LabBook_21_10_2016

Claire Green

Tuesday

I need to do some catching up as a lot of September/October has been theoretical and I haven't got to grips with implementing the investigations I planned way back in August. In the previous analysis I did an experiment to look at the similarities in the expression patterns of my DEGs in the different datasets. Although I had hoped mutation backgrounds would cluster, instead there appears to be a platform effect that is stronger. To confirm this finding, I conducted an experiment to include the controls - the aim to be that controls would cluster together and disease separately. The results were not entirely clear, but it does appear there is still a platform effect.

This is the R script I used:

```
#Selecting DEGS from expression matrix

#Load list of interesting genes
#setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/GeneExpressionAnalysis/Microarray/")
setwd(dir = "/Users/clairegreen/Desktop/")
Genelist <- read.csv("overlap_ens2hgnc_4RNAseq.csv", header = TRUE)

#load dataset
setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/DEG_Test2/")
exprs <- read.csv("C9rankeduniqueresult.csv")

#Make gene symbol row names
#rownames(exprs) <- exprs$Ensembl
exprspat <- exprs[,49:51]
exprspat[,(length(exprspat)+1)] <- exprs$Ensembl

#Make gene symbol a column
# exprspat <- cbind(exprspat, exprs$Gene.Symbol)
# colnames(exprspat)[length(exprspat)] <- "Gene.Symbol"

#Merge by interesting gene names with expression to form matrix
patgene <- merge(Genelist, exprspat, by.x = "ensembl_gene_id", by.y = "V4")
#patgene <- patgene[!duplicated(patgene[,11]),]
# rownames(patgene) <- patgene$V1
# patgene[,1] <- NULL

setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/Pathways_to_TDP-
write.csv(patgene, file = "VCP_DEG_CON_ens.csv")

#### Cor.test Method ####

library(tictoc)
library(gdata)

##load dataset
setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/Pathways_to_TDP-
```

```
Exprs_val <- read.csv("C9r_DEG_CON_Exprs.csv")
rownames(Exprs_val) <- Exprs_val[,2]
Exprs_val[,1:2] <- NULL

CorExprMat <- t(Exprs_val)

reg <- matrix(0, ncol(CorExprMat), ncol(CorExprMat))

tic()
for (i in 1:ncol(CorExprMat)){
  for (j in 1:ncol(CorExprMat)){
    reg[i,j] <- cor.test(CorExprMat[,i], CorExprMat[,j], method = "spearman")$estimate
  }}

rownames(reg) <- colnames(reg) <- colnames(CorExprMat)
toc()


#Extract R values
corRadj <- reg
corRadj[lower.tri(corRadj, diag = TRUE)] <- NA

#Turn into vector
corRadj <- as.matrix(corRadj)
corRvec <- unmatrix(corRadj)
#Remove NA values
corRvec <- na.omit(corRvec)
corRvec <- as.data.frame(corRvec)

write.csv(corRvec, file = "C9r_CON_CorResults.csv")


#Generate matrix with all Rho values

C9mR <- read.csv("C9m_CorResults.csv")
CHmR <- read.csv("CHMP2B_CorResults.csv")
GRNR <- read.csv("GRN_CorResults.csv")
VCPR <- read.csv("VCP_CorResults.csv")
C9rR <- read.csv("C9r_CorResults.csv")

C9mCON <- read.csv("C9m_CON_CorResults.csv")
C9rCON <- read.csv("C9r_CON_CorResults.csv")
CHCON <- read.csv("CH_CON_CorResults.csv")
GRNCON <- read.csv("FTLD_CON_CorResults.csv")
VCPCON <- read.csv("VCP_CON_CorResults.csv")



RhoValues <- merge(C9mR, CHmR, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, GRNR, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, VCPR, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, C9rR, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, C9mCON, by.x = "X", by.y = "X")
```

```
RhoValues <- merge(RhoValues, C9rCON, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, CHCON, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, GRNCON, by.x = "X", by.y = "X")
RhoValues <- merge(RhoValues, VCPCON, by.x = "X", by.y = "X")

colnames(RhoValues) <- c("GenePair","C9m", "CHMP2B", "GRN", "VCP", "C9r", "C9mCON", "C9rCON",
                         "CHCON", "GRNCON", "VCPCON")

rownames(RhoValues) <- RhoValues$GenePair
RhoValues[,1] <- NULL

Rho <- matrix(0, ncol(RhoValues), ncol(RhoValues))

tic()
for (i in 1:ncol(RhoValues)){
  for (j in 1:ncol(RhoValues)){
    Rho[i,j] <- cor.test(RhoValues[,i], RhoValues[,j], method = "kendall")$p.value
  }}

rownames(Rho) <- colnames(Rho) <- colnames(RhoValues)
toc()


#Extract R values
corRadj <- Rho
corRadj[lower.tri(corRadj, diag = TRUE)] <- NA

#Turn into vector
corRadj <- as.matrix(corRadj)
corRvec <- unmatrix(corRadj)
#Remove NA values
corRvec <- na.omit(corRvec)
corRvec <- as.data.frame(corRvec)

write.csv(corRvec, file = "Alldatasets_Pval.csv")
```

Results table can be found in the file “Alldatasets_Correlation.csv”:

