## Explore Roadmap epigenomes

In this activity we are going to compare whole epigenomes derived from multiple epigenetics modifications and from samples from multiple tissues. The goal is to cluster samples based on their epigenome and see if they inform about the tissue of origin.

Download bed mnenomics files for the 18 states chromatin whole genome segmentation from Roadmap portal:

https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#exp_18state

We are not going to work with all samples, but only select those related to the following tissues:

Brain -- Muscle – Heart -- Digestive

In orther to retrive this information read the file: metadata.roadmap_clean.txt

Create a column names PERIOD to distinguish Fetal and Non-fetal samples.

Create a new column with unique names incorporating the tissue and PERIOD information.

---

For a set of at least 2 of these chromatin states: active enhancers, active promoters, heterocromatin and Polycomb repressed regions:

1) Calculate pairwise Jaccard index

- Use Bedtools in R (BEDtoolsR.R)

2) Visualize Jaccard Index matrix in a heatmap, adding color blocs in columns indicating tissue of origin and PERIOD (use Package pheatmap)

3) Report jaccard index between sample E071 and sample E074.

4) Perform muldimensional scaling on a distance matrix based on 1-jaccardIndex.

5) Plot 1:2 and 3:4 dimensions and color by tissue, shape by fetal/adult.

6) Compute hierarchical clustering in jaccard distance matrix with R function "hclust", cut the dendrogram at different Ks using R function "cutree" and plot the dendrogram coloring the obtained clusters of samples. (You can use the R Package dendextend)

Do you see a logic in the clusters using epigenomes?

Which major sub-divisions can you see?

All states convey similar information?

Report this activity using an RMarkdown document explaing all steps and including your batches of code and figures. Transform the final document into pdf.