LabBook_29_04_2016

Claire Green

Monday

I spent most of the day researching for seattle. The information is contained within the file "Seattle.xlsx". I also rewrote my report for the pcxn example and that can be found in /Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/GSEA/PCxN Example.

Tuesday

Now that I know that the function diffPathways exists in the pathprint package, and have learnt a lot about GWAS enrichment using hyperPathways, I decided to implement these on my data. The diffPathways script is as follows:

```
##### Using pathprint to identify common pathways across multiple TDP-43 pathology-containing data sets

library (pathprint)
options(stringsAsFactors = FALSE)


####C9_LCM ######
setwd ("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/C9orf72_LCM") #set working directory to loc
exp_C9.LCM <- read.csv ("eset_NineP_150612_exprs.csv", header=TRUE) #assign the .csv file to a variable
row.names (exp_C9.LCM) <- exp_C9.LCM[,1] #specify that first column contains gene names
exp_C9.LCM<- exp_C9.LCM[,2:12] #specify that all other columns are gene expression data

C9.LCM_pathprint <- exprs2fingerprint(exp_C9.LCM, platform = "GPL570", species="human", progressBar=T)
vec.c9 <- c(1,1,1,1,1,1,1,1,0,0,0)


####CHMP2B_LCM ######
setwd ("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/CHMP2B")
exp_CHMP2B.LCM <- read.csv ("eset_CHMP2B_250615_exprs nooutlier.csv", header=TRUE)
row.names (exp_CHMP2B.LCM) <- exp_CHMP2B.LCM[,1]
exp_CHMP2B.LCM<- exp_CHMP2B.LCM[,2:10]

CHMP2B.LCM_pathprint <- exprs2fingerprint (exp_CHMP2B.LCM, platform = "GPL570", species="human", progres
vec.ch <- c(1,1,1,0,0,0,0,0,0)


####sals_lcm###

setwd ("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/FUS_SALS_LCM_CELfiles")
exp_SALS.LCM <- read.csv ("eset_SALS_LCM_260615_exprs.csv", header=TRUE)
row.names (exp_SALS.LCM) <- exp_SALS.LCM[,1]
exp_SALS.LCM<- exp_SALS.LCM[,2:11]

SALS.LCM_pathprint <- exprs2fingerprint (exp_SALS.LCM, platform = "GPL570", species="human", progressBa
```

```
vec.sals <- c(0,0,0,1,1,1,1,1,1,1)


####FTLD###

setwd ("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/FTD-U.brain")
FTLD <- read.csv ("FTLD_expr_tdp43.csv", header=TRUE)
row.names (FTLD) <- FTLD[,1]
FTLD <- FTLD[,2:25]

#GPL571 = Affymetrix Human Genome U113A 2.0 array
FTLD_pathprint <- exprs2fingerprint (FTLD, platform = "GPL571", species="human", progressBar=T)
vec.FTLD <- c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0)


####VCP###

setwd ("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/VCP.myopathy")
VCP <- read.csv ("eset_VCP.myopathy_170715_exprs.csv", header=TRUE)
row.names (VCP) <- VCP[,1]
VCP <- VCP[,2:11]

VCP_pathprint <- exprs2fingerprint (VCP, platform = "GPL570", species="human", progressBar=T)
vec.vcp <- c(0,0,0,1,1,1,1,1,1,1)


##DiffPathways##

thres <- 0.1

c9.lcm <- diffPathways(C9.LCM_pathprint, vec.c9, thres)
CHMP2B.lcm <- diffPathways(CHMP2B.LCM_pathprint, vec.ch, thres)
SALS.lcm <- diffPathways(SALS.LCM_pathprint, vec.sals, thres)
FTLD_FCx <- diffPathways(FTLD_pathprint, vec.FTLD, thres)
VCP.m <- diffPathways(VCP_pathprint, vec.vcp, thres)


###INTERSECT###

overlap <- Reduce(intersect, list(c9.lcm, CHMP2B.lcm, SALS.lcm, FTLD_FCx, VCP.m)) #selects pathways tha
print(overlap)
```

Pathways are as follows:

Pentose and glucuronate interconversions (KEGG)
Fructose and mannose metabolism (KEGG)
Lysine degradation (KEGG)
Starch and sucrose metabolism (KEGG)
Pantothenate and CoA biosynthesis (KEGG)
Nitrogen metabolism (KEGG)
ABC transporters (KEGG)
Complement and coagulation cascades (KEGG)
Jak-STAT signaling pathway (KEGG)
Phototransduction (KEGG)

