Lab_book_12_2_16

Claire Green

Monday

I have now completed my second attempt at a literature review. This will be uploaded to the repository. Now that this is finished my next step is to look at creating a statistically robust protocol for generating lists of DEGs.

Notes for meeting with Wenbin

Data: Affy HG Plus 2.0 RNA expression (except FTLD - HG U133A 2.0)

C9orf72 motor neurons: 8 patients, 3 controls CHMP2B motor neurons: 3 patients, 6 controls sALS motor neurons: 7 patients, 3 controls FTLD cortical neurons: 16 patients, 8 controls VCP muscle: 7 patients, 3 controls

R script

1. Run each data set through Wenbin's code individually
2. Take top X most DEGs –> Do I rank by fold change, pvalue or adjusted p value?
3. Find genes that are consensus across all data sets –> which genes are commonly dysregulated?
4. Validate number is more than you would expect by chance by random permutations test

   - say top 2000 from each data set yields 10 common genes
   - take 5 sets of 2000 random genes
   - calculate consensus value
   - repeat 10,000 times
   - number of times value is >10 / 10,000 should equal p value (less than 0.05 is acceptable)

Tuesday

I discussed with Wenbin about the optimal way of ordering DEGs, and he essentially said that different people have different methods. I decided then to compare my output of DEGs ordered by adjusted p value as compared to fold change. Expression was calculated by WenBin DE Gene.R and sorted using the following lines of code:

```
###Write results to CSV files for consensus analysis
setwd("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/DE Genes/TopGenes_2016-02-09/")
# dir.create(paste("TopGenes", Sys.Date(), sep = "_"))
#Take results, remove duplicate rows for genes, order by adjusted p value and take top X number of gene
uniqueresult <- result[!duplicated(result[,15]),]

#For ordering by adjusted p value

genesort <- uniqueresult[order(uniqueresult$adj.P.Val),]
topgene <- genesort[1:1000,]
write.csv(x = topgene, file = "VCP_ap_1000")
topgene <- genesort[1:2000,]
```

```r
write.csv(x = topgene, file = "VCP_ap_2000")
topgene <- genesort[1:3000,]
write.csv(x = topgene, file = "VCP_ap_3000")
topgene <- genesort[1:4000,]
write.csv(x = topgene, file = "VCP_ap_4000")
topgene <- genesort[1:5000,]
write.csv(x = topgene, file = "VCP_ap_5000")

#For ordering by fold change
genesort <- uniqueresult[order(uniqueresult$`Fold Change`),]
topgene <- genesort[1:500,]

genesort <- uniqueresult[order(-uniqueresult$`Fold Change`),]
botgene <- genesort[1:500,]

topFC <- rbind(topgene,botgene)

write.csv(x = topgene, file = "VCP_fc_1000")
```

