RESEARCH

Identify epigenetic alterations associated with Alzheimer's disease and classification of gene expressions between healthy and sick patients Project Week 1 - Group 02

Pushpa Koirala, Carmen Calle Huerta, Christina Kirschbaum and Melika Moradi

Full list of author information is available at the end of the article

1 Summary of the week

1.1 What did we do this week?

This week, all of us get more familiar with the pipelines and analysis methods that were used in our main reference paper. Additionally, more research about Alzheimer's disease was done, to get a deeper insight into the biochemical background, like the relevant pathways.

The ChIP-seq and the RNA-seq data were acquired from GEO. For the RNA-seq data, the preprocessing with STAR was started.

The Quality Control was delayed, because the .sam files for RNA-seq was not there and needed to be created.

Member	Last weeks work
Christina	Data Integration
Pushpa	Research on Quality Control for ChIP-seq, prepare Quality Control
Melika	Research on Quality Control for RNA-seq, prepare Quality Control
Carmen	In-depth research about biochemical background of Alzheimer's disease Python tutorials

1.2 What was planned?

Initially, we decided to hold a kick-off meeting for the project, during which we discussed the data. Despite the fact that we only have 300 samples from 30 patients, the amount of data contained within them is enormous. As a result, we've opted to employ these two approaches.

- We will split the Chip-seq data into four parts depending on the histone modification they belong to.
- we will maybe use university server to store the data.

We need to reach a point where we have a clean dataset, that can be used to analyse ChIP-seq and RNA-seq data. We expect that for the ChIP-seq data since this is preprocessed data. For the RNA-seq data there was only the raw data accessible and the preprocessing will be done using STAR with default parameters to align our RNA-seq reads to the human reference genome (assembly GRCh37.75/hg19) like in the reference paper.

1.3 Motive of this week

Overall, this week we acquired in-depth our knowledge about Alzheimer's disease and methods we want to use, as well as prepared our data and setup our GitHub repository.

1.4 Research

1.4.1 What must be done for Chip-seq quality control?

- number of reads should be noted down
- high peaks, low background
- sequencing depth
- duplicates
- low enrichment in control
- similarity of replicates

1.4.2 Tools for Quality Control

ChIP-seq	RNA-seq
ChipQC	FastQC
SPP package	FastX
Guidelines & practices of the ENCODE	TagCleaner
& modENCODE consortia	PRINSEQ

The most used tool for RNA-seq is FastQC. In our case, we have to do the quality control for each sequence. What we will do is that we will take every seq working out the mean score across all the bases in that particular seq and the plotting out the distribution of those means.

- Good quality: All seq from one very tight distribution with universally high quality and no seq with low quality.

- Bad quality: Keep the set of seq that always have universally high quality and remove the set of lower quality seq.

2 Plans for the upcoming week

- Quality Control for RNA-seq with FastQC
- Quality Control for ChIP-seq with a tool chosen from the research results
- Start of the Data Exploration