Lecturer: Jan Baumbach Term: Winter 2015/2016

Class: Introduction to Bioinformatics

Exercise sheet – TFBS sequence prediction

1. Explain/define: "Transcription Factor Binding Site" and "Position Weight Matrix".

2. Given are the following 18 binding sequences for a transcription factor.

cctccacccc	cctcctcccc
cctacgcccc	cctccttgcc
catcctcccg	catcctcccg
cctccttgcc	cctacgcccc
cctcctcccc	cctccacccc
actcatcatc	cctcctcccc
tatccgcccc	tctcatcctg
actcatccct	gctcaccctt
cctcatcctg	actcctccct

Compute (1) a position count matrix and (2) a position weight matrix. For (2), use four pseudo counts for each position. Assume the following background distribution of nucleotides.

A: 0.2 T: 0.2 C: 0.3 G: 0.3

Additionally (3), compute and give the sequence logo (as PNG).

3. Given the strings "banana" and "ananas", construct suffix trees for them step by step. For each step provide a sketch picture.

4. Given the string "teethermometer", construct the suffix tree and the suffix array (including the three tables *suf*, *lcp*, and *skp*).