LabBook_05_05_2017

Claire Green

This Week

This week I was looking into the signal extraction of TDP vs neurodegeneration. I had used the SOD1 and FUS datasets to generate consensus genes and pathways, and then compared them to the TDP-43 positive results. After removing the overlapping genes The enrichments were reduced in significance but stable (see presentation Presentation1 in FoldChangeResults). I was worried that the significance had dropped because I was losing signal but Dennis said that it could just be because the sample is smaller and that I should check the effect sizes or percentage overlap because it might still be the same, the proportions are just different.

I've been having trouble with a FTLD cerebellum dataset that I was trying to use as a TDP-43- ND control but the problem is that the "FTLD" signal is so strong that it's hard to separate them. They don't separate based on all genes or all pathways, so Dennis said "Well that makes sense because you're not looking at your signal which should separate them" so I need to look at whether the dendrogram separates them if I'm only looking at my TDP-43 model pathways or genes. I have briefly looked at the pathways but it doesn't separate them so I'm not that hopeful. Though they are from the same patient so the liklihood is that the batch effect may be stronger as well. At least it's relativly good at splitting by tissue so there isn't necessarily a tissue effect.

I have also been looking at the difference between defining upregulated and downregulated pathways using the difference in mean or using the consensus pathway method. What I have discovered is essentially that although two means may be quite different, this doesn't necessarily translate into consistency in upregulation or downregulation. For example, the pathway "Complement and coagulation Cascades (KEGG)" has the fourth highest difference in mean of all the pathways. However, the mean for patients is 0.05 and the mean for controls is -0.29. What this tells me is that althought the patients mean is higher than controls, a) its difference from the background is extremely small, and b) this mean could have a lot of variance.

