      "wtaother"
## [11] "wtbmbr"     "wtbother"    "wtcmbr"      "wtcother"   "sd"
## [16] "average"    "cv"

```

```

## Only tf2
genes25 <- genes25[,c(2:8)]

scale_data <- as.matrix(t(scale(t(genes25[c(2:7)]))))
pca <- prcomp(scale_data, scale = TRUE)
summary(pca)

## Importance of components:
##              PC1     PC2     PC3     PC4     PC5     PC6
## Standard deviation   1.468  1.1679  1.0943  0.9296  0.6481 1.773e-15
## Proportion of Variance 0.359  0.2273  0.1996  0.1440  0.0700 0.000e+00
## Cumulative Proportion  0.359  0.5864  0.7860  0.9300  1.0000 1.000e+00

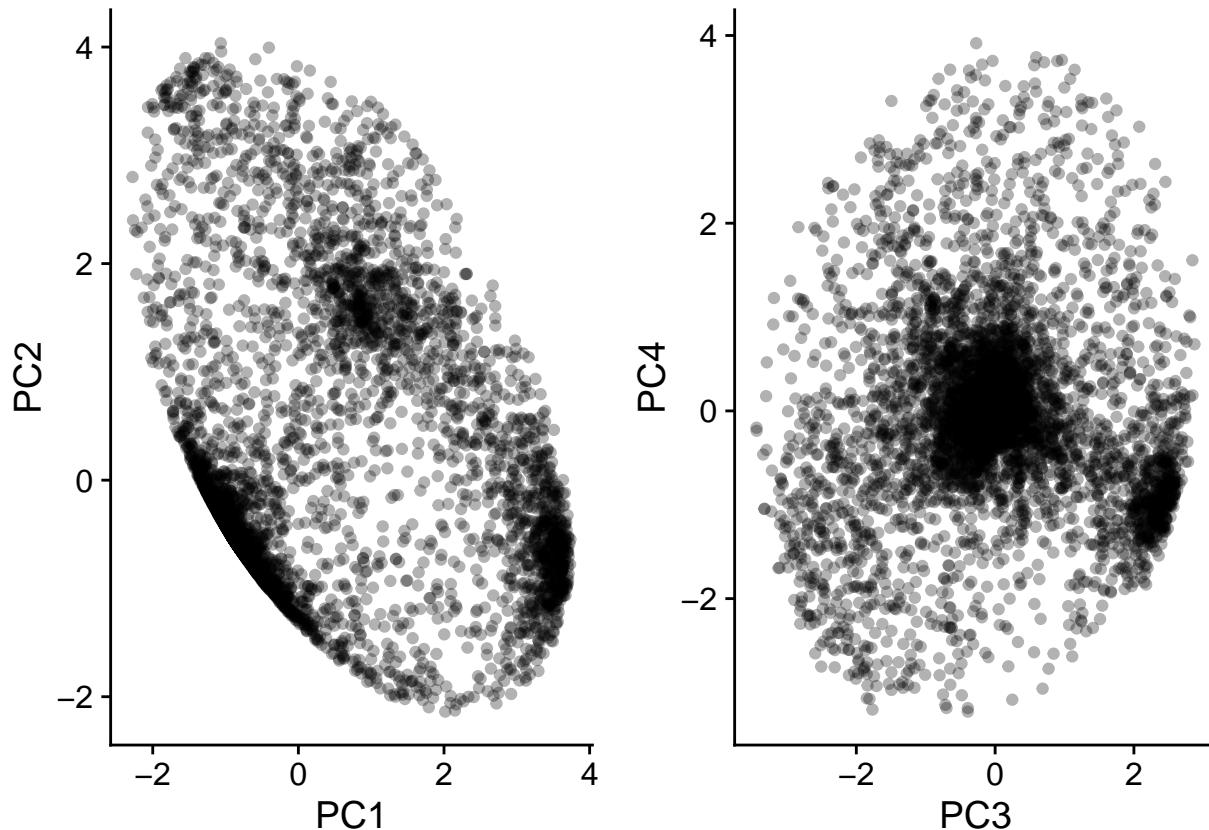
pca.scores <- data.frame(pca$x)
data.val <- cbind(genes25, scale_data, pca.scores)

PCA12 <- ggplot(data.val, aes(PC1, PC2)) +
  geom_point(alpha = .3)

PCA34 <- ggplot(data.val, aes(PC3, PC4)) +
  geom_point(alpha = .3)

plot_grid(PCA12, PCA34, labels = c("", ""), ncol = 2)

```



SOM

```
# subset only the scaled gene expression values
# colnames(scale_data) <- NULL

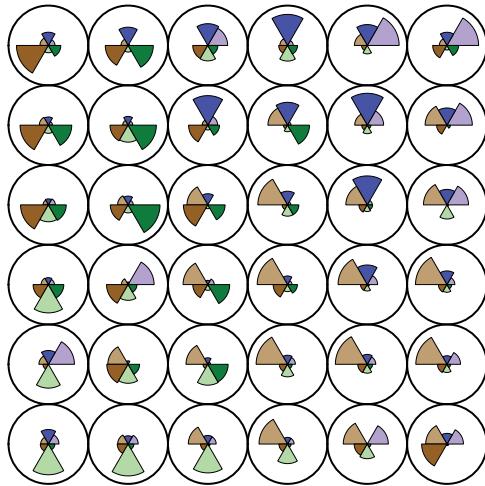
set.seed(6)
som <- som(scale_data, somgrid(6,6)) # This is where you change the size of the map
summary(som)

## SOM of size 6x6 with a rectangular topology and a bubble neighbourhood function.
## Training data included of 6582 objects
## The number of layers is 1
## Mean distance to the closest unit in the map: 0.3675677

lcmPaletteColors <- c( "#b3a2ce", "#4753a4", "#bf9e71", "#956025", "#b2d9a6", "#0f7c3e")
lcmPalette <- function(n, alpha =1){
  lcmPaletteColors[1:n]
}

plot(som, type = "codes", palette.name = lcmPalette)
```

Codes plot



| | | | |
|-----------------|-----------|----------------------|-----------|
| [purple square] | tf2ambr | [brown square] | tf2bother |
| [blue square] | tf2aother | [light green square] | tf2cmbr |
| [tan square] | tf2bmbr | [green square] | tf2cother |

```
## Bind information
data.val2 <- cbind(data.val,som$unit.classif,som$distances)
names(data.val2)

## [1] "gene"                  "tf2ambr"                "tf2aother"
## [4] "tf2bmbr"               "tf2bother"               "tf2cmbr"
## [7] "tf2cother"              "tf2ambr"                "tf2aother"
## [10] "tf2bmbr"                "tf2bother"               "tf2cmbr"
```

```

## [13] "tf2cother"          "PC1"                  "PC2"
## [16] "PC3"                 "PC4"                  "PC5"
## [19] "PC6"                 "som$unit.classif" "som$distances"
##rename
names(data.val2)[20:21] <- c("unitClassIfA", "distanceA")

```

Visualization

```

# Palette
lcmPaletteColors <- c( "#b3a2ce", "#4753a4", "#bf9e71", "#956025", "#b2d9a6", "#0f7c3e")

```

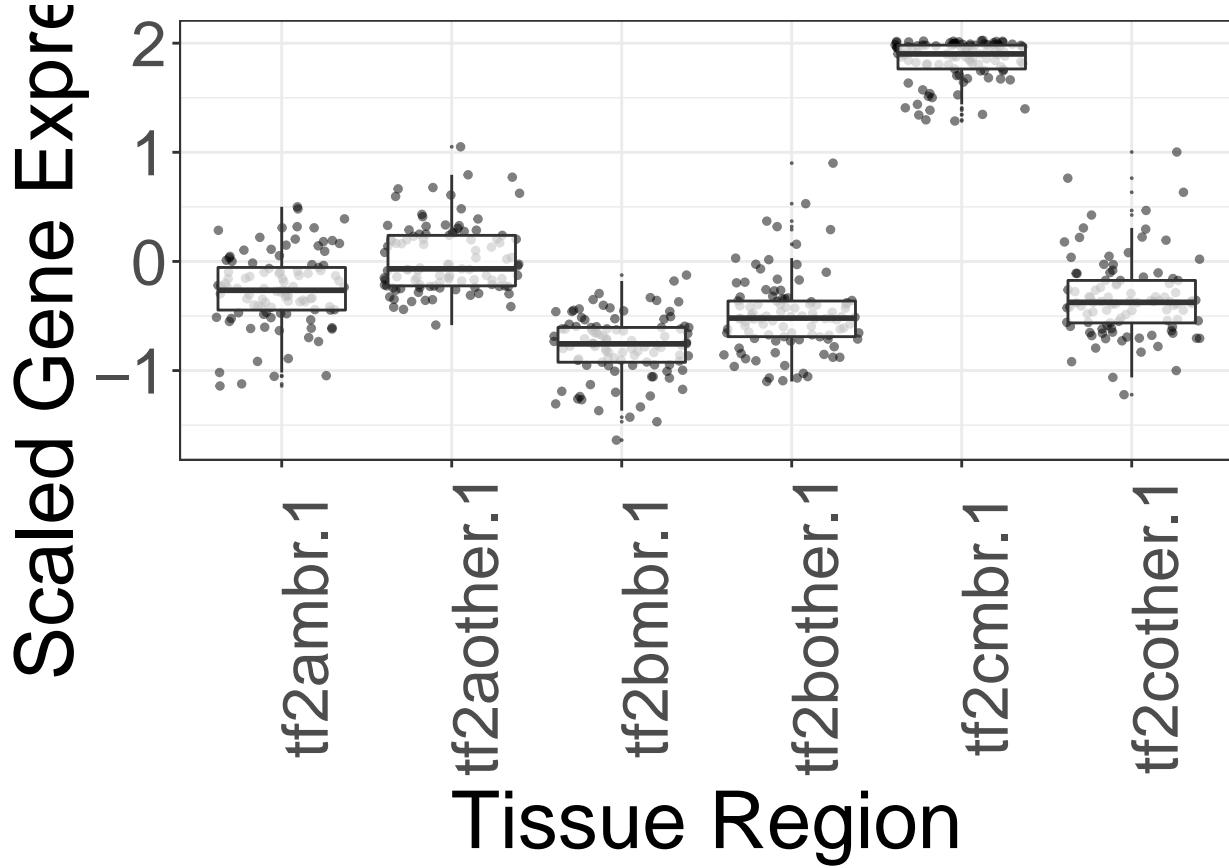
Cluster 1

```

clusterVis(1)

## Using gene as id variables

```



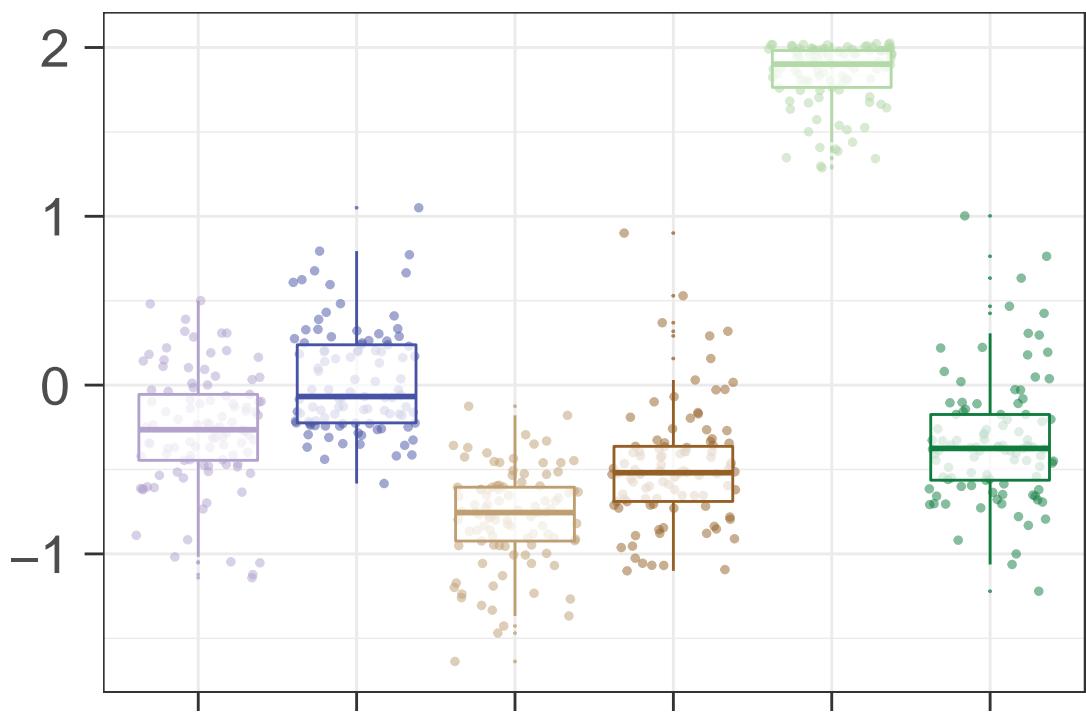
```

## clusterVis_region(1)
clusterVis_color(1)

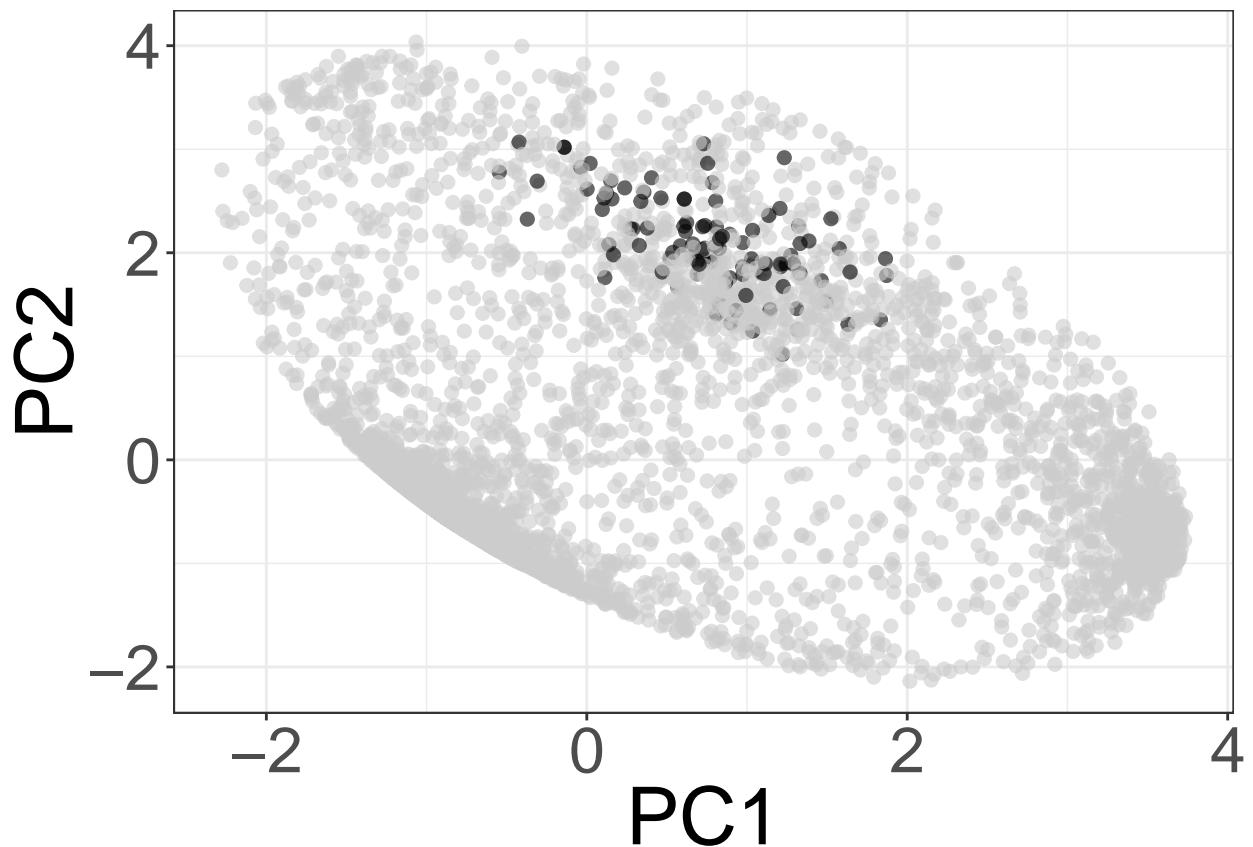
## Using gene as id variables

```

Scaled Gene Expression



clusterVis_PCA(1)

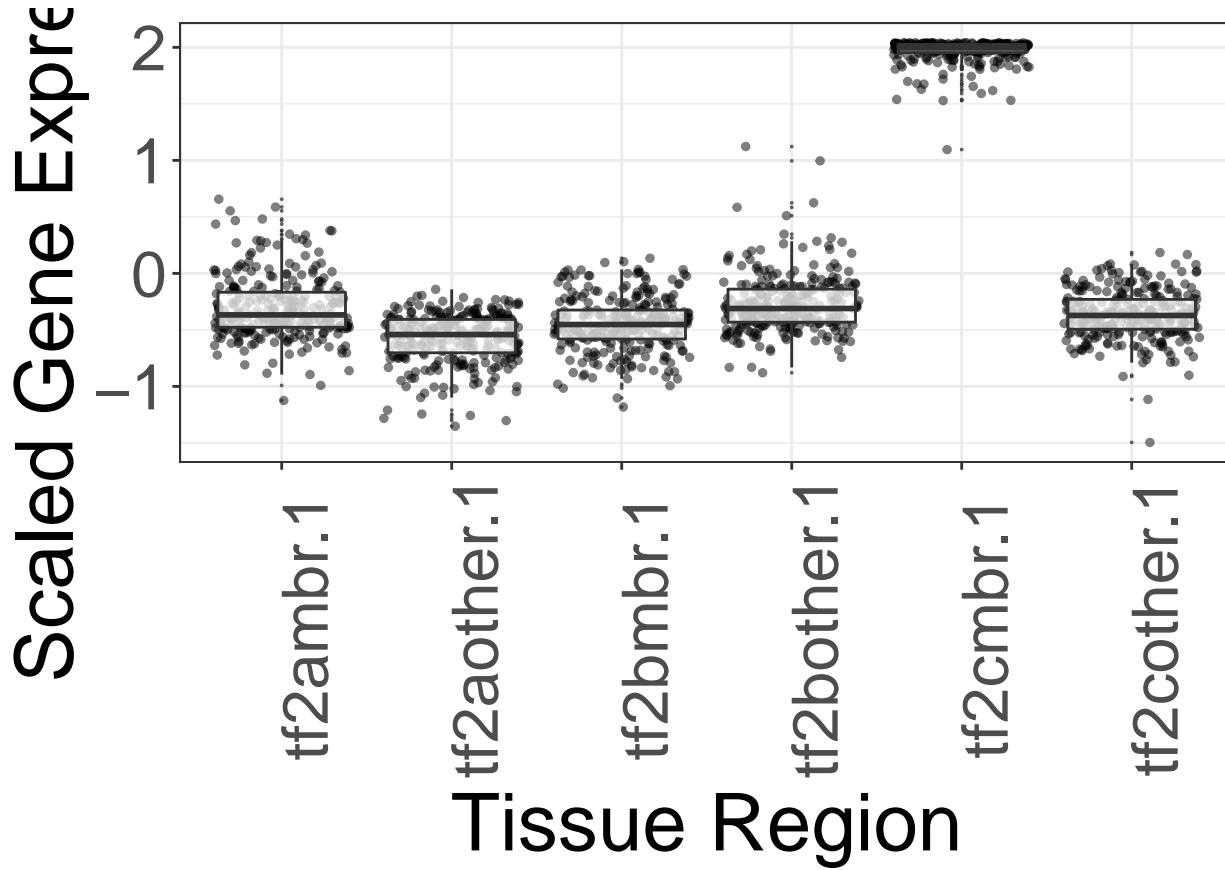


```
## y <- genesInClust(1, data.val2, annotation)
## kable(y, format = "latex", booktabs = TRUE)
```

Cluster 2

```
clusterVis(2)
```

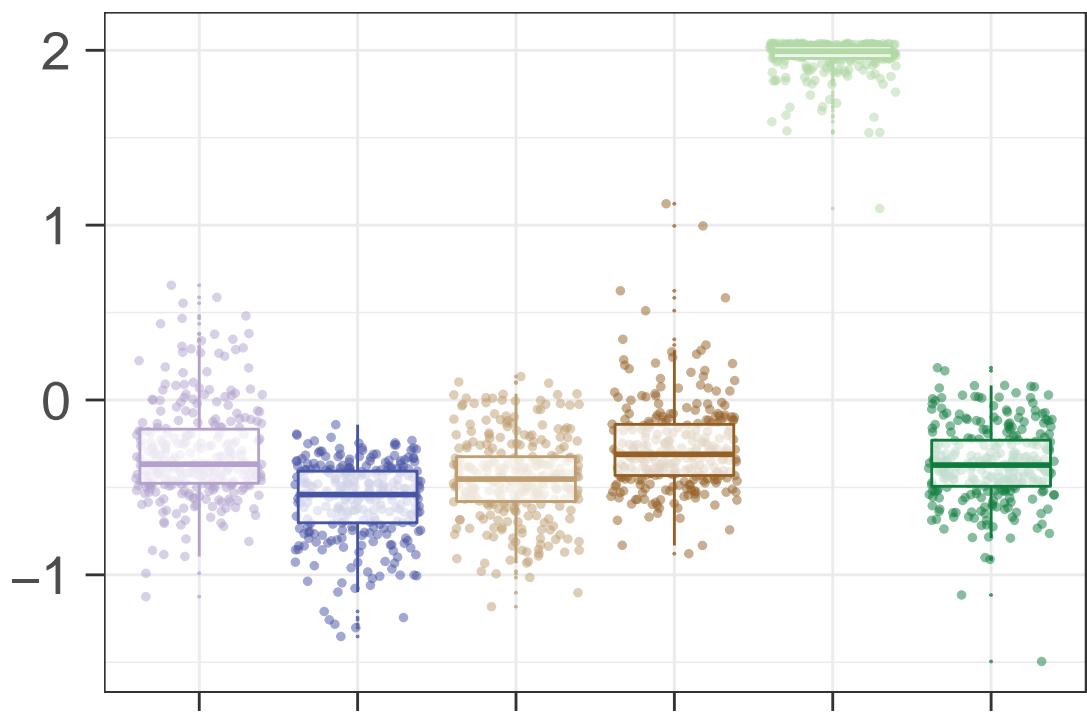
```
## Using gene as id variables
```



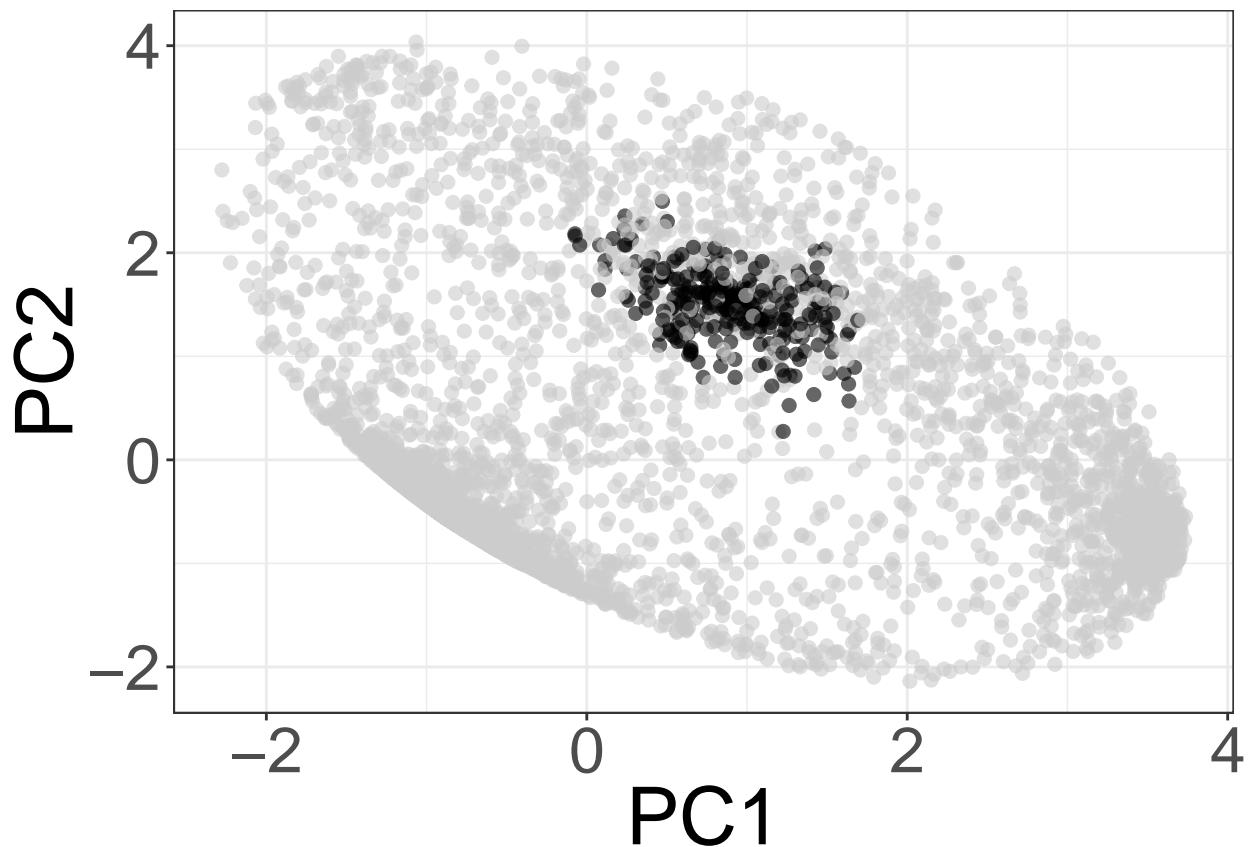
```
clusterVis_color(2)
```

```
## Using gene as id variables
```

Scaled Gene Expression



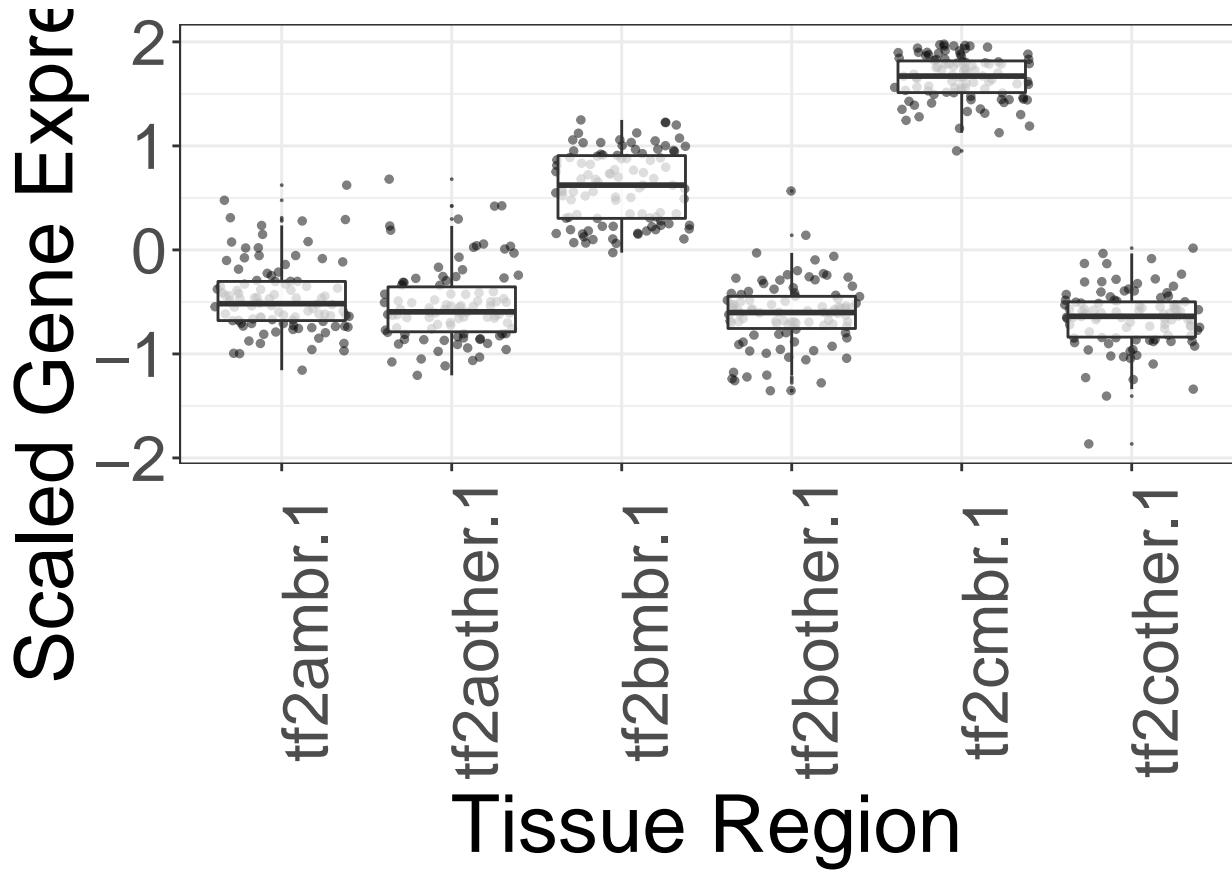
clusterVis_PCA(2)



Cluster 3

```
clusterVis(3)
```

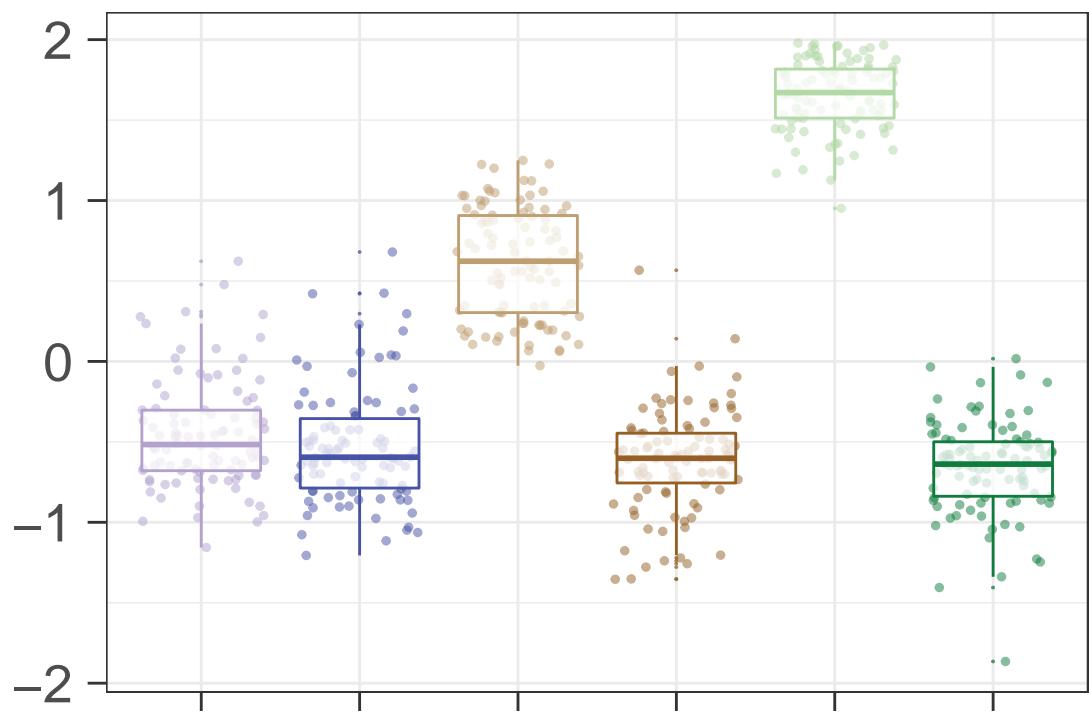
```
## Using gene as id variables
```



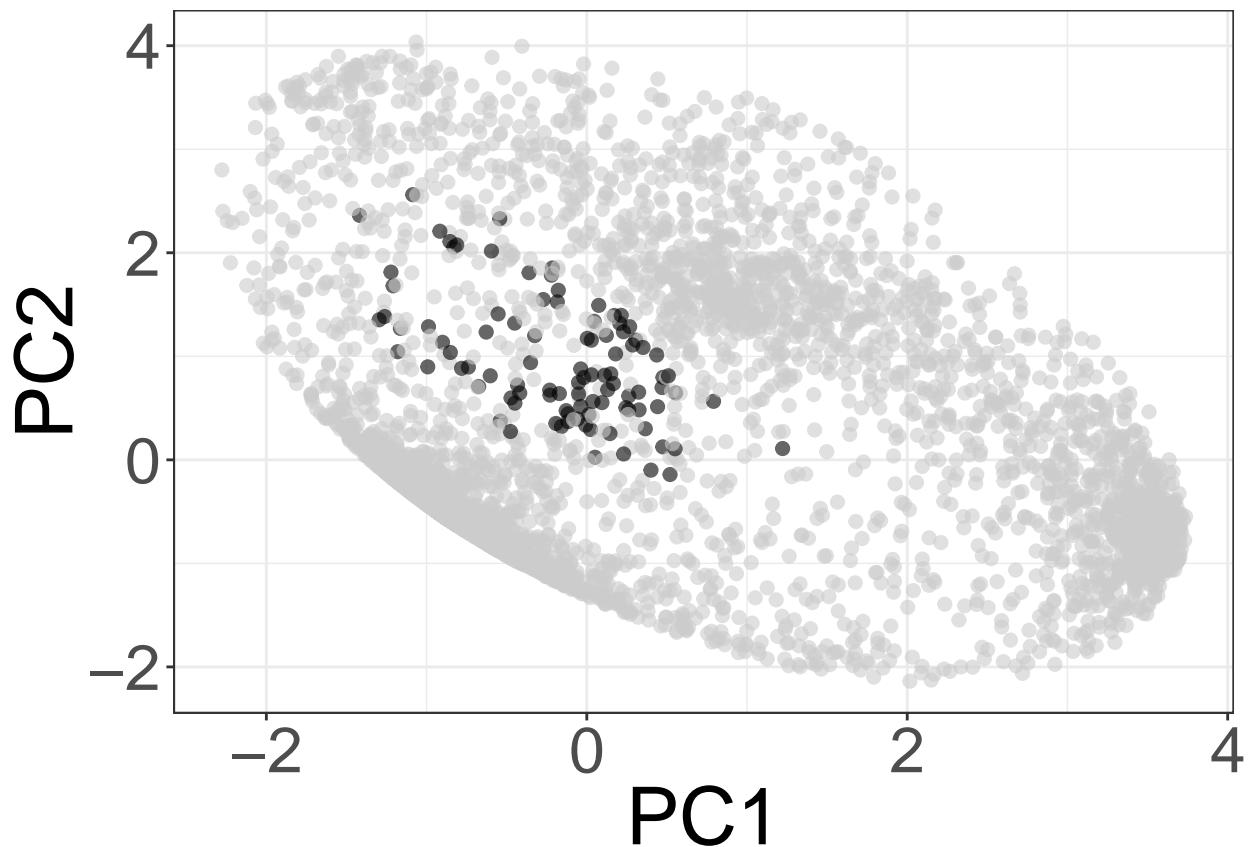
```
clusterVis_color(3)
```

```
## Using gene as id variables
```

Scaled Gene Expression



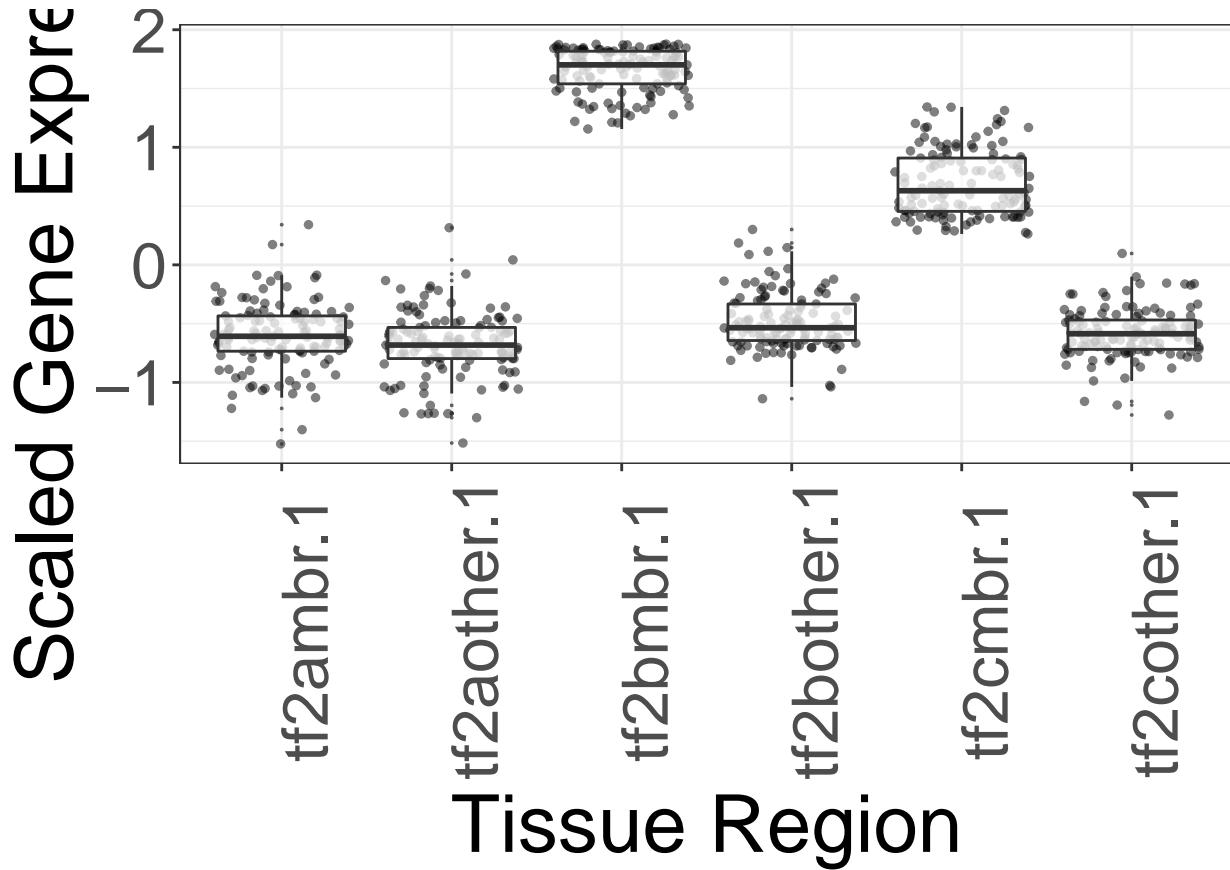
clusterVis_PCA(3)



Cluster 4

```
clusterVis(4)
```

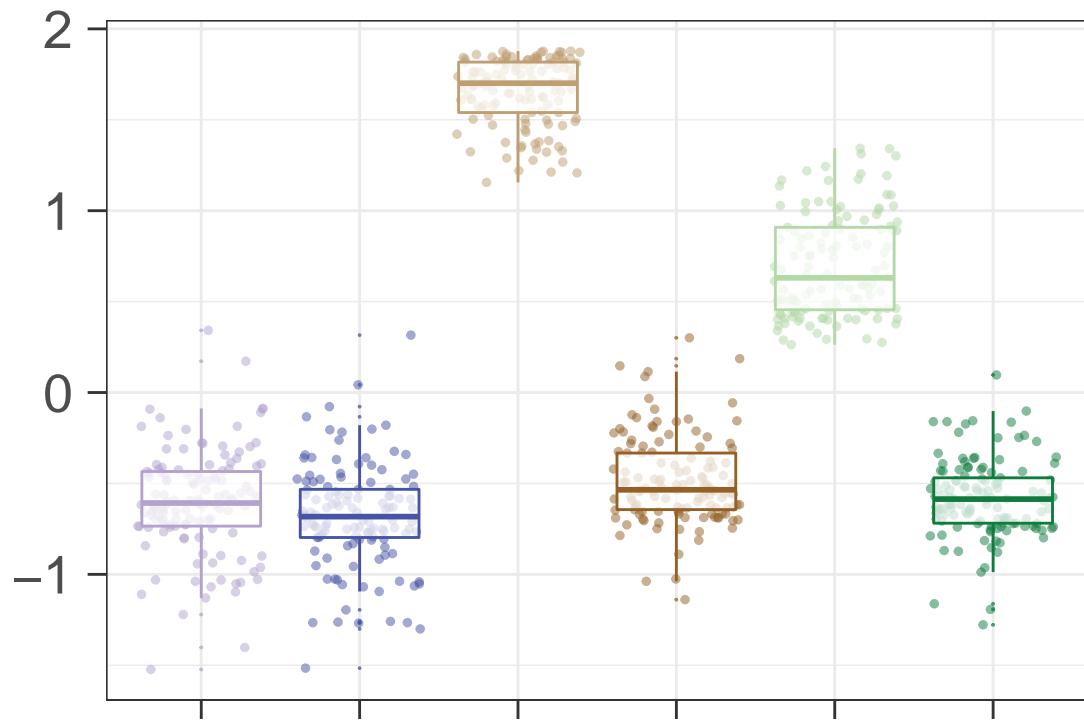
```
## Using gene as id variables
```



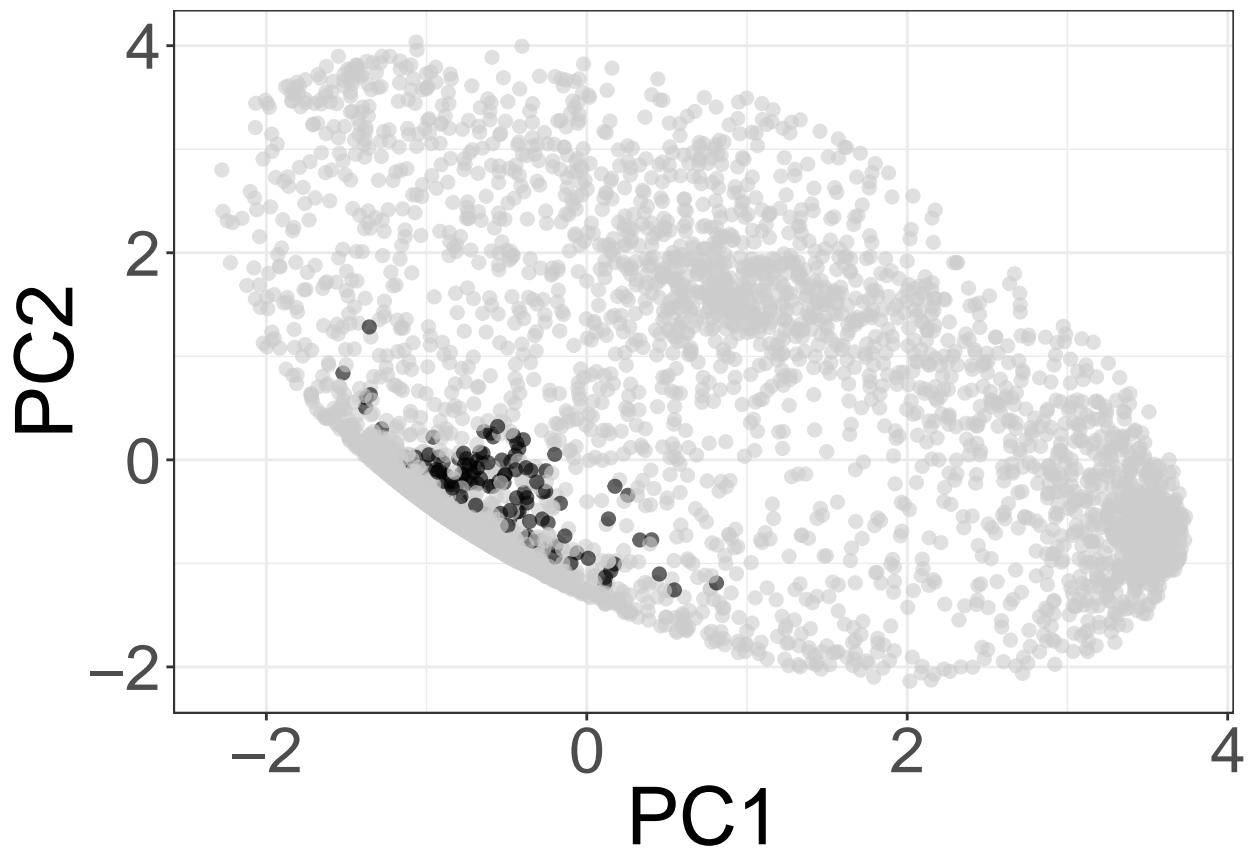
```
clusterVis_color(4)
```

