

BMI882 Homework 1

Reproduce Figure 1b

Grace Yi Chen

07 March, 2023

Reproduce Figure 1b from Baggerly and Coombes (2009)

Steps to reproduce the plot:

1. For the supplementary materials, I went to the link: <https://bioinformatics.mdanderson.org/Supplements/ReproRsch-All/> to download “Everything”.
2. I unzip the “Everything” folder and create this file within the everything folder.
3. To find the code for this plot, I first went to the “Figures” folder to find the plot. After searching the plot title “Identical Columns in Adria_All.txt”, I find file “doxoSummary.pdf” contains the plot.
4. Then I went to the “RNowebSource” folder to find the code producing “doxoSummary.pdf”. I found that file “examining-DoxorubicinInDetail.Rnw” contains the code producing the plot. So I just copy the relevant code here to reproduce the plot.

Challenge:

1. There are two figure folders, one is the “Figures” and the other is “PaperFigures”. It seems the “PaperFigures” folder contains plots ready for publication. But it is confusing as plots are duplicated and I don’t know where is the code for figures in “PaperFigures” folder
2. There is another R file called “examiningDoxorubicinInDetail.R” in “RTangle” folder which also contains the code for reproducing the plot in “chunk number 24: plotDoxoHighCors”. However, we could not directly run this R code as we need to set the directory back to “everything” folder to produce the plot.
3. It also took me some time to extract the relevant code producing the plot as there are a lot of code lines before the plot code and I will need to check which part is needed.

Graph

Dotplot identifying identical columns in the full Adria_ALL.txt matrix. There are no unwanted ties in the training data (lower left corner, separated by solid lines). In the test data (upper right), however, there are tied pairs labeled both consistently (circles) and inconsistently (triangles). One column from the right-hand block from (A) is marked with a dashed line; sample 32 (main diagonal) is labeled “Resp”, but samples 66, 89, and 117 are labeled “NR” (triangles).

```
## Read Data
tempDoxorubicin07Header1 <-
  read.table(file.path("RawData", "PottiNatMed", "Adria_ALL.txt"), sep="\t",
             nrows=1, header=FALSE)
tempDoxorubicin07Header1 <- as.vector(t(tempDoxorubicin07Header1));

tempDoxorubicin07Header2 <-
  read.table(file.path("RawData", "PottiNatMed", "Adria_ALL.txt"), sep="\t",
             skip=1, nrows=1, header=FALSE)
tempDoxorubicin07Header2 <- as.vector(t(tempDoxorubicin07Header2));

## create plot dataset
tempSampleNames <-
  c(paste("Training", c(1:22), sep=""),
    paste("Test", c(1:122), sep=""))

tempGroup <- c(rep("Training", 22), rep("Test", 122))
tempStatus <- tempDoxorubicin07Header2
tempStatus[tempStatus == "Sens"] <- "Sensitive"
tempStatus[tempStatus == "NR"] <- "Resistant"
tempStatus[tempStatus == "Resp"] <- "Sensitive"

doxorubicin07Info <-
  data.frame(sampleGroup = tempGroup,
             status      = tempStatus,
             row.names   = tempSampleNames)

rm(list=ls(pattern="^temp"))

doxorubicin07Numbers <-
  read.table(file.path("RawData", "PottiNatMed", "Adria_ALL.txt"), sep="\t",
             skip=2, header=FALSE)

colnames(doxorubicin07Numbers) <- rownames(doxorubicin07Info)

doxorubicin07Cors <- cor(doxorubicin07Numbers)

doxorubicin07HighCors <- (doxorubicin07Cors > 0.9999)

same07Status <- matrix(rep(doxorubicin07Info[, "status"], 144), 144, 144)
same07Status <- (same07Status == t(same07Status))

temp07Ties <- which(doxorubicin07HighCors & same07Status, arr.ind=TRUE)
temp07Ties <- temp07Ties[temp07Ties[, 1] != temp07Ties[, 2], ]
temp07BadTies <- which(doxorubicin07HighCors & (!same07Status), arr.ind=TRUE)

## plot
plot(1:144, 1:144, pch=".", cex=3,
     xlim=c(0.5, 144.5), ylim=c(0.5, 144.5),
     xaxs="i", yaxs="i",
     xlab="Column in Adria_ALL.txt",
     ylab="Column in Adria_ALL.txt",
```

```

    main="Identical Columns in Adria_ALL.txt",
  )
points(temp07Ties[,1],temp07Ties[,2],pch=21,bg="white",lwd=1)
points(temp07BadTies[,1],temp07BadTies[,2],pch=24,bg="red",lwd=2)
abline(h=22.5, v=22.5)

```

Identical Columns in Adria_ALL.txt

