

Analysis for Ferre et al. (2019)

Claudio Ferre

2019

## Contents

Setup . . . . .	1
<b>Differential Expression Analysis</b>	<b>1</b>
PCA . . . . .	3
PCA Regression . . . . .	4
Model Comparison . . . . .	4

## Setup

## Differential Expression Analysis

```
#read counts from file
d <- DGEList(counts=seg1)
d

#normalize counts (if needed)
d <- calcNormFactors(d)
d

#add groups factor to dataframe
d <- DGEList(counts=seg1,group=factor(seg1Groups))
d

#total counts per sample (i.e. total volume size per individual)
apply(d$counts, 2, sum)
```

```

#normalize data (needed?)
d <- calcNormFactors(d)
d

#Multidimensional Scaling Plot
plotMDS(d, method="bcv", col=as.numeric(d$samples$group), pch=16)
legend("bottomleft", as.character(unique(d$samples$group)), col=1:3, pch=20)

#estimate dispersion
d1 <- estimateCommonDisp(d, verbose=T)

names(d1)

#empircial Bayes
d1 <- estimateTagwiseDisp(d1)
names(d1)

#plotBCV- plots biological coefficient of variation (square root of dispersions) against log2-CPM.
plotBCV(d1)

#glm estimates of dispersion (with TD as reference group)
design.mat <- model.matrix(~d$samples$group)
colnames(design.mat) <- levels(d$samples$group)
d2 <- estimateGLMCommonDisp(d,design.mat)
d2 <- estimateGLMTrendedDisp(d2,design.mat, method="power")
d2 <- estimateGLMTagwiseDisp(d2,design.mat)
plotBCV(d2)

#testing for differentially expressed ROIs using generalized linear model likelihood ratio test

#create comparisons for GLM
fit <- glmFit(d2, design.mat)

#lrt12 - compares MCA group versus TD group (reference)
#lrt13 - compares PV group versus TD group (reference)
#lrt14 - compares MCA group versus PV group

lrt12 <- glmLRT(fit, coef=2)
lrt13 <- glmLRT(fit, coef=3)
lrt14 <- glmLRT(fit, contrast=c(0, 1,-1))

#Determine number of differentially expressed ROIs at FDR<.05
de2 <- decideTestsDGE(lrt12, adjust.method="BH", p.value = 0.05)
summary(de2)

de3 <- decideTestsDGE(lrt13, adjust.method="BH", p.value = 0.05)
summary(de3)

de4 <- decideTestsDGE(lrt14, adjust.method="BH", p.value = 0.05)
summary(de4)

#print the top differeces in ROI volume
DE<-topTags(lrt14, n=50, sort.by="PValue")

```

```

#write results of decide tests to csv
write.table(DE, file="decide4.csv", row.names=FALSE)

#plot of all differentially expressed ROIs for lrt14
plotMD(lrt14, values=c(-1,1), ylim=c(-2,1), hl.col=c("blue4","yellow3"), legend=FALSE, main=NULL, xlab=

```

## PCA

```

#load person-level data
segt2 <- read.table("/Users/Claudio/Desktop/R Data/seg_mnit_v9.csv", header=TRUE, sep=",",)

#create dataframe of DE ROIs
d <- segt2[, c(7:95)]

# PCA on dataframe containing DE ROIs
prin_comp <- prcomp(d, scores=TRUE, cor=TRUE, scale. = T)

summary(prin_comp, loadings=TRUE)

#output the mean of the variables
prin_comp$center
#output the sd of the variables
prin_comp$scale

#output the PC loading
prin_comp$rotation

#plot PC's
biplot(prin_comp, scale=0)

#compute SD of each PC
std_dev <- prin_comp$sdev

#compute variance (i.e. eigenvalues)
pr_var <- std_dev^2

#Check variance of components
pr_var

#proportion of variance explained
prop_varex <- pr_var/sum(pr_var)
prop_varex

#scree plot
plot(prop_varex, xlab = "Principal Component", ylab = "Proportion of Variance Explained", type = "b")

#cumulative scree plot
plot(cumsum(prop_varex), xlab = "Principal Component", ylab = "Cumulative Proportion of Variance Explained")

```

## PCA Regression

```
lm2<-lm(jeb~prin_comp$x[,1]+prin_comp$x[,2]+ prin_comp$x[, 3])
summary(lm2)

lm3<-lm(aha~prin_comp$x[,1]+prin_comp$x[,2]+ prin_comp$x[, 3])
summary(lm3)

lm4<-lm(bb~prin_comp$x[,1]+prin_comp$x[,2]+ prin_comp$x[, 3])
summary(lm4)
```

## Model Comparison

```
lm2b<-lm(jeb~les_type)
str(summary(lm2b))

AIC(lm2)
BIC(lm2)

AIC(lm2b)
BIC(lm2b)

anova(lm2,lm2b)

lm3b<-lm(aha~les_type)
str(summary(lm3b))

AIC(lm3)
BIC(lm3)

AIC(lm3b)
BIC(lm3b)

anova(lm3,lm3b)

lm4b<-lm(bb~les_type)
str(summary(lm4b))

AIC(lm4)
BIC(lm4)

AIC(lm4b)
BIC(lm4b)

anova(lm4,lm4b)
```