```
## Using gene as id variables
```

Scaled Gene Expression



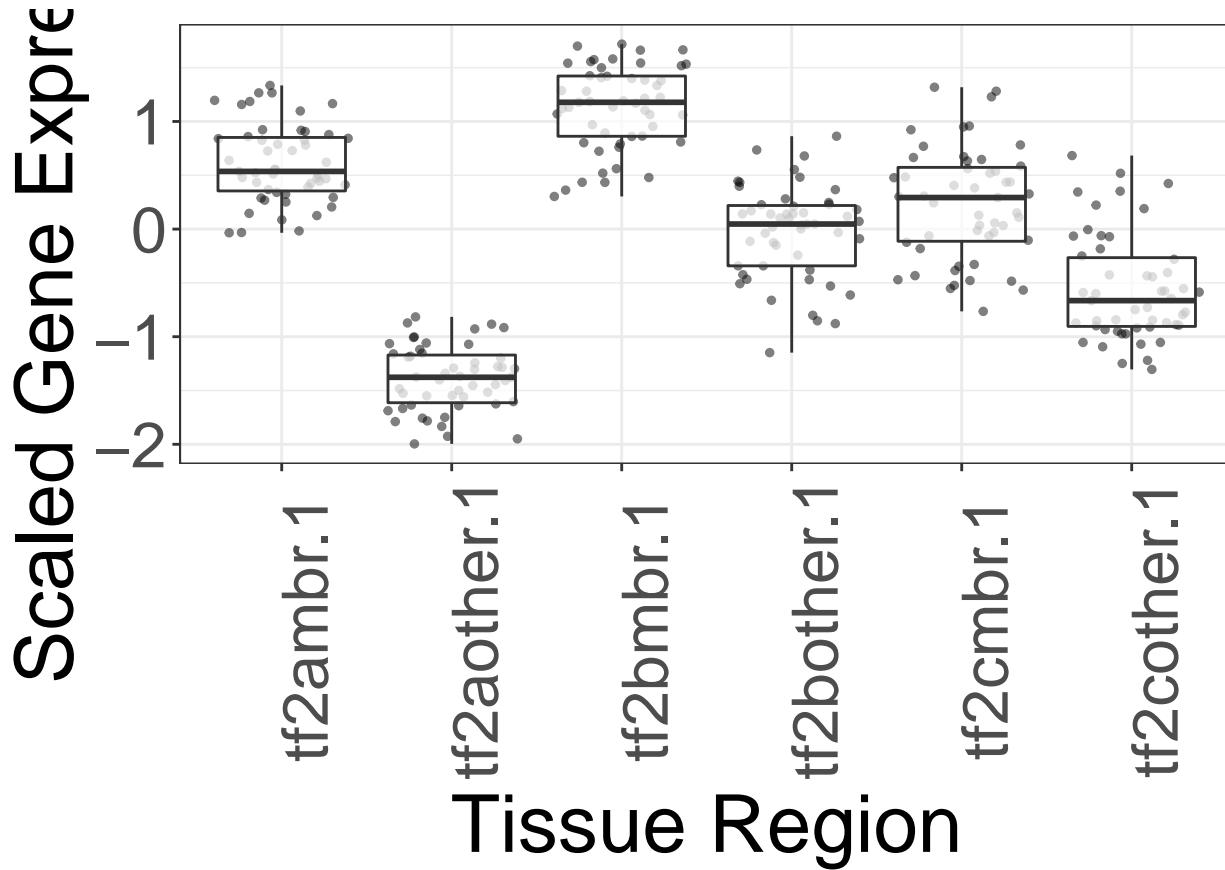
clusterVis_PCA(4)



Cluster 5

```
clusterVis(5)
```

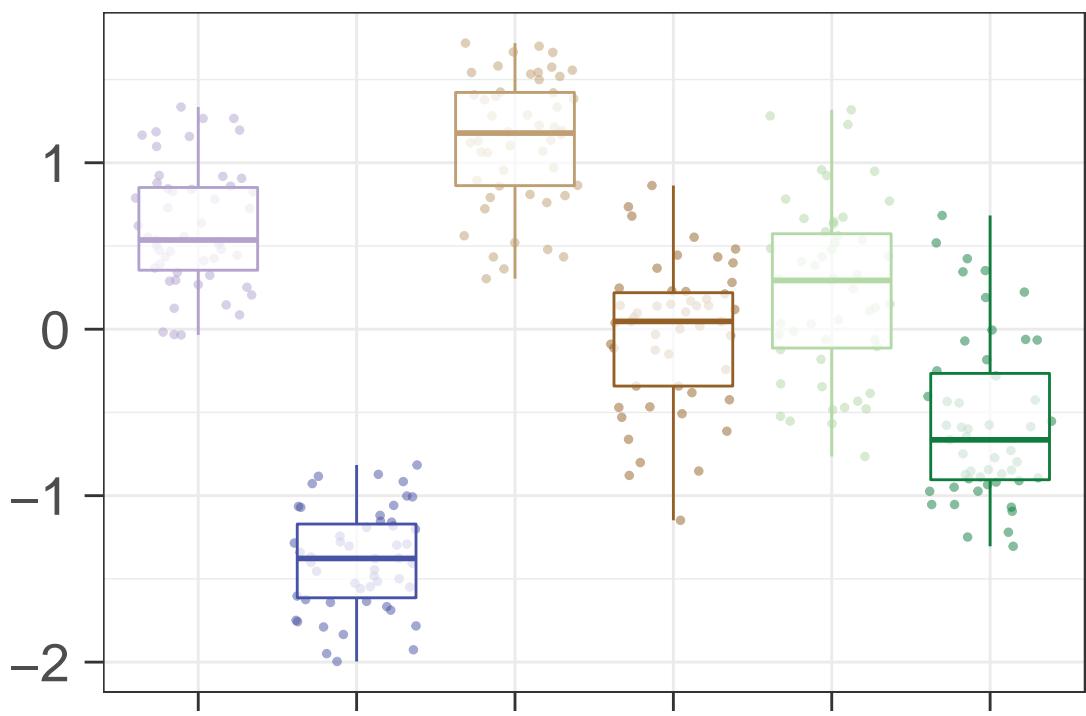
```
## Using gene as id variables
```



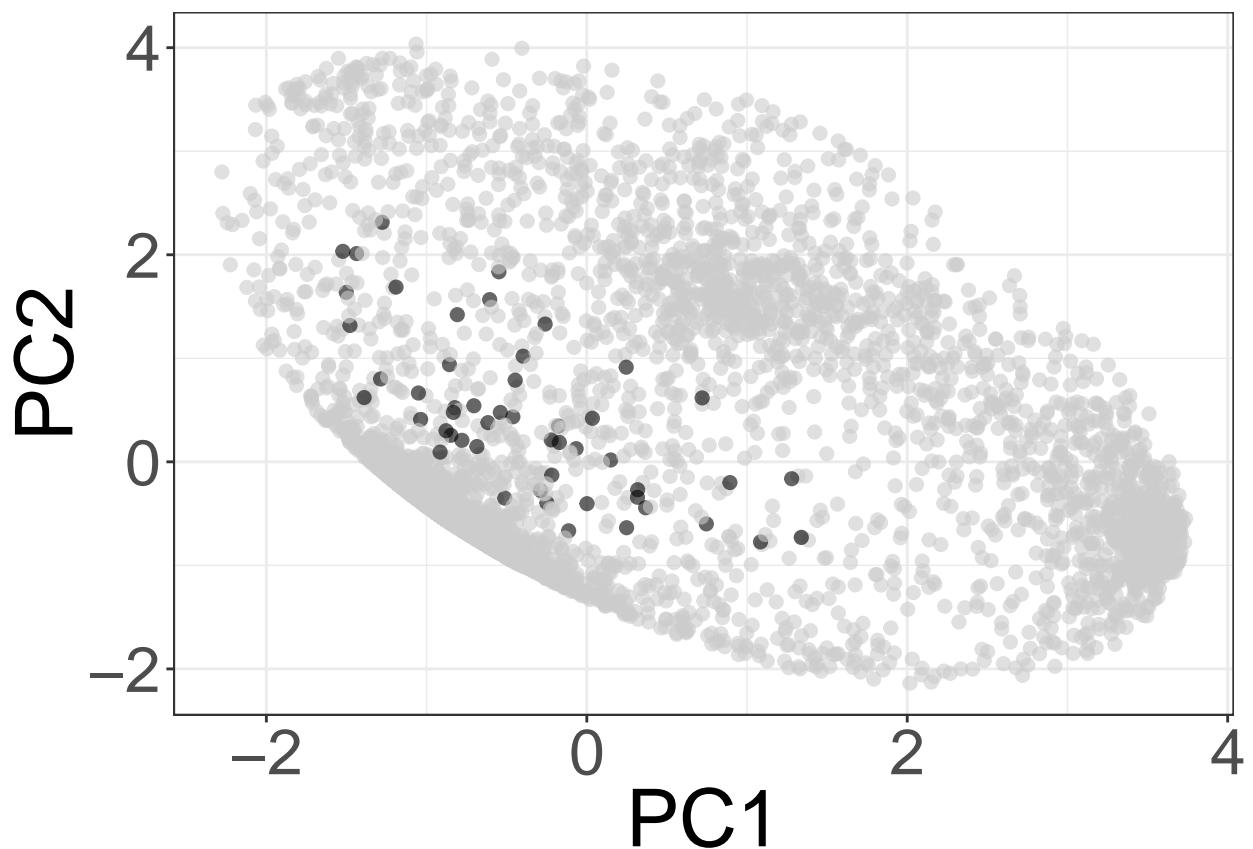
```
clusterVis_color(5)
```

```
## Using gene as id variables
```

Scaled Gene Expression



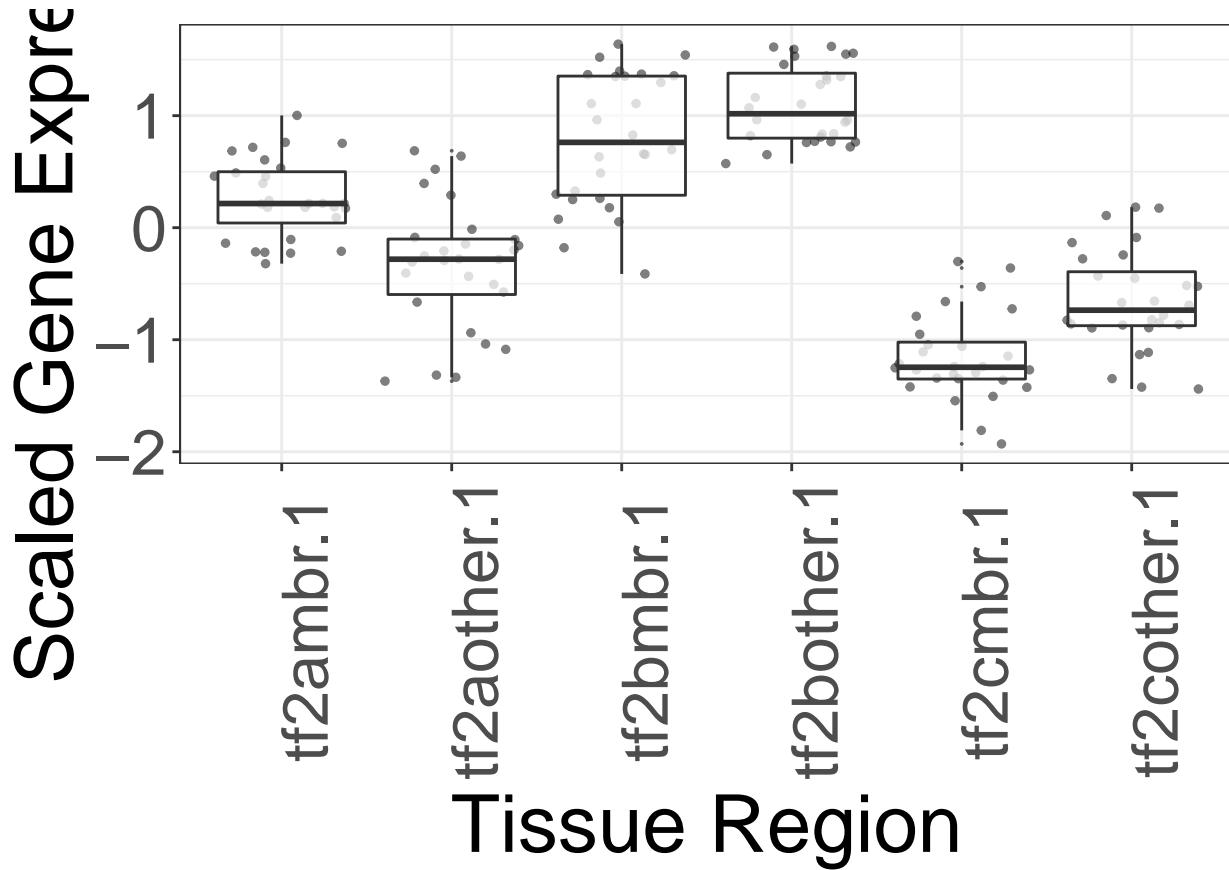
clusterVis_PCA(5)



Cluster 6

```
clusterVis(6)
```

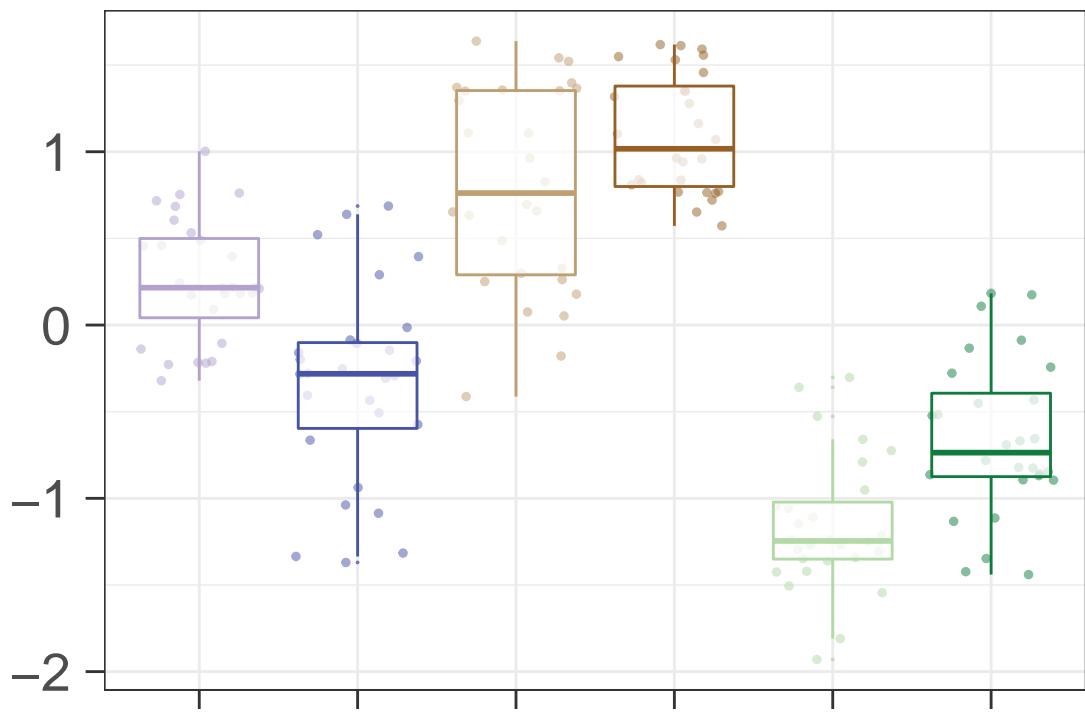
```
## Using gene as id variables
```



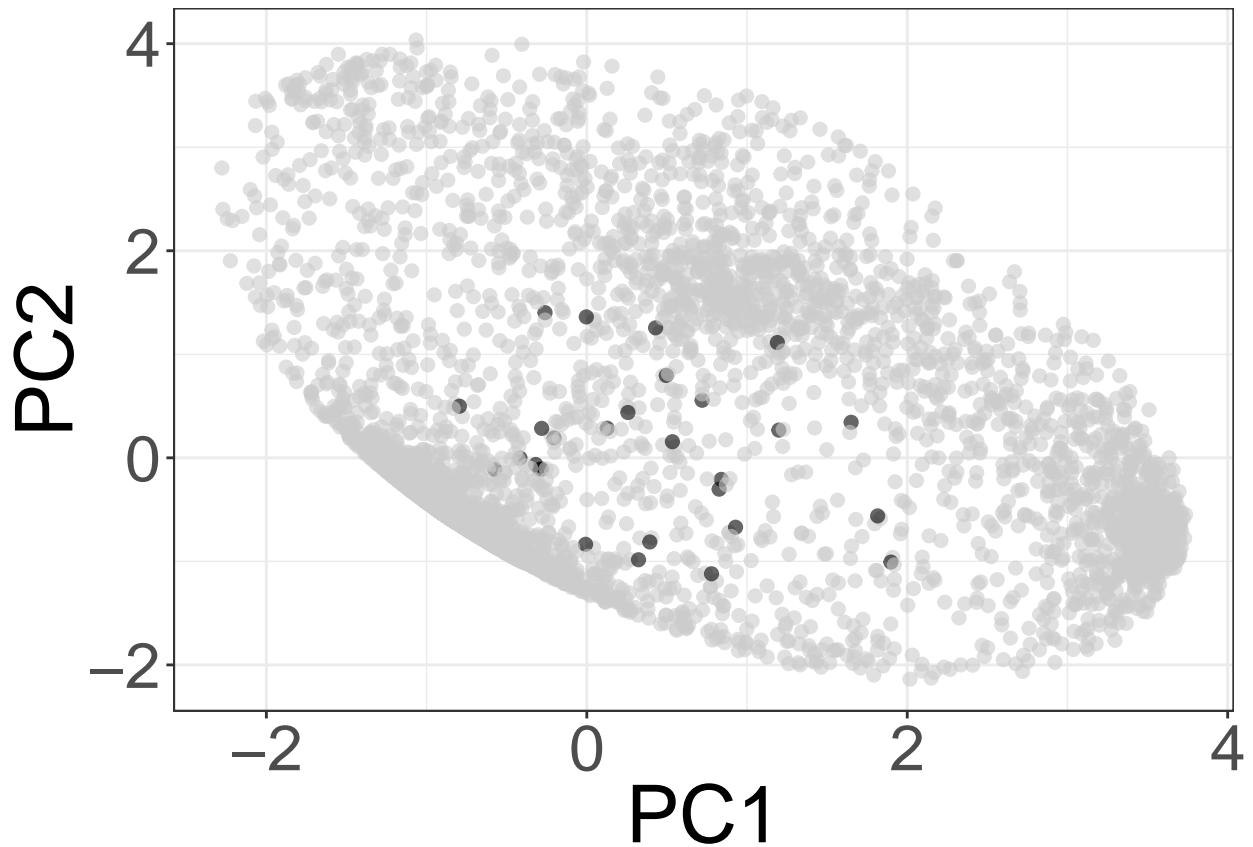
```
clusterVis_color(6)
```

```
## Using gene as id variables
```

Scaled Gene Expression



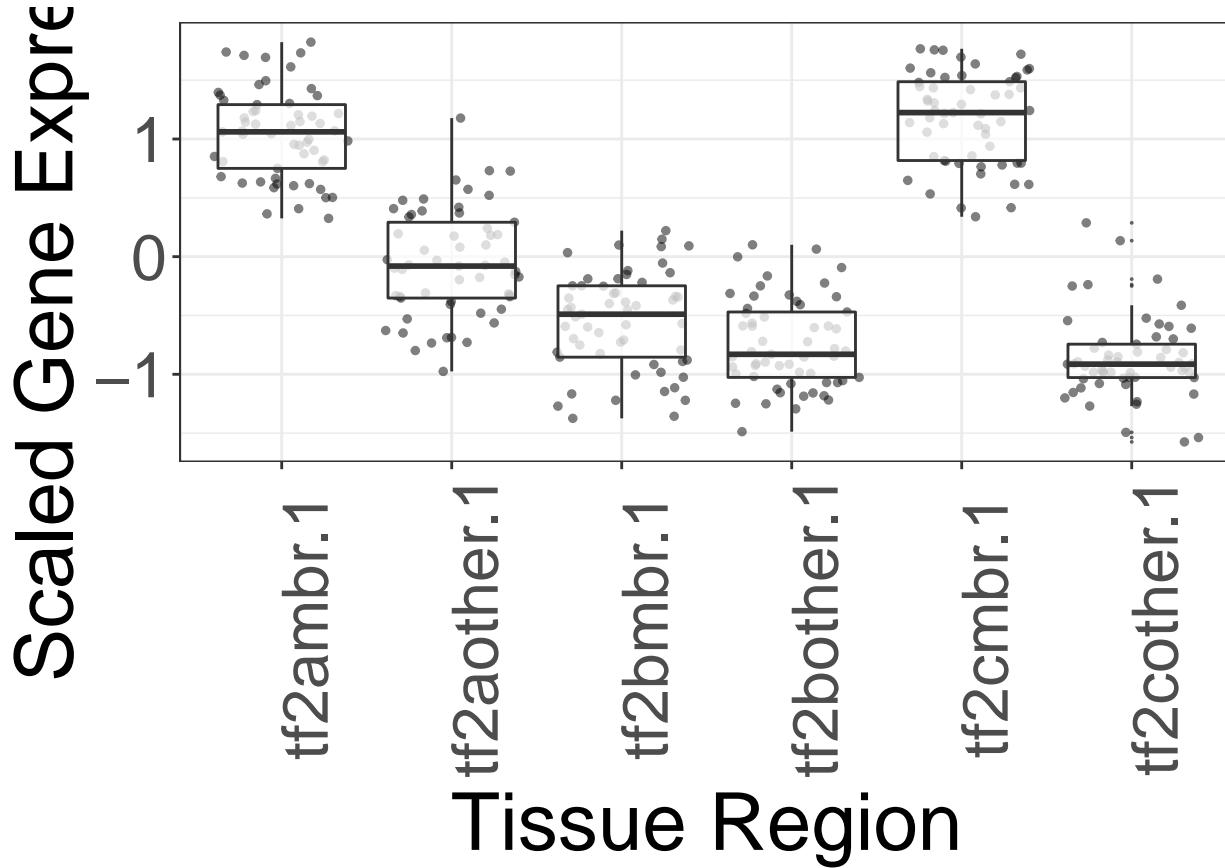
clusterVis_PCA (6)



Cluster 7

```
clusterVis(7)
```

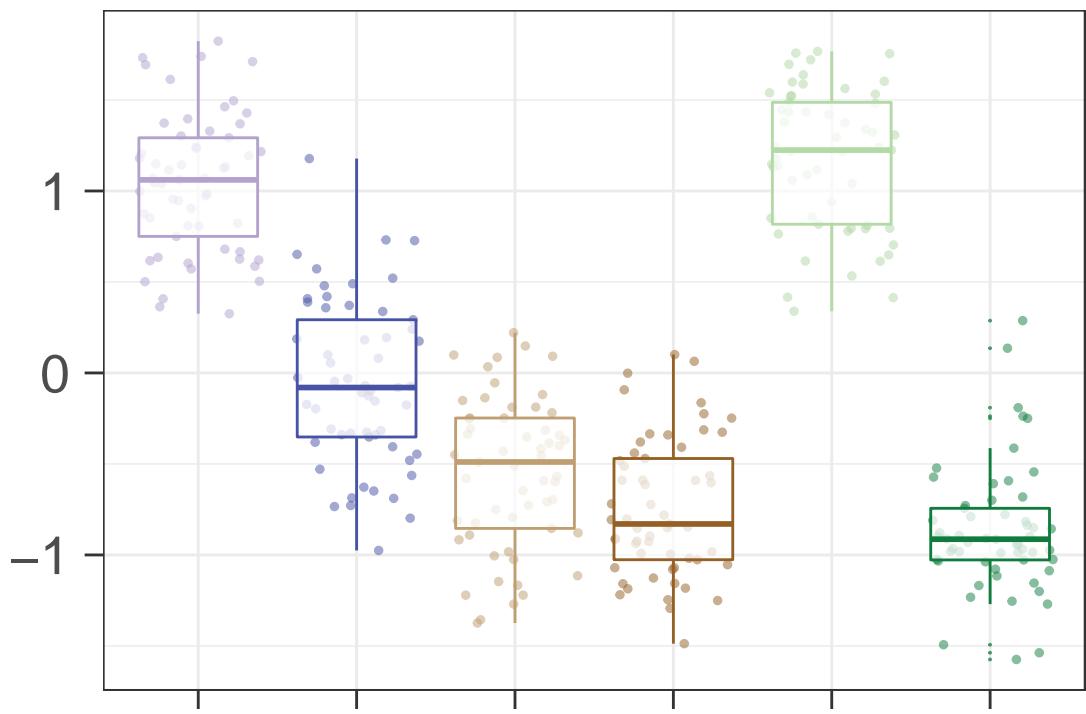
```
## Using gene as id variables
```



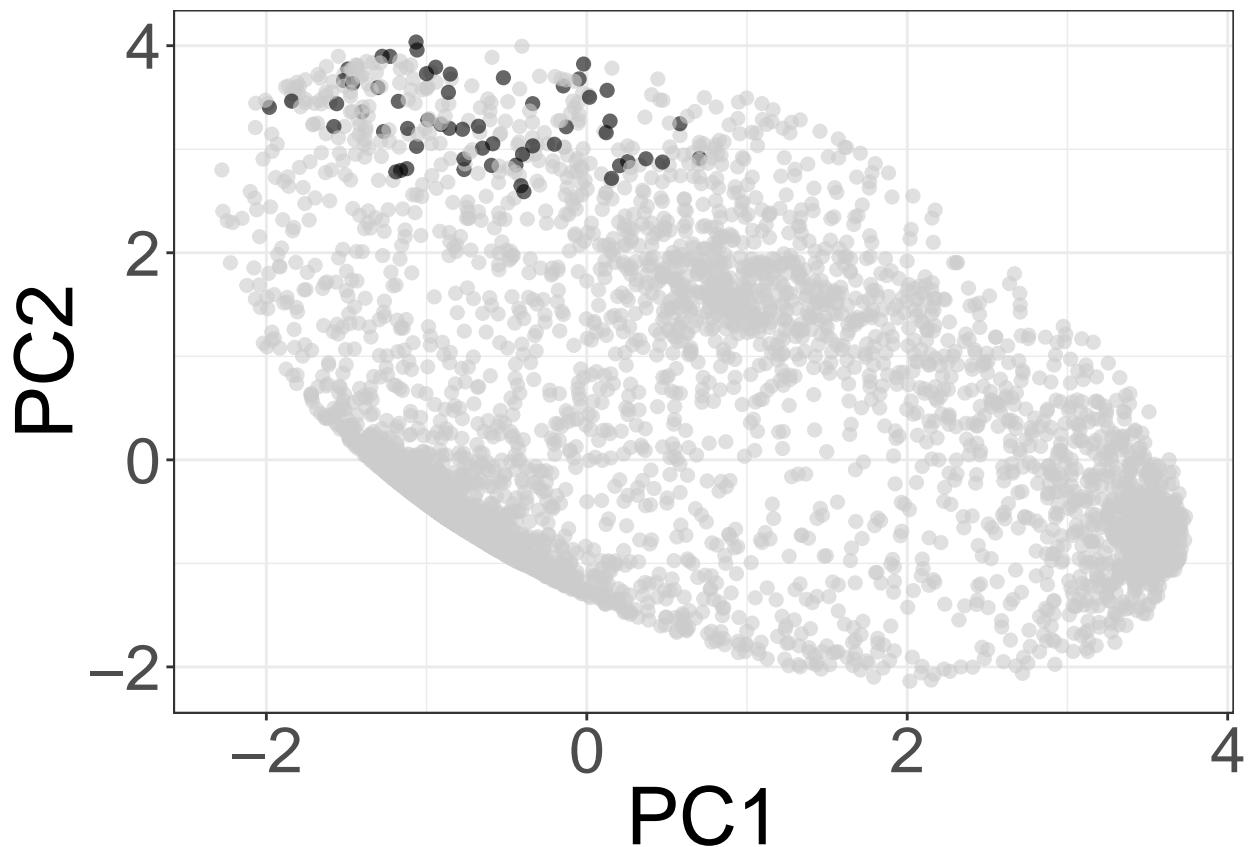
```
clusterVis_color(7)
```

```
## Using gene as id variables
```

Scaled Gene Expression



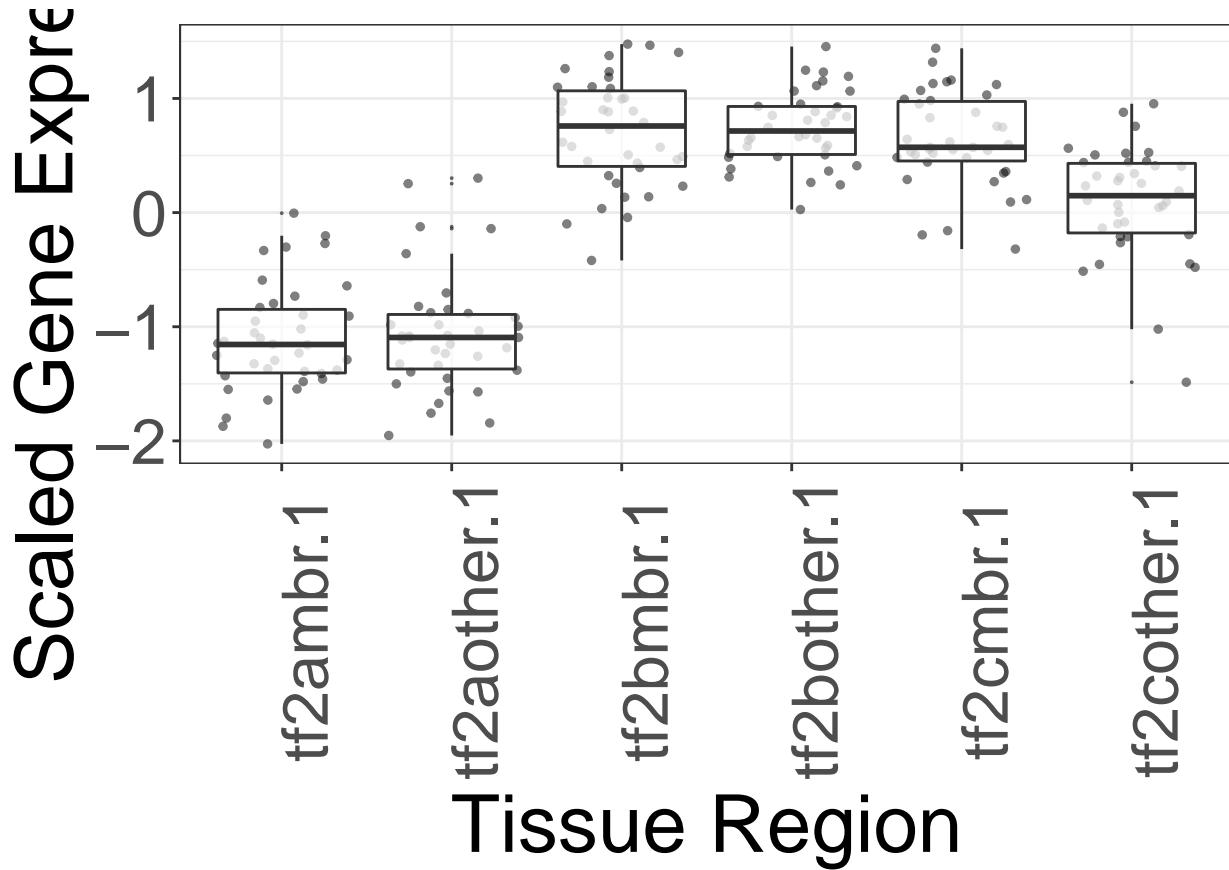
clusterVis_PCA(7)



Cluster 8

```
clusterVis(8)
```

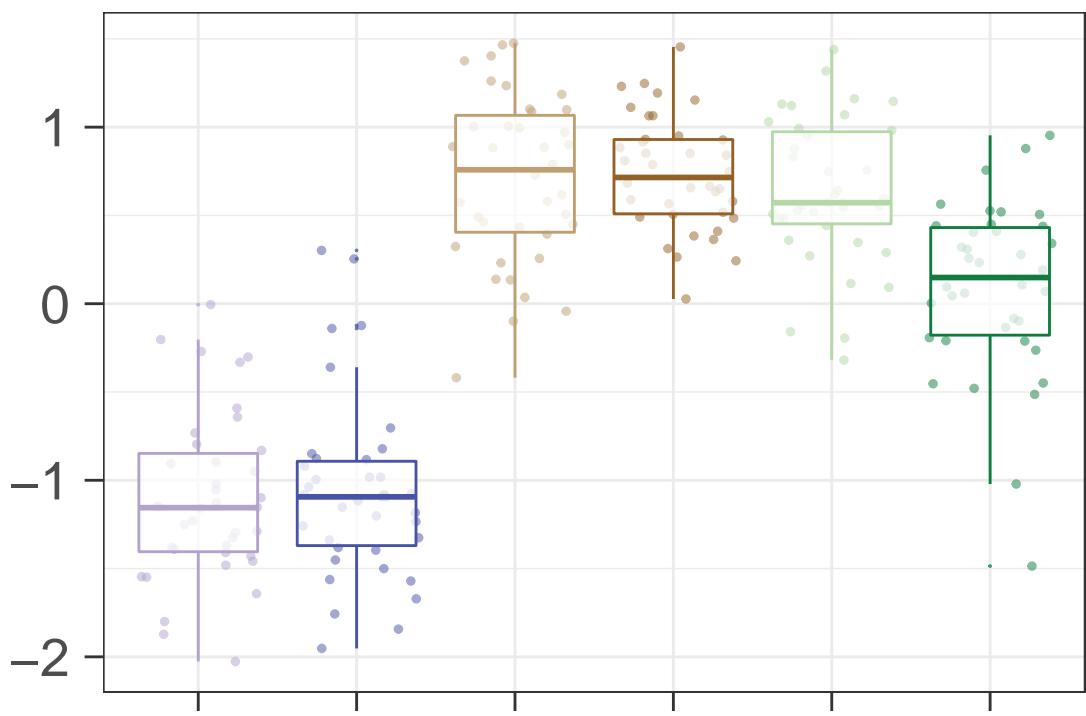
```
## Using gene as id variables
```



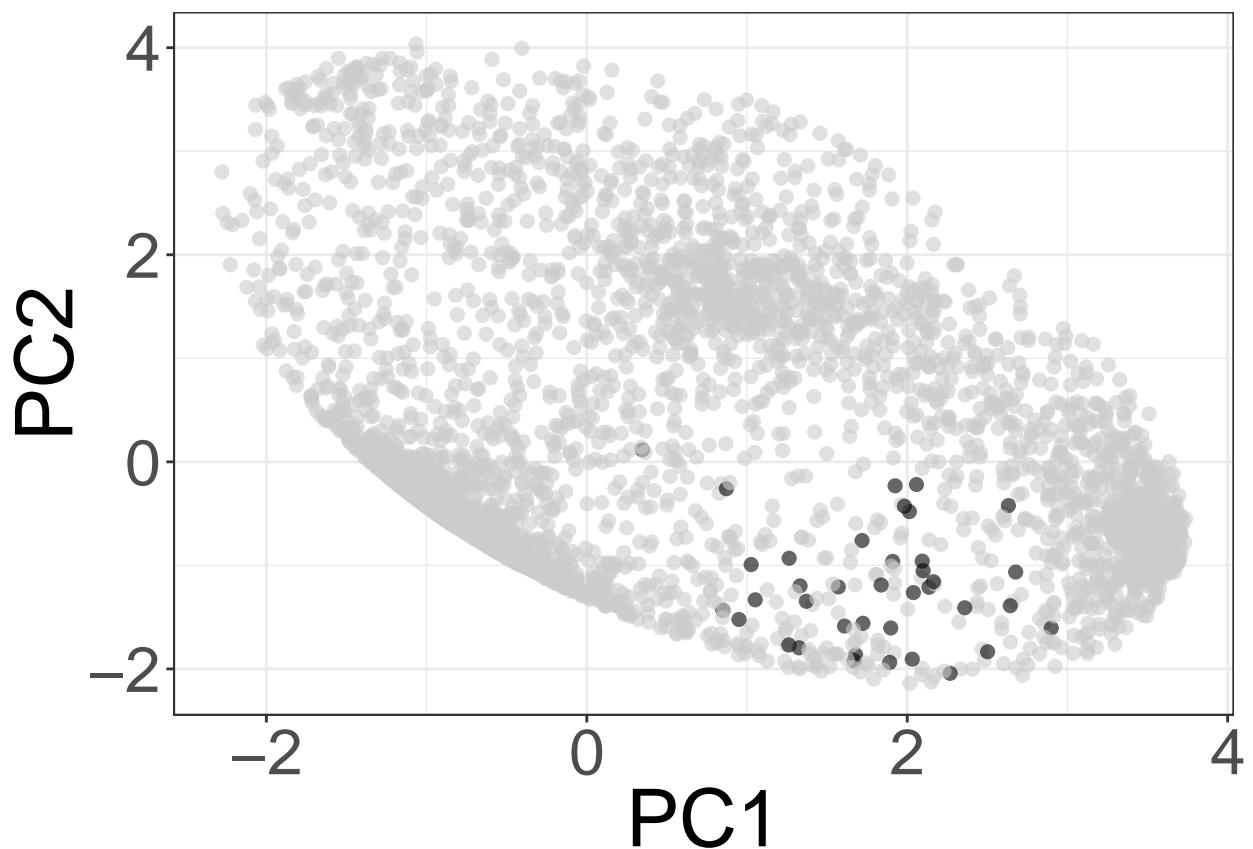
```
clusterVis_color(8)
```

```
## Using gene as id variables
```

Scaled Gene Expression



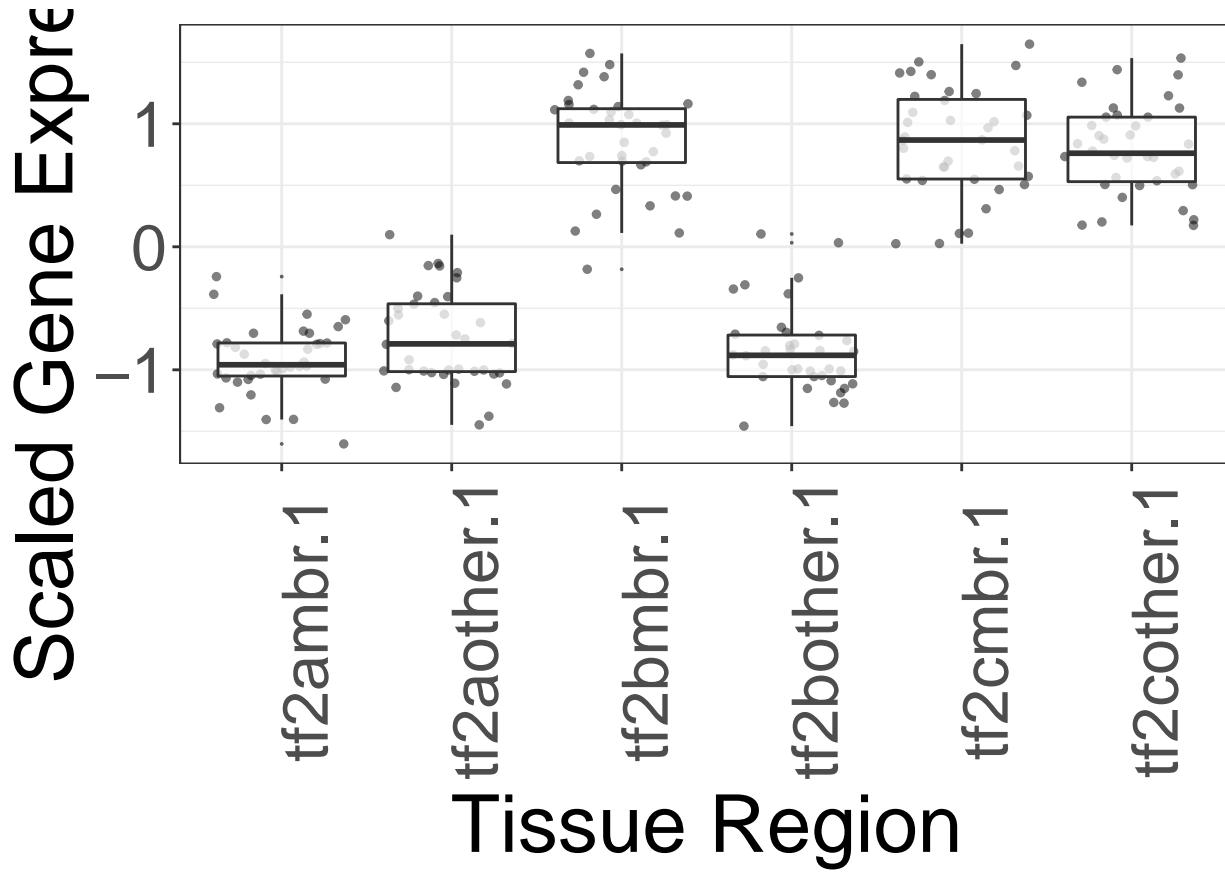
clusterVis_PCA(8)



Cluster 9