I evaluated consensus using the file DE consensus.R. It is as follows:

```r
#####DIFFERENTIAL GENE EXPRESSION INTERSECT
#takes csv files of top X DE genes and identifies any consensus genes

setwd("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Data/DE Genes/TopGenes_2016-02-09/")
A <- read.csv(file = "VCP_ap_1000")

B <- read.csv(file = "VCP_fc_1000")

# sALS<- read.csv(file = "sALS_anno_5000")
#
# FTLD<- read.csv(file = "FTLD_anno_5000")
#
# VCP<- read.csv(file = "VCP_anno_5000")


C9_DE<- A$Gene.Symbol
CHMP2B_DE <- B$Gene.Symbol
# sALS_DE <- sALS$Gene.Symbol
# FTLD_DE <- FTLD$Gene.Symbol
# VCP_DE <- VCP$Gene.Symbol
overlap <- Reduce(intersect, list(C9_DE, CHMP2B_DE))
print(overlap)
```

```
##   [1] "KNSTRN"
##   [2] "CEP72"
##   [3] "GAS2L2"
##   [4] "DTD1"
##   [5] "HAS3"
##   [6] "CNTNAP5"
##   [7] "ERVW-1"
##   [8] "FBXL18"
##   [9] "GPR26"
```

```
##  [10] "CCDC105"
##  [11] "ZDHHC19"
##  [12] "TTC12"
##  [13] "GPRIN2"
##  [14] "ELMO2"
##  [15] "SDC3"
##  [16] "WDR37"
##  [17] "PDZRN3"
##  [18] "FCRL5"
##  [19] "ARHGDIB"
##  [20] "C2orf72"
##  [21] "LEPR"
##  [22] "RP11-384P7.7"
##  [23] "LOC101927604"
##  [24] "LTBR"
##  [25] "LTB"
##  [26] "LOC728353"
##  [27] "LOC100131581"
##  [28] "SYT15"
##  [29] "LOC728099"
##  [30] "LOC101927537"
##  [31] "MRPL23"
##  [32] "LOC101929118"
##  [33] "CTSLP8"
##  [34] "LINC00700"
##  [35] "NUTM1"
##  [36] "LOC100287290"
##  [37] "RP11-36B6.1"
##  [38] "TROAP"
##  [39] "LOC101929631"
##  [40] "LOC101929372"
##  [41] "BIRC5"
##  [42] "TNFAIP2"
##  [43] "GPRC5A"
##  [44] "GFAP"
##  [45] "CYBB"
##  [46] "KLK6"
##  [47] "USH1C"
##  [48] "KAL1"
##  [49] "SLC22A3"
##  [50] "PPP2R2B"
##  [51] "KEL"
##  [52] "SPINK2"
##  [53] "CD19"
##  [54] "F2RL1"
##  [55] "CD5"
##  [56] "STATH"
##  [57] "NTSR2"
##  [58] "HCN4"
##  [59] "NEU3"
##  [60] "TSC22D3"
##  [61] "SH2D2A"
##  [62] "GUCA2B"
##  [63] "OPRK1"
```

```
##  [64] "SHOX"
##  [65] "PI15"
##  [66] "VPS45"
##  [67] "DUX1"
##  [68] "GLP1R"
##  [69] "OR2H2"
##  [70] "SCN10A"
##  [71] "NADK"
##  [72] "IGLC1"
##  [73] "GM2A"
##  [74] "KCNJ1"
##  [75] "FUT7"
##  [76] "HAP1"
##  [77] "IGHG1 /// IGHG2 /// IGHM /// IGHV4-31 /// MIR8071-1 /// MIR8071-2"
##  [78] "RLN1"
##  [79] "PLAC4"
##  [80] "IGLV1-44"
##  [81] "NCKIPSD"
##  [82] "AC007389.3"
##  [83] "UBBP1 /// UBBP1 /// UBBP4 /// UBBP4"
##  [84] "TSPY1"
##  [85] "IGLL5"
##  [86] "LOC100289473"
##  [87] "IGKV1OR2-108 /// IGKV1OR2-108"
##  [88] "BAG4"
##  [89] "DPEP3"
##  [90] "KLK12"
##  [91] "USP29"
##  [92] "SPANXA1 /// SPANXA2 /// SPANXB1 /// SPANXC"
##  [93] "IL1RAPL2"
##  [94] "ASIC4"
##  [95] "TECTA"
##  [96] "CLDN17"
##  [97] "OR10J1"
##  [98] "BAIAP2L1"
##  [99] "NMUR2"
## [100] "ACTR3BP2 /// ACTR3BP5"
## [101] "NOSTRIN"
## [102] "LYPD6B"
## [103] "AHRR"
## [104] "SHISA3"
## [105] "RP11-353N14.1"
## [106] "LOC100505515"
## [107] "TRIM78P"
## [108] "BC130595"
## [109] "TGM6"
## [110] "RP1-31B8.1"
## [111] "KRTAP4-2"
## [112] "GBP4"
## [113] "HHIP-AS1"
## [114] "LOC102723932"
## [115] "LINC01272"
## [116] "TPBGL"
## [117] "AF086288 /// RP11-768F21.1"
```

