

# Figure 1E PCA

RAC

23/04/2020

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(gridExtra)
library(tibble)
theme_set(theme_bw())
library(extrafont)
loadfonts(quiet = T)
library(ggrepel)
library(corrplot)
getwd()
```

```
## [1] "/Users/rosscordiner/Dropbox/OneDrive - Aarhus Universitet/PostDocJensenOD/Documents/xiCLIP Paper"
```

```
library(matrixStats)
`%!in%` = Negate(`%in%`)
library("FactoMineR")
library("factoextra")
```

#load data

Count files for PCA analysis were generated by intersecting bed files from the mapped reads of each tiCLIP sample with a bed file containing all TUs segmented into 10kb blocks i.e. each bed entry = 1 x 10kb segment of a TU

```
# counts of tiCLIP reads from genes which have been segmented into 10kb bins
all10kbbinrRNA <- read.csv("../data/xiCLIP_all.hg38HeLaSoren10kbins.NormTorRNAFactor.counts",
  header = F, sep = "\t")

colnames(all10kbbinrRNA) <- c("Sample", "chr", "start", "end", "ID", "binNumber",
  "strand", "count")
anno10kb <- read.table("../data/hg38_HeLa_Soren.10kbins.200409.bed", header = F,
  sep = "\t")
# 10kb segments are labelled in ascending numbers in col 5
head(anno10kb, n = 30)
```

```
##      V1      V2      V3
## 1     1 629639 630560
## 2     1 631073 631548
## 3     1 633695 634374
## 4     1 778792 787672
## 5     1 827699 837699
## 6     1 837699 842000
## 7     1 904810 914810
## 8     1 914810 918154
```

```

## 9 1 923920 933920
## 10 1 933920 940182
## 11 1 960558 964530
## 12 1 965109 965714
## 13 1 966471 975864
## 14 1 978481 981359
## 15 1 1013475 1014537
## 16 1 1020078 1030078
## 17 1 1030078 1040078
## 18 1 1040078 1050078
## 19 1 1050078 1056112
## 20 1 1064457 1066378
## 21 1 1232237 1234179
## 22 1 1273999 1279393
## 23 1 1280471 1290471
## 24 1 1290471 1292024
## 25 1 1308569 1311676
## 26 1 1324792 1328895
## 27 1 1336209 1336530
## 28 1 1341650 1342543
## 29 1 1347800 1348385
## 30 1 1351457 1355625

```

```

##
## 1 MTND2P28::unprocessed_pseudogene::1::921 1 +
## 2 AC114498.1,MIR6723::unprocessed_pseudogene::1::475 1 +
## 3 MTATP6P1,MTATP8P1,RP5-857K21.11::unprocessed_pseudogene::1::679 1 +
## 4 RP11-206L10.9::lincRNA::1::8880 1 +
## 5 LINC01128::processed_transcript::2::14301 1 +
## 6 LINC01128::processed_transcript::2::14301 2 +
## 7 RP11-5407.1,RP11-5407.16::lincRNA::2::13344 1 +
## 8 RP11-5407.1,RP11-5407.16::lincRNA::2::13344 2 +
## 9 SAMD11::protein_coding::2::16262 1 +
## 10 SAMD11::protein_coding::2::16262 2 +
## 11 KLHL17::protein_coding::5::3972 1 +
## 12 AL645608.1::miRNA::1::605 1 +
## 13 PLEKHN1::protein_coding::4::9393 1 +
## 14 NA.v1::intergenic::1::2878 1 +
## 15 ISG15::protein_coding::2::1062 1 +
## 16 AGRN,RP11-5407.14::protein_coding::36::36034 1 +
## 17 AGRN,RP11-5407.14::protein_coding::36::36034 2 +
## 18 AGRN,RP11-5407.14::protein_coding::36::36034 3 +
## 19 AGRN,RP11-5407.14::protein_coding::36::36034 4 +
## 20 RP11-465B22.3::transcribed_unprocessed_pseudogene::1::1921 1 +
## 21 B3GALT6::protein_coding::1::1942 1 +
## 22 RP5-902P8.10::lincRNA::1::5394 1 +
## 23 SCNN1D::protein_coding::1::11553 1 +
## 24 SCNN1D::protein_coding::1::11553 2 +
## 25 PUSL1::protein_coding::8::3107 1 +
## 26 CPTP::protein_coding::3::4103 1 +
## 27 NA.v2::NA::1::321 1 +
## 28 NA.v3::NA::1::893 1 +
## 29 NA.v4::intragenic::1::585 1 +
## 30 NA.v5::enhancer::1::4168 1 +

```

```

# count file
head(all10kbbinrRNA)

##      Sample chr  start    end
## 1 ALYREF_1_DMSO  1 629639 630560
## 2 ALYREF_1_DMSO  1 631073 631548
## 3 ALYREF_1_DMSO  1 633695 634374
## 4 ALYREF_1_DMSO  1 778792 787672
## 5 ALYREF_1_DMSO  1 827699 837699
## 6 ALYREF_1_DMSO  1 837699 842000

