Your Name: Wai Lam Adele Hong
UCSD Email: w1hong@ucsd.edu
PID: A15999023

To complete this homework you must be working on your own copy in **Google Docs**. Once you have filled in your answers to **Q1-Q4** in the space provided click **File > Download > PDF document** and upload to gradescope (link can be found on the class website).

Q1. [4pts] Consider the following multiple alignment of Transcription Factor Binding site DNA sequences

	1	2	3	4	5
Sequence 1	-	G	A	G	C
Sequence 2	C	T	A	G	A
Sequence 3	C	G	A	-	A
Sequence 4	A	G	C	G	A

Give the (average) profile of the above alignment by filling out the table below. The first position of the first column (i.e. position in the alignment) has been done for you, now complete the rest. You will use this table for answering questions 2 and 3 below.

	1	2	3	4	5
A	0.25	0	0.75	0	0.75
C	0.50	0	0.25	0	0.25
T	0	0.25	0	0	0
G	0	0.75	0	0.75	0
-	0.25	0	0	0.25	0

Q2. [2pts] What is the highest scoring sequence match to your profile above (question 1) and what is it's score?

Sequence: CGAGA

Score: 0.5+0.75+0.75+0.75+0.75= 3.5

Q3. [2pts] Using your completed profile table above (from question 1) score the following two sequences (**S1** and **S2**):

S1. CTGGC → 0.5+0.25+0+0.75+0.25= 1.75

S2. AGCTA → 0.25+0.75+0.25+0+0.75= 2

Q4. [2pts] Following the heuristic threshold for a positive match proposed in Harbison et al. [Nature (2004) 431:99-104.] namely using the threshold for a positive match = 60% x Max Score

Are either of the two sequences in question 3 potential transcription factor binding sites? If so, why?

Heuristic threshold for match = 0.6 x 3.5 = 2.1

S1→ 1.75 < 2.1

S2→ 2 < 2.1

According to Harbison et. al, neither of the two sequences in question 3 are potential transcription factor binding sites because both the S1 and S2 scores are less than 60% of the best possible score.