I decided then to do a combination - set the threshold for consensus at 0.4, and then a minimum mean difference of 0.1. These are the pathways I get:

```
library(pathprint)

load("~/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/Pathprinted.RData")
load("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/Pathprint_biomart.RData")

thresh = 0.4
#Select patient columns
pat_C9 <- pathprint_C9[,4:11]
pat_sals <- pathprint_sals[,4:10]
pat_ftld <- pathprint_ftld[,9:24]
pat_vcp <- pathprint_vcp[,4:10]
#Combine
pat_all_pp <- cbind(pat_C9, pat_sals,pat_ftld, pat_vcp)

#
# keep <- rowSums(pat_all_pp==1) >= ncol(pat_all_pp)/(100/80)
# pat_up<-pat_all_pp[keep,]
# keep <- rowSums(pat_all_pp==-1) >= ncol(pat_all_pp)/(100/80)
# pat_down<-pat_all_pp[keep,]
```

```r
#Run consensus
pat_consensus <- data.frame(consensusFingerprint(pat_all_pp, thresh))

#Select control columns
con_C9 <- pathprint_C9[,1:3]
con_sals <- pathprint_sals[,1:3]
con_ftld <- pathprint_ftld[,1:8]
con_vcp <- pathprint_vcp[,1:3]
#Combine
con_all_pp <- cbind(con_C9, con_sals,con_ftld, con_vcp)

# keep <- rowSums(con_all_pp==1) >= ncol(con_all_pp)/(100/80)
# con_up<-con_all_pp[keep,]
# keep <- rowSums(con_all_pp==-1) >= ncol(con_all_pp)/(100/80)
# con_down<-con_all_pp[keep,]

#Run consensus
con_consensus <- data.frame(consensusFingerprint(con_all_pp, thresh))

# #Split up and downregulated pathways
# pat_up <- subset(pat_consensus, pat_consensus[,1] ==1)
# pat_down <- subset(pat_consensus, pat_consensus[,1]==-1)
#
# con_up <- subset(con_consensus, con_consensus[,1]==1)
# con_down <- subset(con_consensus, con_consensus[,1]==-1)
#
# #Extract pathway names
# up_pat_path <- rownames(pat_up)
# down_pat_path <- rownames(pat_down)
# up_con_path <- rownames(con_up)
# down_con_path <- rownames(con_down)
#
# #Subset patient pathways that are not present in control
# x <- subset(pat_up, !(rownames(pat_up) %in% up_con_path))
# y <- subset(pat_down, !(rownames(pat_down) %in% down_con_path))
#


##########
all_consensus <- merge(pat_consensus, con_consensus, by = 0)
diff_consensus_PP <- subset(all_consensus, !(all_consensus$consensusFingerprint.pat_all_pp..thresh. == a

down_path <- subset(diff_consensus_PP, diff_consensus_PP$consensusFingerprint.pat_all_pp..thresh. < dif
down_pathways <- down_path$Row.names
up_path <- subset(diff_consensus_PP, diff_consensus_PP$consensusFingerprint.pat_all_pp..thresh. > diff_
up_pathways <- up_path$Row.names

dyspathways_PP <- c(down_pathways, up_pathways)




all_pp <- cbind(pat_C9,pat_ftld,pat_sals,pat_vcp, con_C9,con_ftld,con_sals,con_vcp)
```

```
# all_pp <- subset(all_pp, !(rowSums(all_pp)==0))

dys_pp <- all_pp[rownames(all_pp) %in% dyspathways_PP,]

patmean <- apply(pat_all_pp, 1, mean)
names(patmean) <- rownames(pat_all_pp)
conmean <- apply(con_all_pp, 1, mean)
names(conmean) <- rownames(con_all_pp)

mean.df <- data.frame(row.names = rownames(pat_all_pp),
                      Pat.Mean= patmean,
                      Con.Mean = conmean)
mean.df$difference <- (mean.df$Pat.Mean - mean.df$Con.Mean)

dys.mean<- subset(mean.df, rownames(mean.df) %in% dyspathways_PP)
dys.mean$difference <- (dys.mean$Pat.Mean - dys.mean$Con.Mean)
dys.mean <- dys.mean[order(dys.mean$difference, decreasing = TRUE),]
dys.mean <- dys.mean[ which( dys.mean$difference > 0.1 | dys.mean$difference < -0.1) , ]
list <- rownames(dys.mean)


cat(rownames(dys.mean), sep = '\n')
```

```
## IL-2 down reg. targets (Netpath)
## Mitochondrial LC-Fatty Acid Beta-Oxidation (Wikipathways)
## Vitamin B12 Metabolism (Wikipathways)
## Vitamin D synthesis (Wikipathways)
## {BRCA1,28} (Static Module)
## {CD4,14} (Static Module)
## {SEPT2,21} (Static Module)
## Fatty Acid Biosynthesis (Wikipathways)
## {FCGR2B,50} (Static Module)
## {RB1,11} (Static Module)
## Muscle contraction (Reactome)
## Steroid Biosynthesis (Wikipathways)
## Glutathione metabolism (KEGG)
## Hedgehog Signaling Pathway (Wikipathways)
## Fat digestion and absorption (KEGG)
## Aminoacyl-tRNA biosynthesis (KEGG)
## {VCP,17} (Static Module)
## Osteoclast differentiation (KEGG)
## Fluoropyrimidine Activity (Wikipathways)
## IL-7 down reg. targets (Netpath)
## Heme Biosynthesis (Wikipathways)
## {SP1,88} (Static Module)
## Signaling by Notch (Reactome)
## Ubiquinone and other terpenoid-quinone biosynthesis (KEGG)
## Sulfur metabolism (KEGG)
## Oxidative phosphorylation (KEGG)
## Caffeine metabolism (KEGG)
## Phototransduction (KEGG)
## Parkinson's disease (KEGG)
## {PPP2CA,20} (Static Module)
## {ABL1,15} (Static Module)
```

```
## Type II diabetes mellitus (KEGG)
## Signaling by EGFR (Reactome)
## Glucocorticoid &amp;amp; Mineralcorticoid Metabolism (Wikipathways)
## Alzheimer's disease (KEGG)
## Blood Clotting Cascade (Wikipathways)
## Terpenoid backbone biosynthesis (KEGG)
## Signaling by GPCR (Reactome)
```