Dataset1	Dataset2	corRvec	pval	Relationship
GRN	GRNCON	0.230625858	0	Correlation
C9r	C9rCON	0.194266757	7.36E-277	Correlation
CHCON	GRNCON	0.185907259	3.53E-256	Correlation
GRN	CHCON	0.168137597	1.97E-206	Correlation
C9m	CHCON	0.164071069	3.45E-202	Correlation
C9rCON	GRNCON	0.115052331	7.22E-100	Correlation
C9r	CHCON	0.110565899	1.38E-90	Correlation
C9m	C9mCON	0.116612111	1.54E-85	Correlation
C9mCON	CHCON	0.116942577	6.05E-85	Correlation
C9r	GRNCON	0.106495231	2.53E-84	Correlation
CHMP2B	CHCON	0.113398182	1.41E-78	Correlation
C9m	GRNCON	0.101396028	1.41E-78	Correlation
VCP	C9rCON	0.099139702	4.33E-74	Correlation
C9mCON	GRNCON	0.104771445	1.13E-68	Correlation

Dataset1	Dataset2	corRvec	pval	Relationship
GRN	C9rCON	0.093170481	4.65E-65	Correlation
CHMP2B	C9r	0.078650136	3.00E-38	Correlation
C9m	GRN	0.070308169	4.34E-38	Correlation
C9rCON	CHCON	0.069934314	5.62E-38	Correlation
C9m	C9r	0.064566617	1.87E-32	Correlation
VCP	GRNCON	0.061676982	1.18E-29	Correlation
CHMP2B	GRNCON	0.06808106	1.70E-29	Correlation
CHMP2B	C9mCON	0.07348021	2.12E-28	Correlation
C9m	CHMP2B	0.064481371	6.62E-27	Correlation
GRN	C9r	0.058314719	4.36E-26	Correlation
GRN	VCP	0.057055619	3.27E-25	Correlation
C9mCON	VCPCON	0.066445802	3.11E-23	Correlation
C9r	C9mCON	0.059344868	7.09E-23	Correlation
C9r	VCPCON	0.041635816	1.04E-11	Correlation
CHCON	VCPCON	0.040571495	2.56E-11	Correlation
GRN	C9mCON	0.036583101	1.33E-09	Correlation
CHMP2B	C9rCON	0.034647364	8.99E-09	Correlation
VCP	C9mCON	0.032794016	4.74E-08	Correlation
VCP	C9r	0.02915095	1.12E-07	Correlation
VCP	VCPCON	0.021827297	0.000346285	Correlation
GRN	VCPCON	0.021566726	0.000433578	Correlation
VCP	CHCON	0.01886406	0.00054837	Correlation
C9m	VCPCON	0.018218401	0.002575407	Correlation
GRNCON	VCPCON	0.014362344	0.018112961	Correlation
C9m	C9rCON	0.011292094	0.036304906	Correlation
C9rCON	VCPCON	-0.009142945	0.131812358	Correlation
C9mCON	C9rCON	0.007806611	0.191165437	Correlation
CHMP2B	VCPCON	0.005023414	0.456970162	Correlation
CHMP2B	VCP	0.002826308	0.64094965	Correlation
CHMP2B	GRN	-0.000809281	0.894258976	Correlation
C9m	VCP	0.000608703	0.910640153	Correlation

Cytoscape output shows

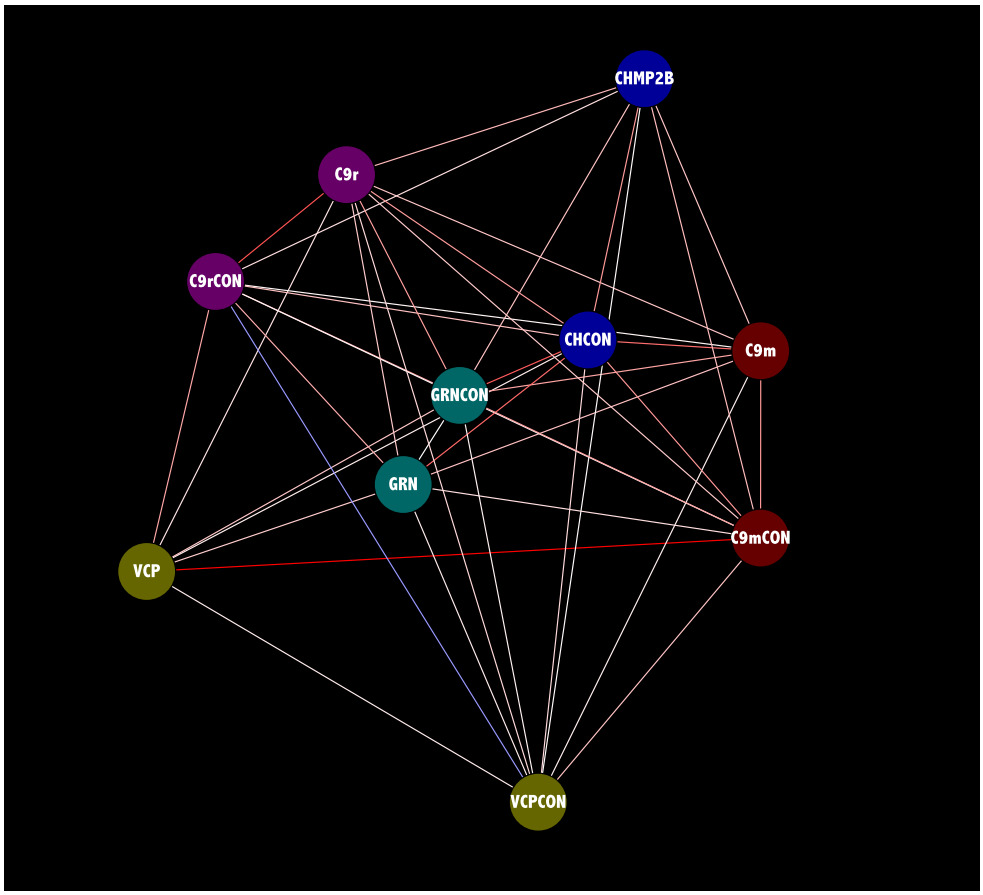

Figure 1: