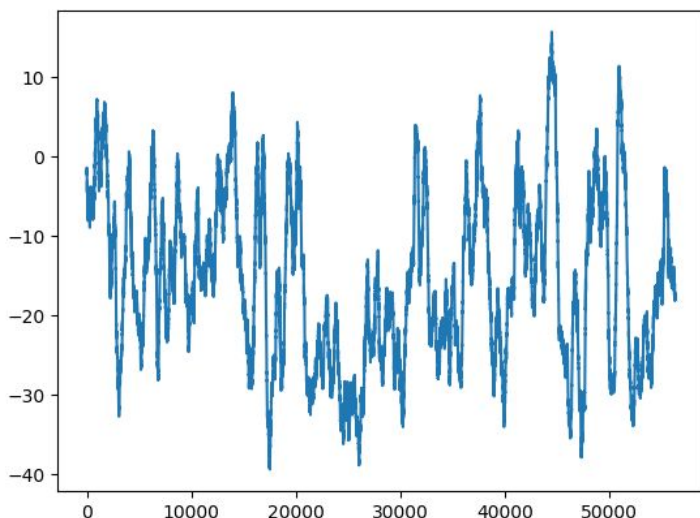


Ergebnisse der Studienleistung ALBI

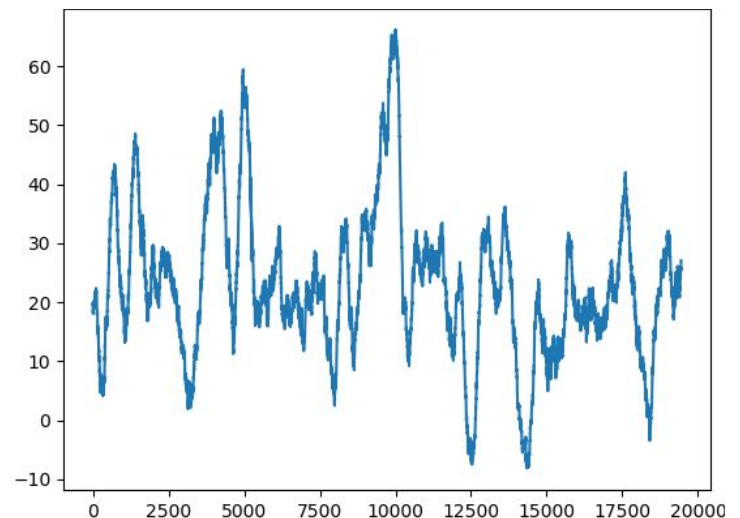
SoSe 2019, Constantin Zackl

Im Rahmen der Studienleistung habe ich mit den von mir geschriebenen Programmen menschliche mitochondriale DNA Sequenzen (**NC_012920**, **GenBank NCBI**) untersucht und die folgenden Plots erstellt. Dabei sind für die Fenstergröße 401 jeweils die log likelihood ratios für die gegebenen Background und CpG Sequenzdateien und für die mitochondriale Sequenz dargestellt.

Background



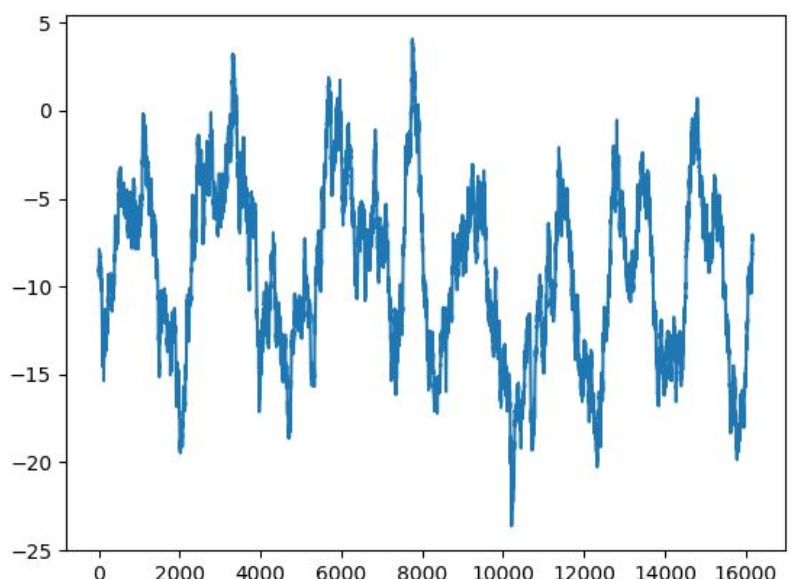
CpG Islands



Auffällig sind die wesentlich höheren Werte im CpG Islands Plot. Die Werte werden kaum negativ. Auch sind klare einzelne Spitzen nach oben zu erkennen während die Background Datei eher einem Hintergrundrauschen gleicht. Ein Wert über null bedeutet eigentlich, dass es wahrscheinlich ist sich in einer CpG Insel zu befinden. Die Background Datei zeigt jedoch auch, dass nur positive Werte alleine nicht ausreichen können um eine CpG Insel klar abzugrenzen.

