

Bioinformatik Übung 4:

Aufgabe 2:

Das Programm „rGADEM“ untersucht ein Set aus mehreren DNA-Strängen nach prädiktiven Motive der Länge k ab, mithilfe der die Konsensus-Sequenz gebildet werden kann. Die Konsensus-Sequenz wird in einem Sequenz-Logo-Diagramm dargestellt, in dem auch die Basen angezeigt werden, die auch mit einer geringeren Wahrscheinlichkeit als die Basen der Konsensus-Sequenz an den einzelnen Positionen vorkommen können.

„rGADEM“ sucht zuerst nach überrepräsentierten Basenabfolgen einer bestimmten Länge in der DNA-Sequenz. Diese Motive werden anschließend in einer Matrix dargestellt, in der die Gewichtung der vier Basen an den verschiedenen Positionen vermerkt wird. Danach verwendet das Programm verschiedene Algorithmen, um die gefundenen Motive zu verbessern. Aus den überarbeiteten Motiven wird die Konsensus-Sequenz gebildet.

Aufgabe 3:

Das mit dem rGADEM Paket erhaltene Sequenz Diagramm weicht stark von den Ergebnissen aus M1 ab. Die beiden Diagramme sind an keiner Position gleich. Außerdem stellt das Diagramm aus M1 eine Sequenz aus 15 Basen dar und das mit dem rGADEM erhaltene nur eine aus 11 Basen.

Für die Ergebnisse vom „FOXA1“ wurde als erstes ein ChIP-seq Experiment durchgeführt, bei dem die FOXA1-DNA-Interaktion untersucht wurde. Die Sequenzen der 500 TF ChIP-seq Peaks wurden dann genutzt, um Motive in der DNA-Sequenz zu finden. Eines der so gefundenen Motive entspricht M1.

Die DNA-Sequenz in der Datei „Test_100.fasta“ besteht aus 49 Strings ($t=49$), die jeweils aus 202 Basen bestehen.