Lab Book: 26/10/15 - 30/10/15

Claire Green

Monday

My aim was to try and incorporate the Ravit's sALS data into the data sets that have already been collected. I was able to download the .CEL files from GEO, but encountered two problems:

1. The array used (Affymetrix Human Exon 1.0 ST Array) does not appear to be supported by the puma & affy libraries, which meant that I could not used the pre-processing protocol used for the other data sets
2. The array type is not one of Pathprint's accepted platforms.

I still attempted to process the data but it was really difficult to find out how. I was able to read in the .CEL files, normalise the data using a package called oligo, and converted the signal values to expression values. So far I have:

```
setwd("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43 Data Sets/Ravits sALS MN/CEL files")

library(oligoClasses)
library(oligo)
library(limma)
library(affyio)
library(makecdfenv)

#locate cell files in directory
celFiles <- list.celfiles("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43 Data Sets/Ravits sALS MN/CEL f
#read these files into a variable
RavitsData <- read.celfiles(celFiles)

#normalise the data
eset <- rma(RavitsData, background=TRUE, normalize=TRUE, subset=NULL, target="core")

#extract expression values
exprs_eset <- exprs(eset)
```

however the expression values are much higher than those I am used to by using mmgmos. I think there may need to be a further step.

Tuesday

1) My first aim was to find out what the common differentially expressed pathways in patients without TDP-43 pathology were. The two datasets I currently have are SOD1 LCM motor neurons (3 patients, 7 controls) and FUS LCM motor neurons (3 patients, 3 controls). The process for identifying these pathways was identical to the protocol in TDP43_Signature. After pre-processing and running through Pathprint, and a threshold of the top 100 DE pathways, the common pathways were as follows:

```
 [1] "{SEPT2,21} (Static Module)"
 [2] "Pantothenate and CoA biosynthesis (KEGG)"
 [3] "Gap junction (KEGG)"
```

```
 [4] "Vitamin digestion and absorption (KEGG)"
 [5] "Interactions of the immunoglobulin superfamily (IgSF) member proteins (Reactome)"
 [6] "{AP1G1,16} (Static Module)"
 [7] "Chagas disease (American trypanosomiasis) (KEGG)"
 [8] "ACE Inhibitor Pathway (Wikipathways)"
 [9] "Muscle contraction (Reactome)"
[10] "{ACY1,11} (Static Module)"
[11] "Metabolism of xenobiotics by cytochrome P450 (KEGG)"
[12] "Sulfur relay system (KEGG)"
[13] "Heart Development (Wikipathways)"
[14] "ID up reg. targets (Netpath)"
[15] "Synthesis and degradation of ketone bodies (KEGG)"
[16] "Pathways in cancer (KEGG)"
[17] "Irinotecan Pathway (Wikipathways)"
[18] "SIDS Susceptibility Pathways (Wikipathways)"
[19] "beta-Alanine metabolism (KEGG)"
[20] "Propanoate metabolism (KEGG)"
[21] "Glycogen Metabolism (Wikipathways)"
[22] "Vitamin A and carotenoid metabolism (Wikipathways)"
[23] "{CBX4,10} (Static Module)"
[24] "{DVL1L1,17} (Static Module)"
[25] "{RYR2,15} (Static Module)"
[26] "Selenium Pathway (Wikipathways)"
[27] "Alanine, aspartate and glutamate metabolism (KEGG)"
[28] "Arachidonic acid metabolism (KEGG)"
[29] "alpha-Linolenic acid metabolism (KEGG)"
```

None of these pathways are found in the TDP43_Signature list, supporting the proposal that the differentially expressed pathways in my signature list are related to TDP-43 dysregulation rather than any other variable in ALS. However, I would like to increase the number of datasets involved in this experiment to gain confidence in this results.

2) Next I tried to identify more datasets however this is proving to be extremely difficult. Of the very few datasets that could be helpful, they are either from Affymetrix Human Exon 1.0 ST Arrays (which I am still struggling to incorporate) or the data has not been made publically available. Will need to discuss how to obtain more data (contacting those with?)

3) Finally, I tried a crude analysis of the ALS KEGG pathway in PCxN to identify if there were any correlations between my list of pathways. I found significant correlations for the following pathways:

1. Prion diseases (KEGG)
2. Collecting duct acid secretion (KEGG)
3. Urea cycle and metabolism of amino groups (Wikipathways)
4. {HSPA8,34} (Static Module)
5. Fructose and mannose metabolism (KEGG)

I take this list with a pinch of salt as it is becoming clear that there is no such thing as a singular "ALS pathway", therefore I will only use it to help prioritise my reading.