```
clusterVis(9)
```

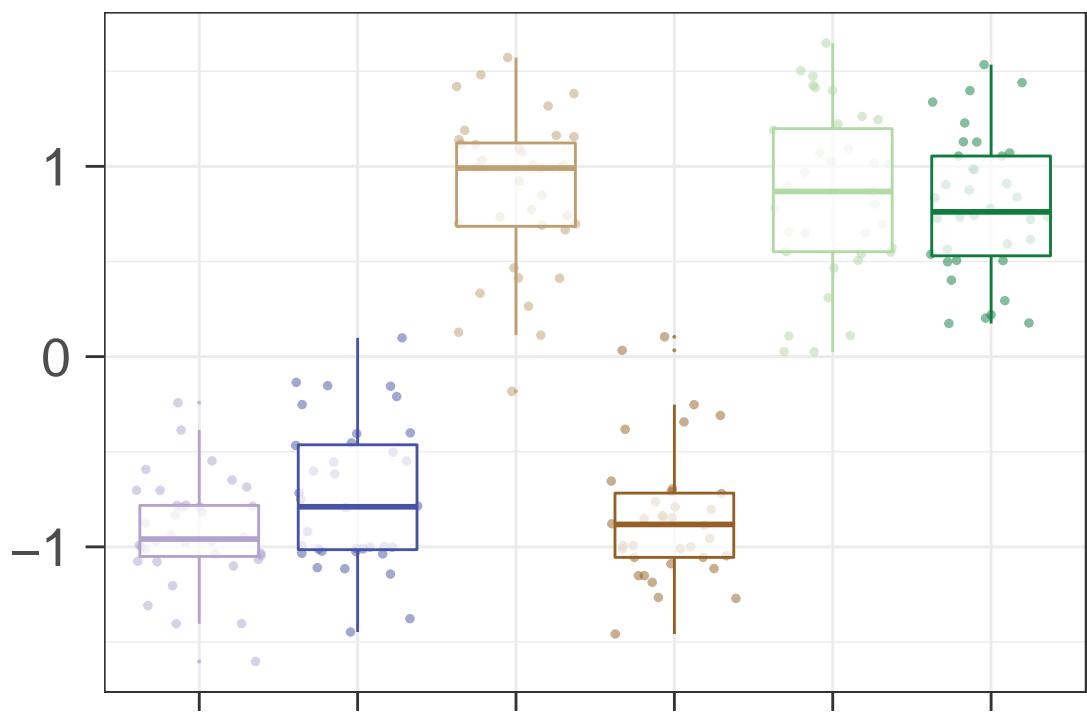
```
## Using gene as id variables
```



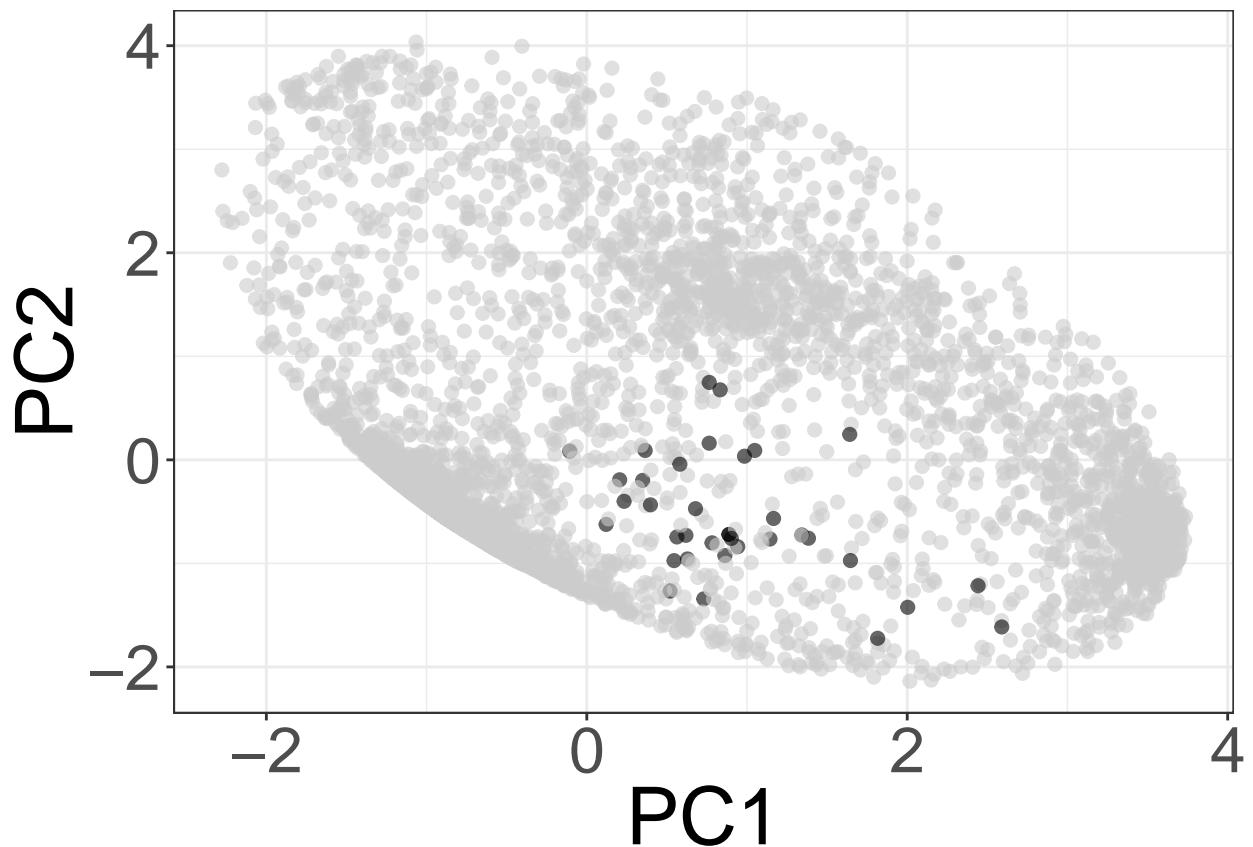
```
clusterVis_color(9)
```

```
## Using gene as id variables
```

Scaled Gene Expression



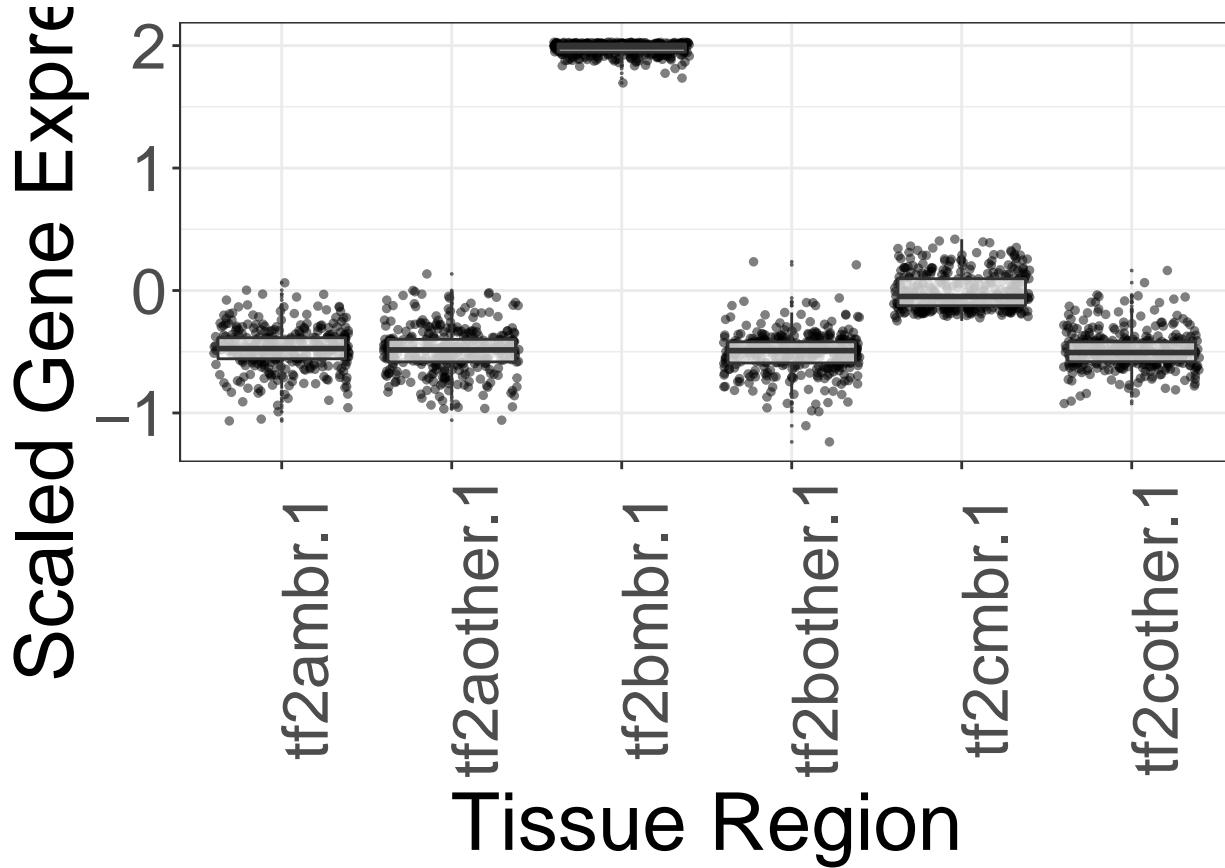
clusterVis_PCA(9)



Cluster 10

```
clusterVis(10)
```

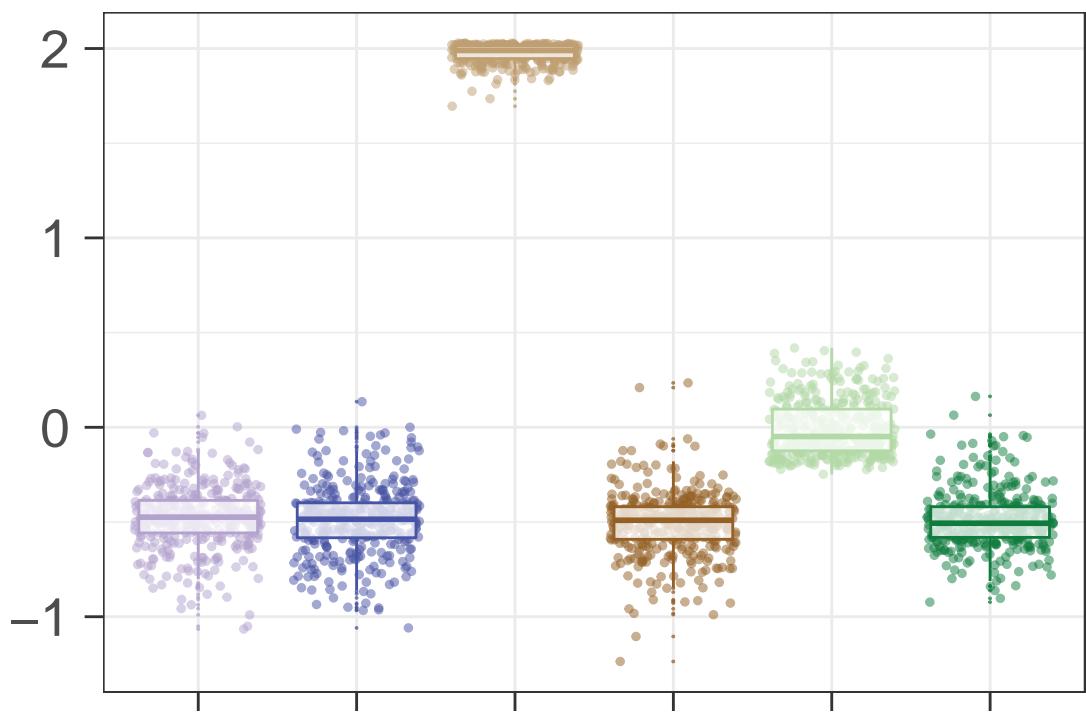
```
## Using gene as id variables
```



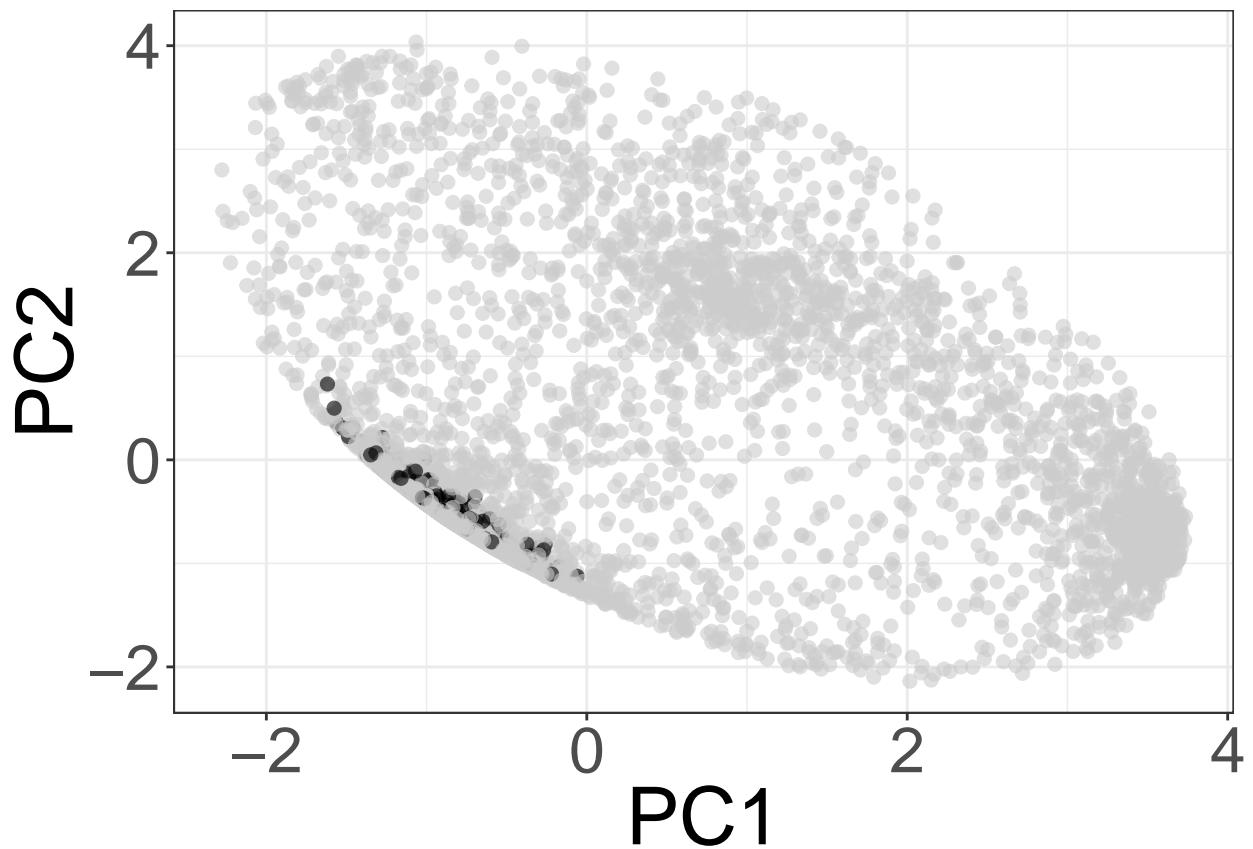
```
clusterVis_color(10)
```

```
## Using gene as id variables
```

Scaled Gene Expression



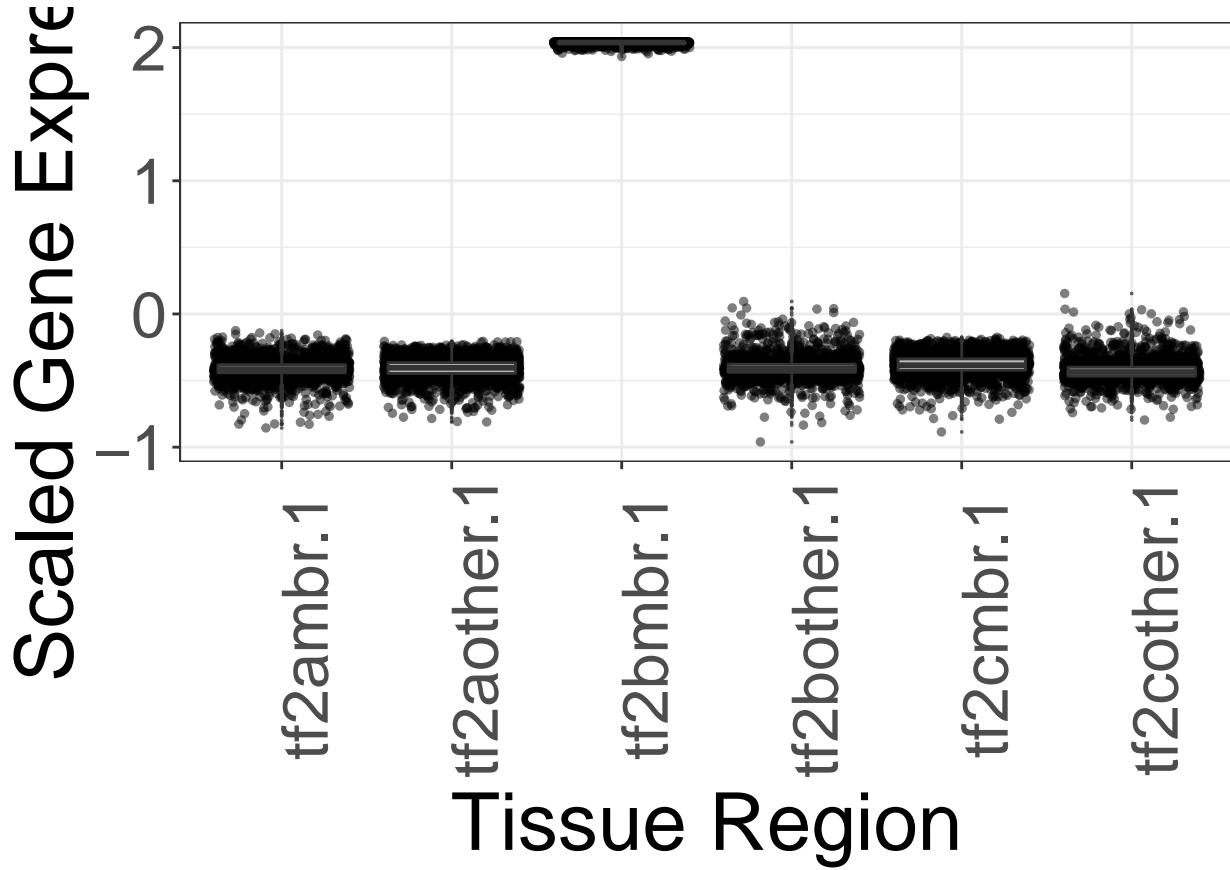
clusterVis_PCA(10)



Cluster 11

```
clusterVis(11)
```

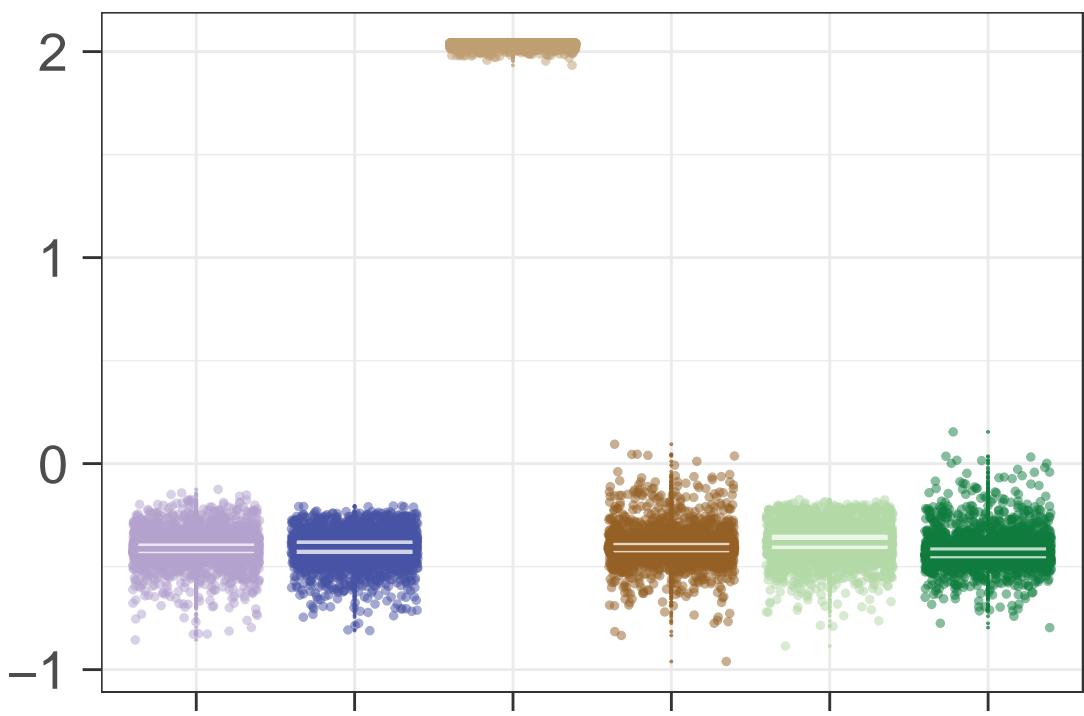
```
## Using gene as id variables
```



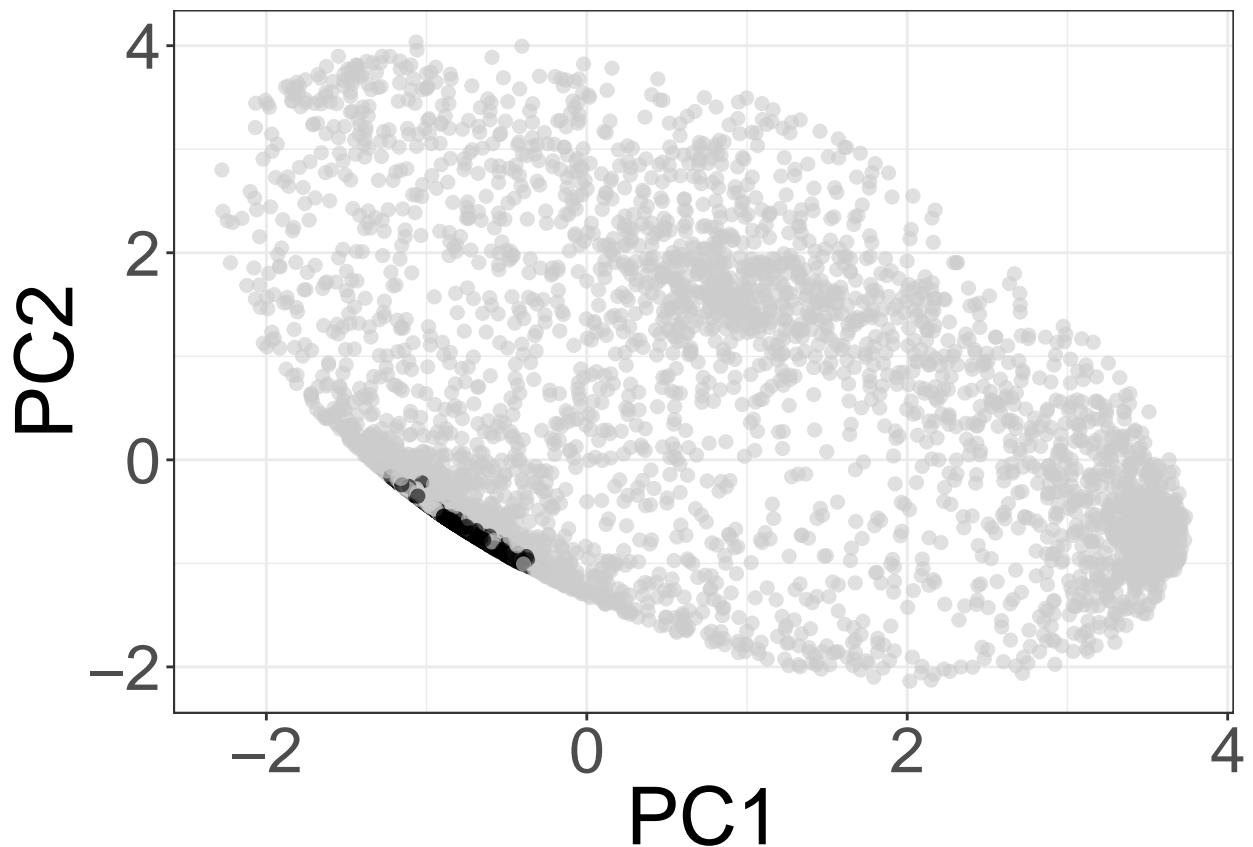
```
clusterVis_color(11)
```

```
## Using gene as id variables
```

Scaled Gene Expression



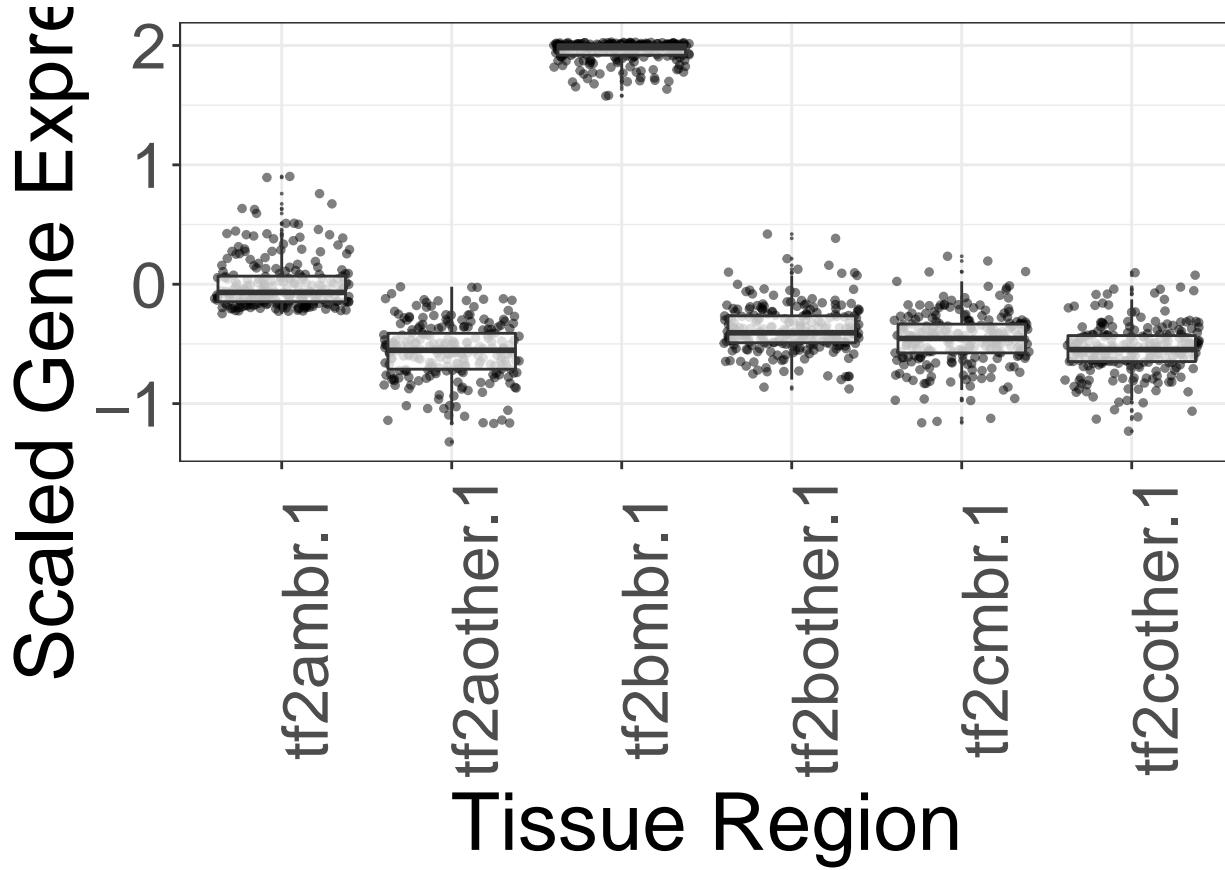
clusterVis_PCA(11)



Cluster 12

```
clusterVis(12)
```

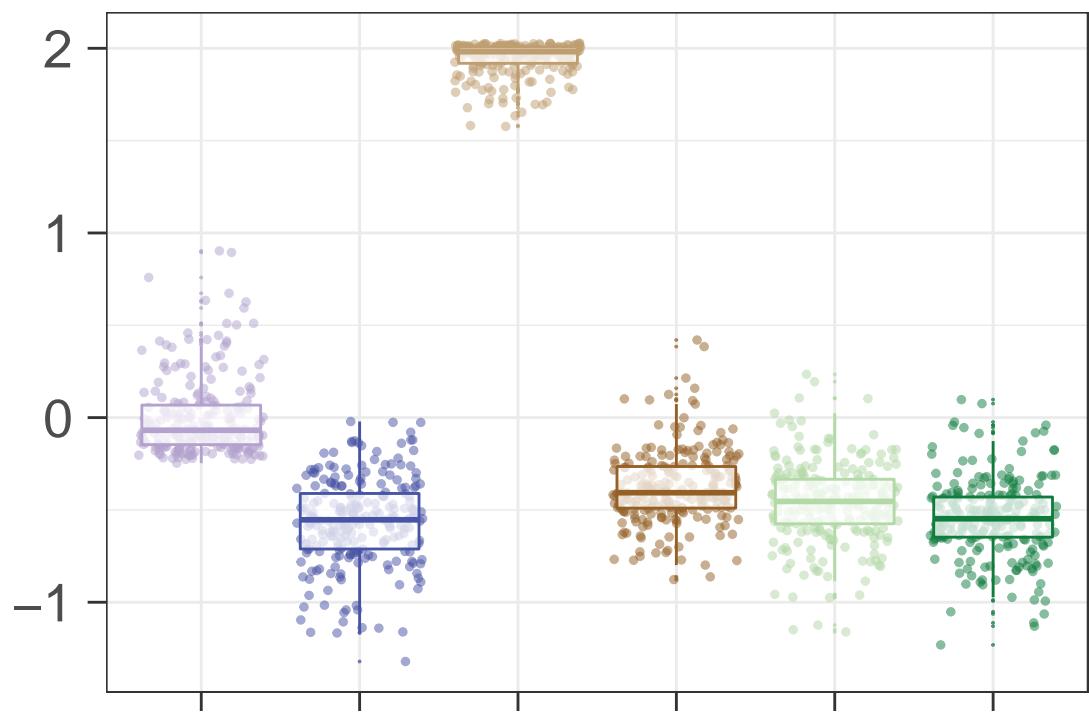
```
## Using gene as id variables
```



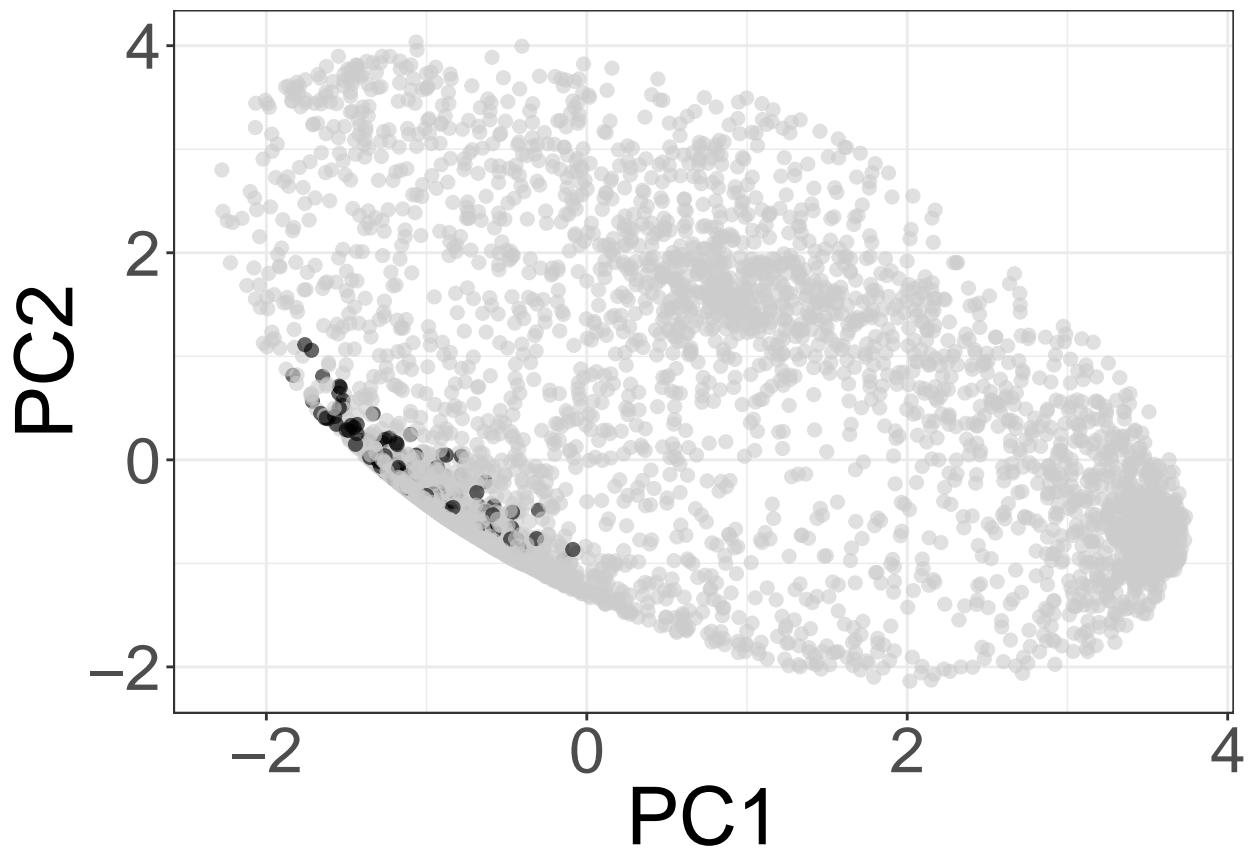
```
clusterVis_color(12)
```

```
## Using gene as id variables
```

Scaled Gene Expression



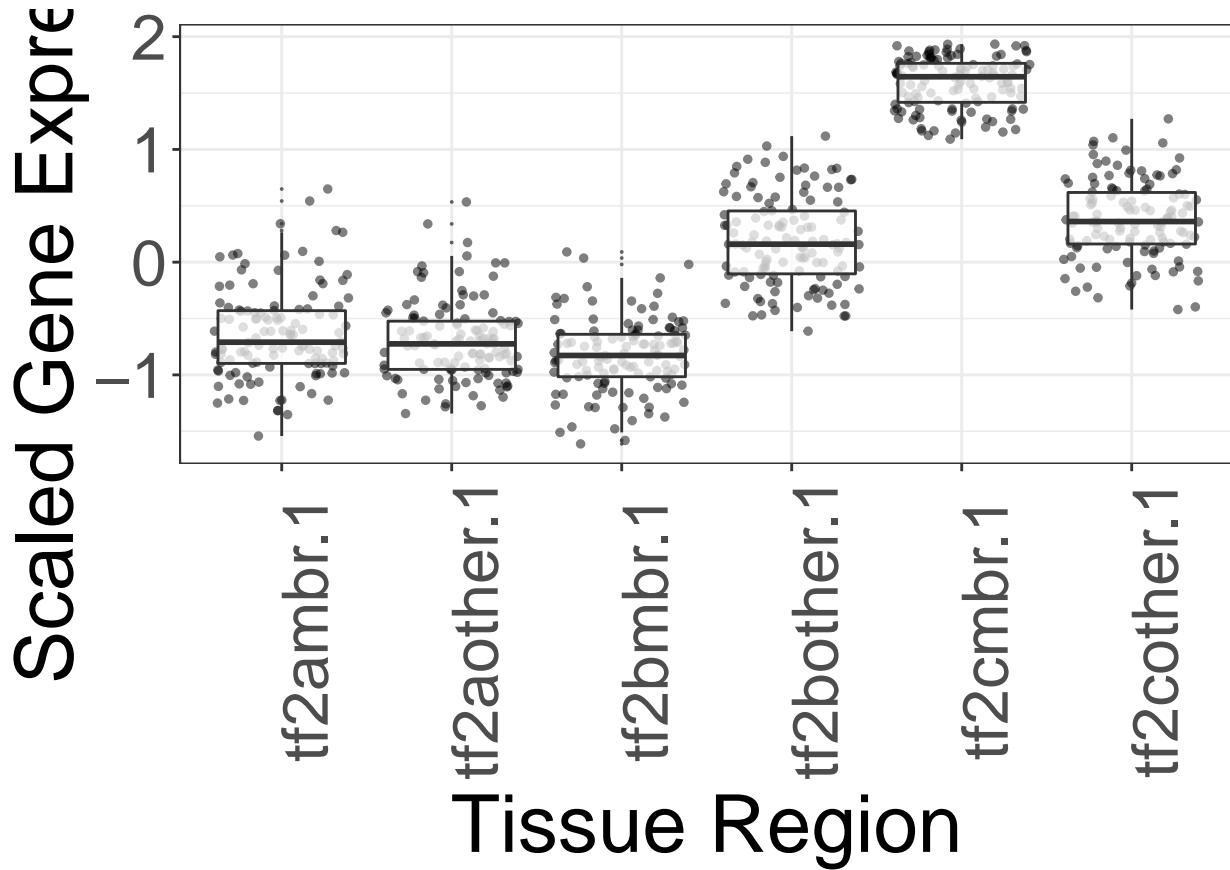
clusterVis_PCA(12)



Cluster 13

```
clusterVis(13)
```

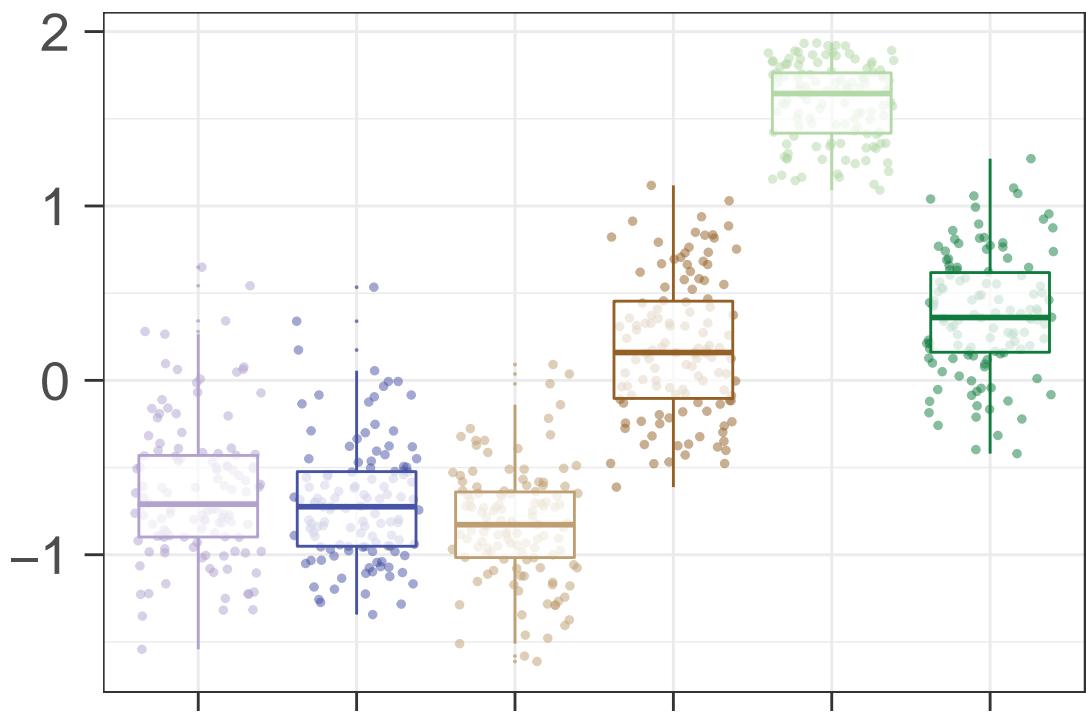
```
## Using gene as id variables
```



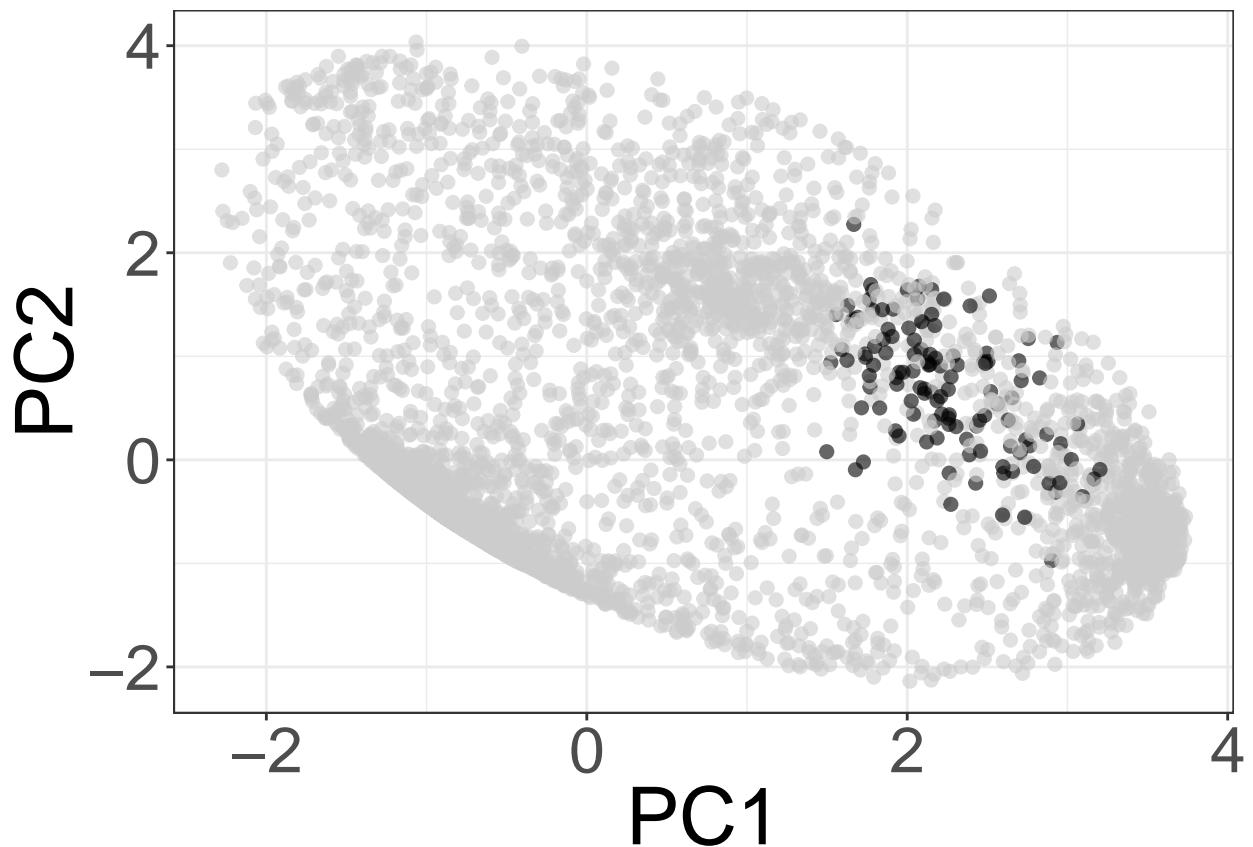
```
clusterVis_color(13)
```

