LabBook_06_05_2016

Claire Green

4 May 2016

Tuesday

I spent the morning reading about eQTLs, particularly how they are used in research pertaining to disease. I also read the FANTOM 5 paper.

Wednesday

I worked on creating an illustrative example of using Genome Translation Commons. The keynote file can be found in the GTX Example folder. David and I also spotted a lot of bugs I also began looking at Win's suggestion about the bidirectional promotor of C9orf72. I used the ZENBU platform to identify the non-coding RNA it's linked to and wrote down some notes on how you would investigate its role in C9orf72's activity. Obviously, as C9orf72 mutations cause TDP-43 pathology, this is interesting as the expansion mutation is in the promotor region. This means the mutation would also affect this ncRNA.

Thursday

I spent the morning talking with Wenbin about PCxN improvements. I undertand at the moment that its functionality is most important as the reviewers will not care about usability, but I do believe that at some point these issues must be addressed. This is likely something the junior programmer would do.

In the afternoon, I worked on looking at my correlation problem. This is the code I have so far:

```
setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/GeneExpressionAnalysis/Microarray/TopG
C9.GeneExpression <- read.csv(file = "C9rankeduniqueresult.csv")

ENS <- C9.GeneExpression[,15]

C9.disease <- C9.GeneExpression[,52:59]
C9.disease <- cbind(ENS, C9.disease)

rownames(C9.disease) <- C9.disease[,1]
C9.disease[,1] <- NULL
tC9.disease <- t(C9.disease)

C9.cor <- cor(tC9.disease, y= NULL, use = "pairwise", method = "pearson", adjust = "fdr")

DEG <- C9.cor[c("ACTN1", "ANXA1", "BBIP1", "BGN", "BPTF", "CDH11", "CREG1", "CSRP1", "CST3", "DCN", "GB
                "KCTD12", "KPNA6", "MPHOSPH9", "MXI1", "NDUFS5 /// RPL10", "NKTR", "NUTF2 /// NUTF2
                "PFDN1", "PLEKHB1", "PLOD2", "POGZ", "PPP1R7", "PPP2CA", "PPP2CB", "PRPF3", "PTPN13
                "SCN1B", "SERBP1", "SF3B1", "SPARC", "STMN1", "TARDBP", "TCF4", "TUBB4B", "TUG1", "
DEG <- t(DEG)

DEG.2 <- DEG[c("ACTN1", "ANXA1", "BBIP1", "BGN", "BPTF", "CDH11", "CREG1", "CSRP1", "CST3", "DCN", "GBA
```

```
               "KCTD12", "KPNA6", "MPHOSPH9", "MXI1", "NDUFS5 /// RPL10", "NKTR", "NUTF2 /// NUTF2P4",
               "PFDN1", "PLEKHB1", "PLOD2", "POGZ", "PPP1R7", "PPP2CA", "PPP2CB", "PRPF3", "PTPN13", "RA
               "SCN1B", "SERBP1", "SF3B1", "SPARC", "STMN1", "TARDBP", "TCF4", "TUBB4B", "TUG1", "VPS13
DEG.2 <- as.data.frame(as.table(DEG.2))
DEG.2 <- subset(DEG.2, subset=(Freq !="1"))
DEG.2 <- DEG.2[!duplicated(DEG.2[,3]),] #check the remaining number is as expected

setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/CorrelationNetwo
write.csv(DEG.2, "SeedGenes.csv")
```

My problem is that the correlation values are okay, but there's too much information. I also need a p value, but I'm not sure how to generate one. I know that corr.test in the psych function exists, but I know Yered doesn't use that so I'm not sure what to do.

Friday

At lab meeting, Win said I should speak to Gabriel about the correlation stuff. I tried the corr.test but it has been running for hours and nothing has happened. There must be a better way...