FEATURE SELECTION

1. Study the Hist1 region of the mouse genome in UCSC genome browser
2. Characterize your **k** clusters according to genomic and epigenomic features:
   1. Features:
      1. Histone genes
      2. Lamina associated domains (LADs)

The Lamina is on the perimeter of the Nucleus

The genome moves around dynamically

Certain regions of the chromosome associate with the **Lamina**

* 1. Feature table
  2. Extend your program to read the feature table and determine if there is there a pattern of features in each cluster. For each cluster do the following:
     1. Calculate the percentage of windows in each NP that contain histone genes (use the **Hist1** column in the feature table).
     2. Show the result of step **i**. in a **boxplot**.
     3. Calculate the percentage of windows in each NP that contain LADs (use the **LAD** column in the feature table).
     4. Show the result of step **iii**. in a **boxplot**.
  3. Do the features allow you to discriminate between the clusters?
     1. Compare the **k** boxplots for the **Hist1** feature.
     2. Compare the **k** boxplots for the **LAD** feature.
     3. What might be the possible meaning of your findings, in the domain of biology?

1. Extend your program to characterize each cluster by radial position. Consider the radial positions of the NPs in the clusters. Is there a pattern? For each cluster do the following:
   1. Consider the previously calculated radial position for each NP, which is an integer in the range [1-5] (1 – strongly apical; 2 – somewhat apical; 3 – neither apical nor equatorial; 4 – somewhat equatorial; and 5 – strongly equatorial).
   2. Calculate the percentage of NPs in each radial position category.
   3. Show the result of step **b**. in a bar graph
      1. show five bars – one bar for each radial position category.
      2. the height of a bar is the number of NPs that were classified as being in a particular radial position category.
   4. Do the clusters differ in terms of the radial positions of the NPs that they contain?
      1. Compare the **k** bar graphs for radial position.
      2. What might be the possible meaning of your findings, in the domain of biology?

# Grading for today’s in class activities

# workon project-related activities assigned in today’s class

# Next Class

* *demo quiz***:** demonstrate completed project-related activities assigned in today’s class