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

GWAS Central (All)

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Pentose.and.glucuronate.interconversions..KEGG.	1	0.431819002	0.736632416	2	27
Fructose.and.mannose.metabolism..KEGG.	2	0.395690918	0.736632416	3	36
Lysine.degradation..KEGG.	3	1	1	0	49
Starch.and.sucrose.metabolism..KEGG.	4	0.745267593	1	2	43
Pantothenate.and.CoA.biosynthesis..KEGG.	5	0.185818362	0.578674418	2	17
Nitrogen.metabolism..KEGG.	6	0.047194945	0.482169181	4	23
ABC.transporters..KEGG.	7	0.330701101	0.736632416	4	43
Complement.and.coagulation.cascades..KEGG.	8	1	1	0	68
Jak.STAT.signaling.pathway..KEGG.	9	0.874012655	1	9	152
Phototransduction..KEGG.	10	0.431819002	0.736632416	2	27
Prion.diseases..KEGG.	11	0.193312545	0.578674418	4	35
Phase.I..non.P450..Wikipathways.	12	1	1	0	7
Ganglio.Sphingolipid.Metabolism..Wikipathways.	13	0.219497193	0.578674418	1	10
Urea.cycle.and.metabolism.of.amino.groups..Wikipathways.	14	0.282687653	0.683161828	2	21
Complement.Activation..Classical.Pathway..Wikipathways.	15	1	1	0	16
Biogenic.Amine.Synthesis..Wikipathways.	16	0.141316007	0.578674418	2	15
Complement.and.Coagulation.Cascades..Wikipathways.	17	1	1	0	50
Glucuronidation..Wikipathways.	18	1	1	0	17
SIDS.Susceptibility.Pathways..Wikipathways.	19	0.058428241	0.482169181	9	65
Signaling.by.Insulin.receptor..Reactome.	20	0.488325203	0.786746161	9	108
Opioid.Signalling..Reactome.	21	0.020688531	0.482169181	12	79
X.ESR1.24. . . Static.Module.	22	0.157448085	0.578674418	3	24
X.F2.46. . . Static.Module.	23	0.923038029	1	1	46
X.HSPA8.34. . . Static.Module.	24	0.757109397	1	1	30
X.NRP1.11. . . Static.Module.	25	0.185932319	0.578674418	1	9
X.POR.15. . . Static.Module.	26	0.387844426	0.736632416	1	15
X.RAN.17. . . Static.Module.	27	1	1	0	16
X.SPTAN1.10. . . Static.Module.	28	0.219497193	0.578674418	1	10
X.SREBF1.11. . . Static.Module.	29	0.066506094	0.482169181	2	11

*

NeuroX (All)*

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Pentose.and.glucuronate.interconversionsKEGG.	1	1	1	0	27
Fructose.and.mannose.metabolismKEGG.	2	1	1	0	36
Lysine.degradationKEGG.	3	0.003084414	0.014463524	2	49
Starch.and.sucrose.metabolismKEGG.	4	0.026952725	0.086847671	1	43
Pantothenate.and.CoA.biosynthesisKEGG.	5	1	1	0	17
Nitrogen.metabolismKEGG.	6	1	1	0	23
ABC.transportersKEGG.	7	0.002120118	0.012296684	2	43
Complement.and.coagulation.cascadesKEGG.	8	0.061822851	0.137912515	1	68
Jak.STAT.signaling.pathwayKEGG.	9	0.228220201	0.472741845	1	152
PhototransductionKEGG.	10	1	1	0	27
Prion.diseasesKEGG.	11	0.001163881	0.011060369	2	35
Phase.Inon.P450Wikipathways.	12	1	1	0	7
Ganglio.Sphingolipid.MetabolismWikipathways.	13	0.001525568	0.011060369	1	10
Urea.cycle.and.metabolism.of.amino.groupsWikipathways.	14	1	1	0	21
Complement.ActivationClassical.PathwayWikipathways.	15	1	1	0	16
Biogenic.Amine.SynthesisWikipathways.	16	0.003491195	0.014463524	1	15
Complement.and.Coagulation.CascadesWikipathways.	17	0.03559842	0.09385038	1	50
GlucuronidationWikipathways.	18	1	1	0	17
SIDS.Susceptibility.PathwaysWikipathways.	19	0.057093973	0.137912515	1	65
Signaling.by.Insulin.receptorReactome.	20	4.52E-05	0.001312125	5	108
Opioid.SignallingReactome.	21	1	1	0	79
.ESR1.24.Static.Module.	22	0.008862604	0.03212694	1	24
.F2.46.Static.Module.	23	0.030540093	0.088566269	1	46
.HSPA8.34.Static.Module.	24	1	1	0	30
.NRP1.11.Static.Module.	25	1	1	0	9
.POR.15.Static.Module.	26	1	1	0	15
.RAN.17.Static.Module.	27	1	1	0	16
.SPTAN1.10.Static.Module.	28	0.001525568	0.011060369	1	10
.SREBF1.11.Static.Module.	29	1	1	0	11

As can be seen from the table, 8 of the 29 pathways are significantly enriched with NeuroX genes.

