COMP364: Working with Matplotlib

Jérôme Waldispühl, McGill University

Analysis of human mRNA

Using CURL download human mRNA file on your account from:

http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/refMrna.fa.gz

Unzip the file unzip `gunzip` : `gunzip refMrna.fa.gz`

Write a parser that reads the file and store the mRNA into a dictionary such that:
- entry : sequence id
- content : sequence in IUPAC code

Using matplotlib, create histograms that plots:
- An histogram of the distribution of mRNA lengths.
- A figure that plots the GC-content (x-axis) and AG-content (y-axis) of each mRNA.
- Same as above but length vs. GC-content

Useful functions

Count the occurrence of one letter: <string>.count(<letter>)

Save a figure: .savefig(<filename>)

Example:

fig = plt.figure()

fig.plot(X,Y,'r--')

fig.savefig('myfigure.pdf',transparent=True)