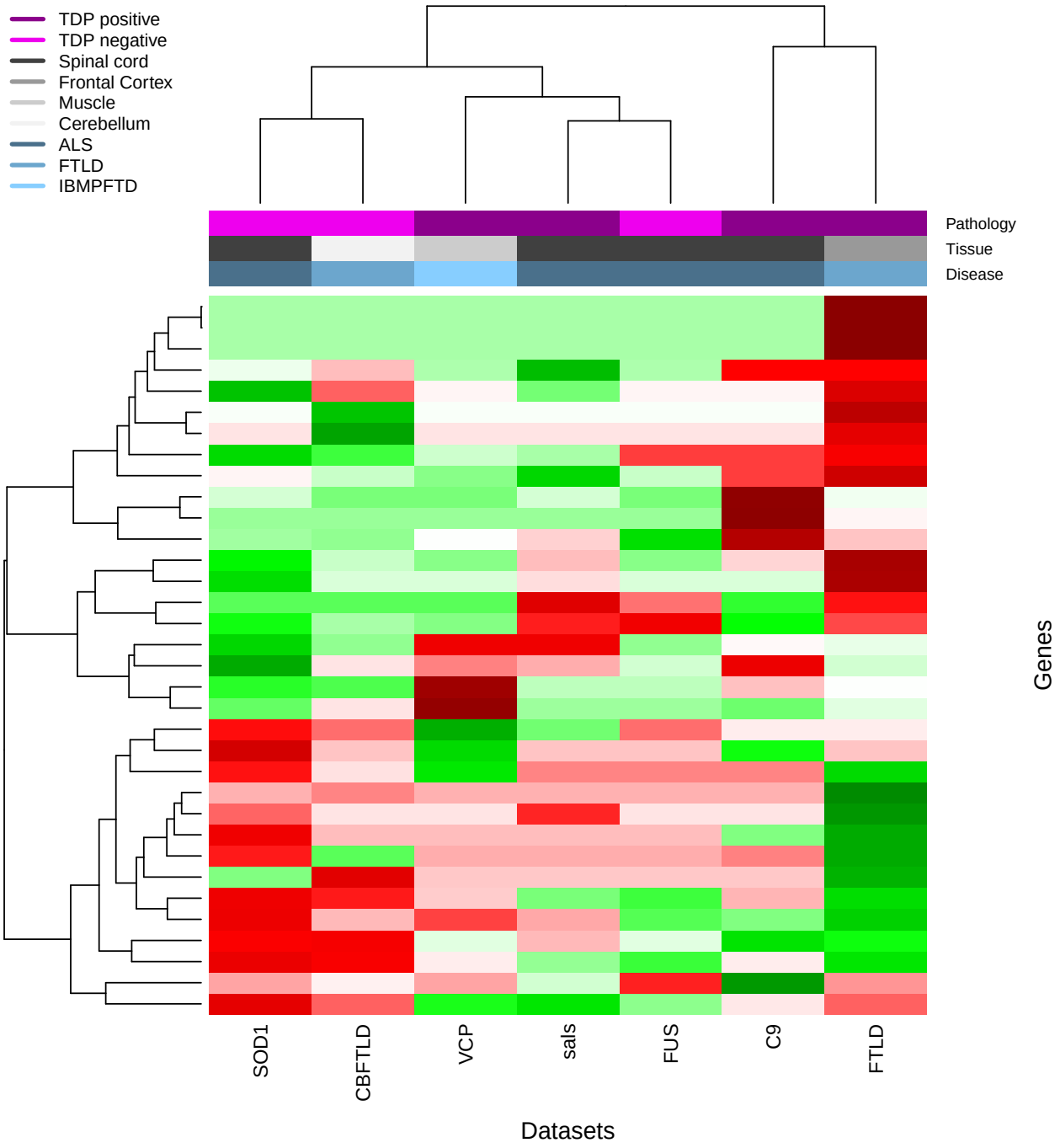
When you run the same code for the non-TDP SOD1 and FUS datasets, the overlap requires the removal of hedgehog signaling pathway and oxidative phosphorylation. There is also an opposite regulation status for BRCA1 SM, and signaling by GPCR.