Mitochondrium:

Betrachtet man den entsprechenden Plot für das menschliche Mitochondrium fällt besonders auf, dass die Log Likelihood Ratios eher der Background Datei gleichen. Die Werte gehen an einigen Stellen über die Null hinaus (Sprich: CpG Insel wahrscheinlich). Jedoch sind die Werte nicht so hoch wie in der CpG Insel (Plus Modell) Datei. Ruft man sich die Definition von CpG Inseln

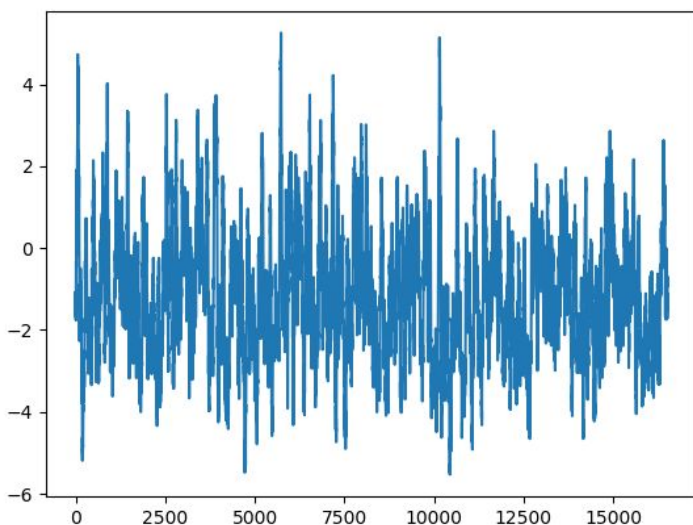


ins Gedächtnis wird dennoch der Unterschied zwischen den oben gezeigten Plots klar. CpG Inseln sind Bereiche mit statistisch höher CpG Zahl, auch in der Background Datei können CG Dinukleotide vorkommen, der sehr sprunghafte Plot deutet aber eher auf einzelnes auftreten dieser hin. Im Vergleich dazu deuten die breiteren, klareren Spitzen im CpG Plot auf größere Bereiche hin. Schaut man sich jetzt den Plot des Mitochondriums an erkennt man trotz der niedrigeren Werte eher breitere Spitzen deren Werte an der höchsten Stelle trotzdem positiv werden. Am auffälligsten ist die ganz klare Spitze zwischen 7000 und 8000 Basenpaaren mit dem höchsten Ausschlag. Meiner Meinung nach enthalten menschliche Mitochondrien CpG Inseln in Bereichen zwischen 3000 und 4000, um die 6000 und zwischen 7000 und 8000 Basenpaaren. Letztendlich bleibt der Test ein statistisches Modell. Auch die Background Datei enthält schließlich CG Dinukleotide. Die Verteilung scheint nur eher zufällig zu sein. Bereiche in denen die Scores positiv werden lassen sich dort also nicht ausschließen.

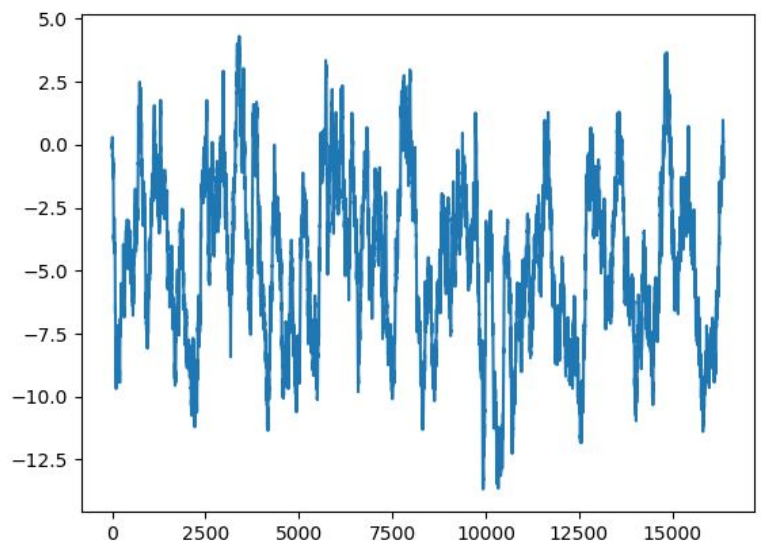
Zur klaren Abgrenzung von CpG Inseln wäre die Definition eines Schwellwertes hilfreich der sich mit Hilfe der Plots erstellen ließe. Dafür müssten die genauen Grenzen jedoch bekannt sein.

Im folgenden noch zwei Diagramme um kleinere Fenstergrößen beispielhaft am Mitochondrium zu betrachten:

Fenster Größe 51:



Fenster Größe 201:



Betrachtet man alle Plots des Mitochondriums fällt auf, dass sich effektiv nur für den Plot der Fenstergröße 401 Aussagen zu CpG Islands treffen lassen, die kleinere Fenstergröße 51 gleicht eher einem Hintergrundrauschen. Am effektivsten wäre ein Scan mit einer Fenstergröße die genau der gesuchten Island Größe entspricht. In einem Plot würde sich dies vermutlich als schmale und hohe Spitze zeigen, eine zu kleine Fenstergröße sorgt vermutlich eher für einen Plateaubereich, während ein zu großes Fenster den Ausschlag ebenfalls dämpft. Mit vielen Scans unterschiedlicher Fenstergrößen könnte man die wahrscheinlichste Islandgröße also theoretisch annähern. Auch der höchste Ausschlag der Spitzen mit verschiedenen Fenstergrößen könnte bei der Bewertung der tatsächlichen CpG Island Größe hilfreich sein.