##                                     ID binNumber
## 1                               MTND2P28::unprocessed_pseudogene::1:::921      1
## 2                   AC114498.1,MIR6723::unprocessed_pseudogene::1:::475      1
## 3 MTATP6P1,MTATP8P1,RP5-857K21.11::unprocessed_pseudogene::1:::679      1
## 4                               RP11-206L10.9::lincRNA::1:::8880      1
## 5                               LINC01128::processed_transcript::2:::14301      1
## 6                               LINC01128::processed_transcript::2:::14301      2

## strand    count
## 1      + 96.50530
## 2      + 74.57230
## 3      + 74.57230
## 4      + 26.31960
## 5      +  0.00000
## 6      +  4.38661

#wrangle data functions
wrangle_data <- function(df) {
  df %>%
    separate(Sample, into = c("Protein", "Rep", "Timepoint")) %>%
    filter(!(Protein == "CBP20" & Rep == "3")) %>%
    separate(ID, into = c("GeneID", "bioType", "TotalExonNumber", "size"), sep = ":::") %>%
    mutate(geneType = case_when(TotalExonNumber > 1 ~ "multiExonic", TRUE ~ "monoExonic")) %>%
    select(-c(chr, start, end, TotalExonNumber, size, strand)) %>%
    mutate(Timepoint = case_when(Timepoint == "PBSDRB" ~ "t0", TRUE ~ Timepoint)) %>%
    unite(Sample, c("Protein", "Timepoint", "Rep")) %>%
    unite(ID, c("GeneID", "bioType", "binNumber"), sep = "|") %>%
    mutate(log2count = log2(count + 1)) %>%
    select(Sample, ID, geneType, log2count) %>%
    separate(ID, into = c("ID", "biotype", "binNumber"), sep = "\\|")
}

data <- wrangle_data(all10kbbinrRNA)

data2 <- data %>%
  filter(biotype != "histone_coding", biotype != "snRNA")

#Figure 1 E - PCA analysis for multi and mono exonic genes with RBM7, CBP20, CBP80 and ALYREF.
PCAPlot <- function(df_log2, GENESTRUCTURE = c("multiExonic", "monoExonic")) {
  m_all <- df_log2 %>%
    filter(geneType == GENESTRUCTURE) %>%
    unite(ID_binN, c("ID", "binNumber", "biotype", "geneType"), sep = ":::") %>%
    spread(Sample, log2count) %>%

```

```

column_to_rownames(var = "ID_binN")

##### chose rows with over 100 counts
m_all <- m_all[rowSums(m_all) > 100, ]
# apply variance to rows and remove nas
myvars <- apply(m_all, 1, var, na.rm = TRUE)
# sort by largest variance, and choose top 2000 bins with most variance
myvars <- sort(myvars, decreasing = TRUE)[1:2000]

# make df of high variance genes.
highVarGenes <- myvars %>%
  names() %>%
  as.data.frame() %>%
  separate(".", into = c("ID", "binNumber", "bioType", "geneType"), sep = ":::") %>%
  select(-binNumber) %>%
  unique() %>%
  select(ID)

###
unique_genes <- myvars %>%
  names() %>%
  as.data.frame() %>%
  separate(".", into = c("ID", "binNumber", "biotype", "geneType"), sep = ":::") %>%
  select(-binNumber) %>%
  group_by(biotype) %>%
  summarise(uniqueID = n_distinct(ID), nonUnique = n()) %>%
  mutate(total_nonUnique = sum(nonUnique), total_unique = sum(uniqueID)) %>%
  select(total_unique) %>%
  unique() %>%
  as.matrix()

# select genes based on criteria above

df_log2_highVarGenes <- df_log2 %>%
  filter(ID %in% highVarGenes$ID) %>%
  unite("ID", c("ID", "binNumber", "biotype", "geneType"), sep = "|")

# all together

df_log2_highVarGenes_T <- df_log2_highVarGenes %>%
  spread(ID, log2count)

row.names(df_log2_highVarGenes_T) <- paste(df_log2_highVarGenes_T$Sample)
df_log2_highVarGenes_T$Sample <- NULL

# preform PCA analysis
res.pca <- PCA(df_log2_highVarGenes_T, graph = FALSE)

# print scree plot to identify elbow and contribution of components.
scree <- fviz_eig(res.pca, addlabels = TRUE, main = paste(GENESTRUCTURE, " scree plot"))
print(scree)
# extract relevant information from res.pca
str(res.pca)

```

```

percentage <- as.data.frame(res.pca[[1]][1:3, 2])
percentage[1, 1]

