

LabBook_26_02_16

Claire Green

Monday

Since the meeting with Jiantao about EdgeRun is not until tomorrow, I decided to spend today trying to run gene expression analysis on the RNA seq data that I have found. Below are the details:

Cell Type	Variant	Pat	Con	Platform	GEO Ref	Contributors
Frontal Cortex (Homo-mogenised)	C9orf72 & sALS	8 C9 & 10 sALS	9	Illumina HiSeq 2000	GSE67196	Prudencio M, Belzil VV, Batra R, Ross CA et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat Neurosci 2015 Aug;18(8):1175-82. PMID: 26192745
Spinal Chord Motor Neurons (LCM)	sALS	13	9	Illumina Genome Analyzer II (Homo sapiens)	GSE76220	Gene Expression Signatures of Sporadic ALS Motor Neuron Populations Ranjan Batra, Kasey Hutt, Anthony Vu, Stuart J Rabin, Michael W Baughn, Ryan T Libby, Shawn Hoon, John Ravits, Gene W Yeo (Under review)

The data for these two data sets comes in both an unprocessed RAW format and as a pre-processed count matrix text file. Each of the samples has two reads, totalling about 6GB of data per sample. Downloading this kind of data over the uni wifi was nigh on impossible. Talking to Sandeep, he said the best way is to use Iceberg - Sheffield's HPC. I am not really familiar with how to use the bash shell to use this and BCBio to do the analysis, so I will need Sandeep to run through with me how to do this. Since I am on holiday at the end of this week, I will probably leave the raw data stuff until next week. In the mean time, Win said that I should use the pre-processed stuff as I can get quick answers through this (and our sources are trustworthy enough that we believe they have processed their data properly).

Tuesday