

## Algorithmische Bioinformatik Übungsblatt 6

Ausgabe: 10. Dezember 2019 · Besprechung: 17. Dezember 2019

**Aufgabe 6.1** Gegeben ist die Zählmatrix zur PWM des Bindemotivs von SOX-Transkriptionsfaktoren (Abbildung 1A), erstellt aus 31 experimentell verifizierten Bindestellen in Knorpelgewebe. Als Hintergrundverteilung werde die Gleichverteilung angenommen.

1. Berechne die Profilmatrix (ein Pseudocount an jeder Stelle) und daraus die PWM in Zehntelbits, ganzzahlig gerundet. Welcher Score ist maximal erreichbar?
2. Berechne die exakten Höhen der Positionen im Sequenzlogo aus der Profilmatrix *ohne* Pseudocounts.

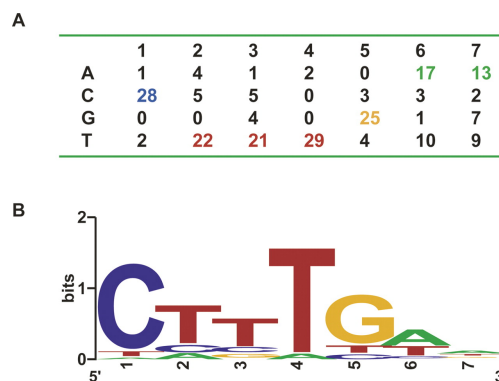


Abbildung 1: A: Zählmatrix; B: Sequenzlogo. [SR Davies et al.: Computational identification and functional validation of regulatory motifs in cartilage-expressed genes. Genome Research 17: 1438–1447 (2007)]

**Aufgabe 6.2** Wir erinnern an die Zählmatrix (siehe Abbildung) und die daraus berechnete PWM (in Zehntelbits, gerundet) der vorigen Aufgabe.

In welcher Reihenfolge und warum sollte man die sieben Positionen der PWM aus der vorigen Aufgabe untersuchen? Bei angenommener Gleichverteilung und einem Schwellenwert von  $T = 56.0$ , nach wie vielen untersuchten Positionen kann man im Erwartungswert abbrechen? Nach wie vielen Positionen ist dies bei der Standard-Reihenfolge  $(1, \dots, 7)$  der Fall?

**Aufgabe 6.3** Betrachte die Muster OMAMA und MAOAM. Das Alphabet ist  $\Sigma = \{A, M, O\}$ ; die Textlänge beträgt 9; es gibt  $3^9 = 19683$  Strings.

Muster	Anzahl			Anteil		
	0 Treffer	1 Treffer	2 Treffer	0 Treffer	1 Treffer	2 Treffer
OMAMA	19278	405	0	0.97942	0.02058	0.00000
MAOAM	19279	403	1	0.97947	0.02047	0.00005

1. Konstruiere ein Minimalbeispiel (binäres Alphabet, möglichst kurze Wörter, kurzer Text) für dasselbe Phänomen.
2. Man könnte argumentieren, dass sich bei MAOAM die Anzahl der Strings mit 0 Treffern gegenüber OMAMA nicht ändern sollte, so dass sich der Vektor  $(19278, 404, 1)$  ergäbe. Benutze ein einfaches Argument, um zu erklären, warum das nicht der Fall sein kann (Erwartungswert).