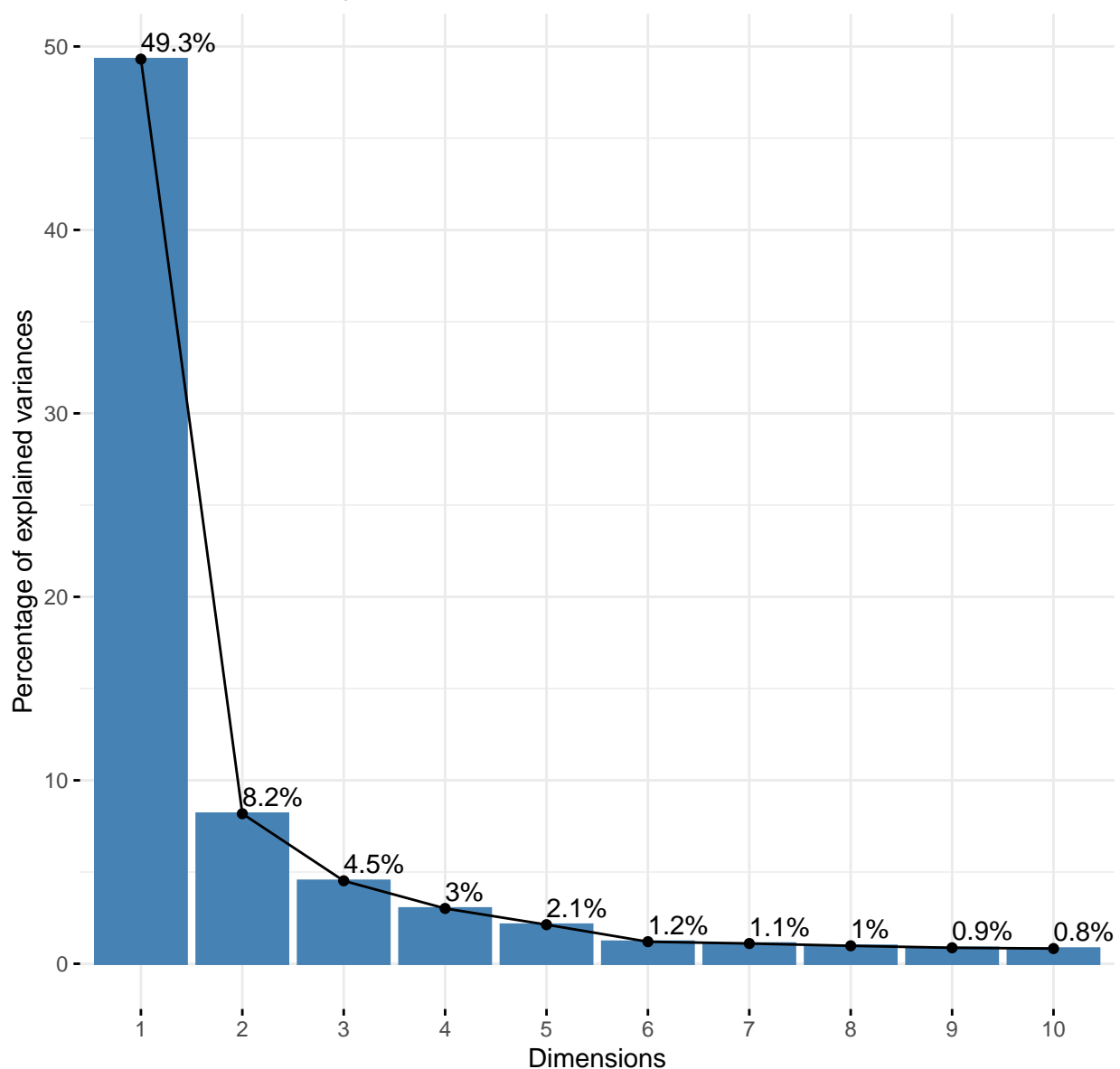
PCAfig <- as.data.frame(res.pca[[3]][1]) %>%
  add_rownames("sample") %>%
  separate(sample, into = c("Protein", "Timepoint", "Rep")) %>%
  ggplot() + geom_point(aes(x = coord.Dim.1, y = coord.Dim.2, color = Protein,
    shape = Rep)) + geom_vline(xintercept = 0, linetype = "dashed", alpha = 0.5) +
  geom_hline(yintercept = 0, linetype = "dashed", alpha = 0.5) + stat_ellipse(aes(x = coord.Dim.1,
    y = coord.Dim.2, color = Protein), type = "t") + xlab(paste0("PC1 (", round(percentage[1,
    1], 2), "%)")) + ylab(paste0("PC2 (", round(percentage[2, 1], 2), "%)")) +
  labs(subtitle = paste0("PCA plot of xiCLIP most variance , ", GENESTRUCTURE,
    ", genes (log2)"), caption = paste0("mostVariation - over 100 counts\nusing PCA()\nn =",
    unique_genes)) + theme(text = element_text(size = 8)) + coord_cartesian(xlim = c(-170,
    225), ylim = c(-90, 65))

print(PCAfig)
}

PCAPlot(data2, "multiExonic")

```

multiExonic scree plot



```
## List of 5
## $ eig : num [1:95, 1:3] 4739 786 434 290 205 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:95] "comp 1" "comp 2" "comp 3" "comp 4" ...
## .. ..$ : chr [1:3] "eigenvalue" "percentage of variance" "cumulative percentage of variance"
## $ var :List of 4
## ..$ coord : num [1:9621, 1:5] 0.433 0.891 0.194 0.715 0.579 ...
## .. .. attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:9621] "AARSD1|1|protein_coding|multiExonic" "AARSD1|2|protein_coding|multiExonic" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cor : num [1:9621, 1:5] 0.433 0.891 0.194 0.715 0.579 ...
## .. .. attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:9621] "AARSD1|1|protein_coding|multiExonic" "AARSD1|2|protein_coding|multiExonic" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cos2 : num [1:9621, 1:5] 0.1871 0.7939 0.0376 0.5111 0.3353 ...
```