```
## [118] "LOC100507277"
## [119] "TMEM71"
## [120] "FIBCD1"
## [121] "ZSCAN12P1"
## [122] "LOC100505912"
## [123] "ARRDC3-AS1"
## [124] "C16orf78"
## [125] "RPL31 /// TBC1D8"
## [126] "ST8SIA6-AS1"
## [127] "RP3-476K8.3"
## [128] "RP11-466P24.7"
## [129] "OASL"
## [130] "TUBA3FP"
## [131] "ASMTL-AS1"
## [132] "CDC42BPG"
## [133] "SPRR4"
## [134] "LOC102725382"
## [135] "CARHSP1"
## [136] "ADORA3"
## [137] "IGLC1 /// IGLJ3 /// IGLV2-14 /// IGLV@"
## [138] "TMEM35"
## [139] "LOC100288152"
## [140] "GSDMC"
## [141] "IQCJ-SCHIP1-AS1"
## [142] "KIRREL3-AS3"
## [143] "SMPDL3B"
## [144] "KLK2"
## [145] "TMEM95"
## [146] "LOC100128840"
## [147] "BMP1"
## [148] "XYLB"
## [149] "IRF5"
## [150] "RDH16"
## [151] "LBX1"
## [152] "CYP2D6"
## [153] "BCOR"
## [154] "FAM115A"
## [155] "UBE2D1"
## [156] "LRRC43"
## [157] "SERPINB8"
## [158] "LOC728175"
## [159] "HRH3"
## [160] "MAPK12"
## [161] "WHAMMP2"
## [162] "LOC101928157"
## [163] "LOC100130456"
## [164] "IGFBP6"
## [165] "CSHL1"
## [166] "CHST7"
## [167] "GALNT10"
## [168] "HNF4A"
## [169] "TSPO2"
## [170] "LRRC2"
## [171] "CYP2S1"
```

```
## [172] "SPESP1"
## [173] "MYLK2"
## [174] "GS1-279B7.1"
## [175] "DEFB106A /// DEFB106B"
## [176] "OR2C1"
## [177] "PAX8"
## [178] "TIGD4"
## [179] "BCL6B"
## [180] "GRID1"
## [181] "LOC100507516"
## [182] "AX748294 /// RP11-1006G14.1"
## [183] "LINC01018"
## [184] "LINC01186"
## [185] "LOC151657"
## [186] "LOC101929584"
## [187] "FLJ33534"
## [188] "DTX2P1-UPK3BP1-PMS2P11"
## [189] "FXYD3"
## [190] "ALOX5"
## [191] "CYP2D6 /// CYP2D7P /// LOC101929829"
## [192] "AIF1"
## [193] "PRM2"
## [194] "USP53"
## [195] "DNAI1"
## [196] "A1CF"
## [197] "HHLA1"
## [198] "PVRL4"
## [199] "FAM167B"
## [200] "SMIM24"
## [201] "GPR133"
## [202] "OR51I1"
## [203] "SHROOM1"
## [204] "PRSS30P"
## [205] "AK130486"
## [206] "GAS8"
## [207] "GRIN2A"
## [208] "VPREB3"
## [209] "LOC101927071"
## [210] "PROL1"
## [211] "TMEM100"
## [212] "MYH15"
## [213] "KCNJ5"
## [214] "CEACAM21"
## [215] "HTR1F"
## [216] "PADI3"
## [217] "MORN5"
## [218] "LMNTD1"
## [219] "LOC100130278"
## [220] "TBATA"
## [221] "BEND7"
## [222] "FILIP1L"
## [223] "CYP4B1"
## [224] "PCOLCE-AS1"
## [225] "LOC100133669"
```

```
## [226] "LOC285638"
## [227] "LOC729173"
## [228] "KRTAP13-1"
## [229] "SETD5"
## [230] "RNF157"
## [231] "LOC646736"
## [232] "MIR6787 /// SLC16A3"
## [233] "PLAGL2"
## [234] "PIGR"
## [235] "HMGCS2"
## [236] "BAI3"
## [237] "LLGL1"
## [238] "BLK"
## [239] "KRT75"
## [240] "EGR4"
## [241] "IFNA14"
## [242] "FGF5"
## [243] "TLX3"
## [244] "CLCN7"
## [245] "CYP4F2"
## [246] "DENND3"
## [247] "KRTAP5-9"
## [248] "SSTR4"
## [249] "SYT2"
```