Lysine degradation (KEGG)
ABC transporters (KEGG)
Prion diseases (KEGG)
Ganglio Sphingolipid Metabolism (Wikipathways)
Biogenic Amine Synthesis (Wikipathways)
Signaling by Insulin receptor (Reactome)
{ESR1,24} (Static Module)
{SPTAN1,10} (Static Module)

I took these pathways and added them into PCxN. I added the 10 most correlated pathways, which were:

IL-3 down reg. targets (Netpath)
TGF beta receptor down reg. targets (Netpath)
IL-4 down reg. targets (Netpath)
IL-6 up reg. targets (Netpath)
Leishmaniasis (KEGG)
Phagosome (KEGG)
Leukocyte transendothelial migration (KEGG)
Malaria (KEGG)
{GRB2,414} (Static Module)
IL-4 up reg. targets (Netpath)

I acquired the genes in these pathways and re-conducted the enrichment analysis. For GWAS central (p<.001)

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Lysine.degradation..KEGG.	1	1	1	0	49
ABC.transporters..KEGG.	2	0.330701101	0.457893832	4	43
Prion.diseases..KEGG.	3	0.193312545	0.359177225	4	35
Ganglio.Sphingolipid.Metabolism..Wikipathways.	4	0.219497193	0.359177225	1	10
Biogenic.Amine.Synthesis..Wikipathways.	5	0.141316007	0.359177225	2	15
Signaling.by.Insulin.receptor..Reactome.	6	0.488325203	0.62784669	9	108
X.ESR1.24. . . Static.Module.	7	0.157448085	0.359177225	3	24
X.SPTAN1.10. . . Static.Module.	8	0.219497193	0.359177225	1	10
IL.3.down.reg..targets..Netpath.	9	0.219497193	0.359177225	1	10
TGF.beta.receptor.down.reg..targets..Netpath.	10	0.03985612	0.143482031	78	735
IL.4.down.reg..targets..Netpath.	11	0.818256368	0.866389096	5	90
IL.6.up.reg..targets..Netpath.	12	0.804160508	0.866389096	4	75
Leishmaniasis..KEGG.	13	0.038961174	0.143482031	10	69
Phagosome..KEGG.	14	0.542475635	0.650970762	12	147
Leukocyte.transendothelial.migration..KEGG.	15	0.004708711	0.042378396	18	113
Malaria..KEGG.	16	0.278894737	0.418342105	5	50
X.GRB2.414. . . Static.Module.	17	0.013879012	0.083274071	47	393
IL.4.up.reg..targets..Netpath.	18	1.36E-05	0.000244606	38	216

here we can see that 2 pathways are now significant - leukocyte transendothelial migration and IL 4 up reg. targets. This does not continue to p<.0001.

For NeuroX, the results are thus:

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Lysine.degradation..KEGG.	1	0.003084414	0.00785519	2	49
ABC.transporters..KEGG.	2	0.002120118	0.006360354	2	43
Prion.diseases..KEGG.	3	0.001163881	0.005492045	2	35
Ganglio.Sphingolipid.Metabolism..Wikipathways.	4	0.001525568	0.005492045	1	10
Biogenic.Amine.Synthesis..Wikipathways.	5	0.003491195	0.00785519	1	15
Signaling.by.Insulin.receptor..Reactome.	6	4.52E-05	0.000814422	5	108
X.ESR1.24. . . Static.Module.	7	0.008862604	0.017725208	1	24
X.SPTAN1.10. . . Static.Module.	8	0.001525568	0.005492045	1	10
IL.3.down.reg..targets..Netpath.	9	1	1	0	10
TGF.beta.receptor.down.reg..targets..Netpath.	10	0.000119654	0.001076883	13	735
IL.4.down.reg..targets..Netpath.	11	1	1	0	90
IL.6.up.reg..targets..Netpath.	12	0.01011211	0.018201798	2	75
Leishmaniasis..KEGG.	13	0.063428009	0.081550297	1	69
Phagosome..KEGG.	14	0.217306489	0.2444698	1	147
Leukocyte.transendothelial.migration..KEGG.	15	0.145137782	0.174165339	1	113
Malaria..KEGG.	16	0.03559842	0.05339763	1	50
X.GRB2.414. . . Static.Module.	17	0.029960126	0.049025661	5	393
IL.4.up.reg..targets..Netpath.	18	0.040003566	0.055389553	3	216

Other than the original 8, 3 other pathways are significant, with 2 reaching significance.