```

## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:9621] "AARSD1|1|protein_coding|multiExonic" "AARSD1|2|protein_coding|multiExonic"
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ contrib: num [1:9621, 1:5] 0.003947 0.016752 0.000793 0.010784 0.007076 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:9621] "AARSD1|1|protein_coding|multiExonic" "AARSD1|2|protein_coding|multiExonic"
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## $ ind :List of 4
## ..$ coord : num [1:96, 1:5] 90.3951 108.2296 -97.4727 -111.8985 0.0814 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cos2 : num [1:96, 1:5] 6.80e-01 5.39e-01 5.24e-01 6.23e-01 7.30e-07 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ contrib: num [1:96, 1:5] 1.80 2.57 2.09 2.75 1.45e-06 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ dist : Named num [1:96] 109.6 147.4 134.7 141.8 95.2 ...
## ..- attr(*, "names")= chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## $ svd :List of 3
## ..$ vs: num [1:95] 68.8 28 20.8 17 14.3 ...
## ..$ U : num [1:96, 1:5] 1.31311 1.57218 -1.41592 -1.62547 0.00118 ...
## ..$ V : num [1:9621, 1:5] 0.00628 0.01294 0.00282 0.01038 0.00841 ...
## $ call:List of 9
## ..$ row.w : num [1:96] 0.0104 0.0104 0.0104 0.0104 0.0104 ...
## ..$ col.w : num [1:9621] 1 1 1 1 1 1 1 1 1 1 ...
## ..$ scale.unit: logi TRUE
## ..$ ncp : num 5
## ..$ centre : num [1:9621] 3.81 3.73 1.41 3.59 3.47 ...
## ..$ ecart.type: num [1:9621] 1.96 2.5 1.58 2.64 3.01 ...
## ..$ X : 'data.frame': 96 obs. of 9621 variables:
## ..$ AARSD1|1|protein_coding|multiExonic : num
## ..$ AARSD1|2|protein_coding|multiExonic : num
## ..$ ABCA2|1|protein_coding|multiExonic : num
## ..$ ABCA2|2|protein_coding|multiExonic : num
## ..$ ABCA2|3|protein_coding|multiExonic : num
## ..$ ABCA7|1|protein_coding|multiExonic : num
## ..$ ABCA7|2|protein_coding|multiExonic : num
## ..$ ABCA7|3|protein_coding|multiExonic : num
## ..$ ABCC2|1|protein_coding|multiExonic : num
## ..$ ABCC2|2|protein_coding|multiExonic : num
## ..$ ABCC2|3|protein_coding|multiExonic : num
## ..$ ABCC2|4|protein_coding|multiExonic : num
## ..$ ABCC2|5|protein_coding|multiExonic : num
## ..$ ABCC2|6|protein_coding|multiExonic : num
## ..$ ABCC2|7|protein_coding|multiExonic : num
## ..$ ABCC4|1|protein_coding|multiExonic : num
## ..$ ABCC4|10|protein_coding|multiExonic : num
## ..$ ABCC4|11|protein_coding|multiExonic : num
## ..$ ABCC4|12|protein_coding|multiExonic : num
## ..$ ABCC4|13|protein_coding|multiExonic : num

```

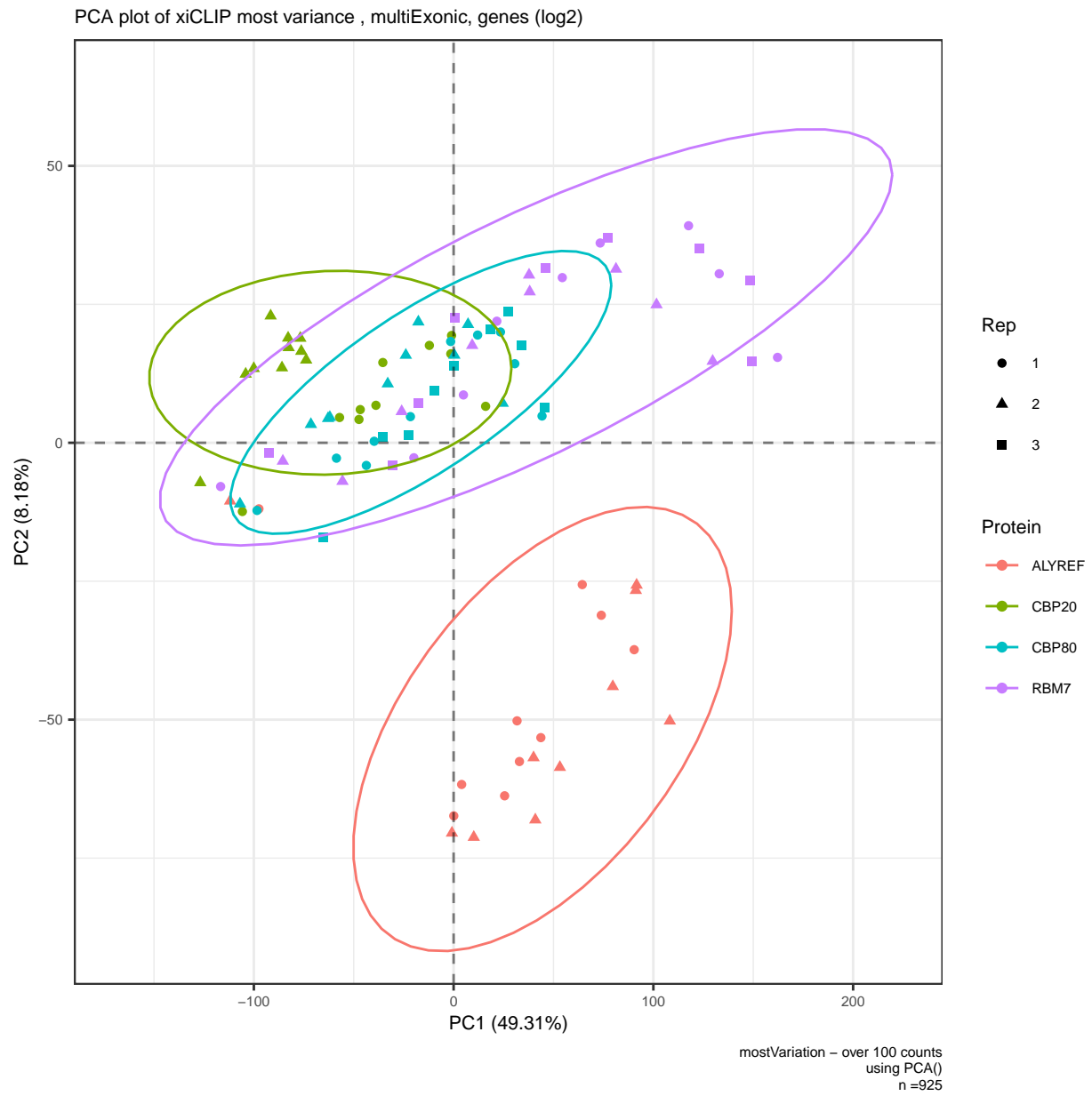
[illegible]



```

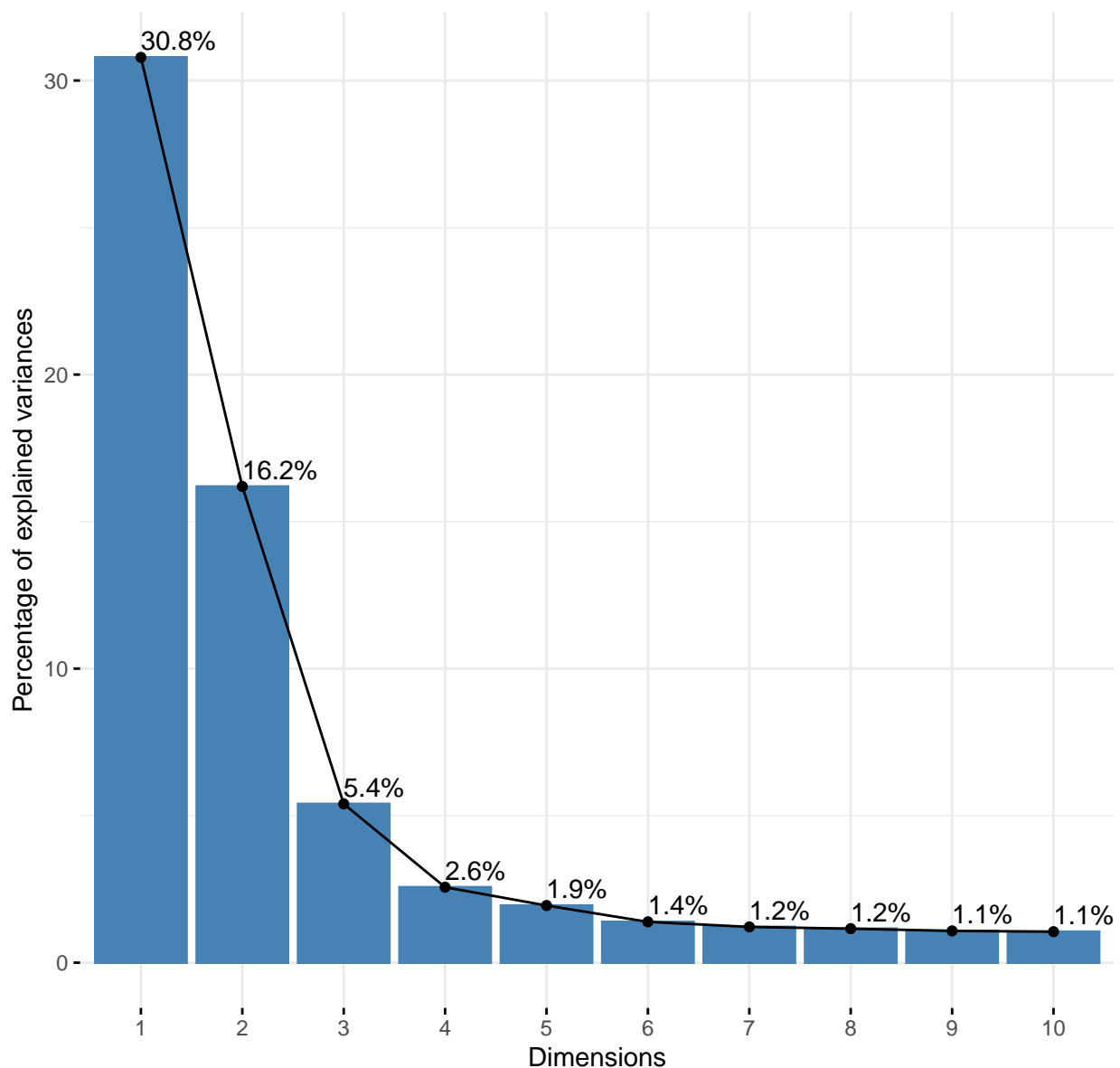
## .. ..$ ABTB2|2|protein_coding|multiExonic : num
## .. ..$ ABTB2|20|protein_coding|multiExonic : num
## .. ..$ ABTB2|21|protein_coding|multiExonic : num
## .. ..$ ABTB2|3|protein_coding|multiExonic : num
## .. ..$ ABTB2|4|protein_coding|multiExonic : num
## .. ..$ ABTB2|5|protein_coding|multiExonic : num
## .. ..$ ABTB2|6|protein_coding|multiExonic : num
## .. ..$ ABTB2|7|protein_coding|multiExonic : num
## .. ..$ ABTB2|8|protein_coding|multiExonic : num
## .. ..$ ABTB2|9|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|1|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|10|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|11|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|12|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|13|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|14|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|15|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|16|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|17|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|18|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|19|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|2|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|20|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|21|protein_coding|multiExonic : num
## .. ..$ AC004943.1,ZFHX3|22|protein_coding|multiExonic : num
## .. .. [list output truncated]
## ..$ row.w.init: num [1:96] 1 1 1 1 1 1 1 1 1 1 ...
## ..$ call : language PCA(X = df_log2_highVarGenes_T, graph = FALSE)
## - attr(*, "class")= chr [1:2] "PCA" "list"

```



