LabBook_21_04_2017

Claire Green

Monday/Tuesday

Whilst I wait for my meeting with Dennis to discuss the bigger picture aspect of my project, I decided to look into the pathprint story. I decided to really investigate the method I was using, particularly as I have to remove a dataset anyway.

After looking through the documentation, I realised I could use the consensusFingerprint function to generate an overall signature for a whole group. So what I decided to do (First, I'll get onto the amendments later) was the following:

1) Run pathprint indiviually on each of the four microarray datasets
2) Split each dataset into patients and controls
3) merge all patients into one group, and all controls into another
4) Run consensusFingerprint on patients and controls using a mean threshold of 0.75
5) Divide pathways into +1 and -1
6) If pathway is present in both patient and control (e.g. is +1 in patient and +1 in control) then it is removed

This is the code I was using (Consensus_pathprint.R):

```
load("~/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/Pathprinted.RData")
thresh = 0.5
#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)

#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)

#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))
```

The problem was threefold - firstly, different threshold produced wildly different results (i.e. completely different lists) I think because of the fact we look for overlap at the end. Secondly, in terms of my up and down genes, you get up genes enriched in down pathways and vice versa - not great for the story. Thirdly, as Win pointed out, if you use this method, you are making the assumption that 0 is unchanged, however, what I should be looking at is the relative difference. I.e. if control is 0 and patient is -1, that is downregulation. However if control is +1 and patient is 0, that is STILL downregulation. So I changed the code in the following way:

```
load("~/Documents/PhD/TDP-43/TDP-43_Code/Results/Pathprint/Pathprinted.RData")
thresh = 0.5
#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)


#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)

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


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

down_path <- subset(diff_consensus_PP, diff_consensus_PP$consensusFingerprint.pat_all_pp..thresh.
                    < diff_consensus_PP$consensusFingerprint.con_all_pp..thresh.)

down_pathways <- down_path$Row.names
```

```
up_path <- subset(diff_consensus_PP, diff_consensus_PP$consensusFingerprint.pat_all_pp..thresh.
                  > diff_consensus_PP$consensusFingerprint.con_all_pp..thresh.)

up_pathways <- up_path$Row.names

dyspathways_PP <- c(down_pathways, up_pathways)
```

The second part of the code takes the consensusfingerprint for each pathway for patients and controls, and then diff_consensus_PP is taking the rows in which the values are not equal. This is then split into down_path and up_path depending on the relative difference between the two groups. This produces the following pathway lists:

Upregulated pathways

{BRCA1,28} (Static Module)
{F2,46} (Static Module)
{FCGR2B,50} (Static Module)
{RB1,11} (Static Module)
{VCP,17} (Static Module)
ABC transporters (KEGG)
Alanine, aspartate and glutamate metabolism (KEGG)
Aminoacyl-tRNA biosynthesis (KEGG)
Basal transcription factors (KEGG)
Codeine and morphine metabolism (Wikipathways)
Endometrial cancer (KEGG)
Fat digestion and absorption (KEGG)
Fatty Acid Biosynthesis (Wikipathways)
Glutathione metabolism (KEGG)
Glycine, serine and threonine metabolism (KEGG)
Id Signaling Pathway (Wikipathways)
IL-2 down reg. targets (Netpath)
Metabolic pathways (KEGG)
Mitochondrial LC-Fatty Acid Beta-Oxidation (Wikipathways)
Muscle contraction (Reactome)
Phagosome (KEGG)
Vitamin B12 Metabolism (Wikipathways)
Vitamin D synthesis (Wikipathways)

Downregulated pathways

{ABL1,15} (Static Module)
{ACY1,11} (Static Module)
{ARRB2,743} (Static Module)
{CHRNA1,13} (Static Module)
{DVL1L1,17} (Static Module)
{HRAS,27} (Static Module)
{TCF3,20} (Static Module)
Aflatoxin B1 metabolism (Wikipathways)
Alzheimer's disease (KEGG)
Blood Clotting Cascade (Wikipathways)
Collecting duct acid secretion (KEGG)
Diabetes pathways (Reactome)