At p<.5e-8 four pathways are significant - prion diseases, Biogenic amine synthesis, signalling by insulin receptor, and TGF beta receptor down reg.targets.

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Lysine.degradation..KEGG.	1	1	1	0	49
ABC.transporters..KEGG.	2	1	1	0	43
Prion.diseases..KEGG.	3	0.000105843	0.001678927	2	35
Ganglio.Sphingolipid.Metabolism..Wikipathways.	4	1	1	0	10
Biogenic.Amine.Synthesis..Wikipathways.	5	0.000696534	0.004179204	1	15
Signaling.by.Insulin.receptor..Reactome.	6	0.000186547	0.001678927	3	108
X.ESR1.24. . . Static.Module.	7	1	1	0	24
X.SPTAN1.10. . . Static.Module.	8	1	1	0	10
IL.3.down.reg..targets..Netpath.	9	1	1	0	10
TGF.beta.receptor.down.reg..targets..Netpath.	10	0.002991262	0.013460681	6	735
IL.4.down.reg..targets..Netpath.	11	1	1	0	90
IL.6.up.reg..targets..Netpath.	12	1	1	0	75
Leishmaniasis..KEGG.	13	1	1	0	69
Phagosome..KEGG.	14	1	1	0	147
Leukocyte.transendothelial.migration..KEGG.	15	1	1	0	113
Malaria..KEGG.	16	1	1	0	50
X.GRB2.414. . . Static.Module.	17	0.276474406	0.829423219	1	393
IL.4.up.reg..targets..Netpath.	18	0.019187682	0.069075657	2	216

The SNP genes are as follows:

Prion diseases - NOTCH1, SOD1

Biogenic emine synthesis - COMT
Signaling by insulin receptor - TSC2, FGFR3, STK11
TGF beta receptor down reg. targets - GRN, EPHA2, FGFR3, APOE, APOC1, SOD1

Wednesday

In the morning I wrote up my feedback on PCxN. This can be found on the basecamp project Hide Laboratory of Computational Biology as a text document.

In the afternoon, I turned back to looking at the pathprint pathways. I have established that four pathways are significaltly enriched with NeuroX SNP genes at genome wide significance. What I went on to do is look for enrichment of my DEGs in that list, using the hyperPathway function.

Pathway	ID	P-value	BHadjP-value	nGenes	nPathwa
Lysine.degradation..KEGG.	1	0.005167366	0.023253148	1	49
ABC.transporters..KEGG.	2	1	1	0	43
Prion.diseases..KEGG.	3	1	1	0	35
Ganglio.Sphingolipid.Metabolism..Wikipathways.	4	1	1	0	10
Biogenic.Amine.Synthesis..Wikipathways.	5	1	1	0	15
Signaling.by.Insulin.receptor..Reactome.	6	1	1	0	108
X.ESR1.24. . . Static.Module.	7	1	1	0	24
X.SPTAN1.10. . . Static.Module.	8	1	1	0	10
IL.3.down.reg..targets..Netpath.	9	1	1	0	10
TGF.beta.receptor.down.reg..targets..Netpath.	10	0.005095754	0.023253148	5	735
IL.4.down.reg..targets..Netpath.	11	1	1	0	90
IL.6.up.reg..targets..Netpath.	12	0.000593837	0.010689063	2	75
Leishmaniasis..KEGG.	13	1	1	0	69
Phagosome..KEGG.	14	0.041198018	0.123594055	1	147
Leukocyte.transendothelial.migration..KEGG.	15	0.025453406	0.091632262	1	113
Malaria..KEGG.	16	1	1	0	50
X.GRB2.414. . . Static.Module.	17	0.2126385	0.546784714	1	393
IL.4.up.reg..targets..Netpath.	18	0.001263182	0.011368634	3	216

This shows that with an adjusted p value, 4 pathways are significantly enriched with DEGs. Lysine degredation, TGF beta receptor down reg targets, IL 6 up reg. targets, and IL 4 up reg targets.