DNAme in iPSC-derived and Post-mortem neurons

TA:
Anna Kononkova

Students:
Aleksandra Gorbunova
Veronika Karpushenkova
Adelere Bamiwo

Goal: to compare DNAme profiles in iPSC-derived and post-mortem neurons and try to establish interconnection with differences in transcription, H3K9me3 modification and chromatin compartments.

Motivation: iPSC-derived neurons differ from post-mortem ones in several features (transcription, epigenetic and chromatin).
What about methylation in CpG and non-CpG context (which is a characteristic of neuronal cell)?

Description of a dataset:

DNAme files (.bam)	iPSC-derived neurons (WT data): GSE230715 (3 replicates) Post-mortem: GSE96615 (3 replicates)
RNA-seq files (salmon, transcriptome - gencode v46 or Ensembl v114)	iPSC-derived: GSE212252 (3 replicates) Post-mortem: GSE96615 (3 replicates)
H3K9me3 ChIP-seq. Signal and input files for merged replicates in .bw.	iPSC-derived: GSE196109 (2 replicates) Post-mortem: GSE211871 (3 replicates)
Hi-C (.cool files, merged, 1 file for post-mortem and 1 for iPSC-derived)	Post-mortem: PMC8233376 iPSC-derived: GSE212252

Methods

<u>DNAme analysis</u>
Having aligned DNAme data, we are going to extract methylation from deduplicated bam files with bismark (bismark_methylation_extractor) and then convert the information in bigwig format as well as to conduct DM analysis providing bismark output to methylKit to obtain CpG and non-CpG contexts.
We would like to convert .bam to .bw for visualization in IGV and further analysis.
We will use methylKit to do differential methylation analysis. Packages biomaRt, and org.Hs.eg.db will be used for annotation and enrichment analysis.

<u>RNAseq analysis</u>
The RNA-seq data will be used to find a correlation between the level of gene expression and methylation patterns. We will use DESeq2 to analyze differential expression, basic

method of normalisation will be used (median-of-ratios). But before that, it will be necessary to make an annotation. Since we have salmon files ready, it will be necessary to match the transcript with the corresponding gene.
For clusterization and further visualisation we will do regularized log normalisation to minimise the noise and stabilize dispersion.

Hi-C analysis
For Hi-C data, we already have files with processed data in cool format. We will need to do several files with different resolutions to work with TADs and compartments. So, for TADs it will be 5-25 kb, for compartents - 250-500 kb. For further analysis, we will use cooler (show, info) for normalization, if necessary. HiCExplorer (hicFindTADs and hicPCA) will be used to search for TADs, and different compartments, and Chromosight will be used to search for loops (and additionally, if time is available, a deeper analysis will be performed with the separation of TADs by size).
In order for the search to be more efficient, we specify the following parameters: --threshold Comparisons, --data and --correctForMultipleTesting fdr.
HiGlass and JuiceBox will be used to visualize the results.
Since we will have the methylation data in bigwig format, we will be able to associate them with Hi-C data and subsequently analyze their correlation with the 3D organization of the genome.

H3K9me3 ChIP-seq analysis
H3K9me3 is tightly linked with DNAme, therefore, we will use the following tools to analyze this correlation:
- deeptools for merging and visualization of enrichment H3K9me3 and DNAme in TADs
- computeMatrix and plotHeatmap to plot signal (.bw) around genes

Areas of responsibilities

Student 1: DNAme analysis: (Aleksandra)
- extract methylation data
- perform differential methylation analysis
- convert methylation data to .bw
- analyze methylation levels in A/B compartments and TADs

Student 2: Hi-C analysis: (Bamiwo)
- process Hi-C .cool files to identify compartments and TADs at 50kb and 100kb, and then check the strength of the compartment.
- classify TADs by size and provide genomic coordinates for integration

Student 3: RNA-seq and H3K9me3 ChIP-seq analysis: (Veronika)
- normalize RNA-seq data, identify DEGs, and correlate expression with methylation changes
- quantify H3K9me3 enrichment in regions of interest and generate visualizations
- integrate histone modification data with methylation and expression results.

Expected results and timeline

By 10th of May: each participant provides a report on their analysis with code, comments and suggestions. By this time, the necessary part of the methylation analysis for all further calculations should be completed.
By 13th of May: intermediate report is provided with combined obtained results and considerations.
By 20th of May: each participant provides a final report on their analysis.
By 22th of May: final report is provided with combined results and conclusions.