```
PCAPlot(data2, "monoExonic")
```

monoExonic scree plot



```
## List of 5
## $ eig : num [1:95, 1:3] 819.2 430.9 143.7 68.3 51.7 ...
##   ..- attr(*, "dimnames")=List of 2
##     ..$. : chr [1:95] "comp 1" "comp 2" "comp 3" "comp 4" ...
##     ..$. : chr [1:3] "eigenvalue" "percentage of variance" "cumulative percentage of variance"
## $ var :List of 4
##   ..$ coord : num [1:2663, 1:5] 0.196 0.799 0.455 0.549 0.38 ...
##   ..- attr(*, "dimnames")=List of 2
##     ..$. : chr [1:2663] "ABBA01017803.1|1|pseudogene|monoExonic" "AC002117.1|1|antisense|monoExonic" ...
##     ..$. : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
##   ..$ cor : num [1:2663, 1:5] 0.196 0.799 0.455 0.549 0.38 ...
##   ..- attr(*, "dimnames")=List of 2
##     ..$. : chr [1:2663] "ABBA01017803.1|1|pseudogene|monoExonic" "AC002117.1|1|antisense|monoExonic" ...
##     ..$. : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
##   ..$ cos2 : num [1:2663, 1:5] 0.0384 0.6377 0.207 0.3013 0.1445 ...
```