Estrogen signaling pathway (Wikipathways)
Glucocorticoid & Mineralcorticoid Metabolism (Wikipathways)
Influenza A (KEGG)
Phototransduction (KEGG)
Serotonin Receptor 4/6/7 and NR3C Signaling (Wikipathways)
Signaling by GPCR (Reactome)
Signaling by Insulin receptor (Reactome)
SNARE interactions in vesicular transport (KEGG)
Statin Pathway (Wikipathways)
Striated Muscle Contraction (Wikipathways)
Tyrosine metabolism (KEGG)
Urea cycle and metabolism of amino groups (Wikipathways)

When I run enrichment analysis on these two datasets, there isn't a huge amount but it does enrich in the way you would expect it to (upgenes enriched in uppathways and vice versa)

Down Pathway/Down Gene

ID	P-value	BHadjP-value	nGenes	nPathway
{ABL1,15} (Static Module)	1	1	0	15
{ACY1,11} (Static Module)	1	1	0	11
{ARRB2,743} (Static Module)	0.704416707	1	1	723
{CHRNA1,13} (Static Module)	1	1	0	13
{DVL1L1,17} (Static Module)	1	1	0	16
{HRAS,27} (Static Module)	1	1	0	21
{TCF3,20} (Static Module)	1	1	0	20
Aflatoxin B1 metabolism (Wikipathways)	1	1	0	8
Alzheimer's disease (KEGG)	0.002588759	0.031065107	3	170
Blood Clotting Cascade (Wikipathways)	1	1	0	24
Collecting duct acid secretion (KEGG)	1	1	0	27
Diabetes pathways (Reactome)	0.086603024	0.692824189	2	310
Estrogen signaling pathway (Wikipathways)	1	1	0	20
Glucocorticoid & Mineralcorticoid Metabolism (Wikipathways)	1	1	0	8
Influenza A (KEGG)	0.119562988	0.717377929	1	177
Phototransduction (KEGG)	1	1	0	32
Serotonin Receptor 4/6/7 and NR3C Signaling (Wikipathways)	1	1	0	18
Signaling by GPCR (Reactome)	0.368220243	1	3	912
Signaling by Insulin receptor (Reactome)	1	1	0	108
SNARE interactions in vesicular transport (KEGG)	1	1	0	36
Statin Pathway (Wikipathways)	1	1	0	29
Striated Muscle Contraction (Wikipathways)	1	1	0	38
Tyrosine metabolism (KEGG)	1	1	0	41
Urea cycle and metabolism of amino groups (Wikipathways)	0.002246715	0.031065107	1	21

. .

Down Pathway/Up Gene

ID	P-value	BHadjP-value	nGenes	nPathway
{ABL1,15} (Static Module)	0.023327525	0.139965151	1	15
{ACY1,11} (Static Module)	1	1	0	11
{ARRB2,743} (Static Module)	0.997297077	1	3	723
{CHRNA1,13} (Static Module)	1	1	0	13
{DVL1L1,17} (Static Module)	1	1	0	16
{HRAS,27} (Static Module)	0.004350774	0.104418576	2	21
{TCF3,20} (Static Module)	1	1	0	20
Aflatoxin B1 metabolism (Wikipathways)	1	1	0	8
Alzheimer's disease (KEGG)	0.289125168	0.991286291	3	170
Blood Clotting Cascade (Wikipathways)	1	1	0	24
Collecting duct acid secretion (KEGG)	1	1	0	27
Diabetes pathways (Reactome)	0.062280113	0.249120451	8	310
Estrogen signaling pathway (Wikipathways)	1	1	0	20
Glucocorticoid & Mineralcorticoid Metabolism (Wikipathways)	1	1	0	8
Influenza A (KEGG)	1	1	0	177
Phototransduction (KEGG)	1	1	0	32
Serotonin Receptor 4/6/7 and NR3C Signaling (Wikipathways)	0.032942058	0.158121879	1	18
Signaling by GPCR (Reactome)	0.958215261	1	8	912
Signaling by Insulin receptor (Reactome)	1	1	0	108
SNARE interactions in vesicular transport (KEGG)	0.019592171	0.139965151	2	36
Statin Pathway (Wikipathways)	0.010880629	0.130567547	2	29
Striated Muscle Contraction (Wikipathways)	1	1	0	38
Tyrosine metabolism (KEGG)	1	1	0	41
Urea cycle and metabolism of amino groups (Wikipathways)	1	1	0	21