However, the problem is that these pathways aren't that good at clustering TDP and nonTDP datasets.

```
library(heatmap3)

load("~/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/allpathprint.RData")

dyspathways_PP <- read.table("~/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/Consensus_fingerprin
dyspathways_PP <- dyspathways_PP$V1
# dyspathways_PP <- overlap

#Select patient columns
pat_C9 <- pathprint_C9[,4:11]
pat_C9_dys <- pat_C9[rownames(pat_C9) %in% dyspathways_PP,]
pat_sals <- pathprint_sals[,4:10]
pat_sals_dys <- pat_sals[rownames(pat_sals) %in% dyspathways_PP,]
pat_ftld <- pathprint_ftld[,9:24]
pat_ftld_dys <- pat_ftld[rownames(pat_ftld) %in% dyspathways_PP,]
pat_vcp <- pathprint_vcp[,4:10]
pat_vcp_dys <- pat_vcp[rownames(pat_vcp) %in% dyspathways_PP,]
# pat_ch <- pathprint_chmp2b[,7:9]
pat_sod1 <- pathprint_sod1[,8:10]
pat_sod1_dys <- pat_sod1[rownames(pat_sod1) %in% dyspathways_PP,]
pat_fus <- pathprint_fus[,4:6]
pat_fus_dys <- pat_fus[rownames(pat_fus) %in% dyspathways_PP,]
pat_CBFTLD <- pathprint_cbftld[,8:16]
pat_CBFTLD_dys <- pat_CBFTLD[rownames(pat_CBFTLD) %in% dyspathways_PP,]


#Select control columns
con_C9 <- pathprint_C9[,1:3]
con_C9_dys<- con_C9[rownames(con_C9) %in% dyspathways_PP,]
con_sals <- pathprint_sals[,1:3]
con_sals_dys<- con_sals[rownames(con_sals) %in% dyspathways_PP,]
con_ftld <- pathprint_ftld[,1:8]
con_ftld_dys<- con_ftld[rownames(con_ftld) %in% dyspathways_PP,]
con_vcp <- pathprint_vcp[,1:3]
con_vcp_dys<- con_vcp[rownames(con_vcp) %in% dyspathways_PP,]
# con_ch <- pathprint_chmp2b[,1:6]
con_sod1 <- pathprint_sod1[,1:7]
con_sod1_dys<- con_sod1[rownames(con_sod1) %in% dyspathways_PP,]
con_fus <- pathprint_fus[,1:3]
con_fus_dys<- con_fus[rownames(con_fus) %in% dyspathways_PP,]
con_CBFTLD <- pathprint_cbftld[,1:7]
```

```
con_CBFTLD_dys<- con_CBFTLD[rownames(con_CBFTLD) %in% dyspathways_PP,]




C9_pat_mean <- apply(pat_C9_dys, 1, mean)
C9_con_mean <- apply(con_C9_dys, 1, mean)
C9diff <- C9_pat_mean - C9_con_mean

sals_pat_mean <- apply(pat_sals_dys, 1, mean)
sals_con_mean <- apply(con_sals_dys, 1, mean)
salsdiff <- sals_pat_mean - sals_con_mean

ftld_pat_mean <- apply(pat_ftld_dys, 1, mean)
ftld_con_mean <- apply(con_ftld_dys, 1, mean)
ftlddiff <- ftld_pat_mean - ftld_con_mean

vcp_pat_mean <- apply(pat_vcp_dys, 1, mean)
vcp_con_mean <- apply(con_vcp_dys, 1, mean)
vcpdiff <- vcp_pat_mean - vcp_con_mean

sod1_pat_mean <- apply(pat_sod1_dys, 1, mean)
sod1_con_mean <- apply(con_sod1_dys, 1, mean)
sod1diff <- sod1_pat_mean - sod1_con_mean

fus_pat_mean <- apply(pat_fus_dys, 1, mean)
fus_con_mean <- apply(con_fus_dys, 1, mean)
fusdiff <- fus_pat_mean - fus_con_mean

cbftld_pat_mean <- apply(pat_CBFTLD_dys, 1, mean)
cbftld_con_mean <- apply(con_CBFTLD_dys, 1, mean)
cbftlddiff <- cbftld_pat_mean - cbftld_con_mean


PP_diffs <- data.frame(row.names = rownames(pat_C9_dys),
                       C9 = C9diff,
                       sals=salsdiff,
                       FTLD=ftlddiff,
                       VCP=vcpdiff,
                       SOD1=sod1diff,
                       FUS=fusdiff,
                       CBFTLD=cbftlddiff)



colsidecolorsLFC <- cbind(Disease=c(rep("skyblue4",2),
                                    "skyblue3",
                                    "skyblue1",
                                    rep("skyblue4",2),
                                    "skyblue3"),
                          Tissue=c(rep("gray25",2),
                                   "gray60",
                                   "gray80",
```

```
                                    rep("gray25",2),
                                    "gray95"),
                            Pathology=c(rep("magenta4",4),
                                        rep("magenta2", 3)))

hmcols<- colorRampPalette(c("green4","green","white", "red","red4"))(256)

heatmap3(PP_diffs,
         col = hmcols,
         Rowv = TRUE,
         Colv = TRUE,
         distfun = dist,
         # hclustfun = hclust,
         scale = "row",
         labCol = colnames(PP_diffs),
         ColSideColors = colsidecolorsLFC,
         ColSideWidth = 1,
         cexCol = 1.3,
         cexRow = 0.01,
         margins = c(7,7),
         xlab = "Datasets",
         ylab = "Genes",
         legendfun = function() showLegend(legend = c("TDP positive", "TDP negative",
                                                      "Spinal cord", "Frontal Cortex", "Muscle", "Cereb
                                                      "ALS", "FTLD", "IBMPFTD"),
                                           col = c("magenta4","magenta2",
                                                   "gray25","gray60","gray80","gray95",
                                                   "skyblue4","skyblue3","skyblue1")))
```


TDP positive
TDP negative
Spinal cord
Frontal Cortex
Muscle
Cerebellum
ALS
FTLD
IBMPFTD
Pathology
Tissue
Disease
Genes
SOD1
CBFTLD
VCP
sals
FUS
C9
FTLD
Datasets