I519, Introduction to bioinformatics

R in Bioinformatics

Yuzhen Ye (yye@indiana.edu)
School of Informatics and Computing, IUB

Example 1: data matrix

```
gene1 <- c(1.00, 1.50, 1.25)
gene2 <- c(1.35, 1.55, 1.30)
gene3 <- c(-1.10, -1.50, -1.25)
gene4 <- c(-1.20, -1.30, -1.00)
rowcolnames <- list(c("g1","g2","g3","g4"), c("1h", "2h", "3h"))
genedat <- matrix(c(gene1,gene2,gene3,gene4), nrow=4, ncol=3,
   byrow=TRUE, dimnames = rowcolnames)

datafile = "/Users/yye/courses/Rexamples/genedat.Rdata"
#save matrix to a file
write.table(genedat, file=datafile)
#can also load from a file
#genedat2 <- read.table(datafile)

apply(genedat,2,mean) #specify matrix, rows (1) or columns (2),
   function
```

Example 2: plotting (golub dataset)

```
#gene expression data collected by Golub et al. Science, Vol.
   286:531-537. 1999
data(golub, package = "multtest")
dim(golub)
golub.gnames
golub.cl
gol.fac <- factor(golub.cl,levels=0:1, labels= c("ALL","AML"))
golub.gnames[1042,]
hist(golub[1042, gol.fac=="ALL"]
boxplot(golub[1042,] ~ gol.fac)
qqnorm(golub[1042, gol.fac=="ALL"])
qqline(golub[1042, gol.fac=="ALL"])
#boxplot shows ALL of CCND3 Cyclin D3 are positive
#one sample t-Test
t.test(golub[1042,gol.fac=="ALL"], mu=0, alternative = c
   ("greater"))
#two sample t-Test
t.test(golub[1042,] ~ gol.fac, var.equal=FALSE)
```

Example 3: clustering (golub dataset)

```
data(golub, package="multtest")
clusdata <- data.frame(golub[1042,],golub[2124,])
colnames(clusdata)<-c("CCND3 Cyclin D3","Zyxin")
gol.fac <- factor(golub.cl,levels=0:1, labels= c("ALL","AML"))
par(mfrow=c(1,2))
plot(clusdata, pch=as.numeric(gol.fac))
legend("topright",legend=c("ALL","AML"),pch=1:2)
plot(hclust(dist(clusdata,method="euclidian"),method="single"))
```

Example 4: ROC (golub dataset)

```
library(ROCR)
gol.true <- factor(golub.cl,levels=0:1,labels= c
   ("TRUE","FALSE"))
par(mfrow=c(1,2))
pred <- prediction(golub[1042,], gol.true)
perf <- performance(pred, "tpr", "fpr" )
plot(perf)
pred <- prediction(golub[2000,], gol.true)
perf <- performance(pred, "tpr", "fpr" )
plot(perf)
```

Example 5: working with genomic sequences

```
install.packages(c("ape"),repo="http://cran.r-project.org",dep=TRUE)
library(ape)
table(read.GenBank(c("X94991.1"), as.character=TRUE))
pie(table(read.GenBank(c("X94991.1"), as.character=TRUE)))
```

References

- "Applied Statistics for Bioinformatics using R", by Wim P. Krijnen
- Online documents and tutorials on R website, http://www.r-project.org