CSE527 Homework 3

Anthony Adkins ([anadkins@uw.edu](mailto:anadkins@uw.edu))

# 1.1 Most significantly expressed genes per cell type

Cell Type : Astrocytes

1. Aqp4

2. Acsbg1

3. Lcat

Cell Type : CA1 Pyrimidal

1. Crym

2. Cpne6

3. Neurod6

Cell Type : Endothelial

1. Ly6c1

2. Cldn5

3. Flt1

Cell Type : Ependymal

1. 1700001C02Rik

2. Dynlrb2

3. Tmem212

Cell Type : Interneurons

1. Gad1

2. Gad2

3. Slc32a1

Cell Type : Microglia

1. C1qb

2. C1qa

3. Fcgr3

Cell Type : Mural

1. Acta2

2. Tagln

3. Tpm2

Cell Type : Oligodentrocytes

1. Mog

2. Ugt8a

3. Mobp

Cell Type : S1 Pyrimidal

1. Gm11549

2. Mef2c

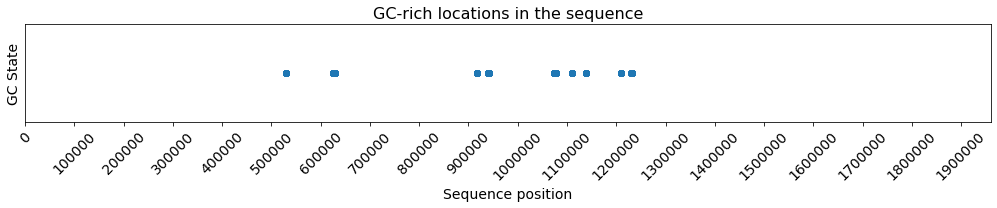
3. Car10

# 2.1 Viterbi decoding

## 2.1.1 Joint log likelihood

P(e1, ..., eT , s1, ..., sT ) = -2640006.40

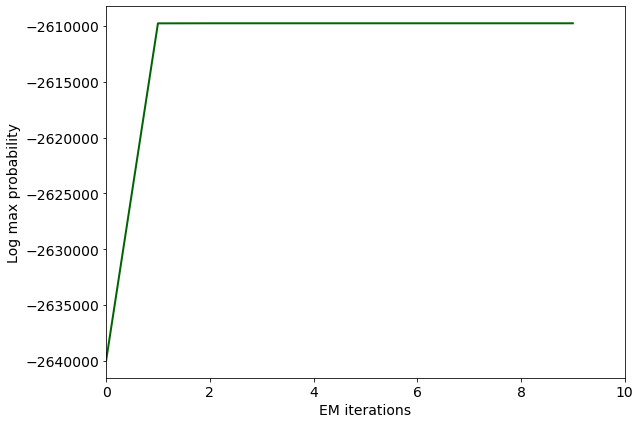
## 2.1.2 Plot G+C rich regions



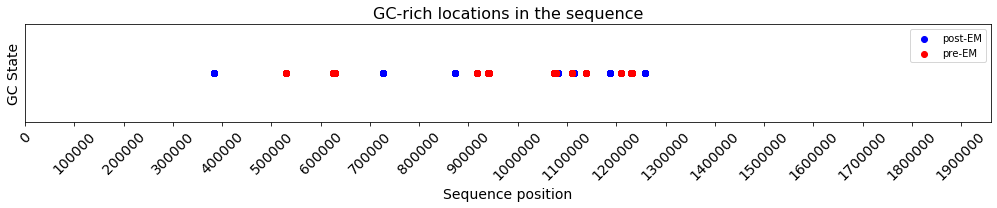
# 2.2 HMM Parameter Re-estimation

## 2.2.1 Plot EM iteration vs Log-likelihood

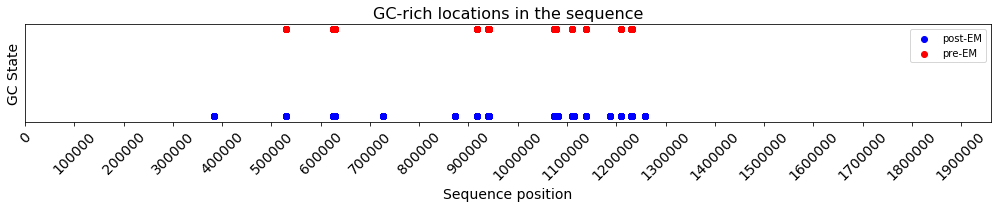
**Log-likelihood after each EM iteration**



## 2.2.2 New G+C rich regions



Or to put them so that they are not overlayed:



After EM, I did detect more G+C-rich regions. Before, there were 10,961 GC states reported and afterwards there were 14,181. This ~30% increase is a significant increase and it is visible in the above graphs. Also, the log max likelihood increased from -2,640,006 to -