Up Pathway/Up Gene

ID	P-value	BHadjP-value	nGenes	nPathway
{BRCA1,28} (Static Module)	0.008916476	0.068359646	2	27
{F2,46} (Static Module)	0.167356816	0.549886682	1	46
{FCGR2B,50} (Static Module)	1	1	0	14
{RB1,11} (Static Module)	1	1	0	11
{VCP,17} (Static Module)	0.002332058	0.02681867	2	17
ABC transporters (KEGG)	0.156065913	0.549886682	1	44
Alanine, aspartate and glutamate metabolism (KEGG)	1	1	0	32
Aminoacyl-tRNA biosynthesis (KEGG)	1	1	0	41
Basal transcription factors (KEGG)	1	1	0	45
Codeine and morphine metabolism (Wikipathways)	1	1	0	8
Endometrial cancer (KEGG)	1	1	0	53
Fat digestion and absorption (KEGG)	1	1	0	46
Fatty Acid Biosynthesis (Wikipathways)	1	1	0	22
Glutathione metabolism (KEGG)	1	1	0	50
Glycine, serine and threonine metabolism (KEGG)	1	1	0	32
Id Signaling Pathway (Wikipathways)	0.001423569	0.02681867	4	52
IL-2 down reg. targets (Netpath)	0.093700029	0.431020134	7	289
Metabolic pathways (KEGG)	0.368210814	1	19	1141
Mitochondrial LC-Fatty Acid Beta-Oxidation (Wikipathways)	1	1	0	16
Muscle contraction (Reactome)	0.0434919	0.250078427	2	49
Phagosome (KEGG)	0.447884856	1	2	154
Vitamin B12 Metabolism (Wikipathways)	1	1	0	52
Vitamin D synthesis (Wikipathways)	1	1	0	8

Up Pathway/Down Gene

ID	P-value	BHadjP-value	nGenes	nPathway
{BRCA1,28} (Static Module)	1	1	0	27
{F2,46} (Static Module)	1	1	0	46
{FCGR2B,50} (Static Module)	1	1	0	14
{RB1,11} (Static Module)	1	1	0	11
{VCP,17} (Static Module)	1	1	0	17
ABC transporters (KEGG)	1	1	0	44
Alanine, aspartate and glutamate metabolism (KEGG)	1	1	0	32
Aminoacyl-tRNA biosynthesis (KEGG)	1	1	0	41
Basal transcription factors (KEGG)	1	1	0	45
Codeine and morphine metabolism (Wikipathways)	1	1	0	8
Endometrial cancer (KEGG)	1	1	0	53
Fat digestion and absorption (KEGG)	1	1	0	46
Fatty Acid Biosynthesis (Wikipathways)	1	1	0	22
Glutathione metabolism (KEGG)	0.012309093	0.094369713	1	50
Glycine, serine and threonine metabolism (KEGG)	1	1	0	32
Id Signaling Pathway (Wikipathways)	1	1	0	52
IL-2 down reg. targets (Netpath)	0.002888097	0.053507943	4	289
Metabolic pathways (KEGG)	0.004652865	0.053507943	9	1141
Mitochondrial LC-Fatty Acid Beta-Oxidation (Wikipathways)	1	1	0	16
Muscle contraction (Reactome)	1	1	0	49
Phagosome (KEGG)	1	1	0	154
Vitamin B12 Metabolism (Wikipathways)	1	1	0	52
Vitamin D synthesis (Wikipathways)	1	1	0	8