

## Aufgabe 2

- *Führen Sie das Programm aus und erklären Sie kurz, was das Programm macht und welches Ergebnis es erzeugt.*

Das Programm liest mehrere gleich lange Sequenzen, findet darin häufig vorkommende Motive, mit minimaler Abweichung voneinander. Findet damit die höchste Wahrscheinlichkeit, dass eine bestimmte Base an einer Position des Motivs vorhanden ist, und erstellt aus den Daten einen Plot, der die Wahrscheinlichkeiten grafisch darstellt.

## Aufgabe 3

- *Vergleichen Sie ihr Programm-Ergebnis (qualitativ) mit dem M1 Ergebnissen des Abschnittes „Motifs Enriched in the Top 500 ChIP-seq Peaks“ auf der Seite: <http://v1.factorbook.org/mediawiki/index.php/FOXA1> Erklären Sie dabei kurz, wie das Ergebnis auf dieser Seite erzeugt wurden.*

Die Länge der Sequenz ist in der Online Ressource um 4 Basen Länger. Die Häufigkeiten bzw. die Abfolge der Sequenz stimmen an keiner Position eindeutig überein.

Das Ergebnis auf der Seite wurde durch eine Analyse von 500 Sequenzen, die potentielle Peaks für die Motive enthalten. Anhand von den Motiven bzw. Consensus Sequenzen werden dann die Sequenzen nach Häufigkeit durchsucht und sortiert. 499/500 zeigt die Anzahl der Peaks, die das Motiv enthalten.

- *Erklären Sie kurz, welche Art von Sequenzen in der in ihrem Programm verwendeten Datei „Test\_100.fasta“ enthalten sein müssen.*

DNA-Sequenzen, aus dem gleichen Genom mit möglichst gleicher Länge, die alle eine hohe Chance aufweisen, den gesuchten Transkriptionsfaktoren bzw. das gesuchte Motiv zu enthalten.

## Aufgabe 4

- *Erklären Sie kurz das Resultat.*

Das Programm zeigt das theoretisch erwartete Ergebnis. Die Consensus Sequenz entspricht der in der Vorlesung gezeigten Sequenz. Die Positionen mit der größten Genauigkeit sind am höchstem im Plot erschienen.