```
## Using gene as id variables
```

Scaled Gene Expression



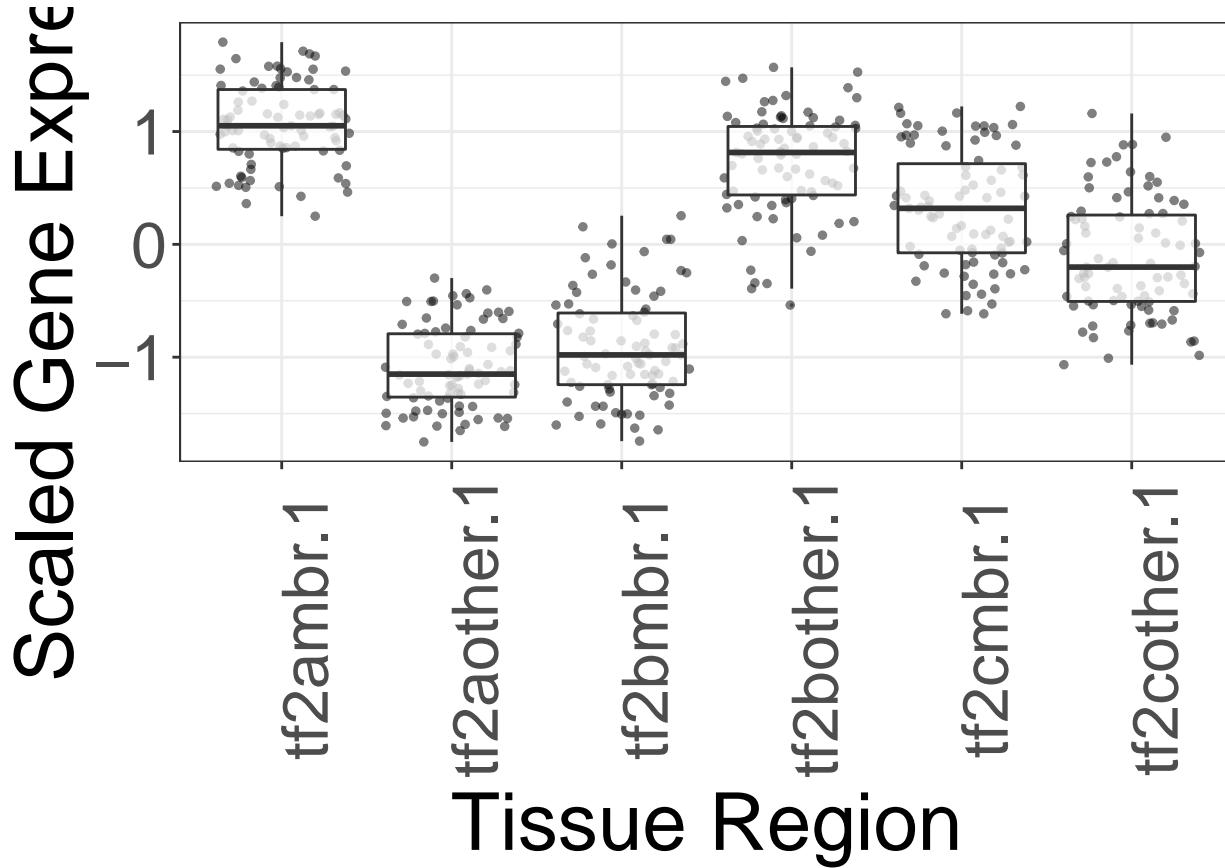
clusterVis_PCA(13)



Cluster 14

```
clusterVis(14)
```

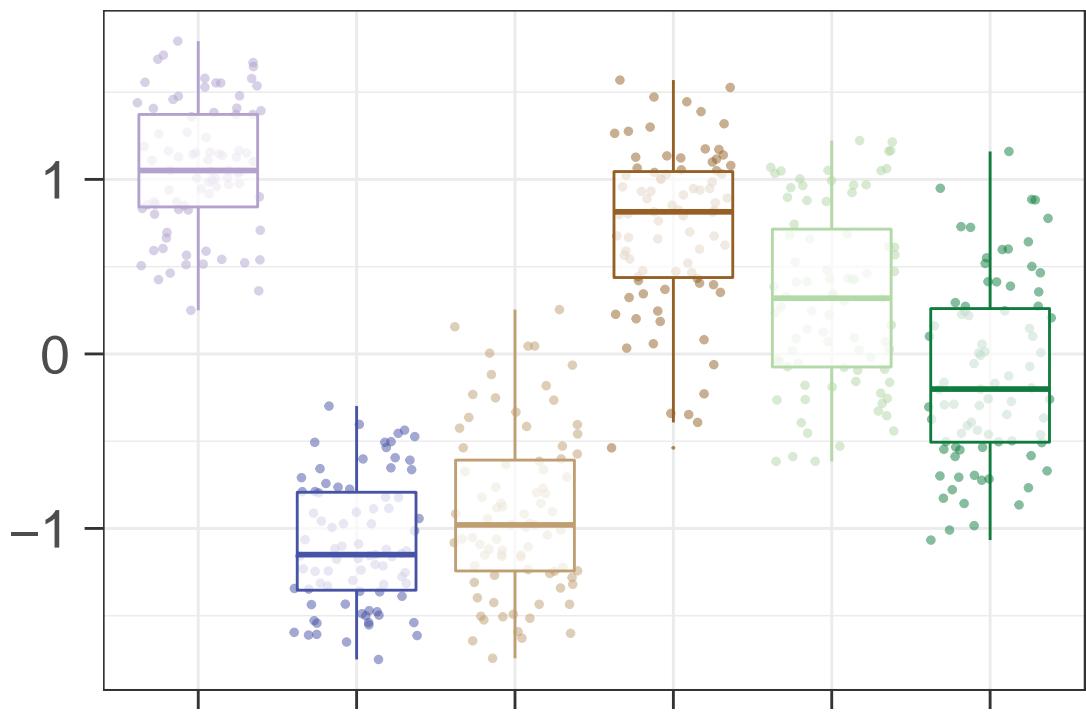
```
## Using gene as id variables
```



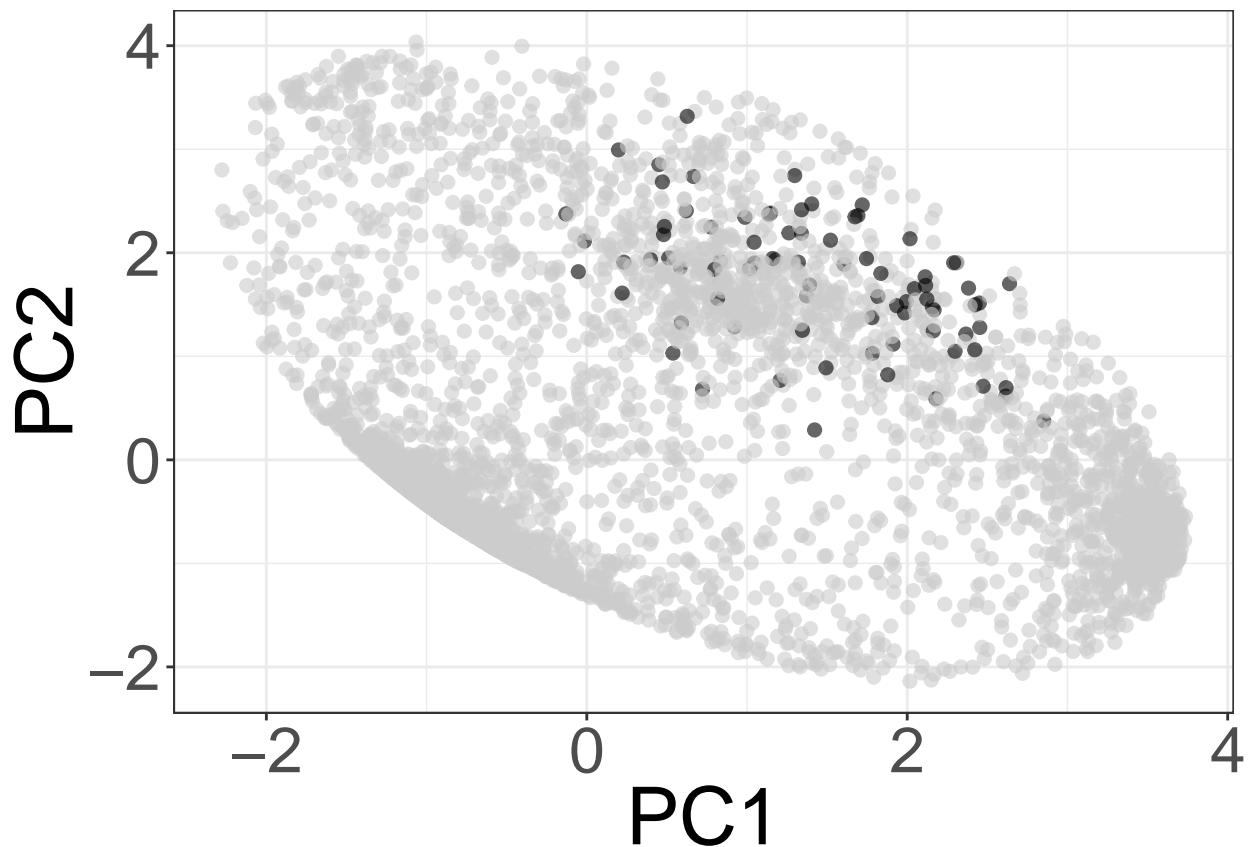
```
clusterVis_color(14)
```

```
## Using gene as id variables
```

Scaled Gene Expression



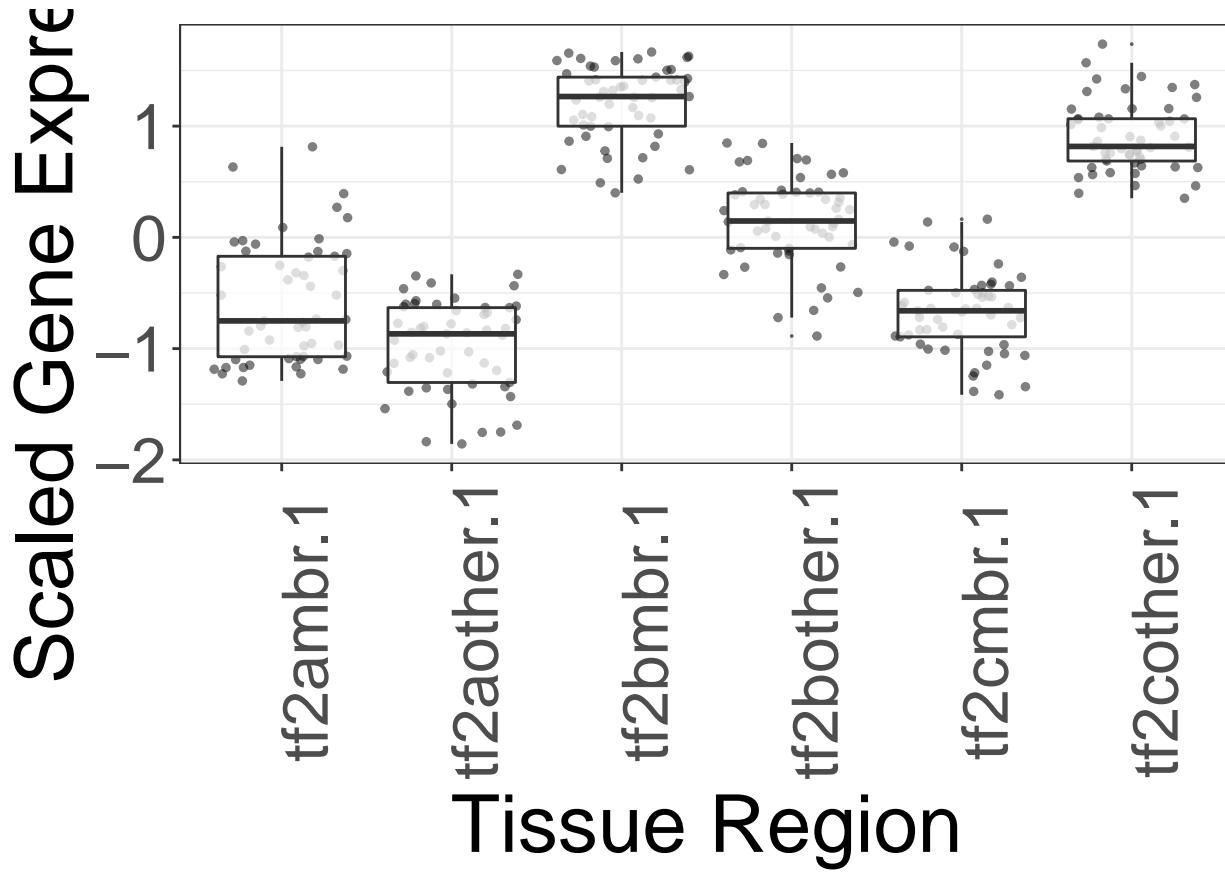
clusterVis_PCA(14)



Cluster 15

```
clusterVis(15)
```

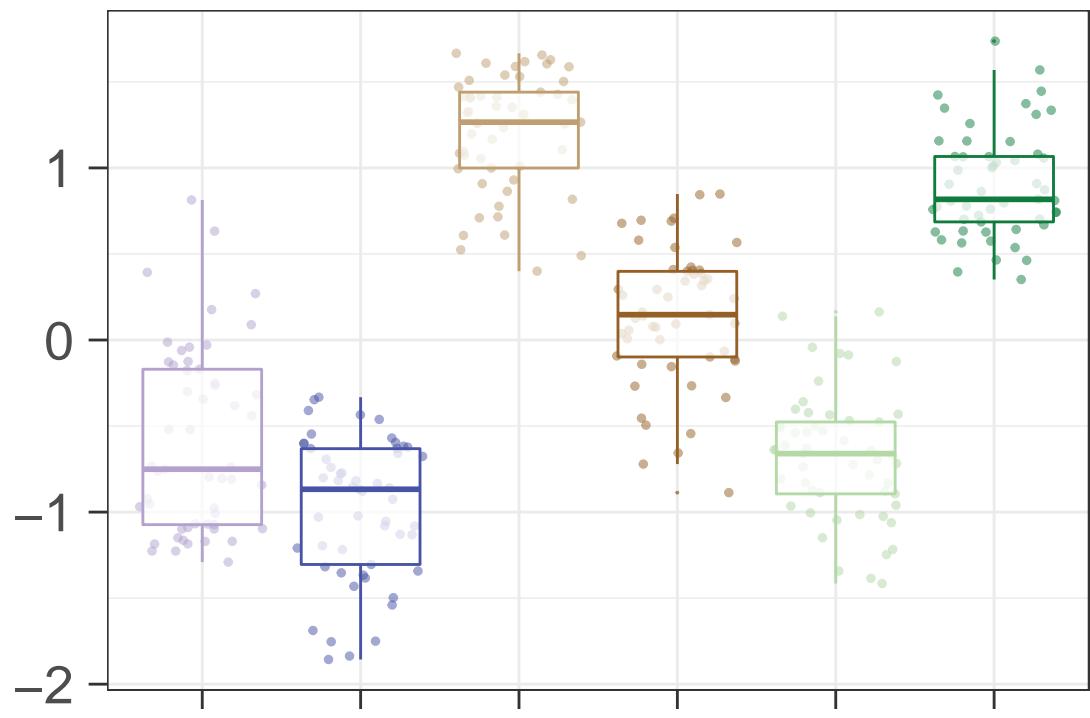
```
## Using gene as id variables
```



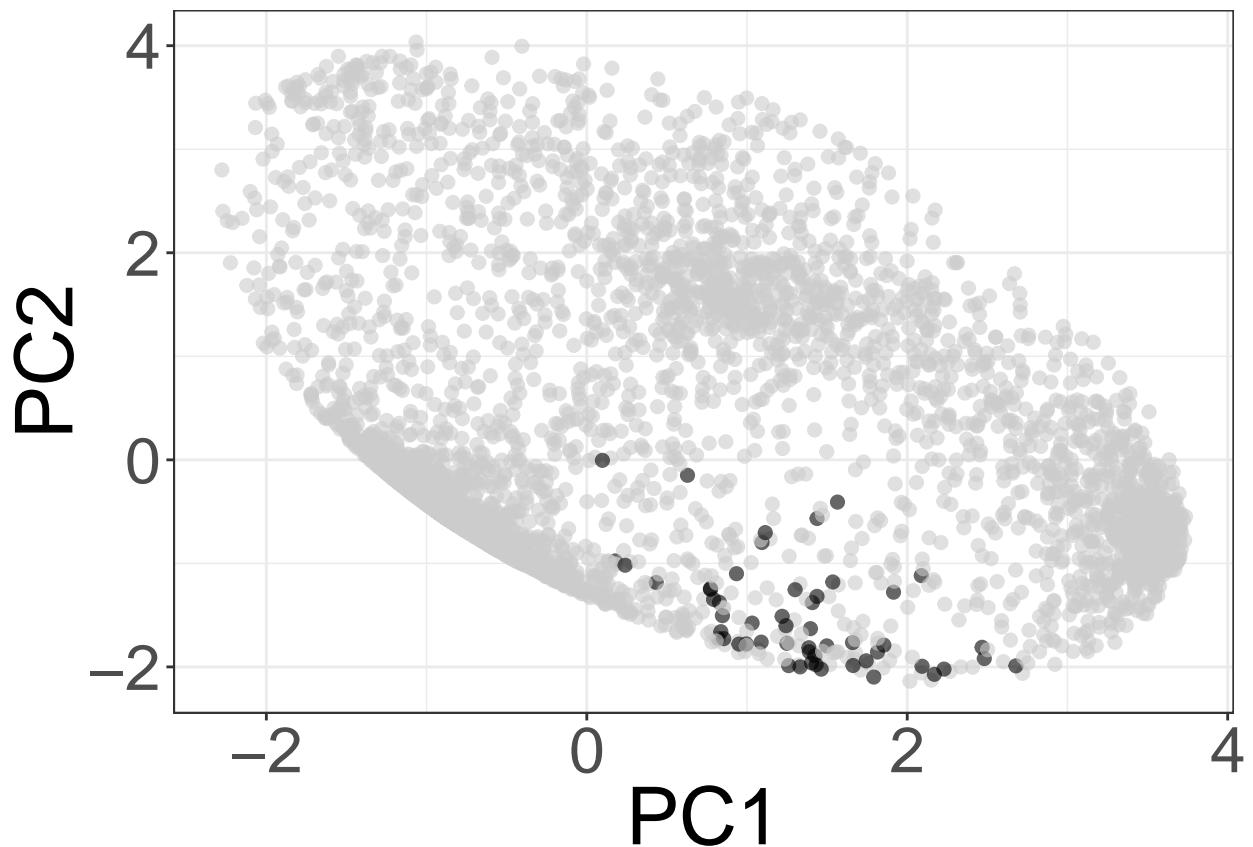
```
clusterVis_color(15)
```

```
## Using gene as id variables
```

Scaled Gene Expression



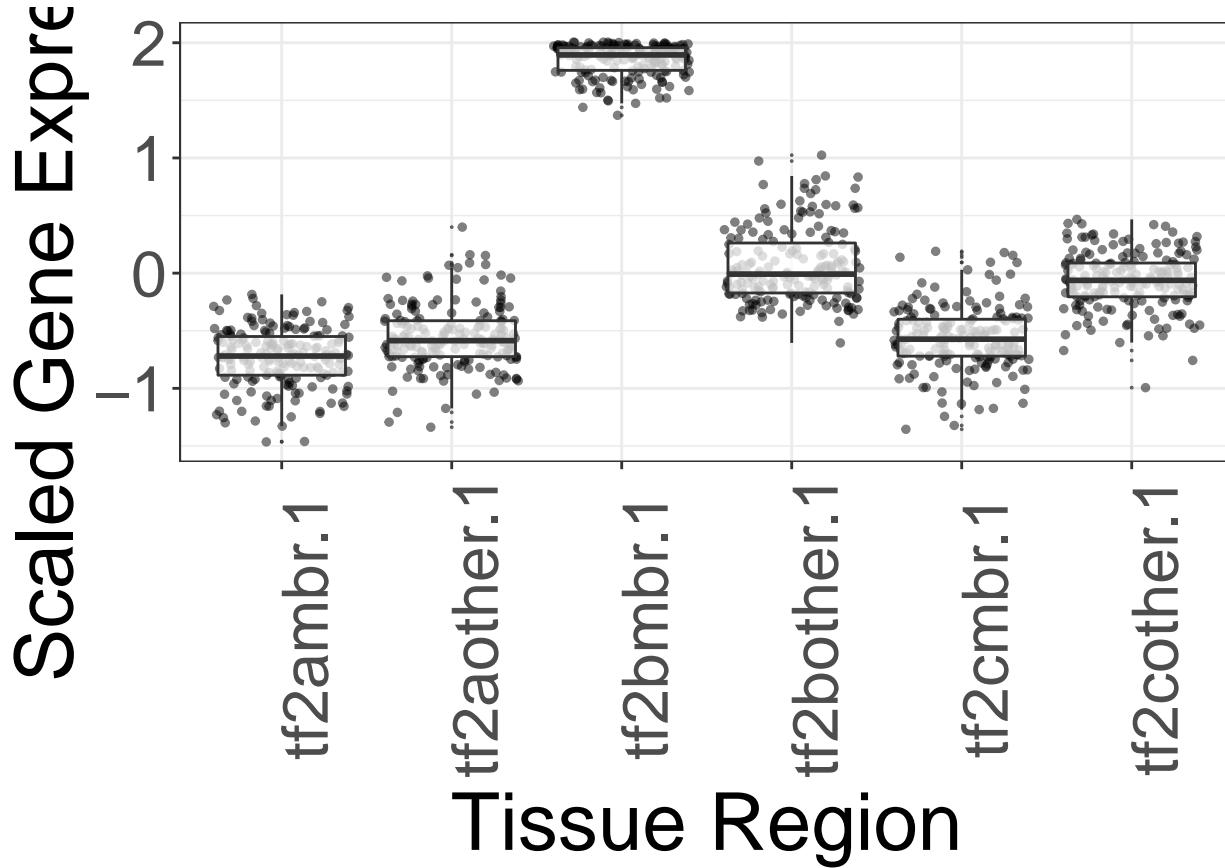
clusterVis_PCA(15)



Cluster 16

```
clusterVis(16)
```

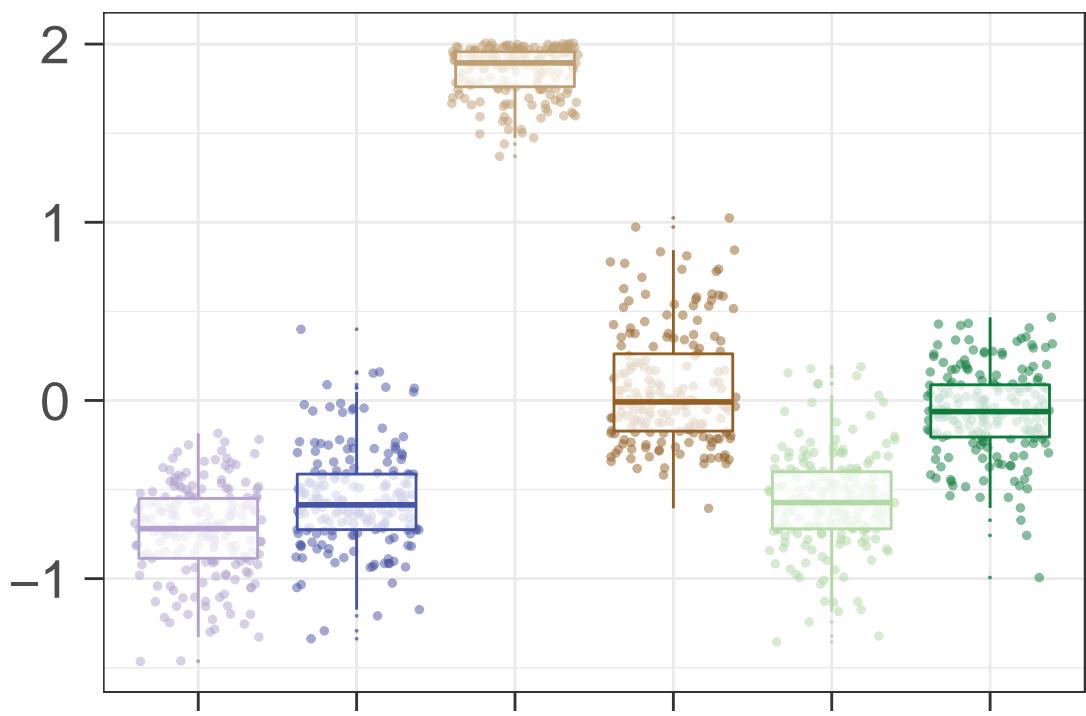
```
## Using gene as id variables
```



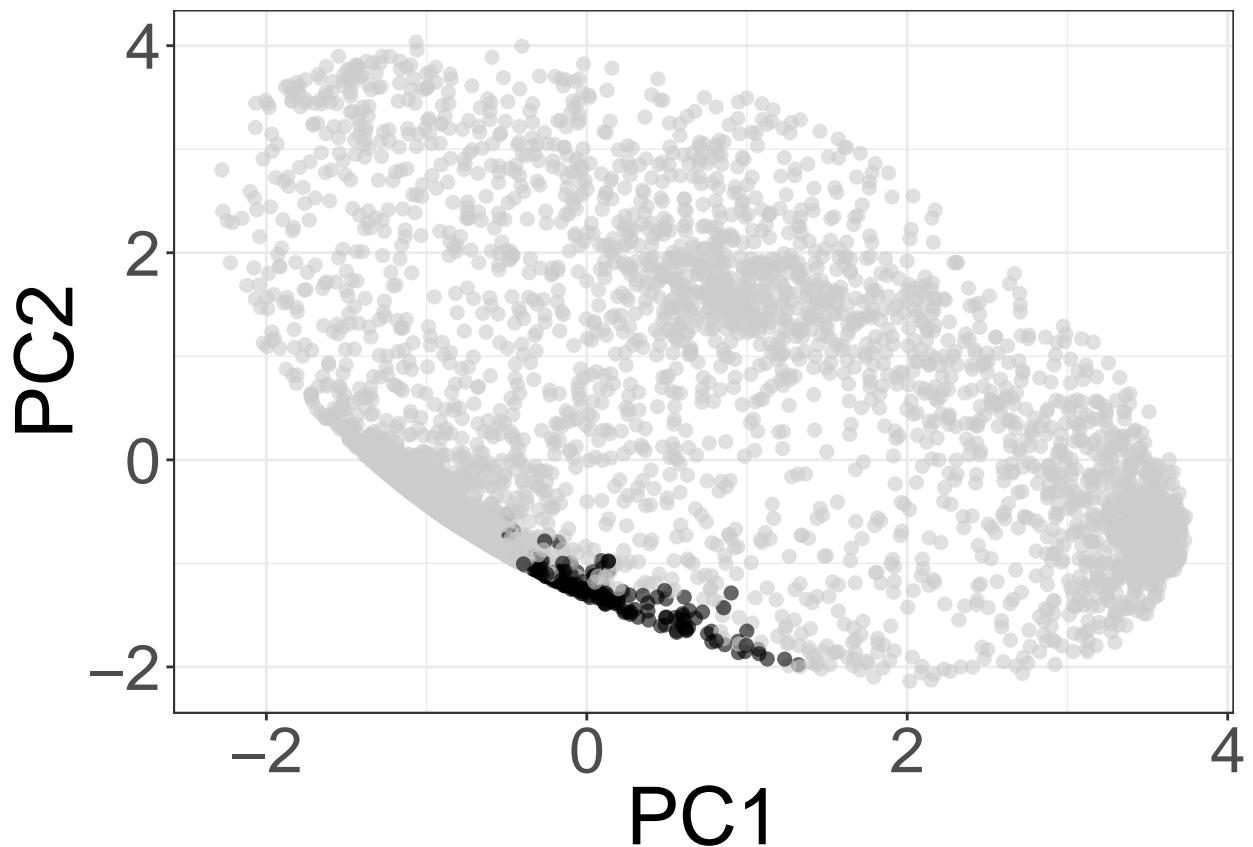
```
clusterVis_color(16)
```

```
## Using gene as id variables
```

Scaled Gene Expression



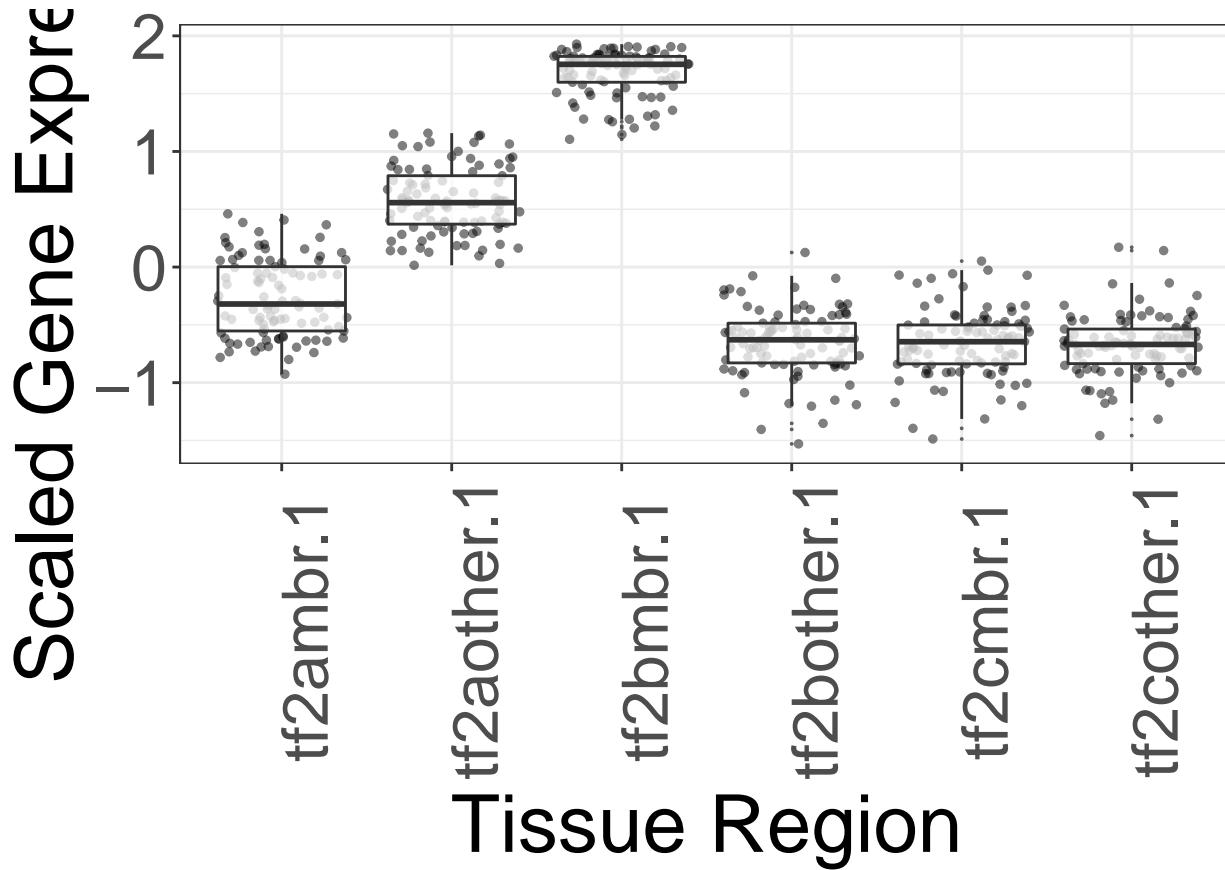
clusterVis_PCA(16)



Cluster 17

```
clusterVis(17)
```

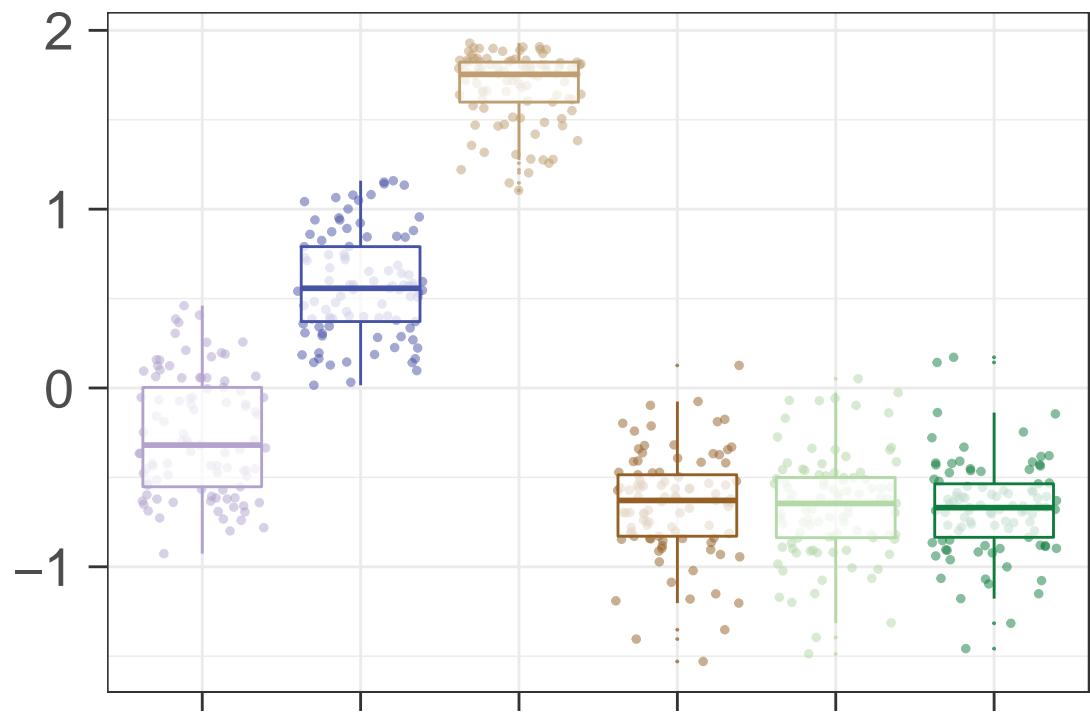
```
## Using gene as id variables
```



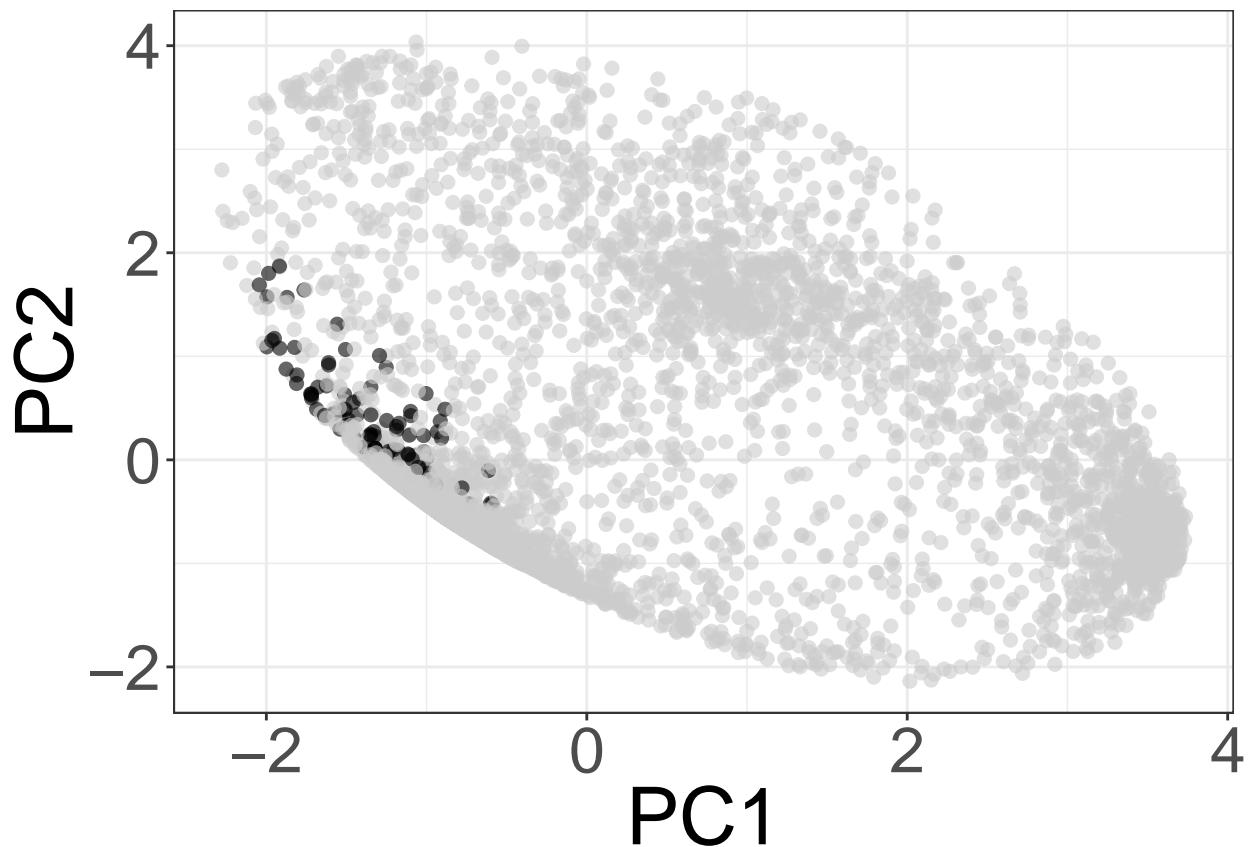
```
clusterVis_color(17)
```

```
## Using gene as id variables
```

Scaled Gene Expression



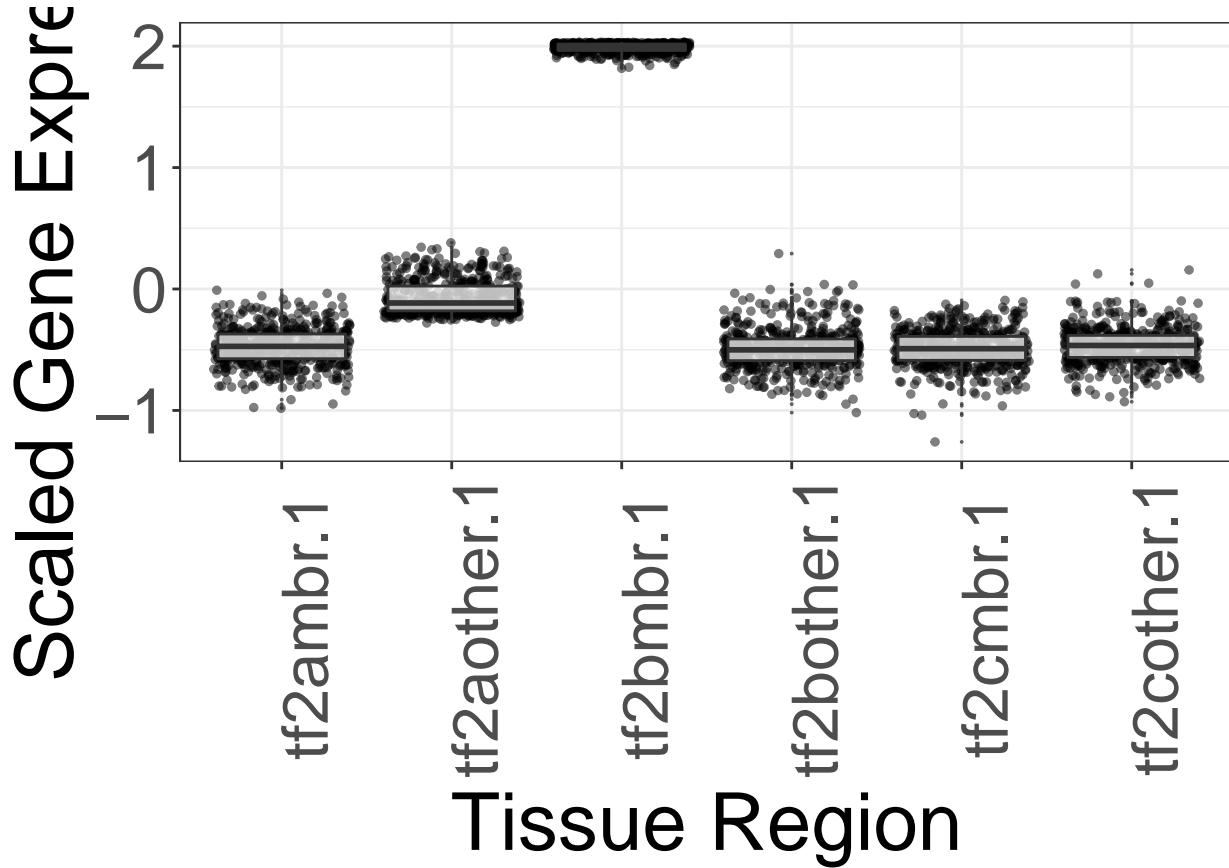
clusterVis_PCA(17)



Cluster 18

