Lecture 3 Homework
http://thegrantlab.org/
Dr. Barry Grant

Your Name: Anh Tran
UCSD Email: att003@ucsd.edu
PID: A15581788

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 (frequency matrix) 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 your, 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	0.5	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 its score?

Sequence: CGAGA

Score: 4.6

-1.6, 0, 0.6, 1

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

S1. CTGGC   score: 0

S2. AGCTA   score: 0.4

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?

Following method in Harbison et al. (2004) Nature 431:99-104

Heuristic threshold for match = 60% x Max Score = (0.6 x 4.6 = 2.76); Neither scores (0 and 0.4) in Q3 are larger 2.76; Therefore **neither** of the sequences are potential TFBS.