```

## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:2663] "ABBA01017803.1|1|pseudogene|monoExonic" "AC002117.1|1|antisense|monoExonic"
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ contrib: num [1:2663, 1:5] 0.00468 0.07784 0.02527 0.03678 0.01764 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:2663] "ABBA01017803.1|1|pseudogene|monoExonic" "AC002117.1|1|antisense|monoExonic"
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## $ ind :List of 4
## ..$ coord : num [1:96, 1:5] -14.9 -19.7 -64.8 -66.3 -29.1 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cos2 : num [1:96, 1:5] 0.0799 0.0829 0.7967 0.8133 0.2917 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ contrib: num [1:96, 1:5] 0.28 0.492 5.331 5.584 1.079 ...
## ..- attr(*, "dimnames")=List of 2
## ..$ : chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ dist : Named num [1:96] 52.5 68.3 72.5 73.5 53.9 ...
## ..- attr(*, "names")= chr [1:96] "ALYREF_DMSO_1" "ALYREF_DMSO_2" "ALYREF_negative_1" "ALYREF_negative_2" ...
## $ svd :List of 3
## ..$ vs: num [1:95] 28.62 20.76 11.99 8.26 7.19 ...
## ..$ U : num [1:96, 1:5] -0.519 -0.687 -2.262 -2.315 -1.018 ...
## ..$ V : num [1:2663, 1:5] 0.00684 0.0279 0.0159 0.01918 0.01328 ...
## $ call:List of 9
## ..$ row.w : num [1:96] 0.0104 0.0104 0.0104 0.0104 0.0104 ...
## ..$ col.w : num [1:2663] 1 1 1 1 1 1 1 1 1 1 ...
## ..$ scale.unit: logi TRUE
## ..$ ncp : num 5
## ..$ centre : num [1:2663] 2.66 4.43 1.78 2.25 2.55 ...
## ..$ ecart.type: num [1:2663] 1.98 2.21 1.87 1.96 2.49 ...
## ..$ X : 'data.frame': 96 obs. of 2663 variables:
## ..$ ABBA01017803.1|1|pseudogene|monoExonic : num [1:96] 5.1
## ..$ AC002117.1|1|antisense|monoExonic : num [1:96] 3.8
## ..$ AC002117.1|2|antisense|monoExonic : num [1:96] 0 0
## ..$ AC002310.12|1|antisense|monoExonic : num [1:96] 0 0
## ..$ AC003101.1,MIR365B|1|lincRNA|monoExonic : num [1:96] 4.9
## ..$ AC004453.8|1|processed_pseudogene|monoExonic : num [1:96] 3.8
## ..$ AC005037.3|1|antisense|monoExonic : num [1:96] 0 0
## ..$ AC005363.11|1|lincRNA|monoExonic : num [1:96] 2.4
## ..$ AC006273.4|1|lincRNA|monoExonic : num [1:96] 3.2
## ..$ AC006273.5|1|lincRNA|monoExonic : num [1:96] 6.5
## ..$ AC006486.10|1|TEC|monoExonic : num [1:96] 0 0
## ..$ AC007318.5|1|processed_pseudogene|monoExonic : num [1:96] 4.7
## ..$ AC007773.2|1|antisense|monoExonic : num [1:96] 3.2
## ..$ AC007773.2|2|antisense|monoExonic : num [1:96] 2.4
## ..$ AC007792.1|1|TEC|monoExonic : num [1:96] 0 0
## ..$ AC007842.1|1|miRNA|monoExonic : num [1:96] 3.8
## ..$ AC007842.1|2|miRNA|monoExonic : num [1:96] 2.4
## ..$ AC007969.5|1|processed_pseudogene|monoExonic : num [1:96] 4.7
## ..$ AC009133.12|1|antisense|monoExonic : num [1:96] 0 0
## ..$ AC009133.12|2|antisense|monoExonic : num [1:96] 0 0

```

##	.. ..\$	AC009245.3 1 processed_pseudogene monoExonic	:	num	[1:96]	4.7
##	.. ..\$	AC012640.1 1 miRNA monoExonic	:	num	[1:96]	4.7
##	.. ..\$	AC016700.5 1 lincRNA monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AC021087.1 1 miRNA monoExonic	:	num	[1:96]	4.2
##	.. ..\$	AC023347.1 1 lincRNA monoExonic	:	num	[1:96]	3.8
##	.. ..\$	AC023347.1 2 lincRNA monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AC023347.1 3 lincRNA monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AC025171.1 1 antisense monoExonic	:	num	[1:96]	5.6
##	.. ..\$	AC027763.2,RP11-589P10.5 1 antisense monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AC027763.2,RP11-589P10.5 2 antisense monoExonic	:	num	[1:96]	4.2
##	.. ..\$	AC079250.1 1 processed_pseudogene monoExonic	:	num	[1:96]	3.2
##	.. ..\$	AC079780.3 1 processed_pseudogene monoExonic	:	num	[1:96]	4.2
##	.. ..\$	AC091729.9 1 antisense monoExonic	:	num	[1:96]	3.8
##	.. ..\$	AC093627.9 1 lincRNA monoExonic	:	num	[1:96]	4.5
##	.. ..\$	AC097724.3 1 antisense monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AC099850.1 1 antisense monoExonic	:	num	[1:96]	4.7
##	.. ..\$	AC099850.1 2 antisense monoExonic	:	num	[1:96]	0.4
##	.. ..\$	AC114271.2 1 antisense monoExonic	:	num	[1:96]	3.2
##	.. ..\$	AC142472.6 1 lincRNA monoExonic	:	num	[1:96]	3.2
##	.. ..\$	ADAMTSL4-AS1 1 processed_transcript monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AGAP1-IT1 1 sense_intronic monoExonic	:	num	[1:96]	5.3
##	.. ..\$	AGAP1-IT1 2 sense_intronic monoExonic	:	num	[1:96]	4.2
##	.. ..\$	AHCTF1P1 1 processed_pseudogene monoExonic	:	num	[1:96]	8.9
##	.. ..\$	AL121932.1 1 miRNA monoExonic	:	num	[1:96]	6.3
##	.. ..\$	AL139099.1 1 miRNA monoExonic	:	num	[1:96]	4.7
##	.. ..\$	ANKRD20A18P,AP001347.6 1 unprocessed_pseudogene monoExonic	:	num	[1:96]	0.6
##	.. ..\$	ANKRD20A18P,AP001347.6 2 unprocessed_pseudogene monoExonic	:	num	[1:96]	3.2
##	.. ..\$	ANKRD20A18P,AP001347.6 3 unprocessed_pseudogene monoExonic	:	num	[1:96]	2.4
##	.. ..\$	ANKRD20A18P,AP001347.6 4 unprocessed_pseudogene monoExonic	:	num	[1:96]	3.8
##	.. ..\$	ANKRD20A18P,AP001347.6 5 unprocessed_pseudogene monoExonic	:	num	[1:96]	0.4
##	.. ..\$	ANKRD20A18P,AP001347.6 6 unprocessed_pseudogene monoExonic	:	num	[1:96]	3.2
##	.. ..\$	AP000442.4 1 lincRNA monoExonic	:	num	[1:96]	5.7
##	.. ..\$	AP000487.5 1 antisense monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AP000640.2 1 processed_transcript monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AP000662.4 1 sense_overlapping monoExonic	:	num	[1:96]	2.4
##	.. ..\$	AP000692.10 1 antisense monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AP000783.1 1 transcribed_unprocessed_pseudogene monoExonic	:	num	[1:96]	0.6
##	.. ..\$	AP001187.11 1 processed_pseudogene monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AP001412.1 1 antisense monoExonic	:	num	[1:96]	0.0
##	.. ..\$	AP003900.6 1 lincRNA monoExonic	:	num	[1:96]	5.7
##	.. ..\$	AP003900.6 2 lincRNA monoExonic	:	num	[1:96]	7.7
##	.. ..\$	APOLD1 1 protein_coding monoExonic	:	num	[1:96]	4.7
##	.. ..\$	APOLD1 2 protein_coding monoExonic	:	num	[1:96]	2.4
##	.. ..\$	APOLD1 3 protein_coding monoExonic	:	num	[1:96]	0.0
##	.. ..\$	APOLD1 4 protein_coding monoExonic	:	num	[1:96]	0.0
##	.. ..\$	APOLD1 5 protein_coding monoExonic	:	num	[1:96]	2.4
##	.. ..\$	APOLD1 6 protein_coding monoExonic	:	num	[1:96]	2.4
##	.. ..\$	APOLD1 7 protein_coding monoExonic	:	num	[1:96]	2.4
##	.. ..\$	ARHGAP5-AS1 1 antisense monoExonic	:	num	[1:96]	3.2
##	.. ..\$	ARHGEF9 1 protein_coding monoExonic	:	num	[1:96]	2.4
##	.. ..\$	ASB16-AS1 1 antisense monoExonic	:	num	[1:96]	2.4
##	.. ..\$	ASB16-AS1 2 antisense monoExonic	:	num	[1:96]	6.2
##	.. ..\$	ASH1L-AS1 1 antisense monoExonic	:	num	[1:96]	4.5
##	.. ..\$	ATP5EP2 1 protein_coding monoExonic	:	num	[1:96]	2.4

