Meeting Notes: 3/5/16

Expansion of Consensus DEG set

It is likely that immediately outside my core set of DEGs, there are correlated genes that may also contribute to the disease process. This expansion can be done in many ways, but the optimal method is likely to be as follows:

1. Take each individual data set, and calculate correlation coefficient values between all genes and the DEGs
2. Using some criteria, select which genes will be added to the network. These criteria can include;
   1. A minimum p value
   2. A minimum coefficient value
   3. The number of data sets in which a significant correlation is evident

This may have to be adjusted during experimentation depending on the number of correlated genes that are identified

Enrichment in DEG network

Although the current DEG network (45 genes) is not enriched for GWAS SNPs or genes from John’s network 28, there are strong correlations. It could be that expanding the network might result in significant enrichment. This could either be as a whole, or after sub-networking using WGCNA.

In addition, I need to investigate looking for enrichment of eQTLs in this gene list. It is likely that the list of eQTLs required would ideally be tissue and disease specific. John has recommended I look into Sherlock for this.

My main aim for this week is to investigate eQTLs, in particular

* What methodologies have others used?
* How have the utilised eQTLs best in terms of a) identifying important functions in disease and b) prioritising candidate genes
* What is FANTOM, and how was it developed \*read paper\*
* Investigate ZENBU, in particular the C9orf72 gene
* Investigate the non-coding region that shares a promoter with C9orf72
  + What is it?
  + What does it contain?
  + Does this have any relevance to what I have discovered so far i.e. can it be related to TDP-43 pathology?
  + If this is unknown, how could I find out

Investigation of rewiring

To understand the way in which these genes are affected in disease, you can investigate the way in which the structure of the network changes between disease and background. This would involve building a coexpression network based on local data (i.e. my datasets) and comparing to a network built of non-disease data. Thoughts I have are:

* How to consolidate expression from different data sets (as I believe the direction of coexpression changes between certain data sets)
* What data set to build the control from – the control samples of my data sets or a global data set such as CoexprsDB (preferably something tissue specific, but we have multiple tissues)
* Potentially unlikely to be one network of disease – may have to develop one per data set and find a way of summarising results.

Investigating relationship of pathology-causing genes and DEGs

To further the relationship between DEGs and disease, I can investigate how related the network of genes surrounding known pathology-causing genes may overlap or at least correlate with my DEGs. I can do this quickly through GeneMANIA (Global coexpression), through my own data, or a global coexpression data set that is tissue specific (must research coexprsDB for this).

Working with David on an example for genometranslationcommons.org

1. Get an instance of toy.genometransltioncommons.org
2. Fill in the details for JCKs C9orf72 data set – with illustrative screenshots of the process with descriptions
3. Discuss a way of linking to GEO so that data is extractable through GTC