Prion diseases (KEGG)
Phase I, non P450 (Wikipathways)
Ganglio Sphingolipid Metabolism (Wikipathways)
Urea cycle and metabolism of amino groups (Wikipathways)
Complement Activation, Classical Pathway (Wikipathways)
Biogenic Amine Synthesis (Wikipathways)
Complement and Coagulation Cascades (Wikipathways)
Glucuronidation (Wikipathways)
SIDS Susceptibility Pathways (Wikipathways)
Signaling by Insulin receptor (Reactome)
Opioid Signalling (Reactome)
{ESR1,24} (Static Module)
{F2,46} (Static Module)
{HSPA8,34} (Static Module)
{NRP1,11} (Static Module)
{POR,15} (Static Module)
{RAN,17} (Static Module)
{SPTAN1,10} (Static Module)
{SREBF1,11} (Static Module)

GWAS gene enrichment is conducted as follows:

```
# set working directory
setwd(dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/GSEA/PCxN Example/probesets/")

#Load individual gene names for each significance threshold
A <- read.table(file = "threegenes.txt")
a <- A$V1

B <- read.table(file = "fourgenes.txt")
b <- B$V1

C <- read.table(file = "fivegenes.txt")
c <- C$V1

D <- read.table(file = "sixgenes")
d <- D$V1

setwd (dir = "/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint")

Z <- read.csv(file = "gseagenes.csv", na.strings = c("", "NA)"))
Z <- as.list(Z)
Z <- lapply(Z, function(x) x[!is.na(x)])

#Load file with all genes
library(hgu133plus2.db)
sym <- hgu133plus2SYMBOL
sym1 <- mappedkeys(sym)
sym2 <- as.list (sym[c(sym1)])
sym3 <- data.frame (sym2)
sym.probes <- names (sym2)
sym.genes <- sym3[1,]

sym.genes <- t(sym.genes)

allgenes <- sym.genes[!duplicated(sym.genes),]

pathwayEnrichment <- hyperPathway(
                genelist = a,
                                geneset = Z,
                                Nchip = length(allgenes)
                                )
write.csv(pathwayEnrichment, file = "GPnd.P<.000001.csv")
```

Enrichment is not very good with the GWAS central SNPs, but is much better with the neuroX list

Pathway	ID	P-value	BHadjP-value
Pentose.and.glucuronate.interconversionsKEGG.	1	1	1
Fructose.and.mannose.metabolismKEGG.	2	1	1
Lysine.degradationKEGG.	3	0.003084414	0.014463524
Starch.and.sucrose.metabolismKEGG.	4	0.026952725	0.086847671
Pantothenate.and.CoA.biosynthesisKEGG.	5	1	1
Nitrogen.metabolismKEGG.	6	1	1
ABC.transportersKEGG.	7	0.002120118	0.012296684
Complement.and.coagulation.cascadesKEGG.	8	0.061822851	0.137912515
Jak.STAT.signaling.pathwayKEGG.	9	0.228220201	0.472741845
PhototransductionKEGG.	10	1	1
Prion.diseasesKEGG.	11	0.001163881	0.011060369
Phase.Inon.P450Wikipathways.	12	1	1
Ganglio.Sphingolipid.MetabolismWikipathways.	13	0.001525568	0.011060369
Urea.cycle.and.metabolism.of.amino.groupsWikipathways.	14	1	1
Complement.ActivationClassical.PathwayWikipathways.	15	1	1
Biogenic.Amine.SynthesisWikipathways.	16	0.003491195	0.014463524
Complement.and.Coagulation.CascadesWikipathways.	17	0.03559842	0.09385038
GlucuronidationWikipathways.	18	1	1
SIDS.Susceptibility.PathwaysWikipathways.	19	0.057093973	0.137912515
Signaling.by.Insulin.receptorReactome.	20	4.52E-05	0.001312125
Opioid.SignallingReactome.	21	1	1
.ESR1.24.Static.Module.	22	0.008862604	0.03212694
.F2.46.Static.Module.	23	0.030540093	0.088566269
.HSPA8.34.Static.Module.	24	1	1
.NRP1.11.Static.Module.	25	1	1
.POR.15.Static.Module.	26	1	1
.RAN.17.Static.Module.	27	1	1
.SPTAN1.10.Static.Module.	28	0.001525568	0.011060369
.SREBF1.11.Static.Module.	29	1	1