```

## .. ..$ ATXN7L3B|1|protein_coding|monoExonic : num [1:96] 6.7
## .. ..$ BCL6|1|protein_coding|monoExonic : num [1:96] 3.2
## .. ..$ BCL7C|1|protein_coding|monoExonic : num [1:96] 3.2
## .. ..$ BCL7C|2|protein_coding|monoExonic : num [1:96] 2.4
## .. ..$ BCL7C|3|protein_coding|monoExonic : num [1:96] 2.4
## .. ..$ BLOC1S4|1|protein_coding|monoExonic : num [1:96] 4.7
## .. ..$ BTBD19|1|protein_coding|monoExonic : num [1:96] 3.2
## .. ..$ C11orf48.v2|1|protein_coding|monoExonic : num [1:96] 2.4
## .. ..$ C11orf65|1|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|2|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|3|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|4|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|5|protein_coding|monoExonic : num [1:96] 3.2
## .. ..$ C11orf65|6|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|7|protein_coding|monoExonic : num [1:96] 3.2
## .. ..$ C11orf65|8|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf65|9|protein_coding|monoExonic : num [1:96] 0 0
## .. ..$ C11orf71|1|protein_coding|monoExonic : num [1:96] 0 4
## .. ..$ C14orf169|1|protein_coding|monoExonic : num [1:96] 4.9
## .. ..$ C15orf37|1|antisense|monoExonic : num [1:96] 0 0
## .. ..$ C19orf83|1|protein_coding|monoExonic : num [1:96] 4.5
## .. ..$ C1orf145|1|protein_coding|monoExonic : num [1:96] 4.2
## .. ..$ C1orf213|1|antisense|monoExonic : num [1:96] 0 5
## .. ..$ C21orf119|1|lincRNA|monoExonic : num [1:96] 0 0
## .. ..$ C5orf34|1|protein_coding|monoExonic : num [1:96] 5.1
## .. .. [list output truncated]
## ..$ row.w.init: num [1:96] 1 1 1 1 1 1 1 1 1 1 ...
## ..$ call : language PCA(X = df_log2_highVarGenes_T, graph = FALSE)
## - attr(*, "class")= chr [1:2] "PCA" "list"

```