```
clusterVis(18)
```

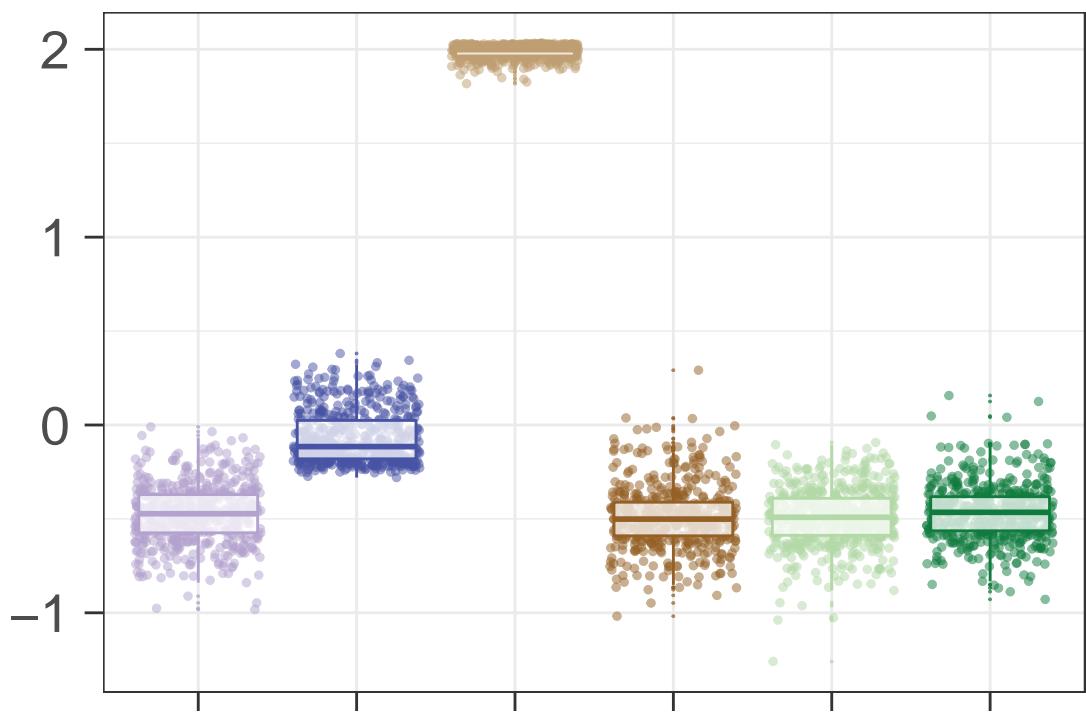
```
## Using gene as id variables
```



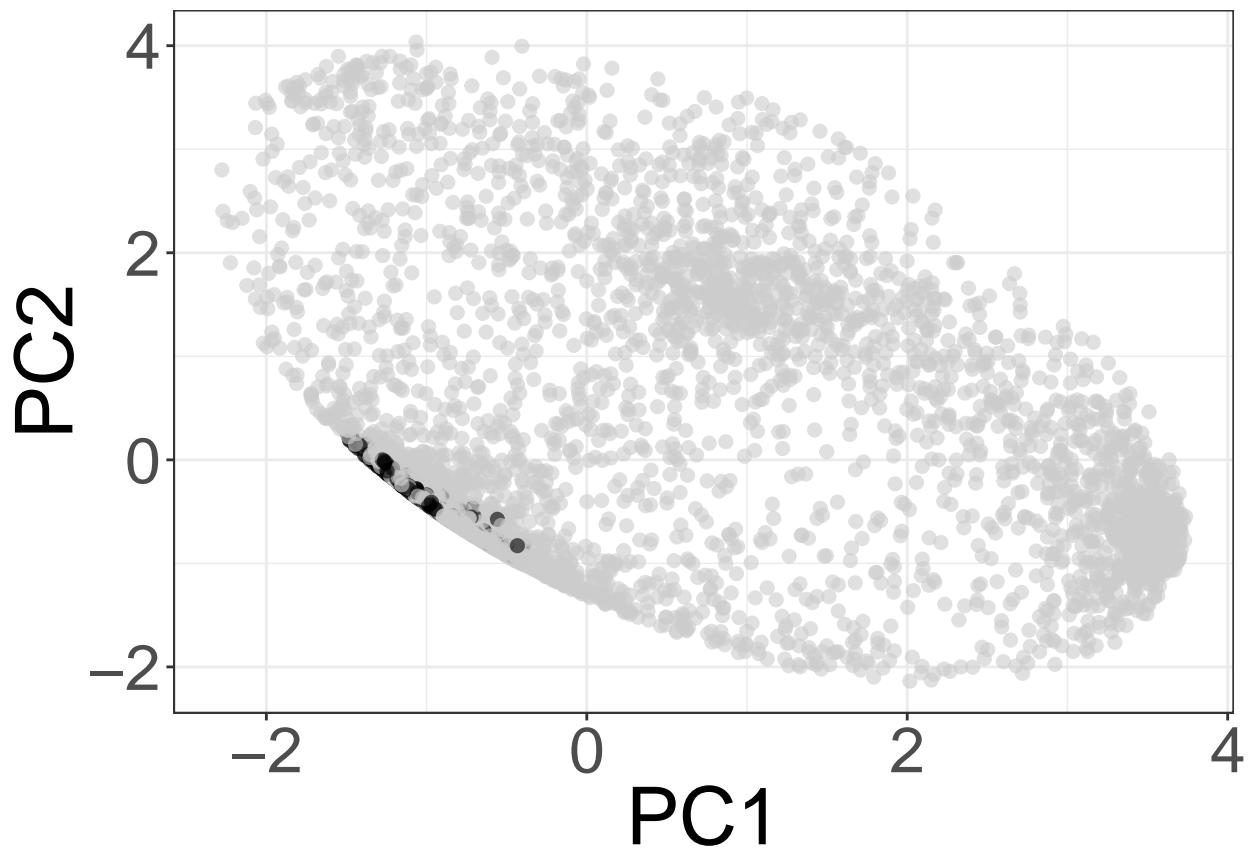
```
clusterVis_color(18)
```

```
## Using gene as id variables
```

Scaled Gene Expression



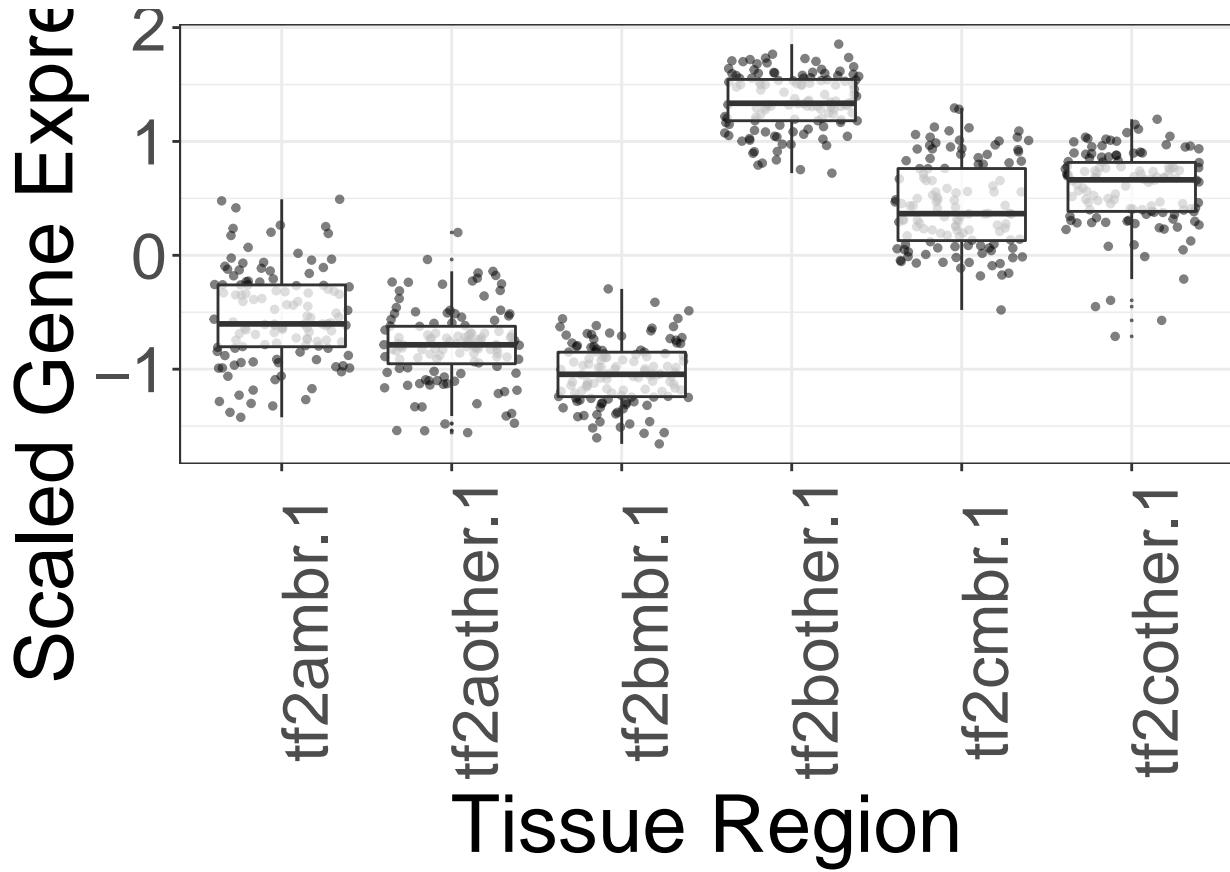
clusterVis_PCA(18)



Cluster 19

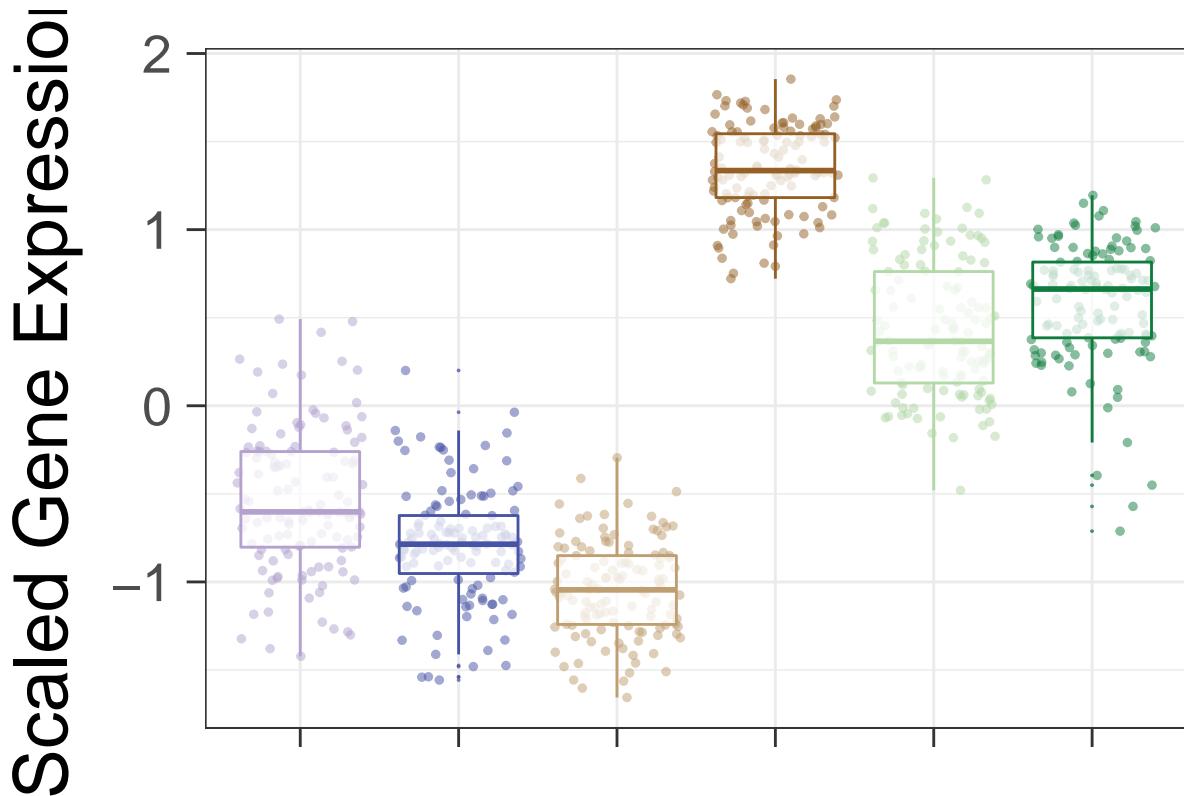
```
clusterVis(19)
```

```
## Using gene as id variables
```

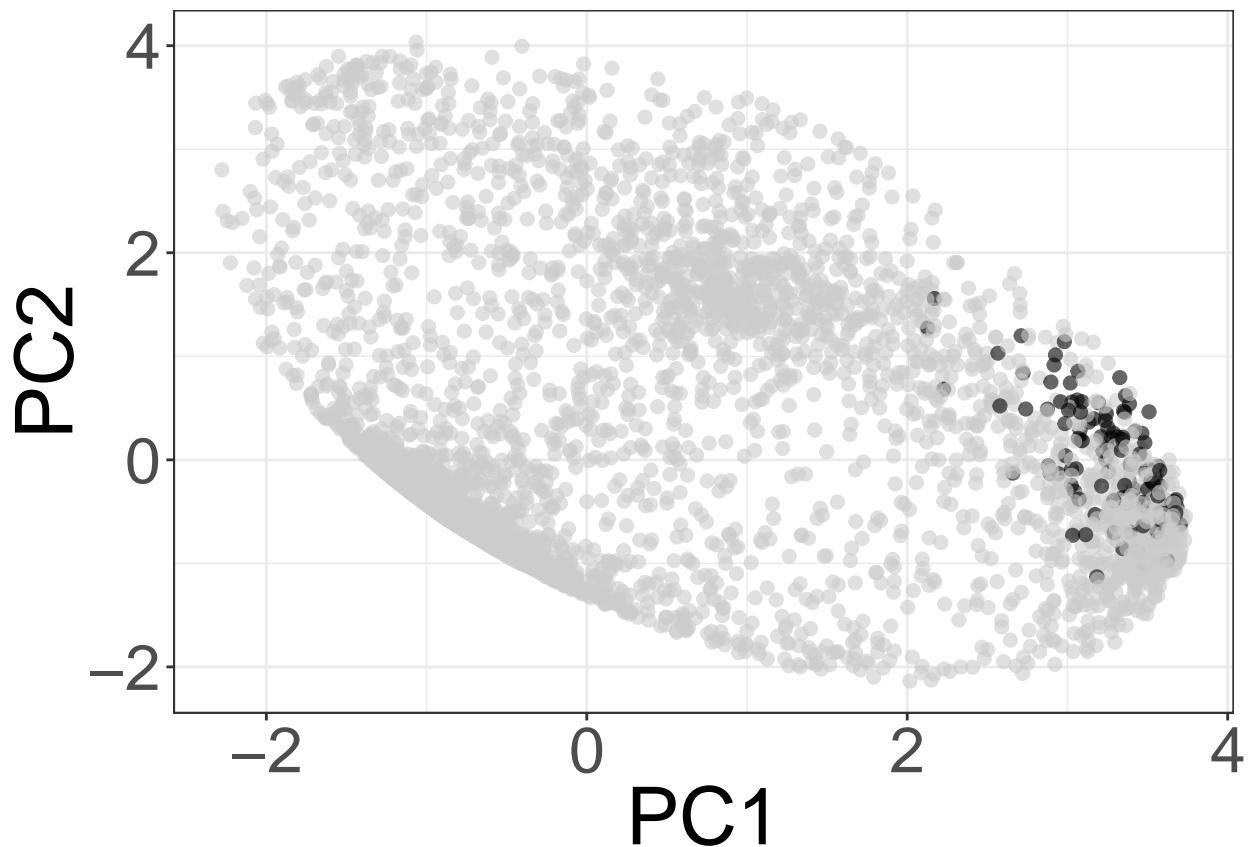


```
clusterVis_color(19)
```

```
## Using gene as id variables
```



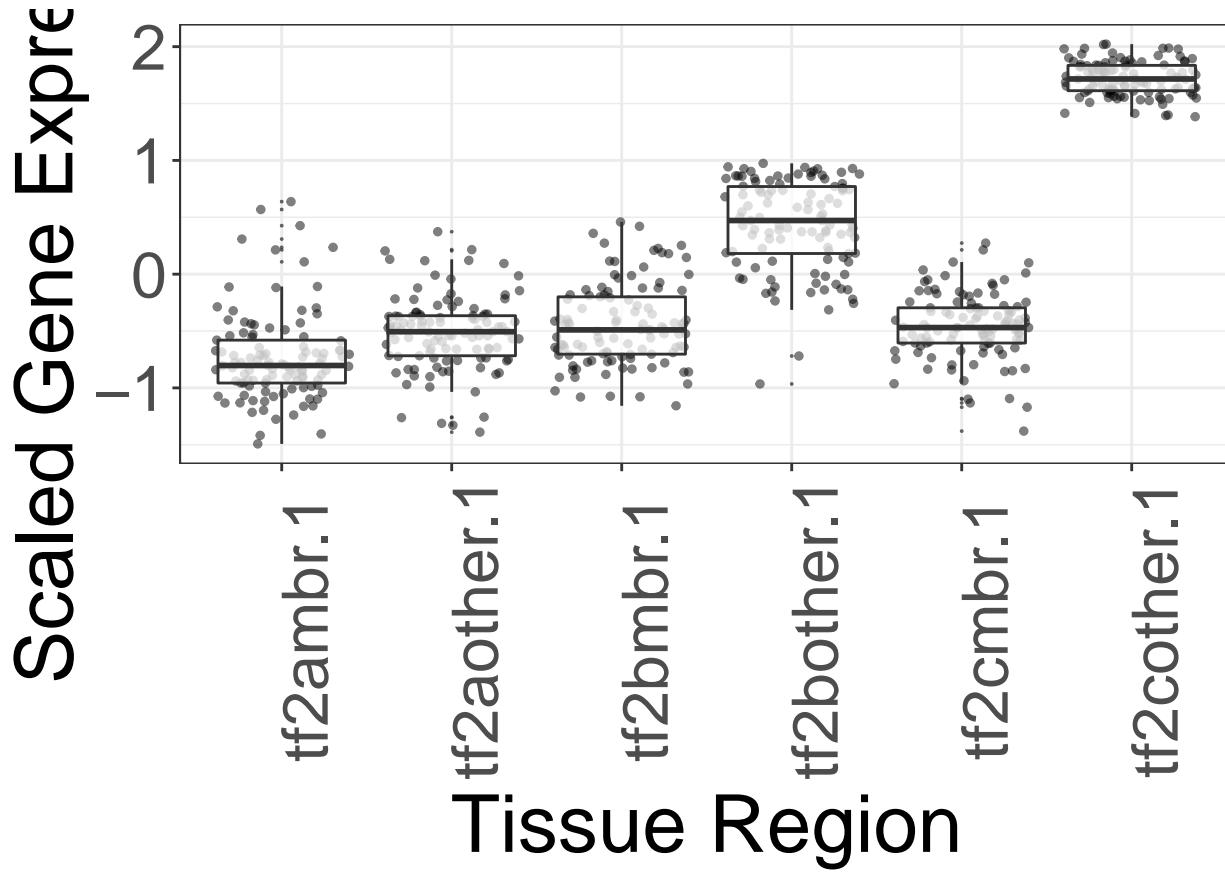
clusterVis_PCA(19)



Cluster 20

```
clusterVis(20)
```

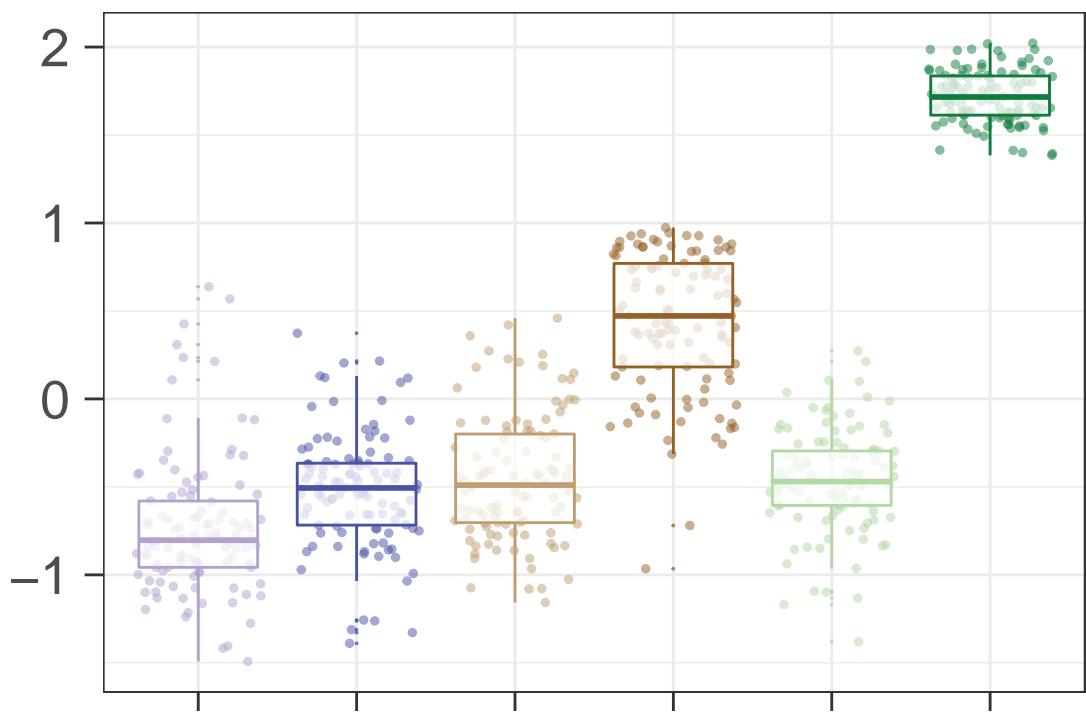
```
## Using gene as id variables
```



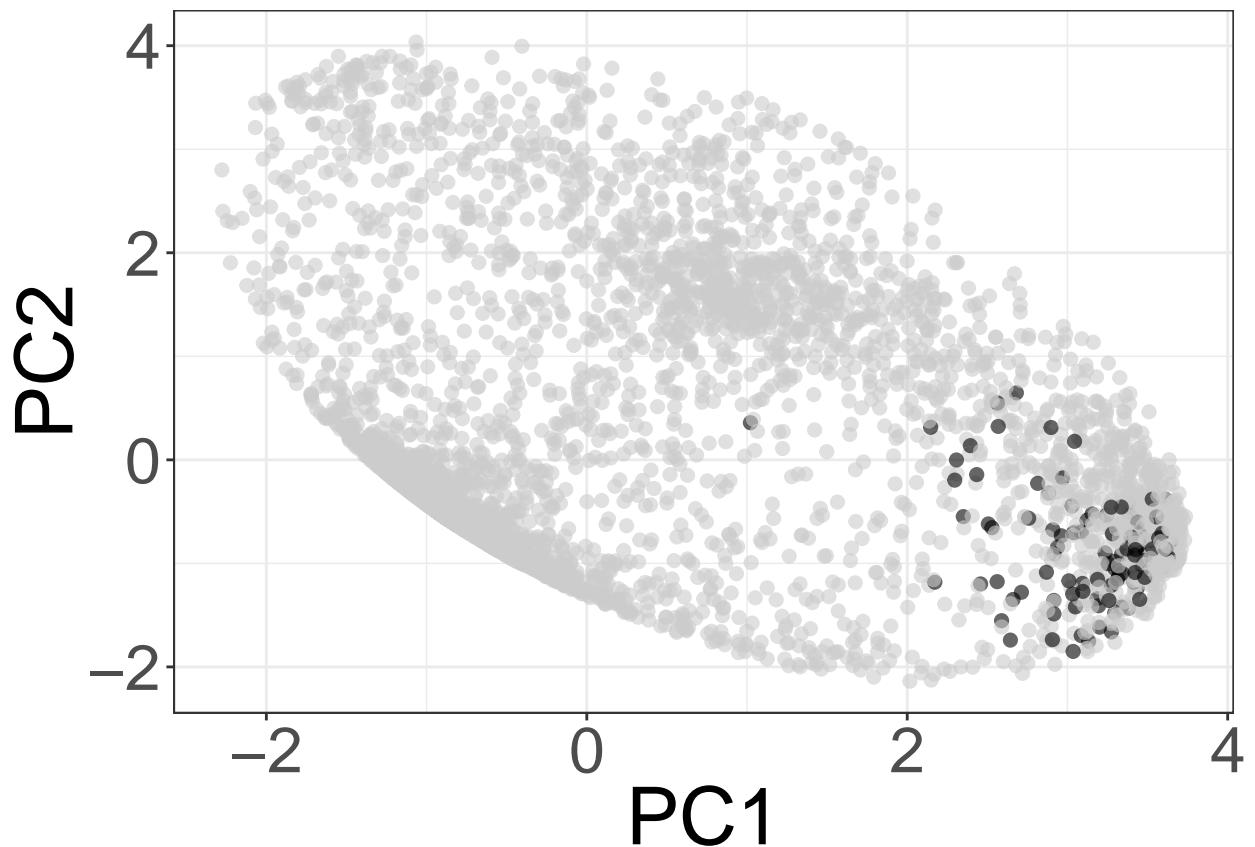
```
clusterVis_color(20)
```

```
## Using gene as id variables
```

Scaled Gene Expression



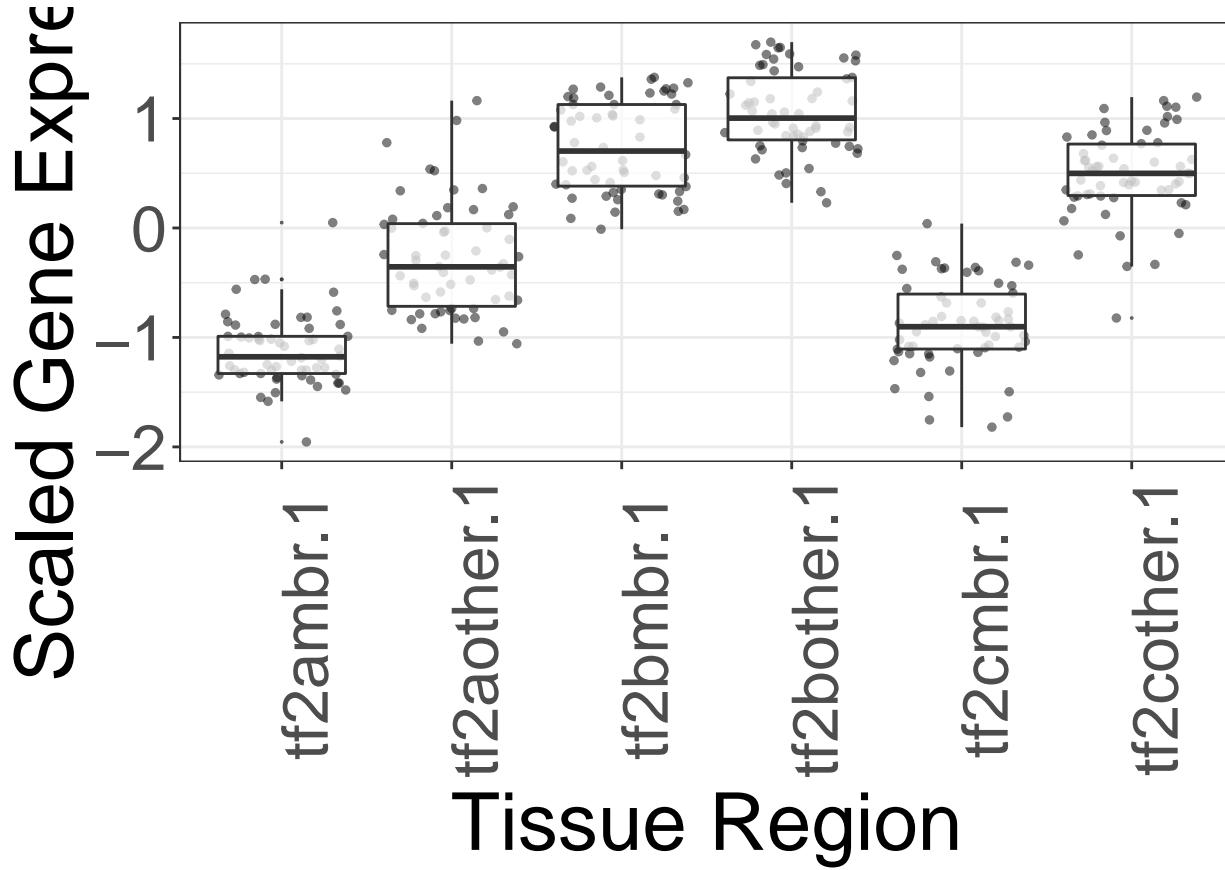
clusterVis_PCA(20)



Cluster 21

```
clusterVis(21)
```

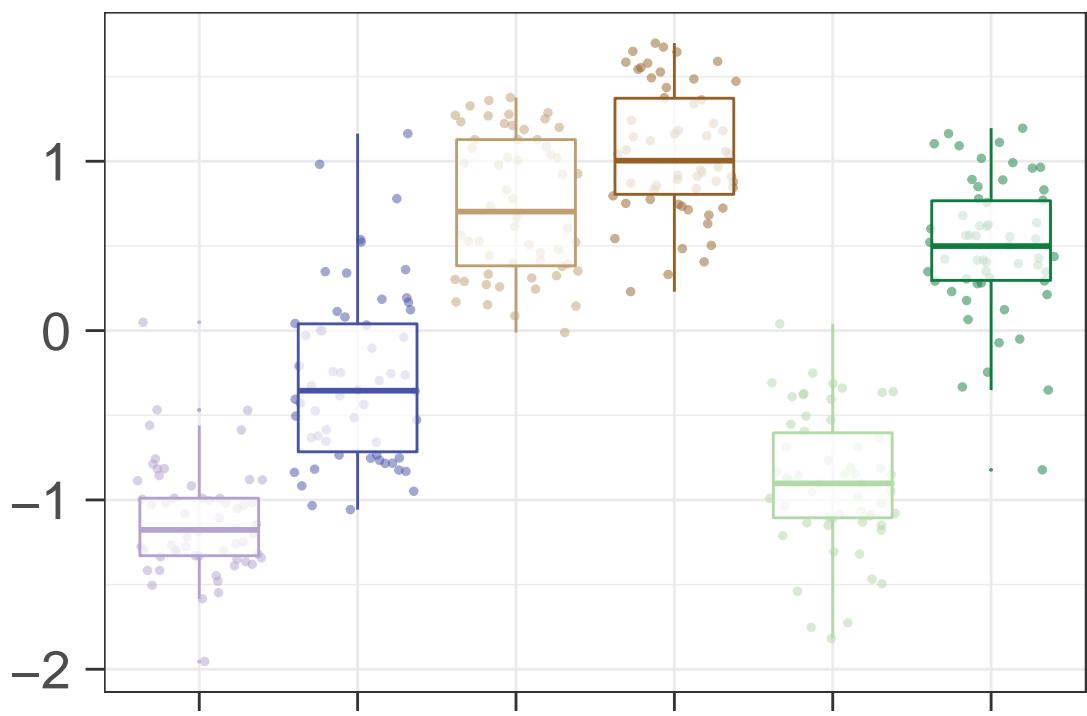
```
## Using gene as id variables
```



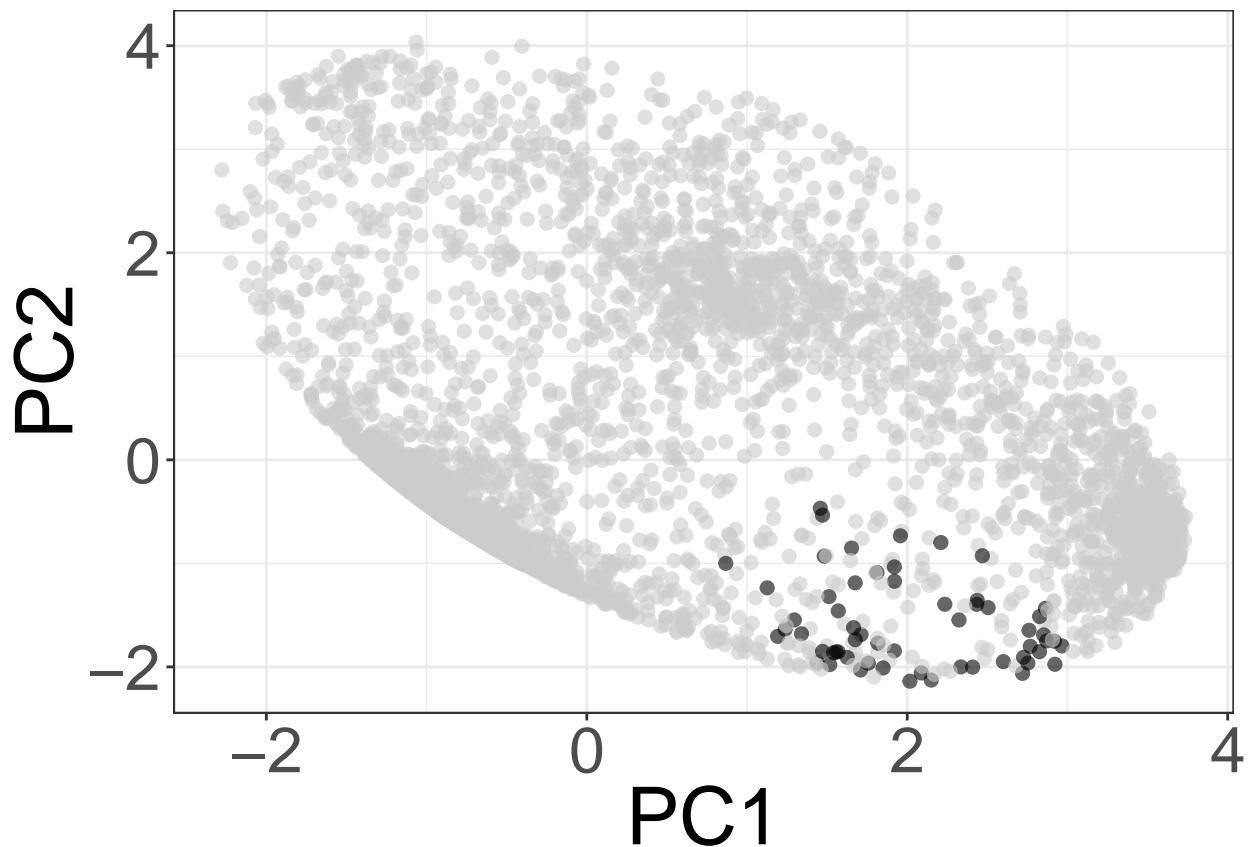
```
clusterVis_color(21)
```

```
## Using gene as id variables
```

Scaled Gene Expression



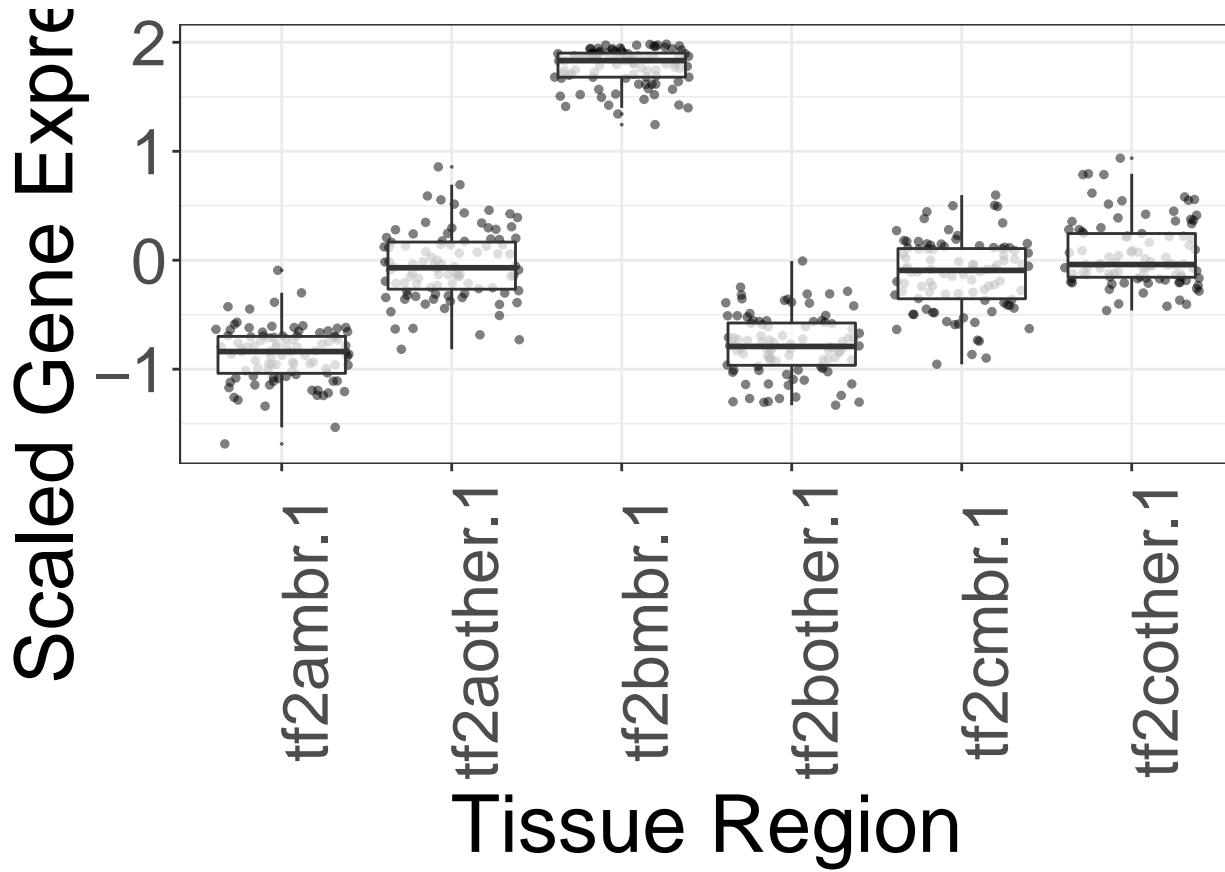
clusterVis_PCA(21)



Cluster 22

```
clusterVis(22)
```

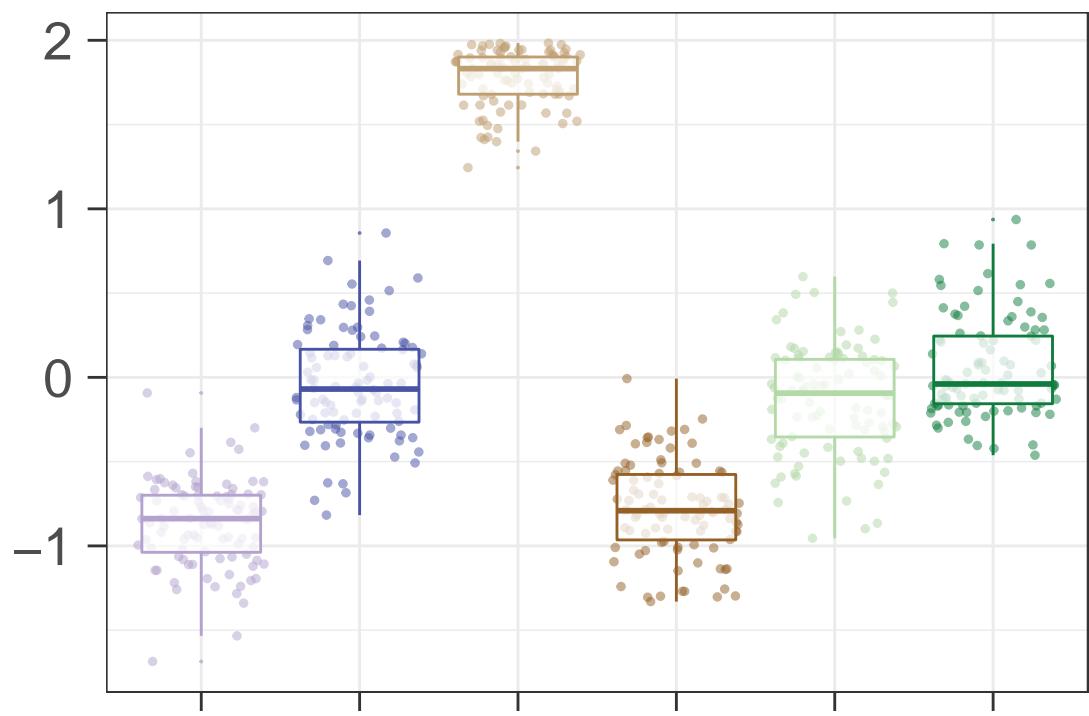
```
## Using gene as id variables
```



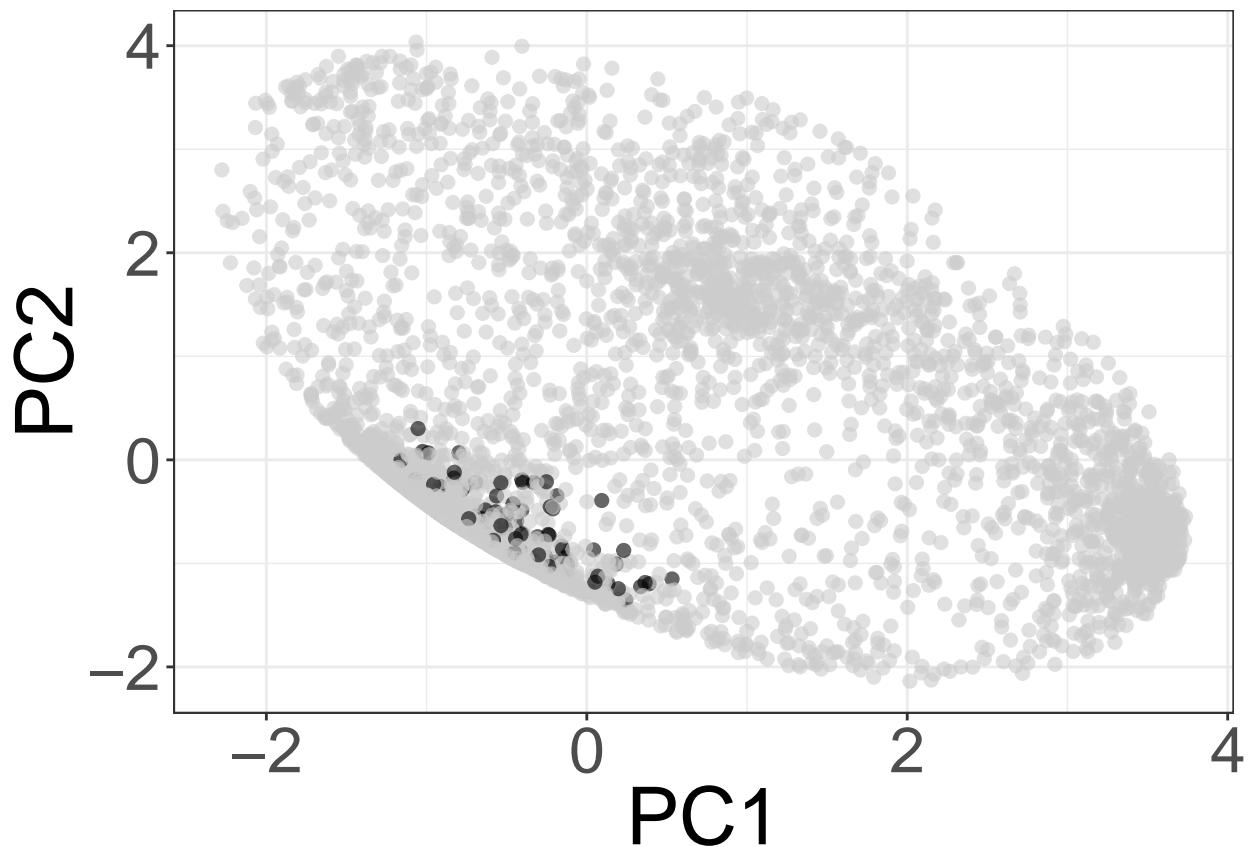
```
clusterVis_color(22)
```

