LabBook_10_06_16

Claire Green

Monday

I'm trying to conduct the same experiment again but just with the 2 RNA-seq data sets and the FTLD data set. Unfortunately the trick I used on friday to turn the RNA-seq matrices into numeric data frames didn't work (and I have no idea why). In my monday meeting with Win we talked about it and decided that it wasn't worth torturing the data to try and extract something that might not be that useful. Particularly as Pathprint does something very similar.

Tuesday

Today I worked on my Mini Thesis plan. I have written it up and so far it goes:

MINI THESIS PLAN - DEVELOPMENT OF AN IN SILICO MOLECULAR MODEL FOR TDP-43 PATHOLOGY

INTRODUCTION (MAJORITY FROM LIT REVIEW)

- TDP-43 structure and function
- TDP-43 associated diseases o ALS o FTLD o MSP o Alexander's disease o AD and Lewy body dementia o CTE
- TDP-43 dysfunction
- Project rationale o What is a model? Explain combination of 'omics and pathways → structure combined with function o Why a model? – multiple muts/tiss/disease etc. In silico is more dynamic. o How have other models of disease been used beneficially? o What new information can a model of TDP-43 pathology provide? I.e. new upstream mechanisms, drivers, targets for therapeutics o Brief overview of achievements?

MATERIALS AND METHODS

- Data – describe criteria for selection and provide table of information for each data set
- Normalisation and Quality Control – explain normalisation procedures for both microarray (rma) and RNA-seq (bcbio). Describe QC process for microarray (PCA, hierarchical clustering, boxplots) and RNA-seq (bcbio)
- Pathprinting – explain process as depicted by the code. Describe all parameters used and justifications. Explain use of microarray and not RNA-seq.
- Differential Expression Analysis – same as above. Limma, thresholds etc.
- Enrichment – Provide table of gene lists and sources (full list in appendix). Describe hyperpathway code and parameters.

RESULTS

- Quality Control – produce examples of PCA plots, dendrograms and box plots. Explain why CHMP2B control was removed

- Pathprint – briefly introduce the concept of Pathprinting. Show table containing list of pathways. Show PCxN diagram of interaction. Do I need to show it's more inter-correlated than expected by chance? Would have to work out how to access PCxN data and run random permutation tests
- Differential Expression Analysis – Show results at differing thresholds, as well as results of random permutation tests. Explain why the threshold of 5500 was chosen. Show Genemania results of correlation and physical interaction. Again, does this need RPTs? Discuss presence of TARDBP and other associations with TDP-43 Path. Diseases.
- Enrichment Analysis – Describe concept of what enrichment suggests, e.g. enrichment of SNP containing genes suggests upstream processes, direct association with known pathogenic contributors, etc. Present results of enrichment in DEG lists and pathways.

DISCUSSION

- Brief overview of project aim
- Brief overview of main results
- Explain significance of common pathways – discuss the fact that pathways that you would expect are present (e.g. prion diseases) and new ones might be implicated - discuss the enrichment results and how that relates certain pathways to disease (particularly lysine degradation)
- Explain why DEGs did not enrich as well (biological bias towards certain genes/processes, still unknown variants, could be detecting downstream etc) but discuss enrichment of TDP-43 interactors.
- Discuss pitfalls – low sample sizes, no TARDBP mutations. Explain need for properly controlled sampling from in vitro/vivo experiments
- Future plans – cross-validate in external data set.