Babraham Bioinformatics

Exercises: Analysing High Throughput Sequencing Data with SeqMonk

Version 2023-06

Licence

Exercise 1: Creating a Project and Importing Data

- Create a new project based on the *Mus musculus* GRCm38_v100 genome.
- Use the BAM/SAM import to read in the GSM307618.bam and GSM307619.bam files - extending each read by 250bp.

- Right click on the data sets in the data view to rename the GSM307618 DataSet to ES.H3K4me3 and GSM307619 to ES.H3K27me3 – note that the names also update in the chromosome view.

- Change the annotation tracks in the chromosome view to just show gene and CpG island features
- Change the view of the data so that the reads are packed with “Medium” density so you can see more of them.

- Create a data group called “Everything” which contains the reads from both of the imported datasets and add this track to the chromosome view.

- Try out the movement and visualisation controls in the chromosome view
    - Try zooming in and out of some sequence clusters to see how they look
    - Move to some different chromosomes
    - Find the Xist gene and look at the data around it

Exercise 2: Quantitation and Plotting

- Generate a probe set of promoter probes. This will use the Feature Probe Generator. Make probes **upstream** of all **gene** features from **-1000bp** to **+1000bp**. Call the new probe set “Promoters”

- Do a simple read count quantitation using default values, but **without** log transforming.

- Look at the probes and check they are appearing in the right place in the genome.
- Double click on the probe set in the data view and read the description of the probe generation and quantitation to check you did what you intended.

- Adjust the scaling of the quantitations to an appropriate level.

- Select the K4 data set in the Data View and have a look at how the quantitations spread out over the whole genome in the genome view.

- Select Plotting > Probe Value Histogram to look at the distribution of values in the K4 data. You can drag a box within this plot to zoom in on parts of the quantitation range (right click to zoom out again).

- Go back to Data > Quantitate existing probes and requantitate the data as before, but this time use a log transformed read count.

- View the modified histogram to see if it's clearer.

- Draw a scatterplot of the K4 vs the K27 datasets to see how they compare. Double click on interesting looking points to change the chromosome view to show that region of the genome.

Exercise 3: Filtering

- Look at the probe value histogram to pick a cut-off value which will filter out only promoters showing enrichment in H3K4me3 (those in the right half of the bimodal distribution).

- Use Filtering > Filter by Values > Individual probes to create a probe list from just probes which have higher values in K4 than the cutoff you selected. Call the list “High in K4”

- Find the new list in the data view and look at the description to check it matches what you intended to do.

- Select the list and draw a scatterplot again. Note that only the probes in the filtered list are used.

- Draw a beanplot of all of the data tracks for both the Promoters and High in K4 lists.

- Filter from this enriched list promoters which do and don’t overlap with CpG islands. Draw trend plots for both of these and see if there’s a difference.

- Select the “High in K4” list then create a new list using Filtering > Filter by position to separate out the probes on the mitochondrion (MT)

- Draw a scatterplot of the High in K4 probes, then use the button at the top to highlight the Mitochrondrial probes on the plot.

Exercise 4: Saving things

- Save your seqmonk project to your home directory. Call it “course_example.smk”
- Draw a bean plot of all probes for all data stores and then save the plot as both a PNG and an SVG file.
- Find the “nanog” gene and export a picture of the chromosome view for the region around that gene (File > Export Current View > Chromosome View)
- Generate a Data Store Summary Report and save it to a text file.

- Have a look at your vistory and see if you can track all of the events it has recorded.
- Put a title and some explanatory text into the vistory to describe what you’ve done.
- Try embedding some plots into the vistory so you know how that works.
- Save your Data Store Summary Report into the vistory.
- Save the vistory and export it as an HTML file.