```
## Using gene as id variables
```

Scaled Gene Expression



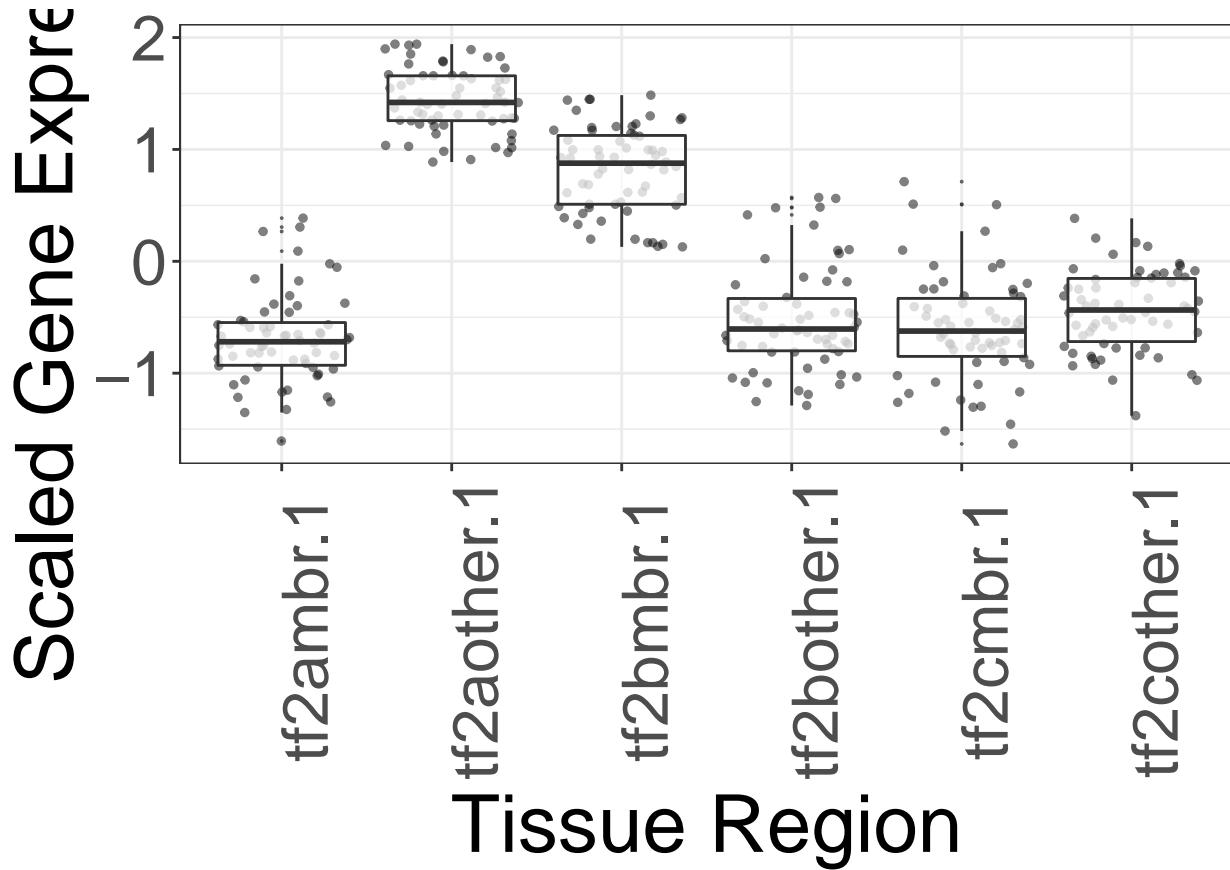
clusterVis_PCA(22)



Cluster 23

```
clusterVis(23)
```

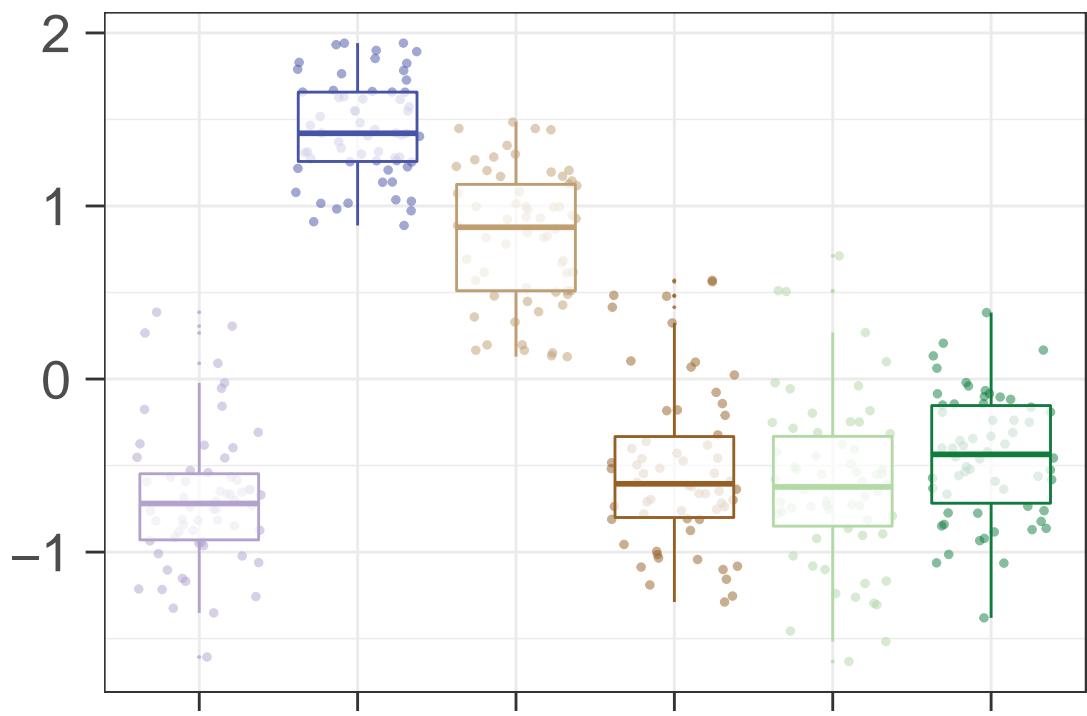
```
## Using gene as id variables
```



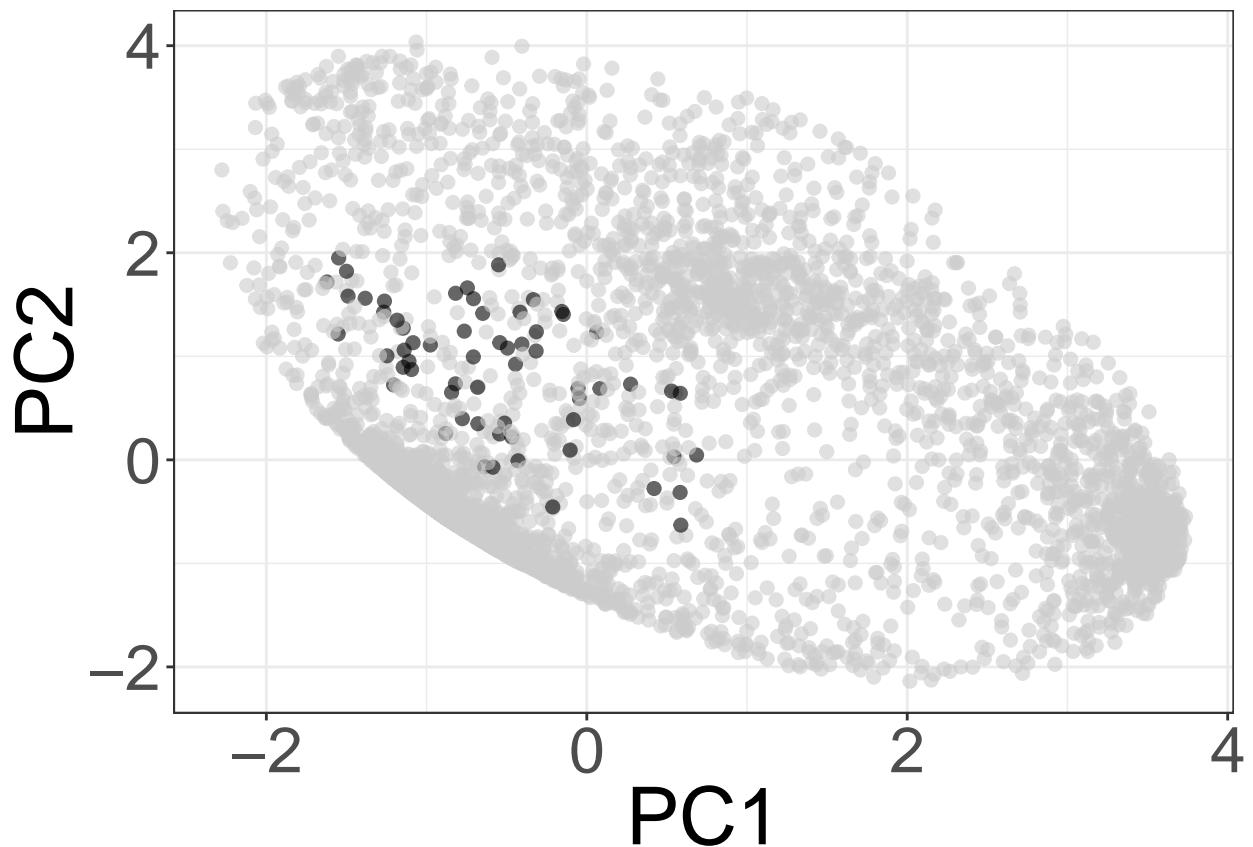
```
clusterVis_color(23)
```

```
## Using gene as id variables
```

Scaled Gene Expression



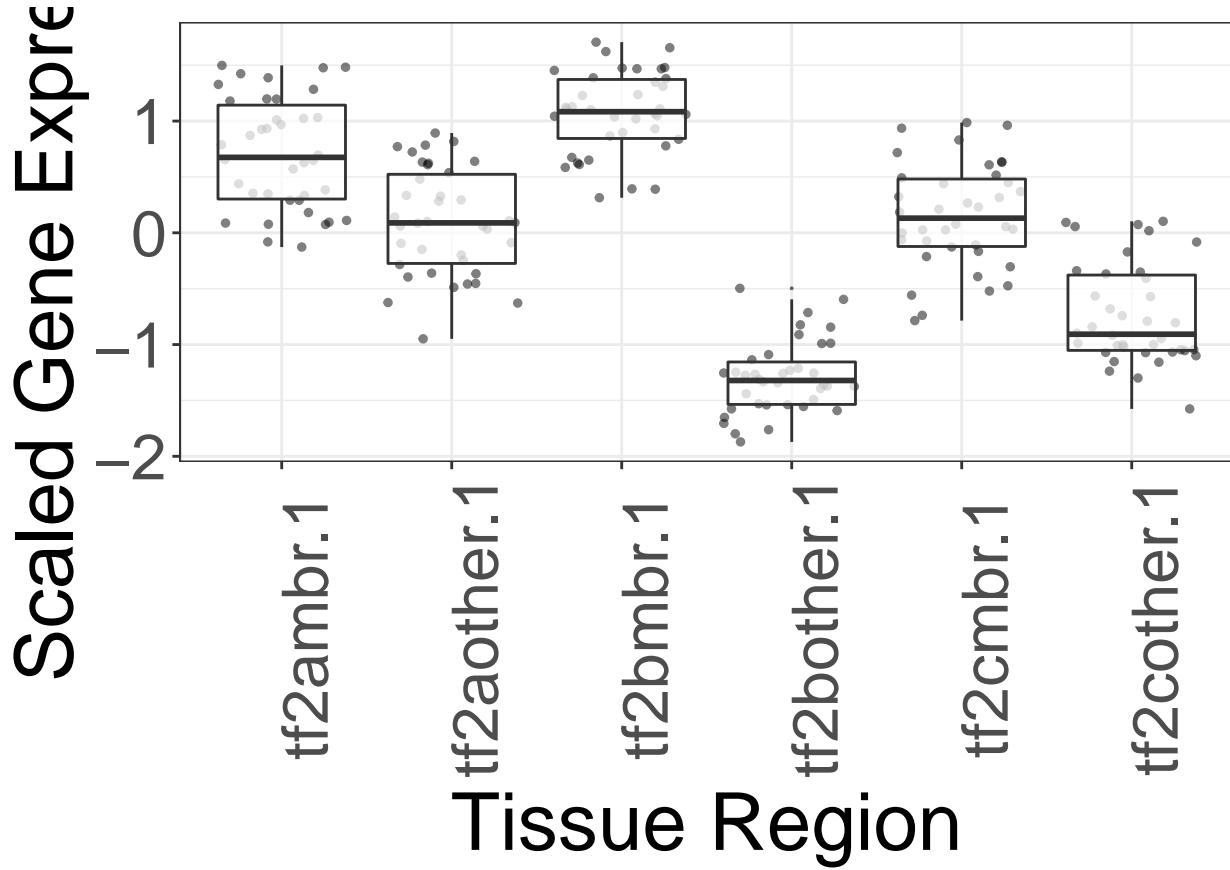
clusterVis_PCA(23)



Cluster 24

```
clusterVis(24)
```

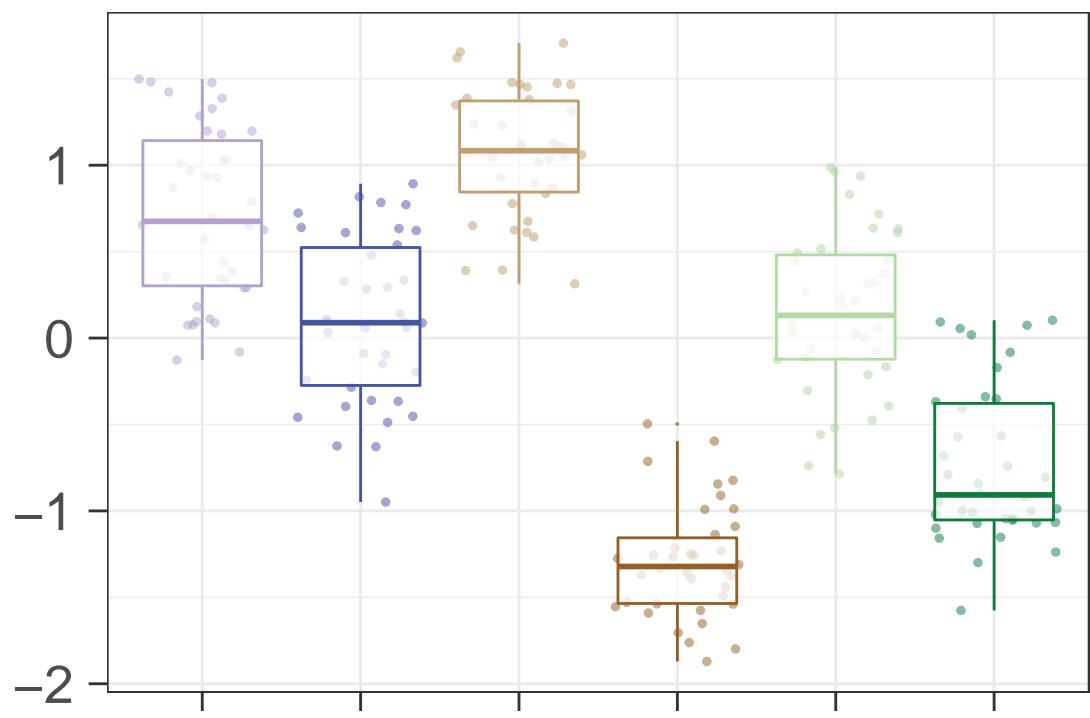
```
## Using gene as id variables
```



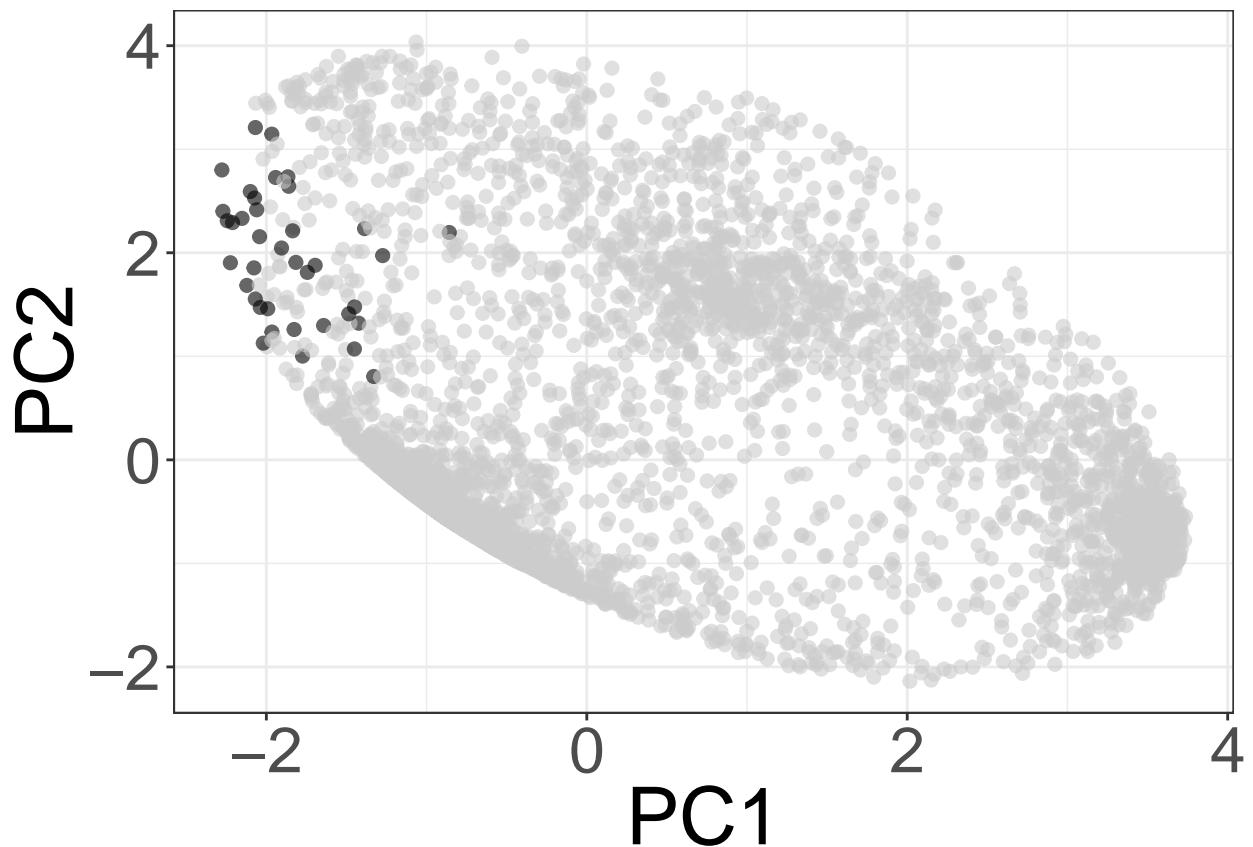
```
clusterVis_color(24)
```

```
## Using gene as id variables
```

Scaled Gene Expression



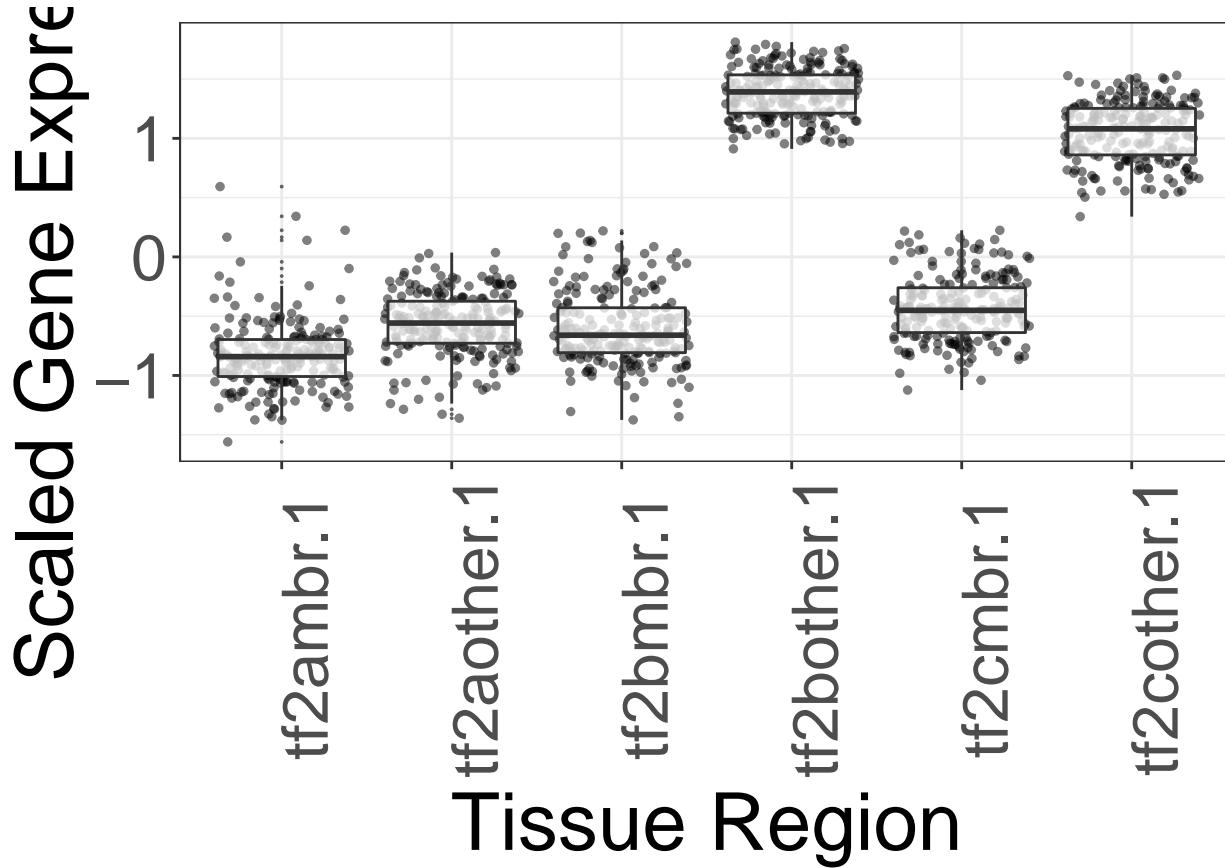
clusterVis_PCA(24)



Cluster 25

```
clusterVis(25)
```

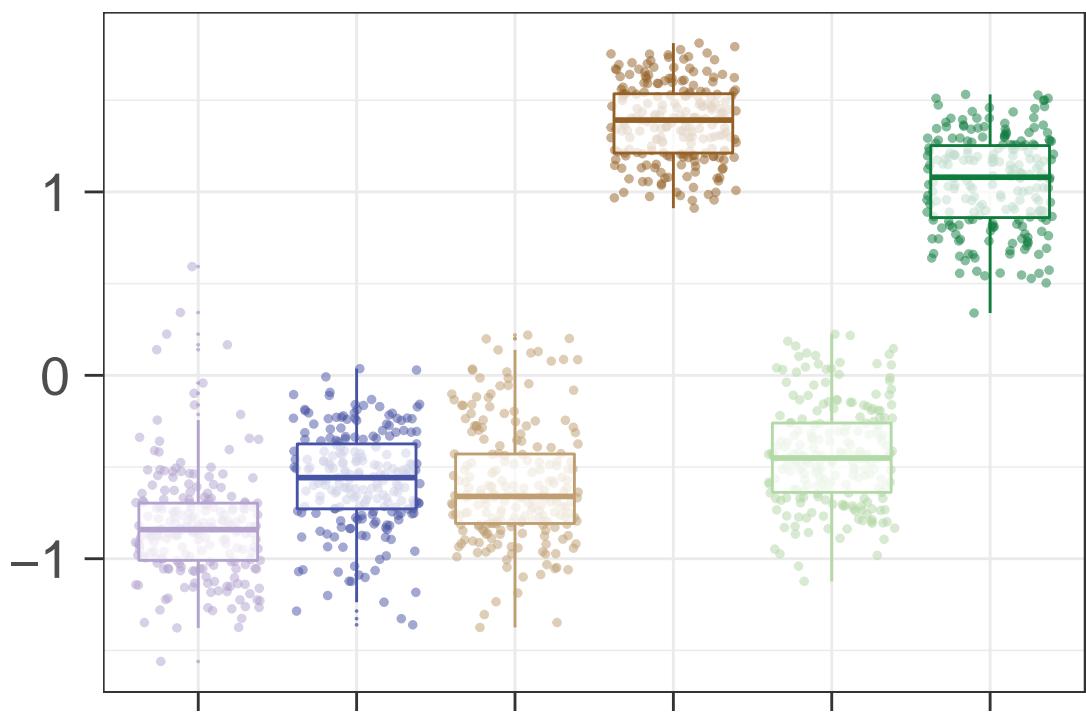
```
## Using gene as id variables
```



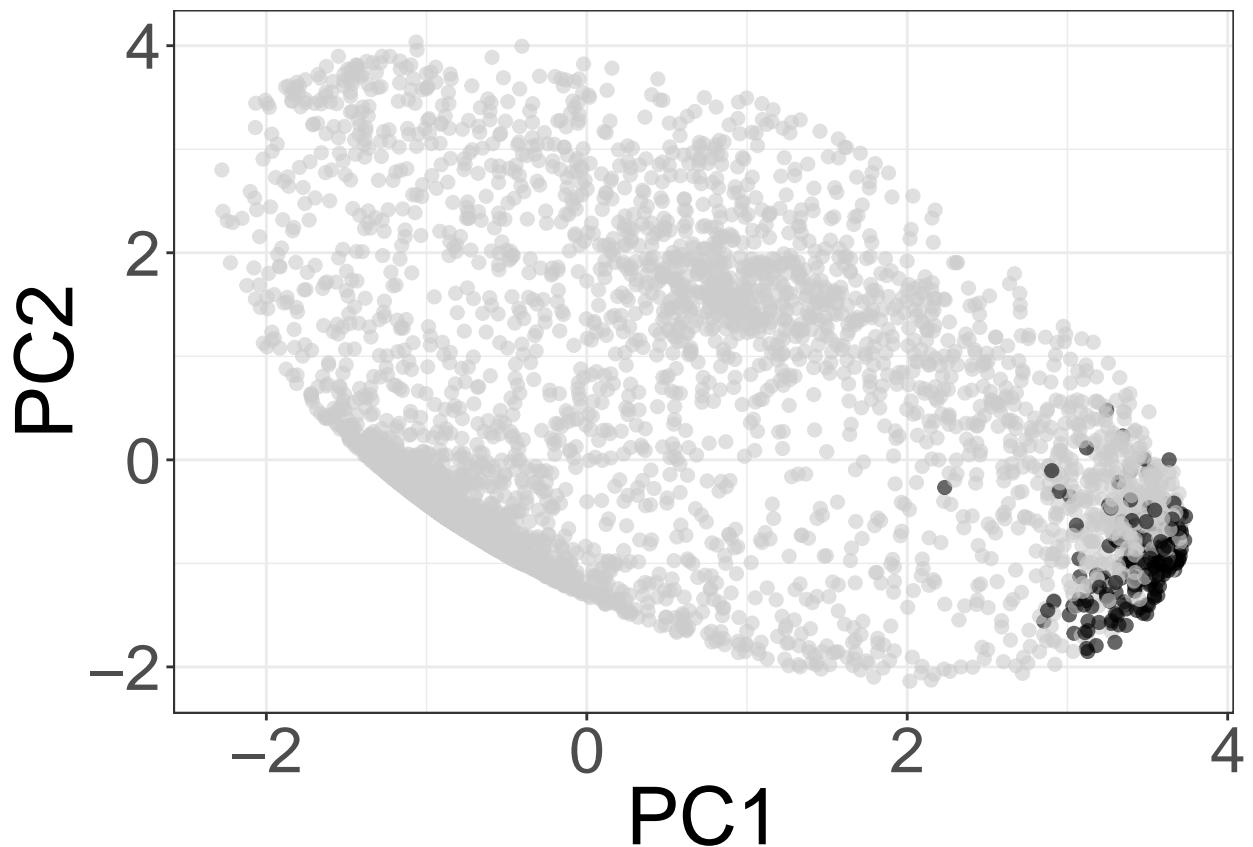
```
clusterVis_color(25)
```

```
## Using gene as id variables
```

Scaled Gene Expression



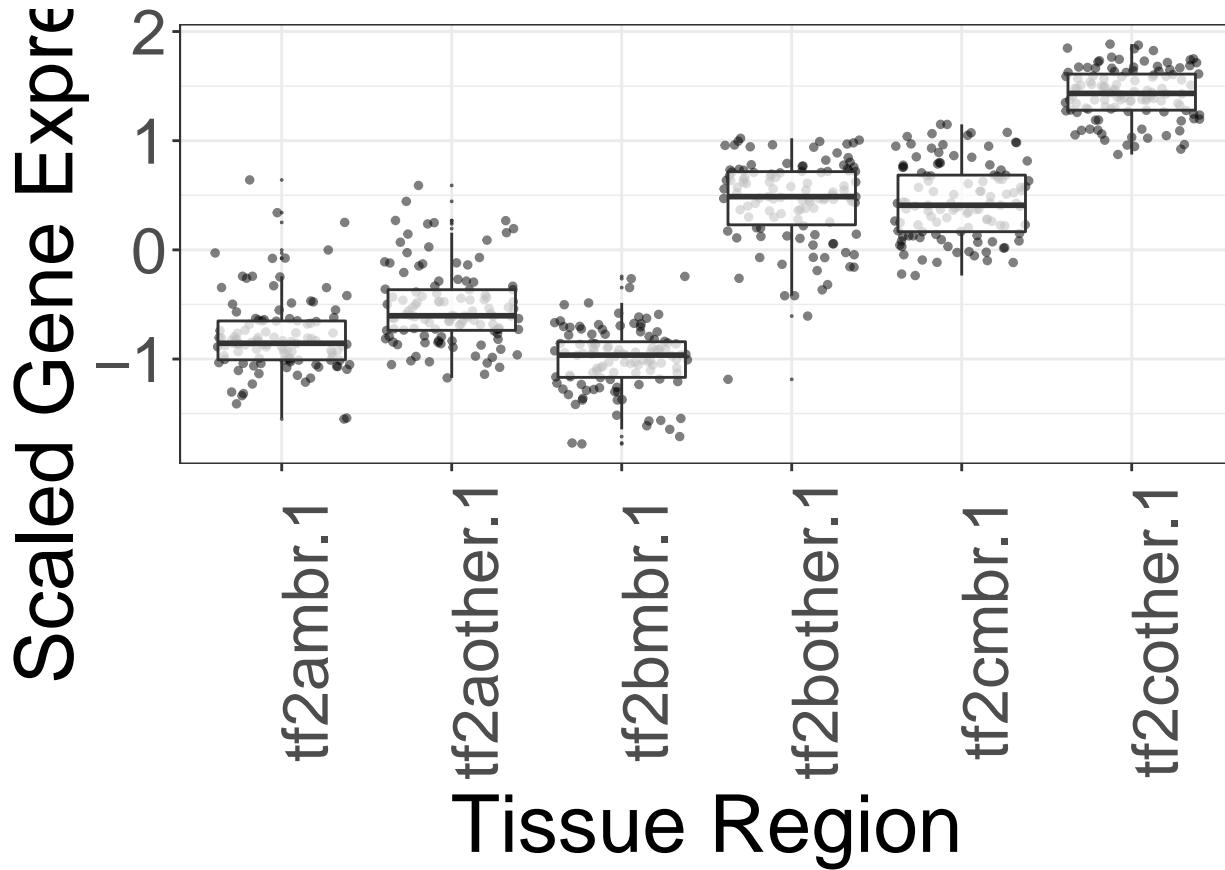
clusterVis_PCA(25)



Cluster 26

```
clusterVis(26)
```

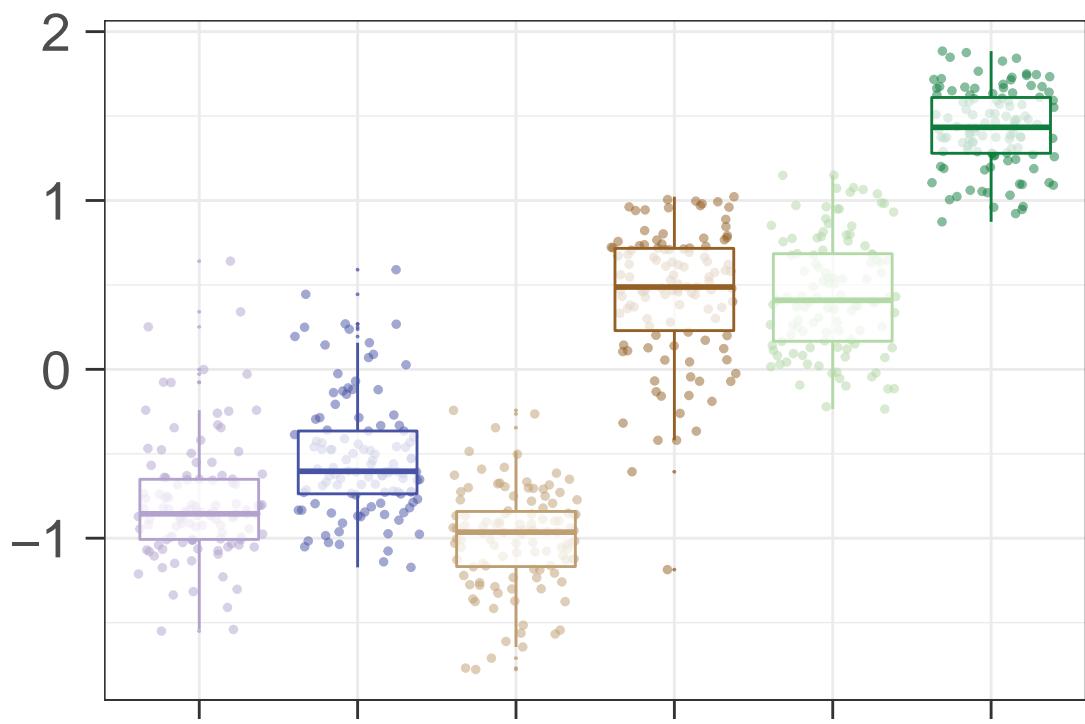
```
## Using gene as id variables
```



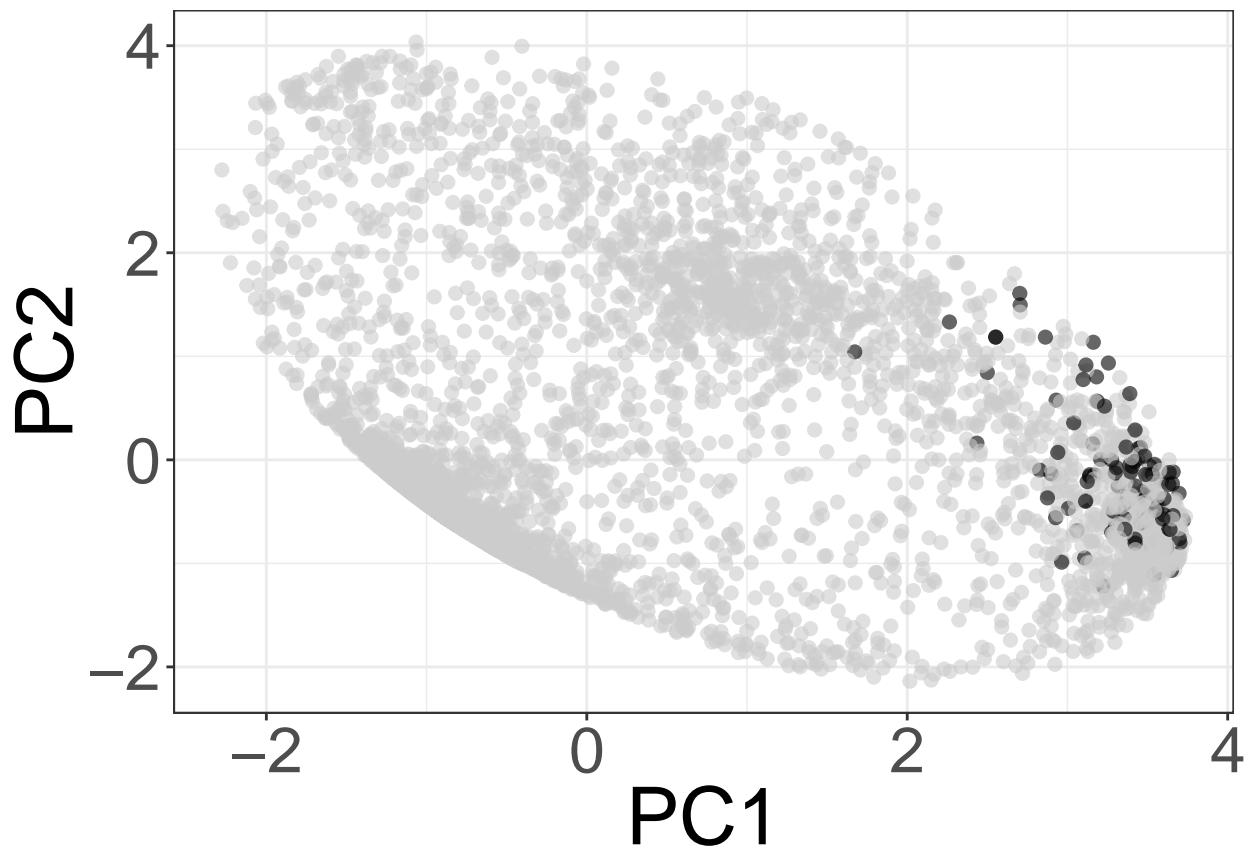
```
clusterVis_color(26)
```

```
## Using gene as id variables
```

Scaled Gene Expression



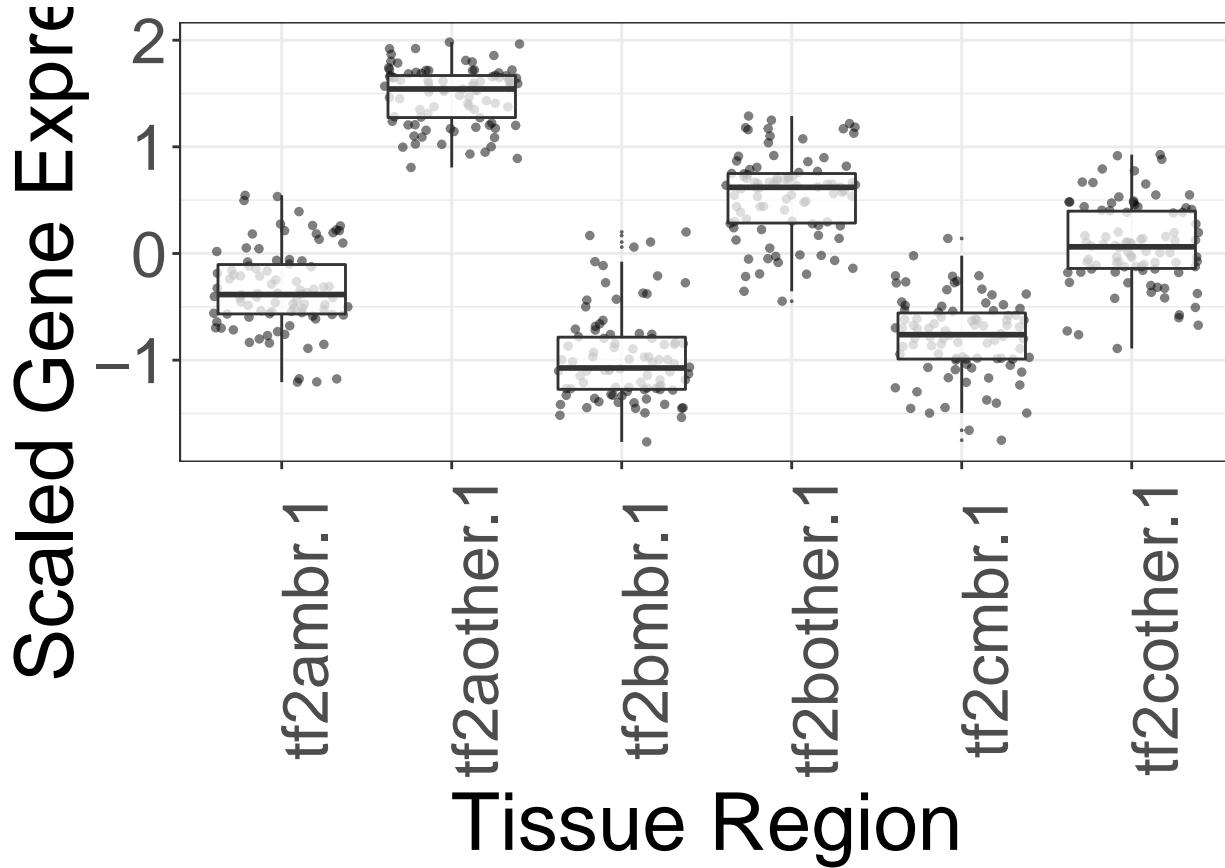
clusterVis_PCA(26)



Cluster 27

```
clusterVis(27)
```

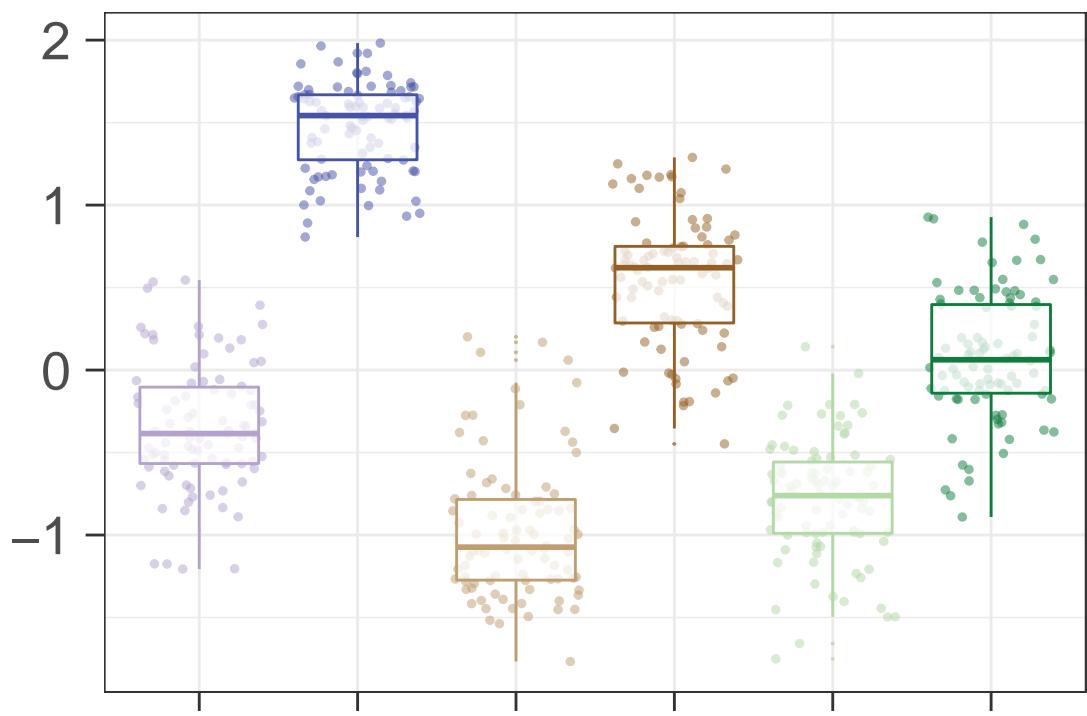
```
## Using gene as id variables
```



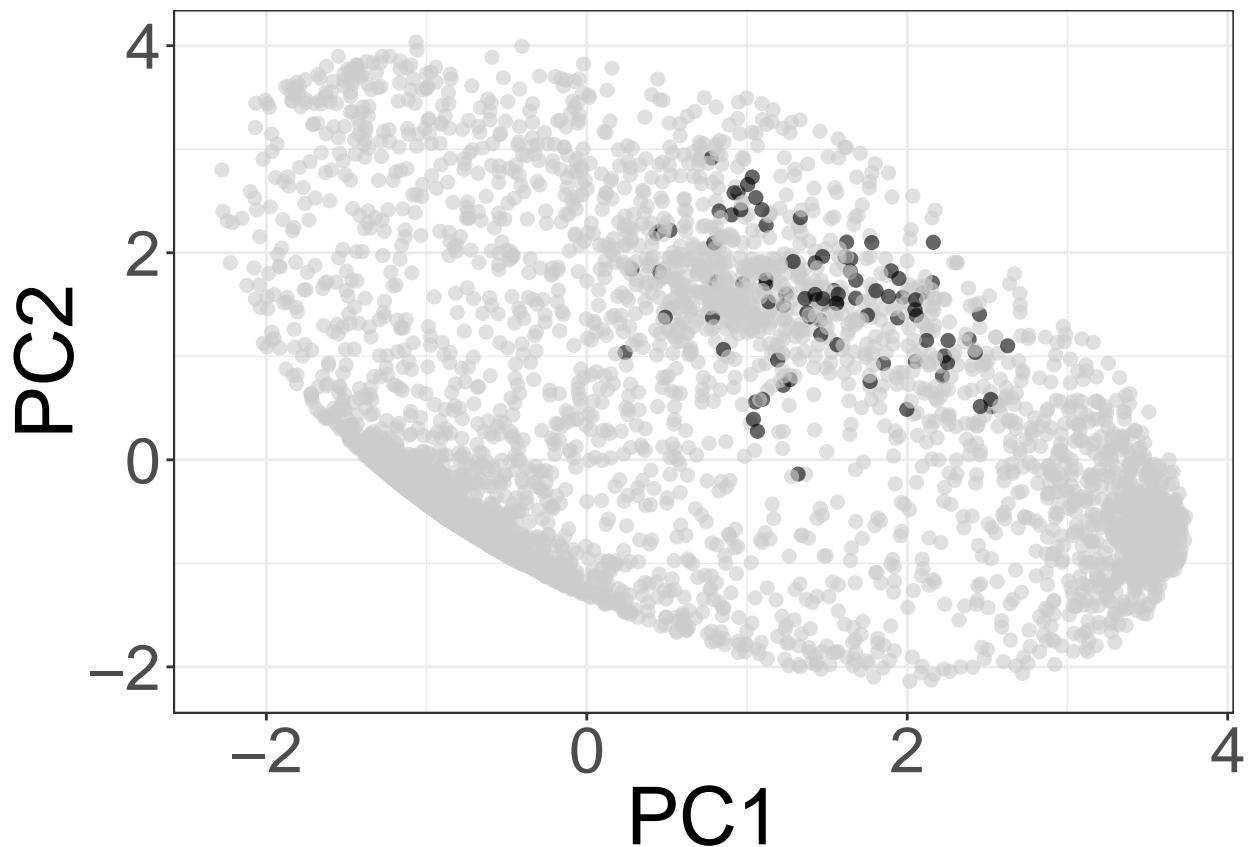
```
clusterVis_color(27)
```

```
## Using gene as id variables
```

Scaled Gene Expression



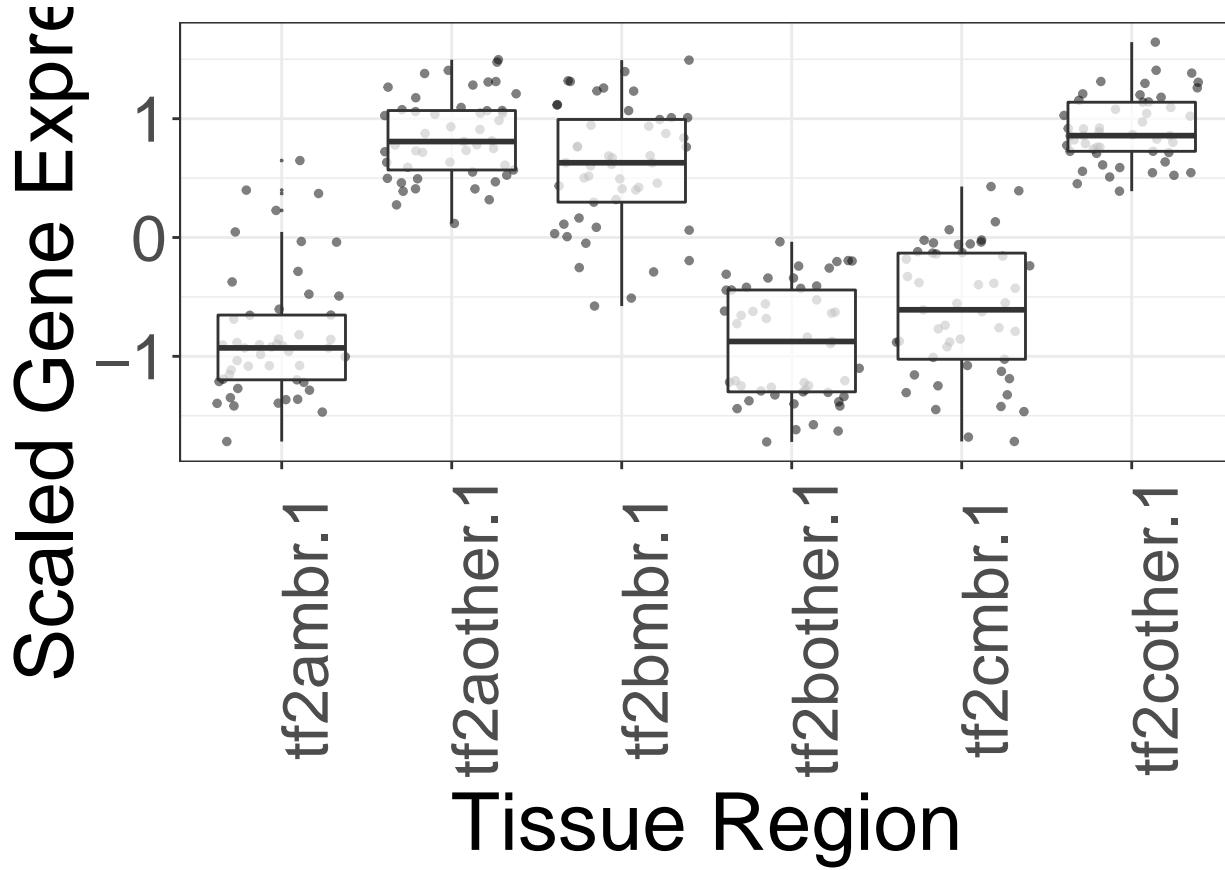
clusterVis_PCA(27)



Cluster 28

```
clusterVis(28)
```

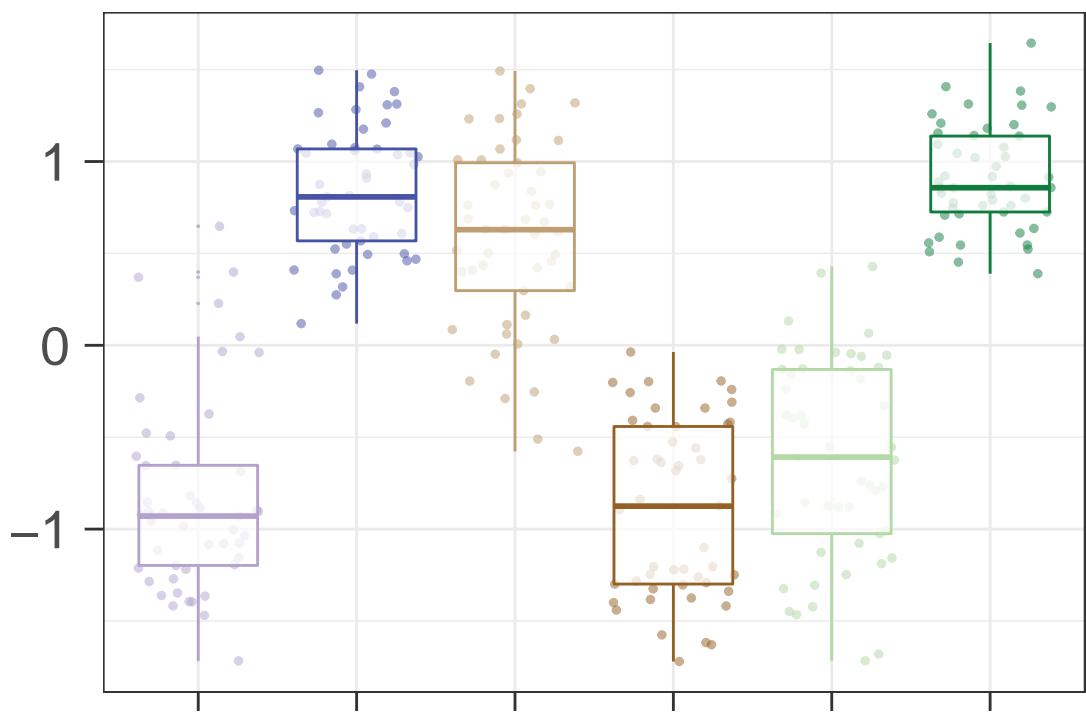
```
## Using gene as id variables
```



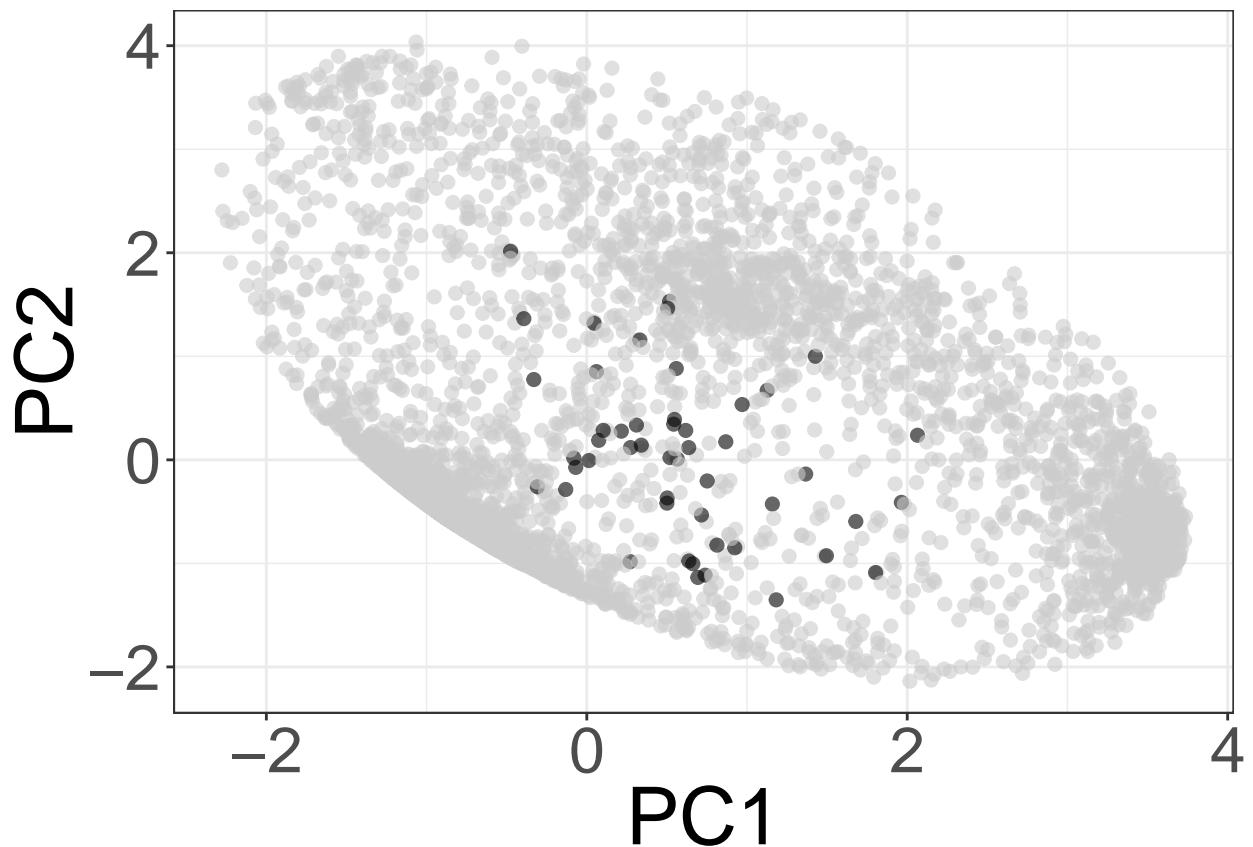
```
clusterVis_color(28)
```

```
## Using gene as id variables
```

Scaled Gene Expression



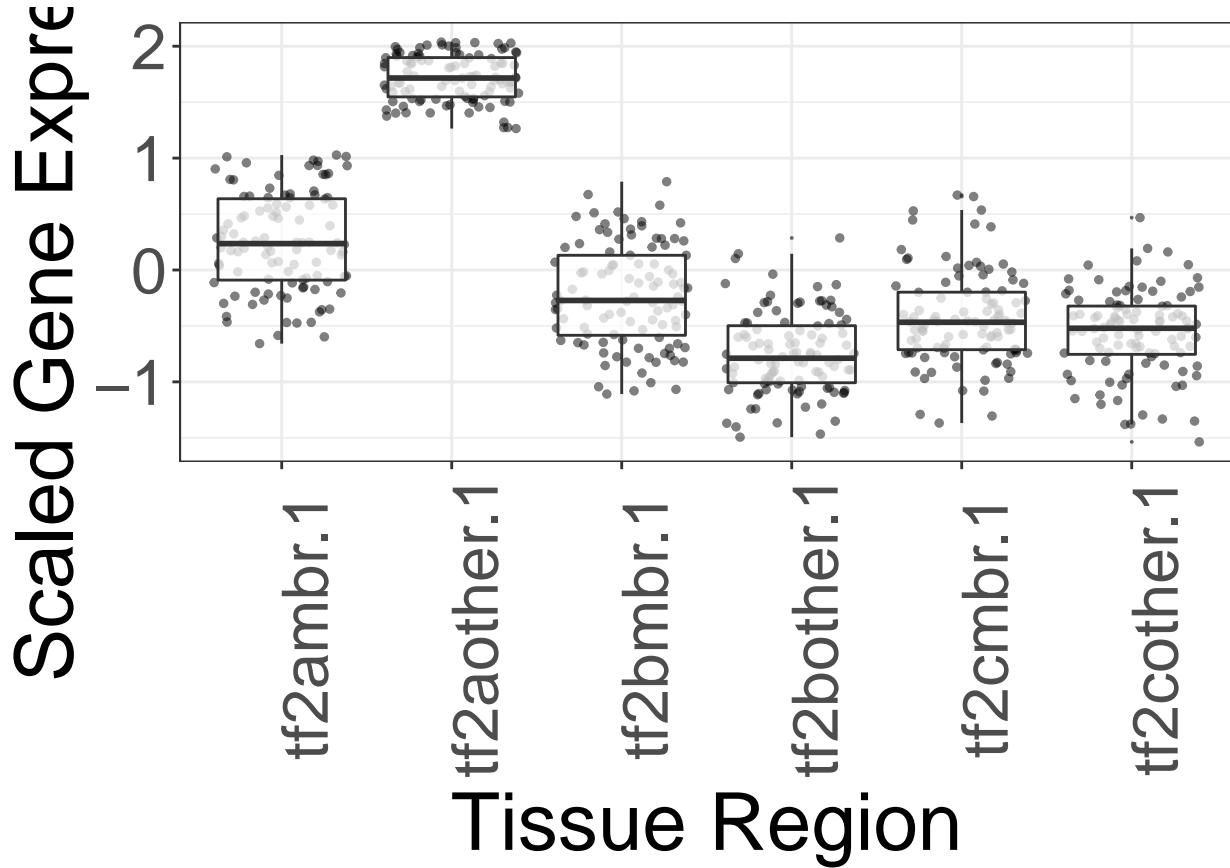
clusterVis_PCA(28)



Cluster 29

```
clusterVis(29)
```

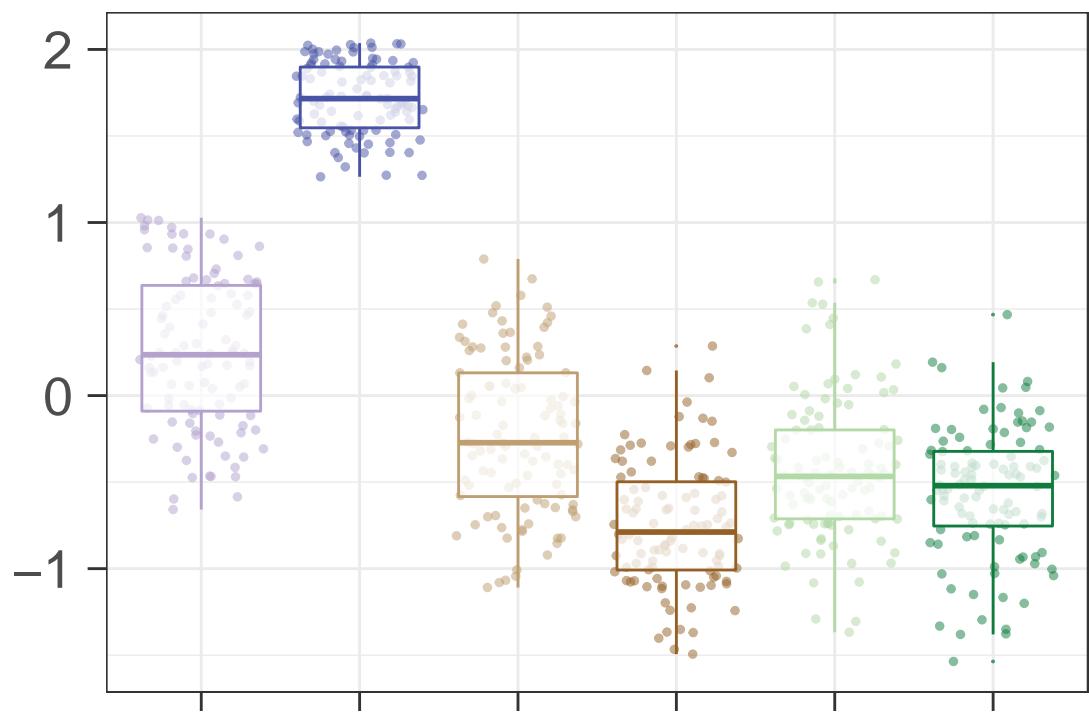
```
## Using gene as id variables
```



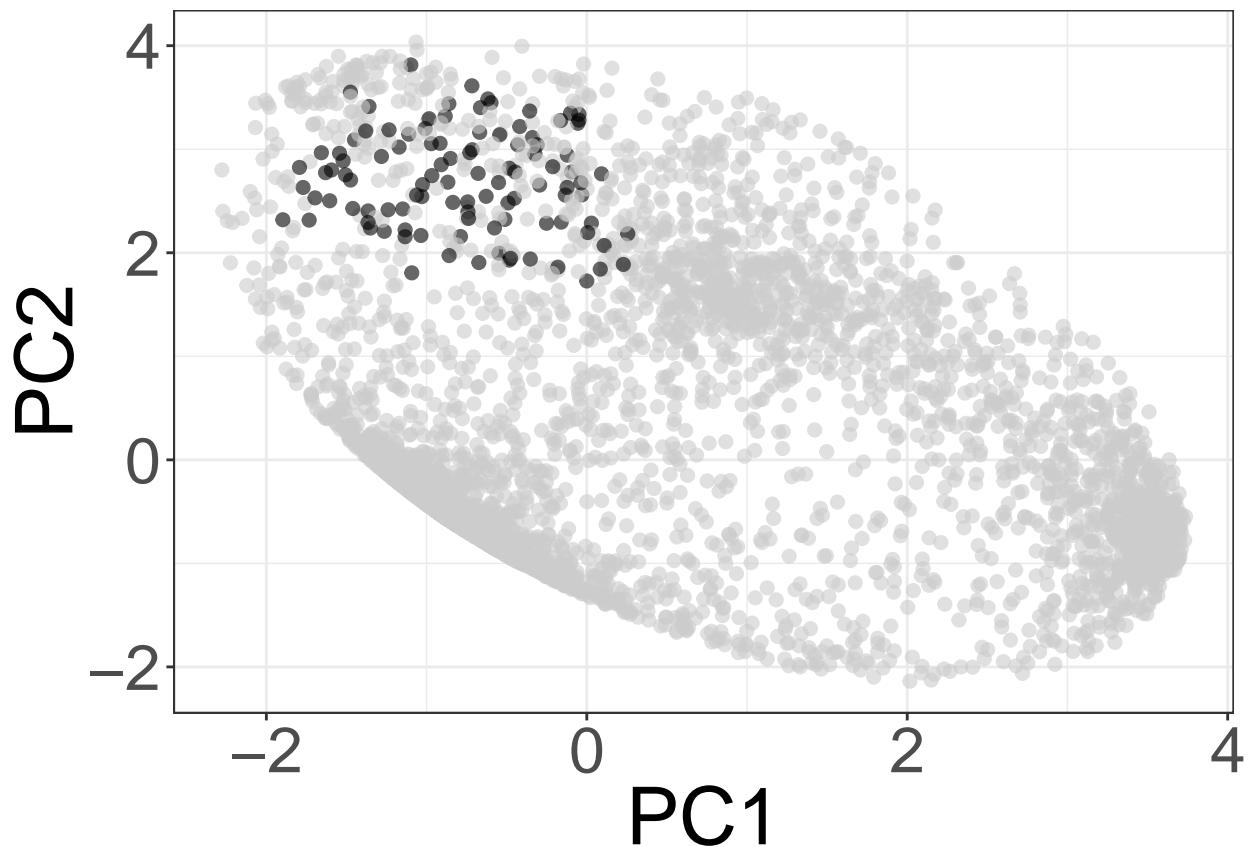
```
clusterVis_color(29)
```

```
## Using gene as id variables
```

Scaled Gene Expression



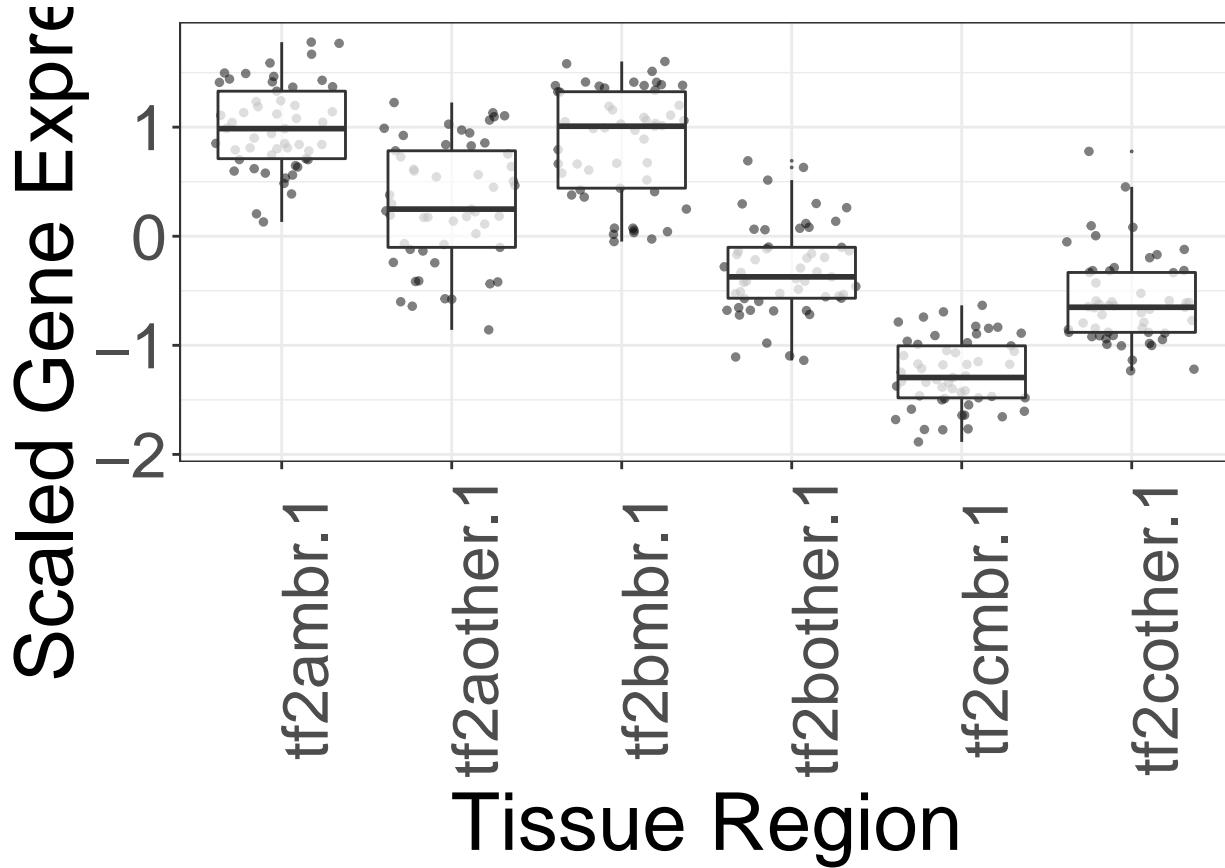
clusterVis_PCA(29)



Cluster 30

```
clusterVis(30)
```

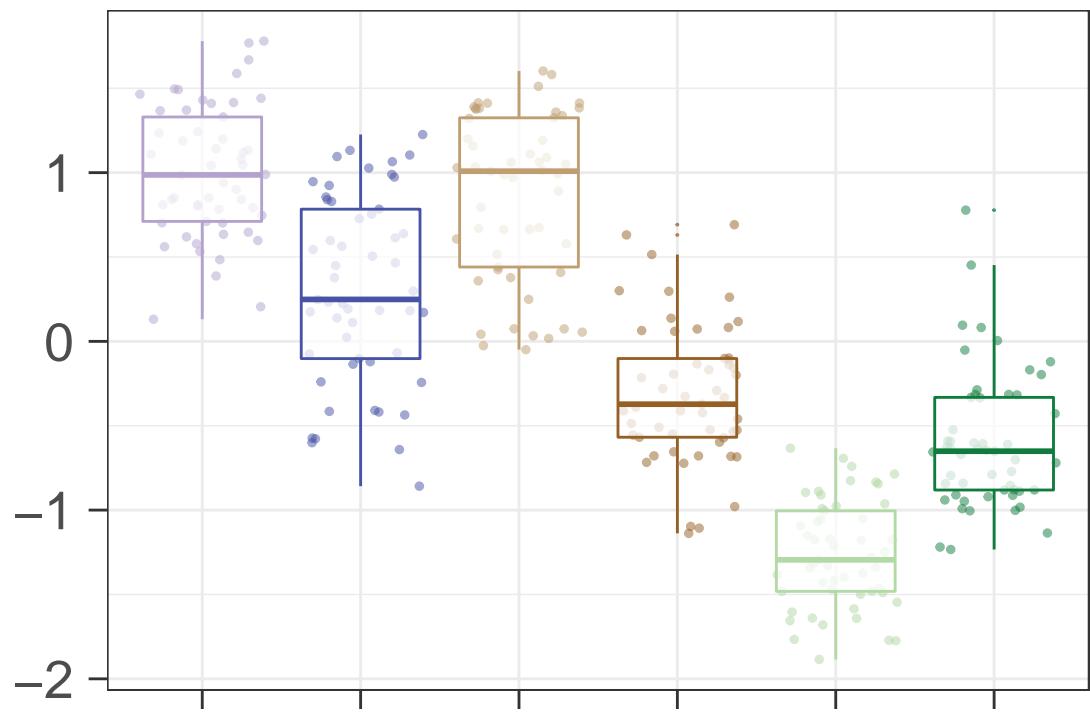
```
## Using gene as id variables
```



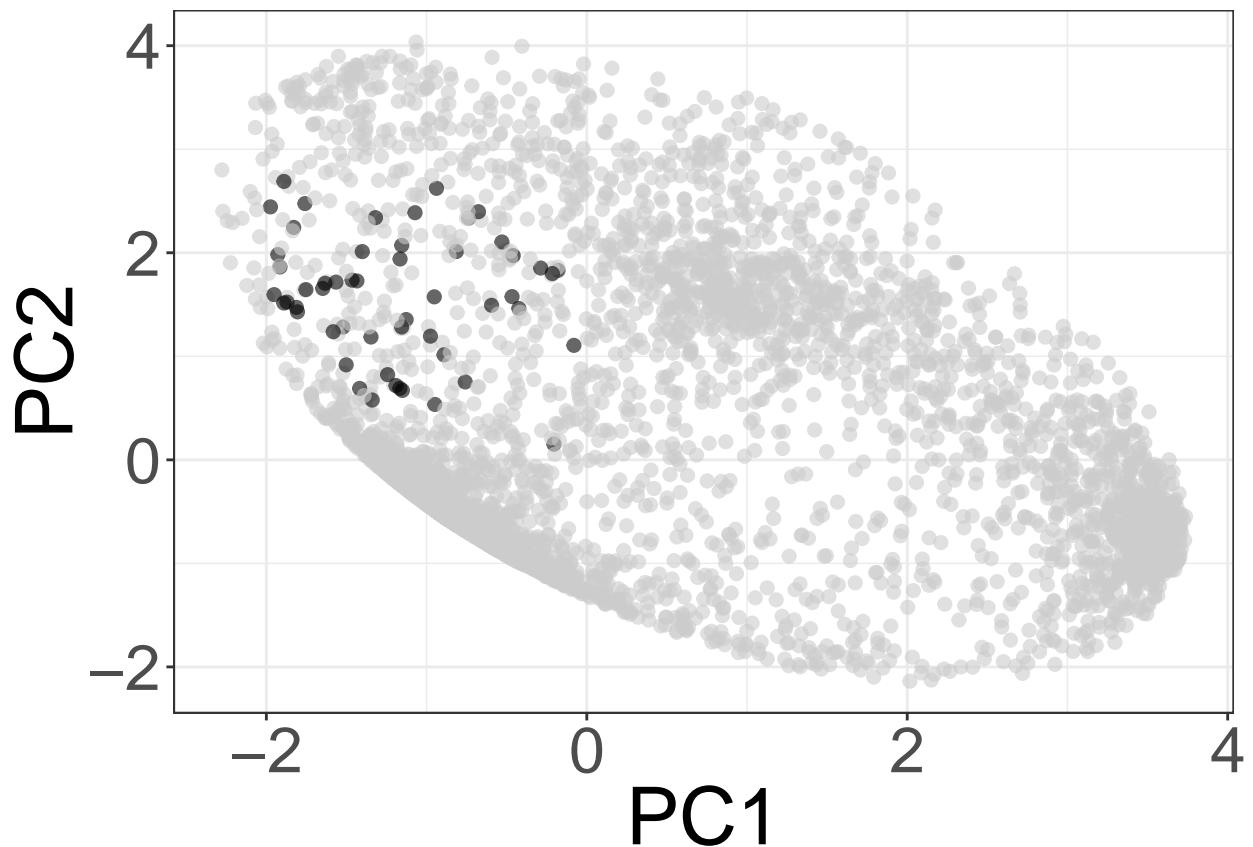
```
clusterVis_color(30)
```

```
## Using gene as id variables
```

Scaled Gene Expression



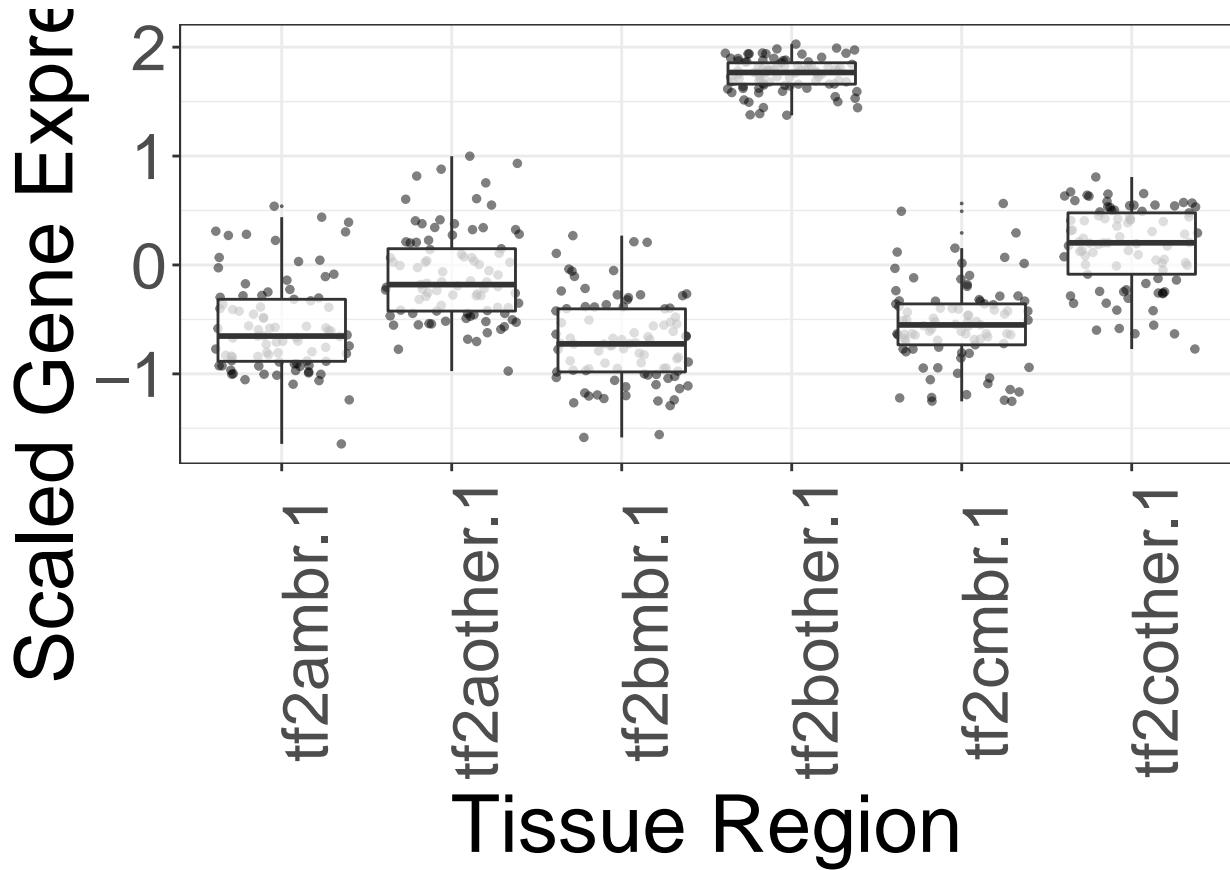
clusterVis_PCA(30)



Cluster 31

```
clusterVis(31)
```

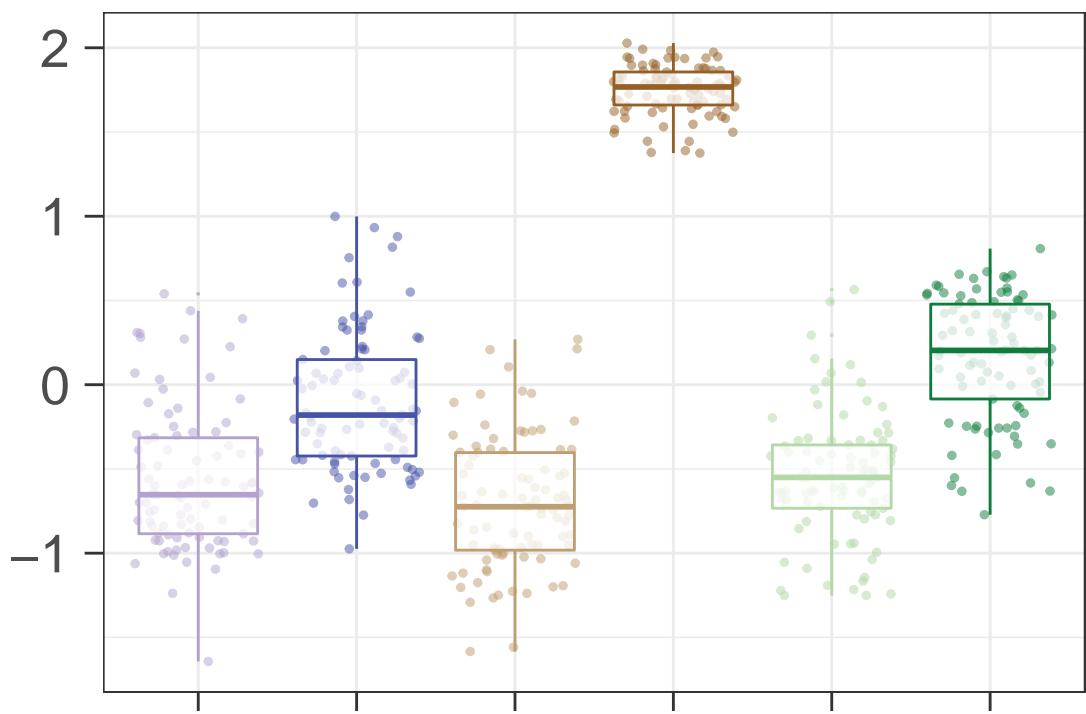
```
## Using gene as id variables
```



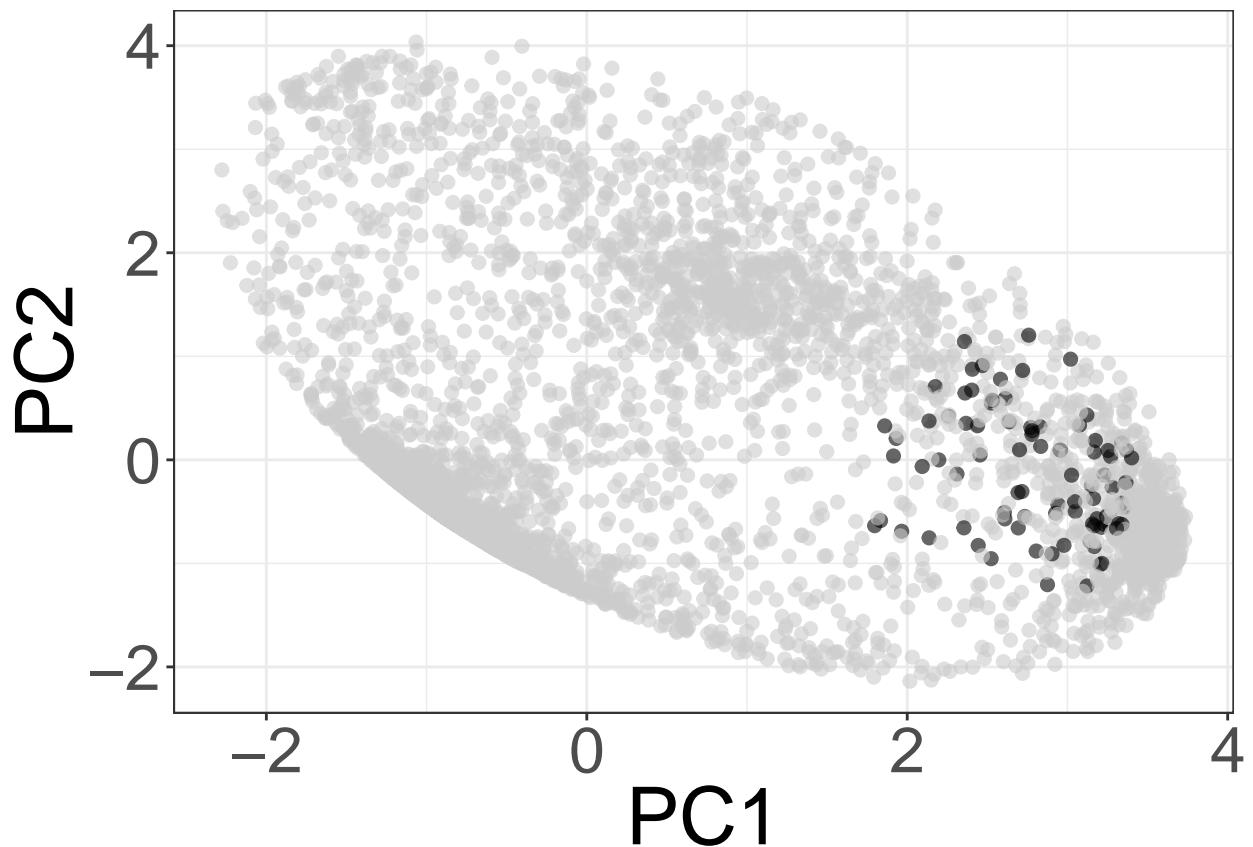
```
clusterVis_color(31)
```

```
## Using gene as id variables
```

Scaled Gene Expression



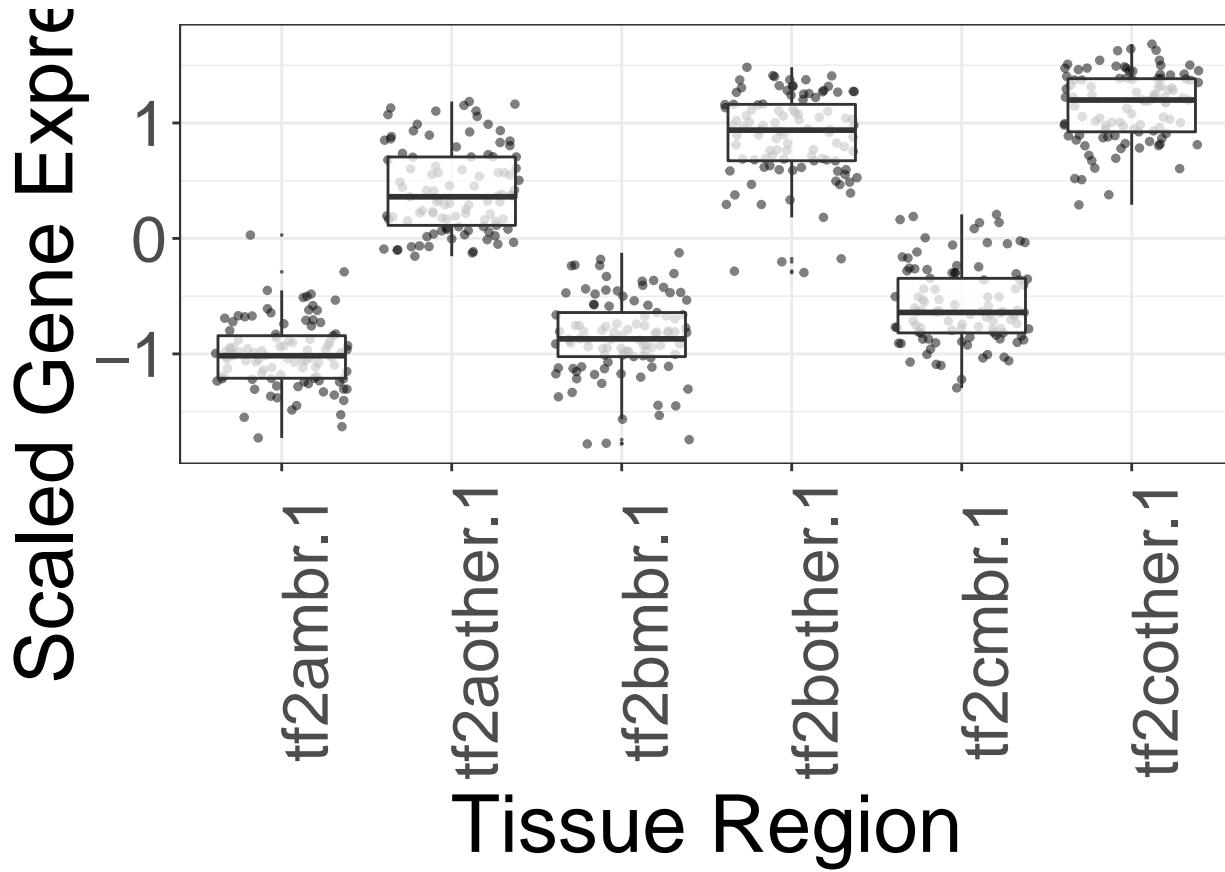
clusterVis_PCA(31)



Cluster 32

```
clusterVis(32)
```

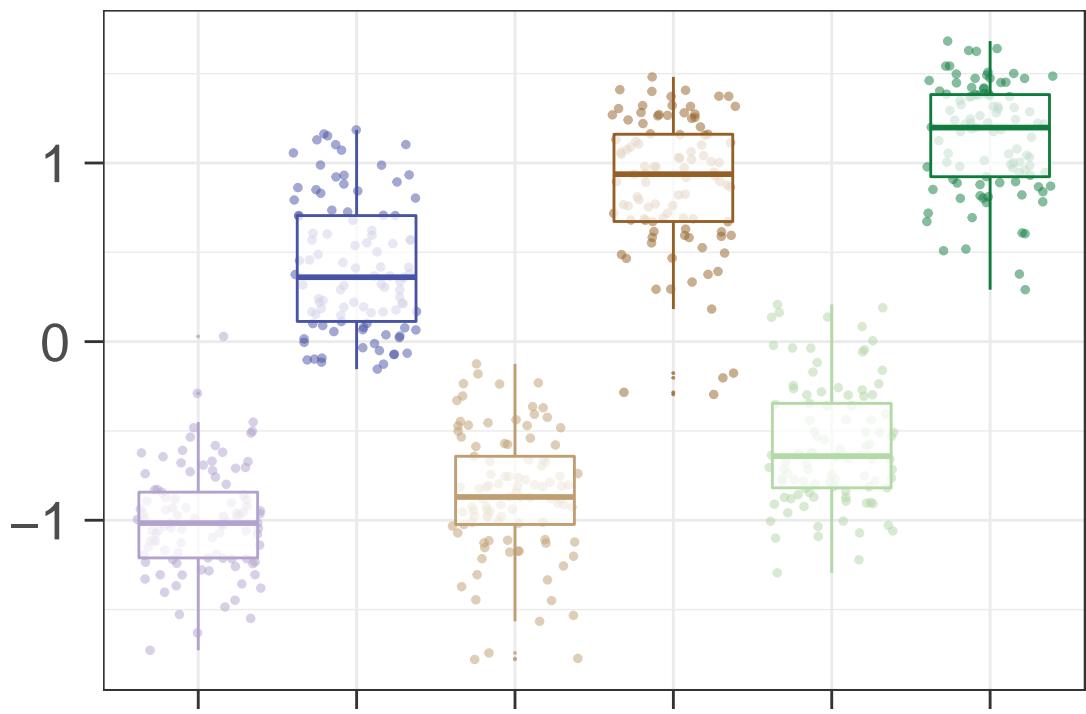
```
## Using gene as id variables
```



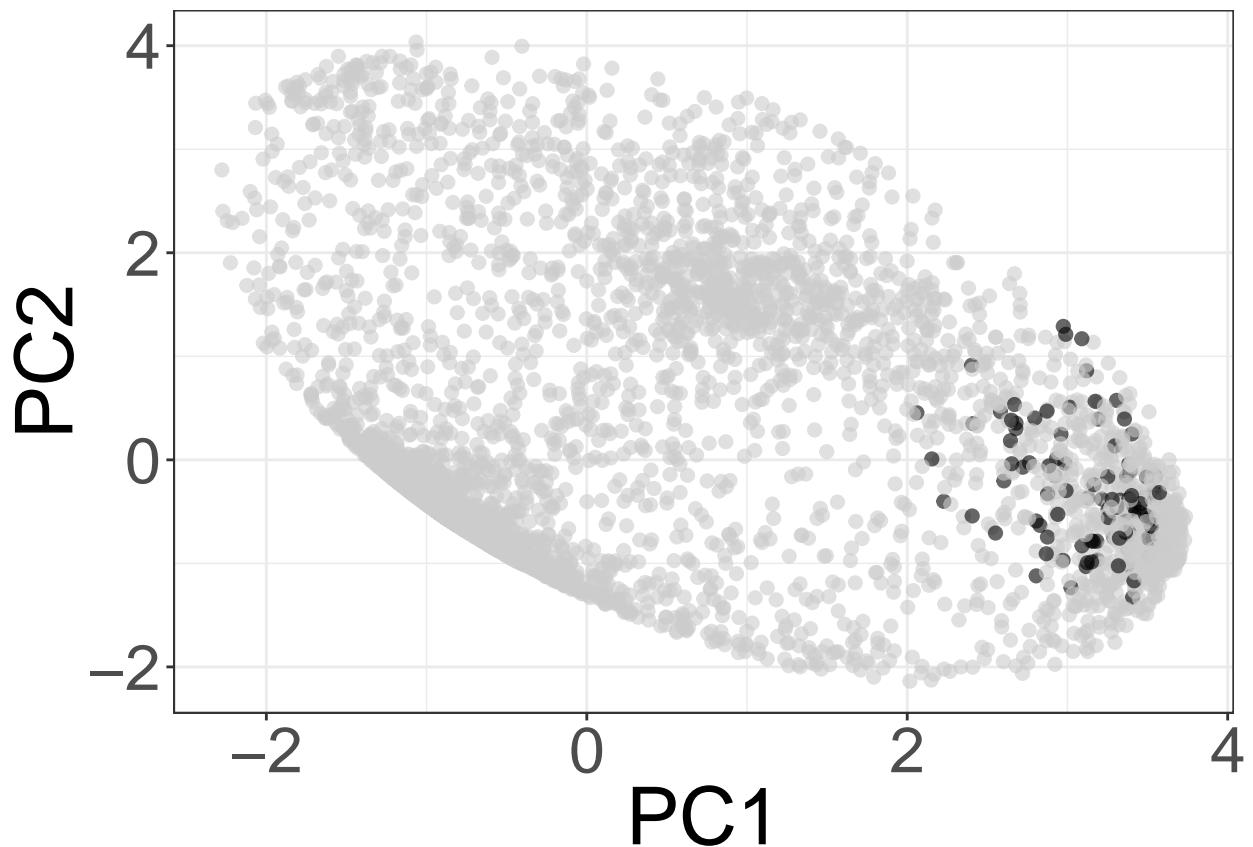
```
clusterVis_color(32)
```

```
## Using gene as id variables
```

Scaled Gene Expression



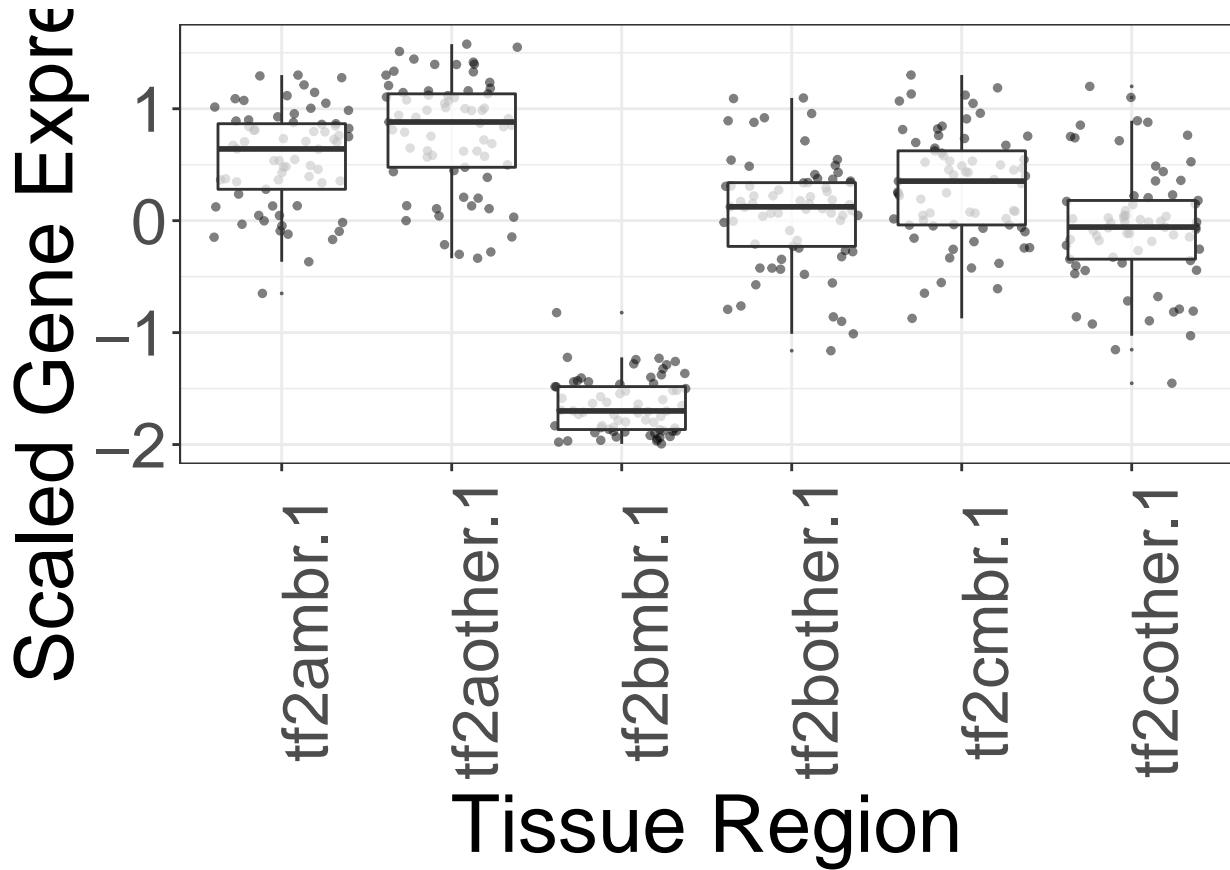
clusterVis_PCA(32)



Cluster 33

```
clusterVis(33)
```

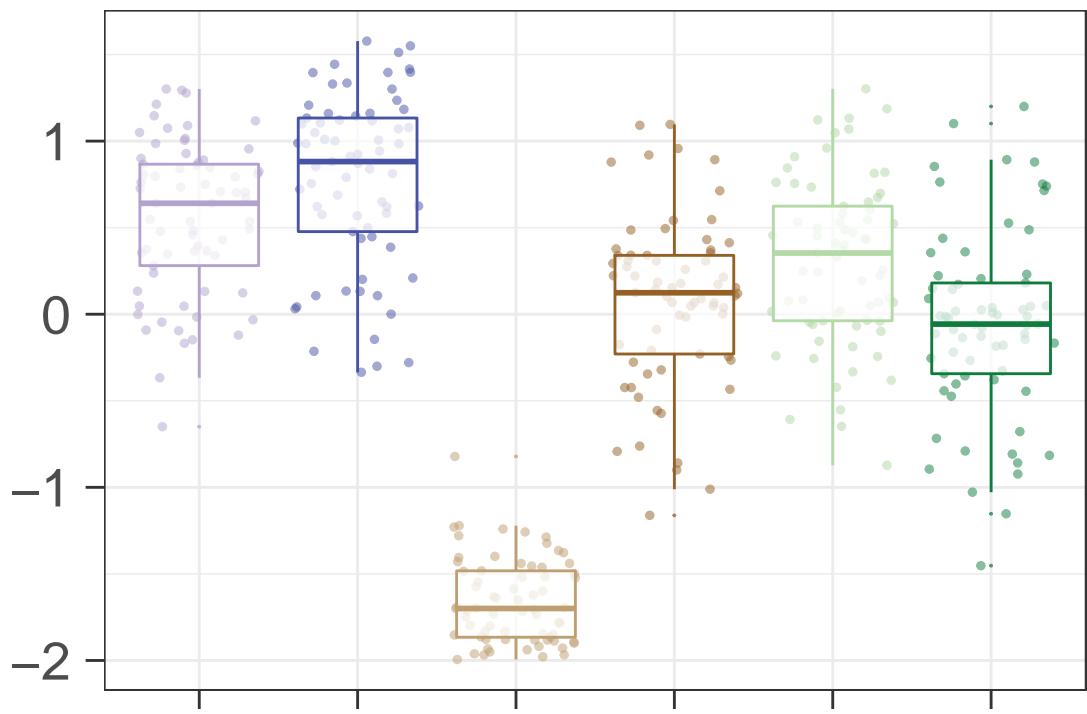
```
## Using gene as id variables
```



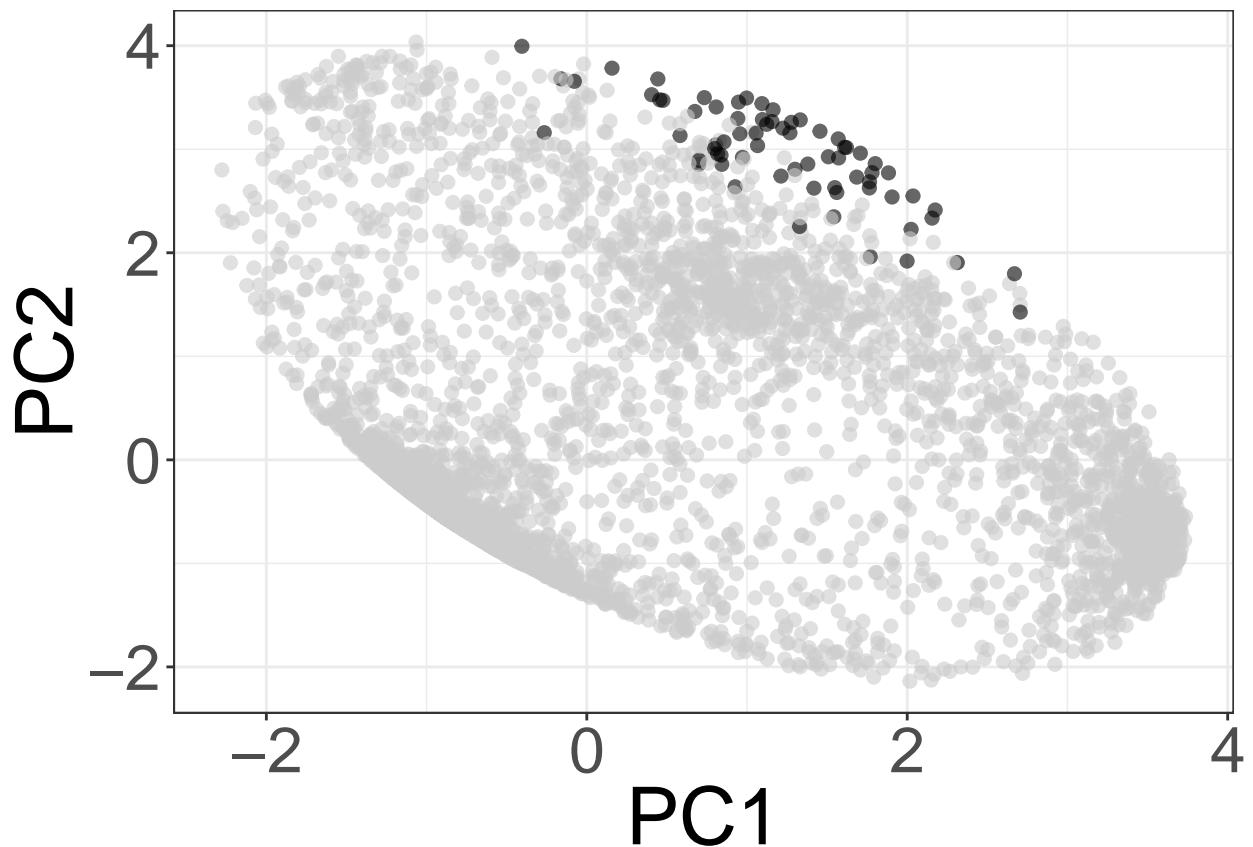
```
clusterVis_color(33)
```

```
## Using gene as id variables
```

Scaled Gene Expression



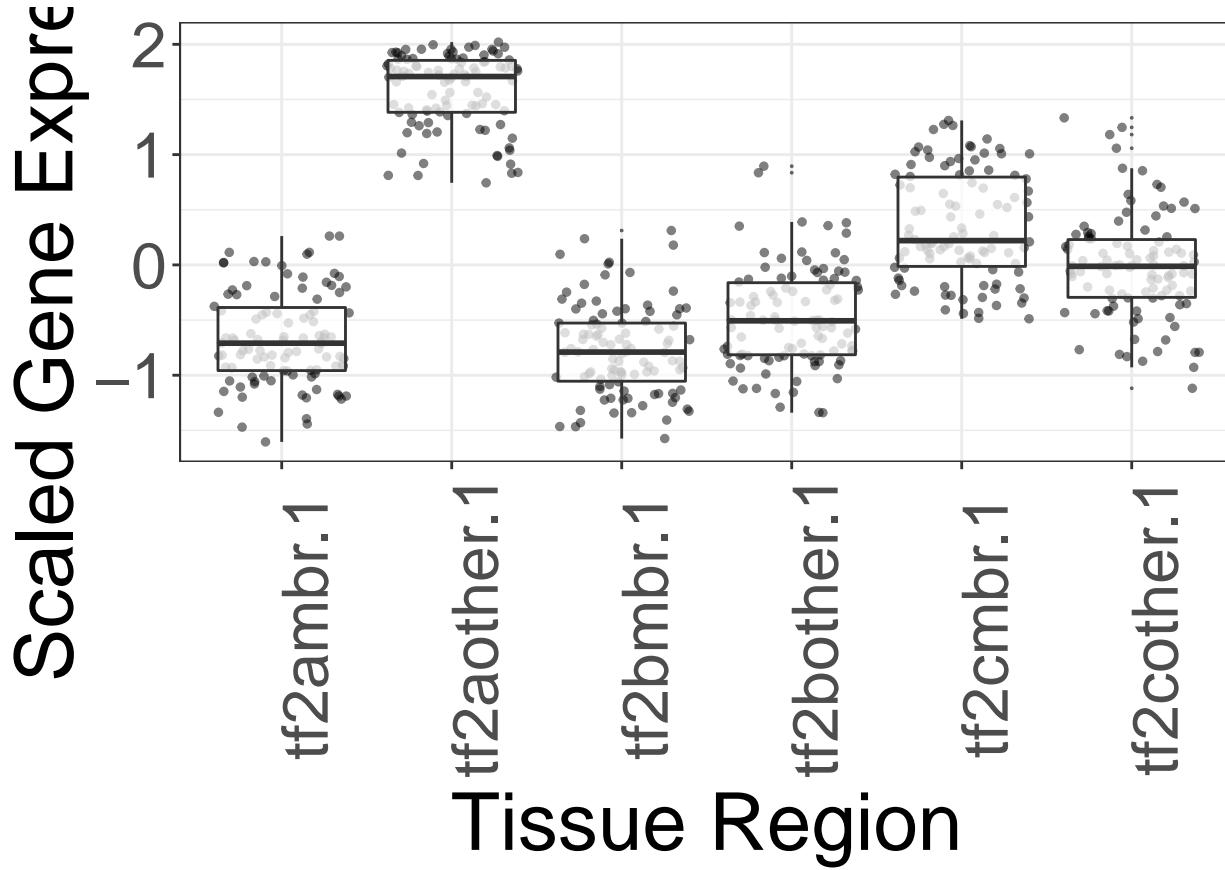
clusterVis_PCA(33)



Cluster 34

```
clusterVis(34)
```

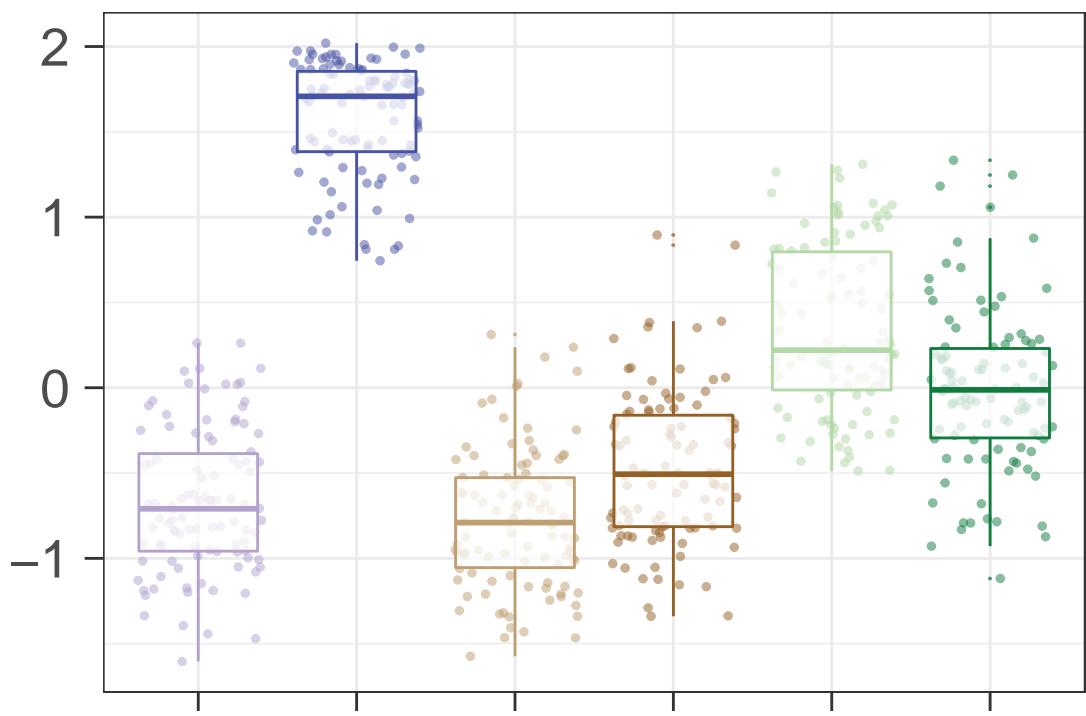
```
## Using gene as id variables
```



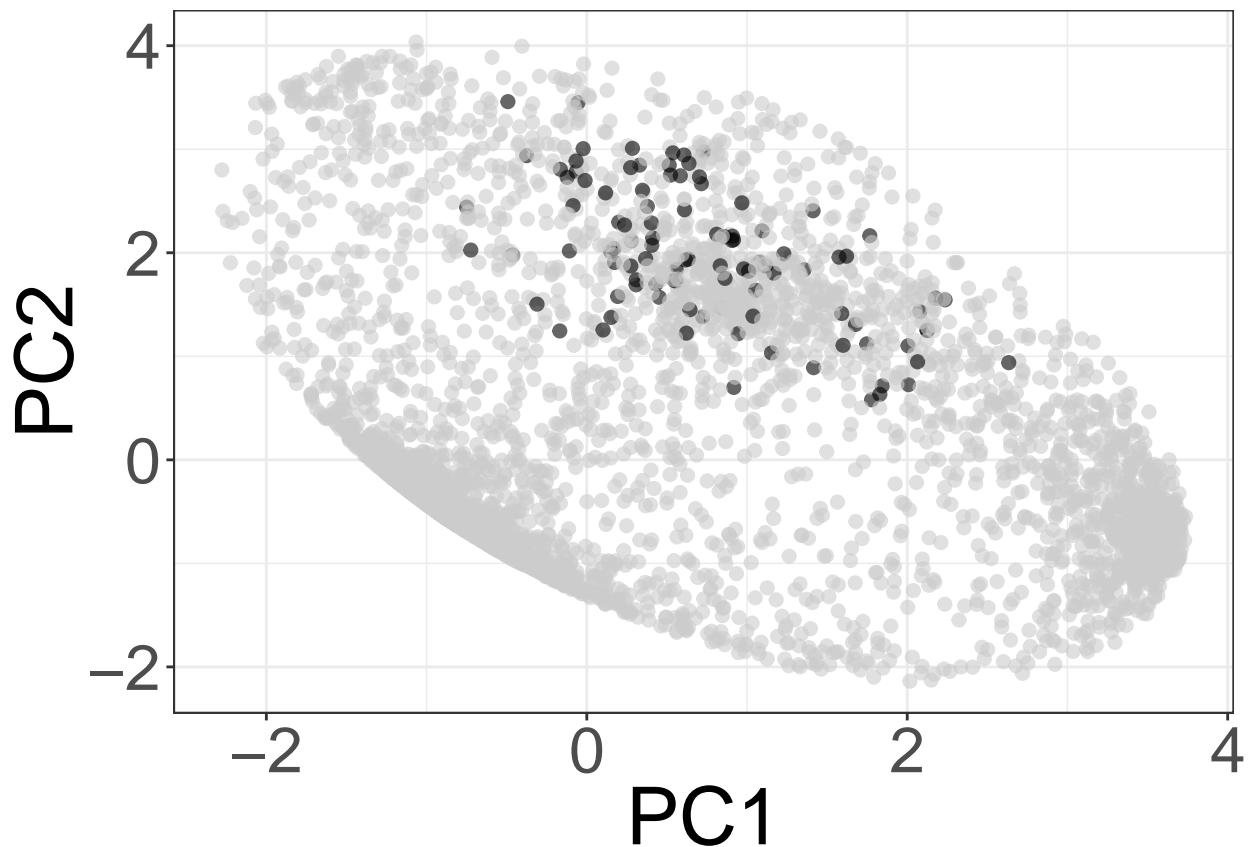
```
clusterVis_color(34)
```

```
## Using gene as id variables
```

Scaled Gene Expression



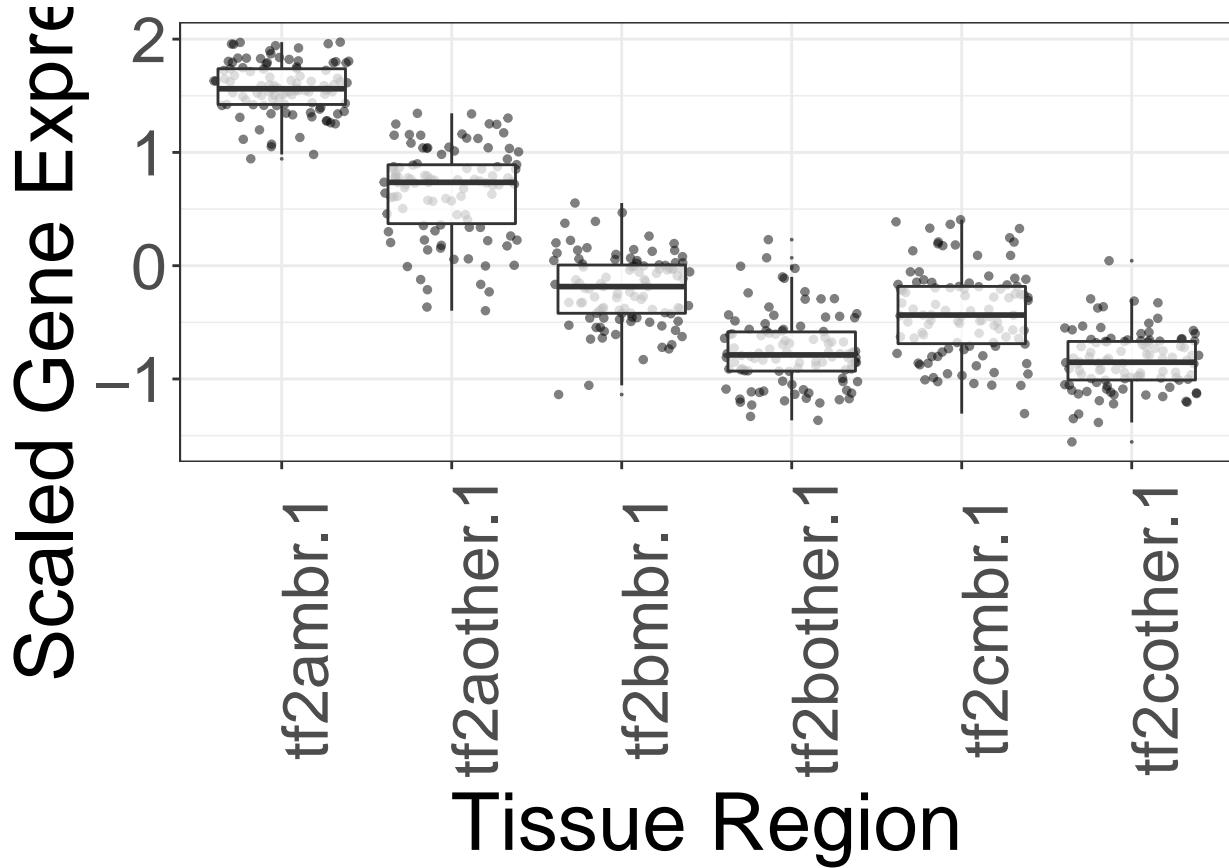
clusterVis_PCA(34)



Cluster 35

```
clusterVis(35)
```

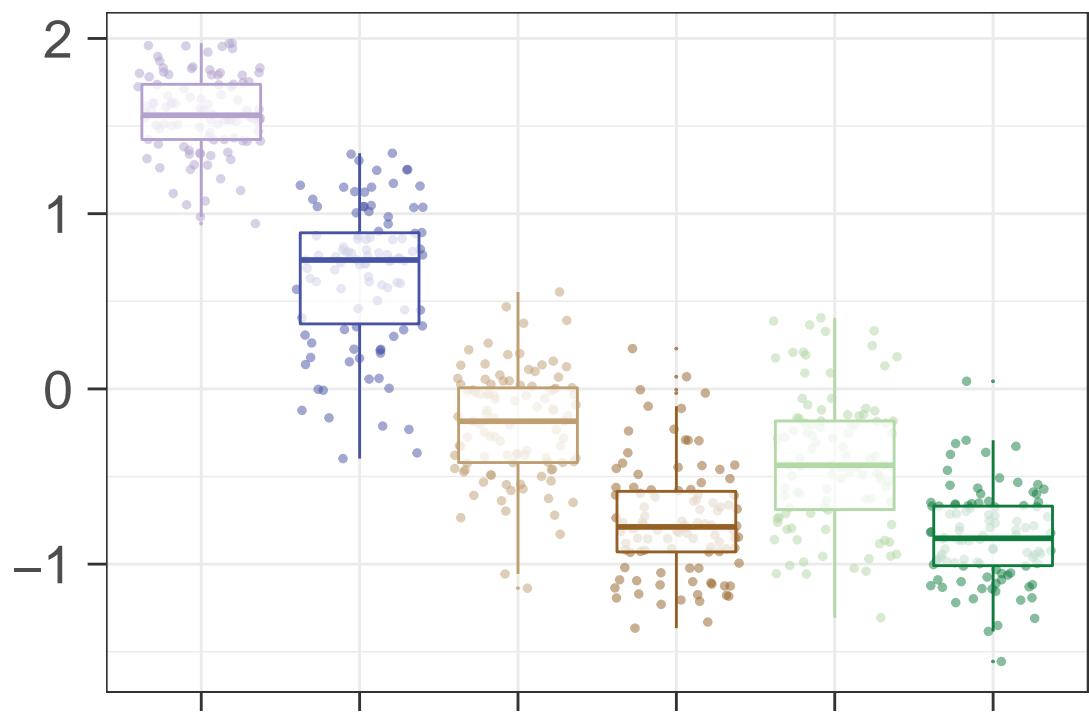
```
## Using gene as id variables
```



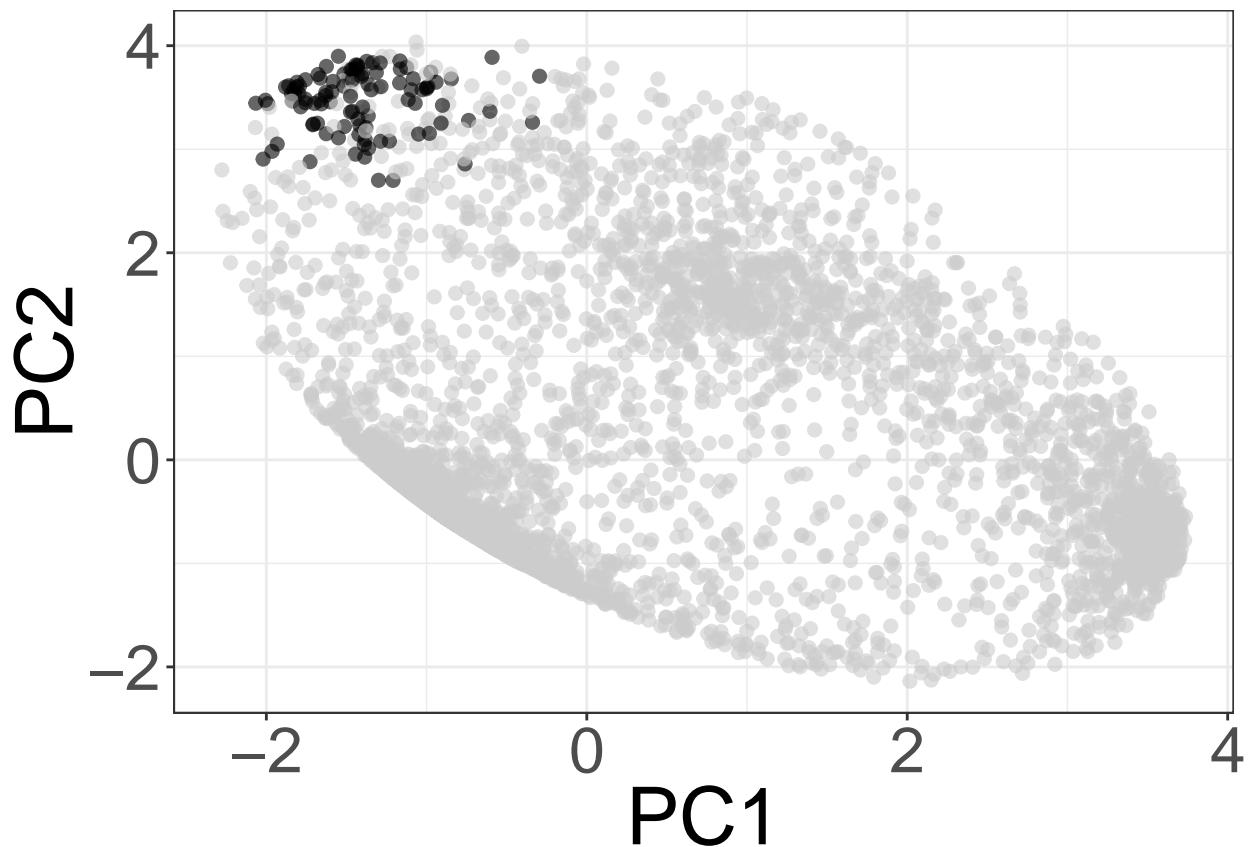
```
clusterVis_color(35)
```

```
## Using gene as id variables
```

Scaled Gene Expression



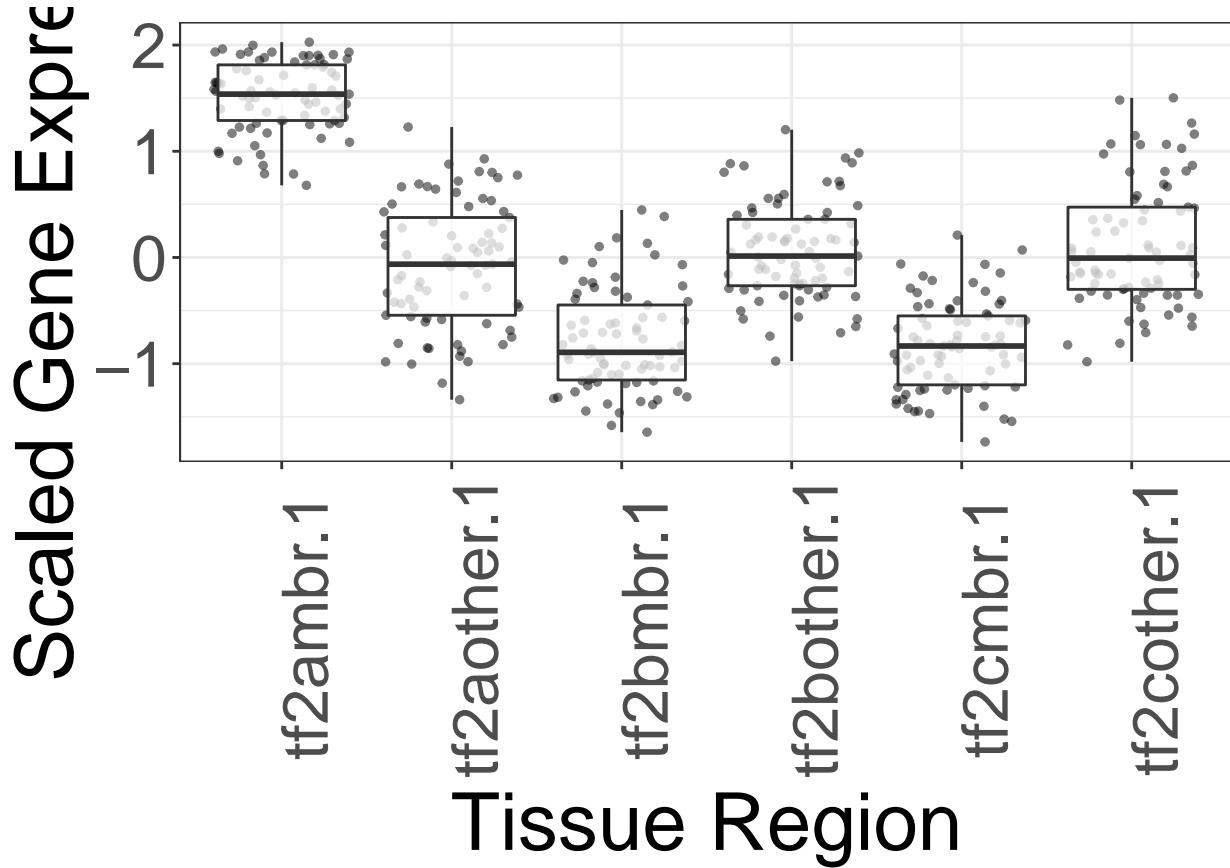
clusterVis_PCA(35)



Cluster 36

```
clusterVis(36)
```

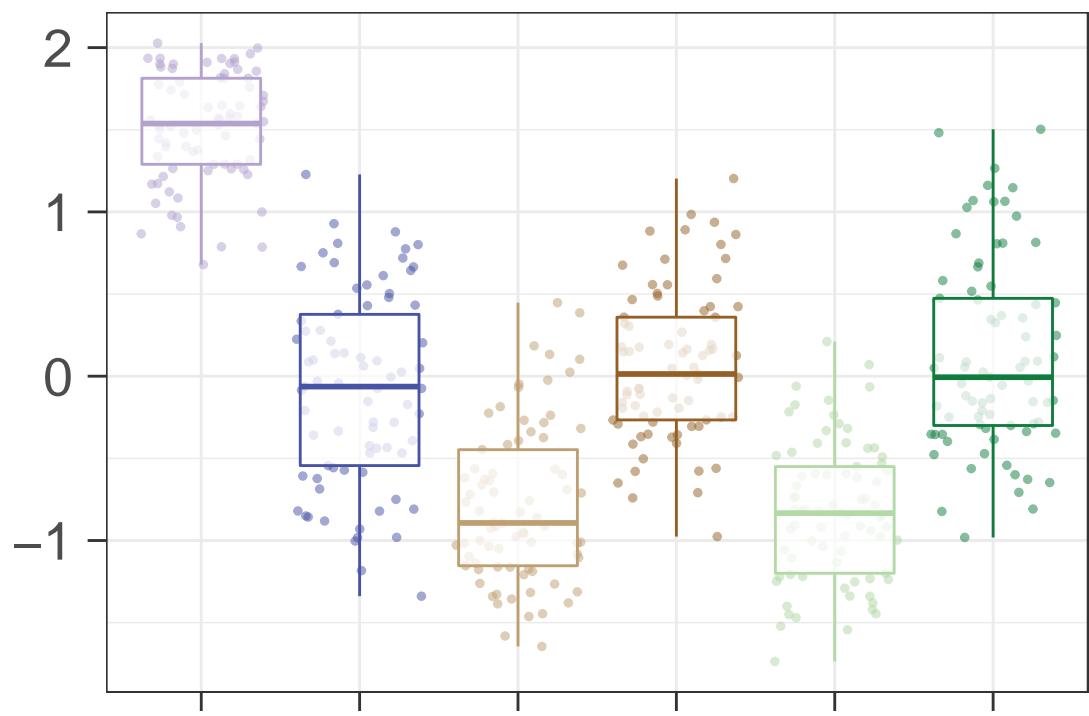
```
## Using gene as id variables
```



```
clusterVis_color(36)
```

```
## Using gene as id variables
```

Scaled Gene Expression



clusterVis_PCA(35)

