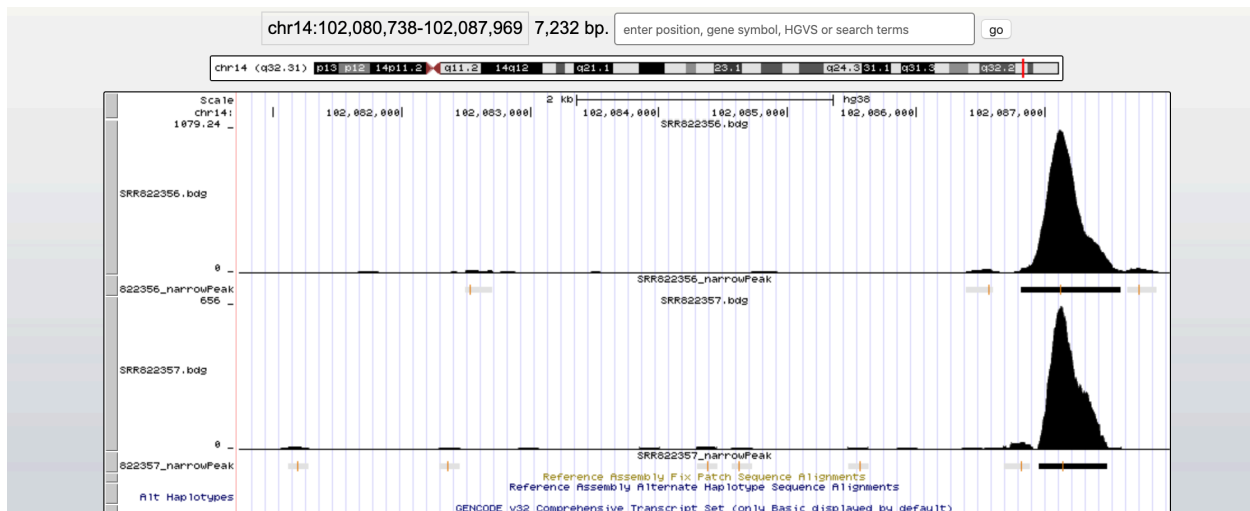


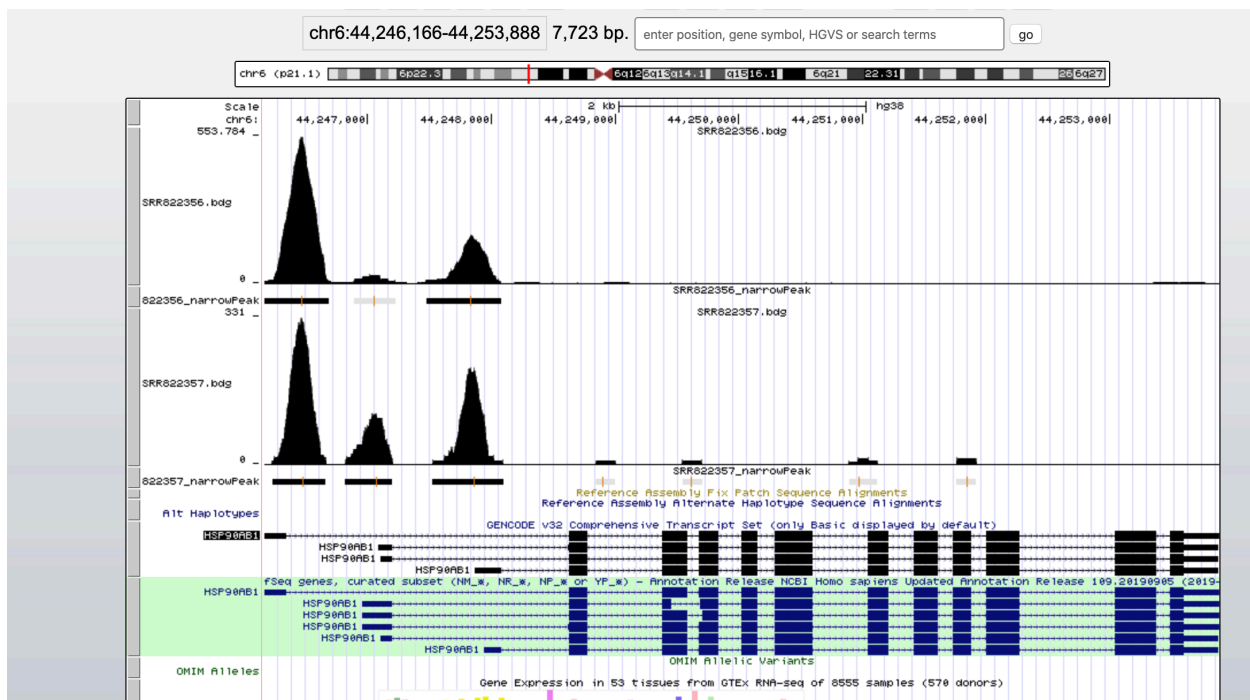
## Homework 3

1.



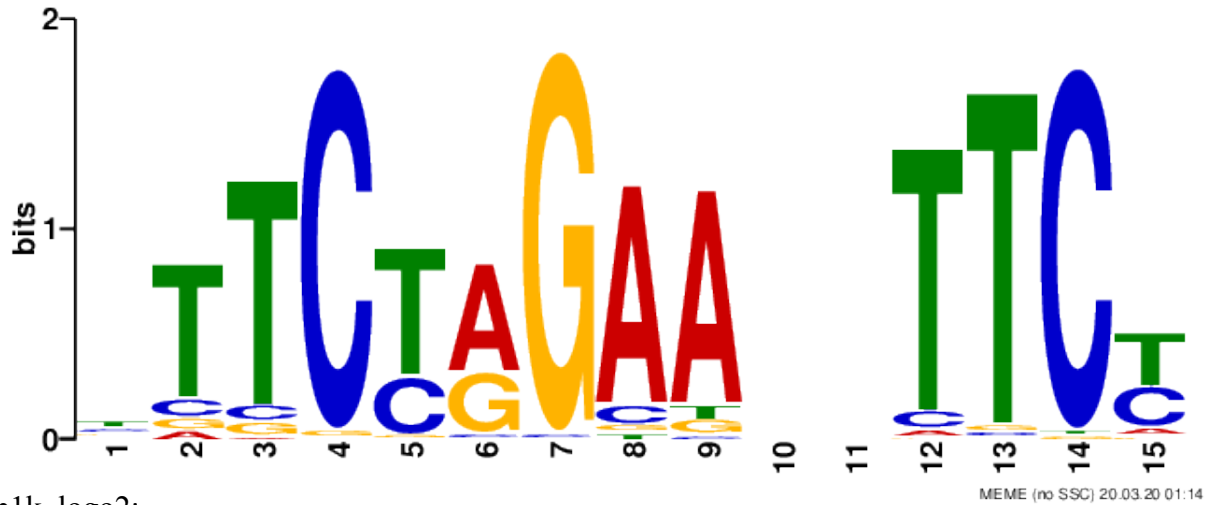
2. The peaks appear to be very similar since they both align in essentially the same area. The top peak is slightly more extended than the bottom peak. Because the peaks overlap, meaning that this is a set of high confidence peaks. Since these peaks are similar in both replicates, we can assume that these peak calls are reliable and that these protein:DNA binding sites are accurately defined.

3.

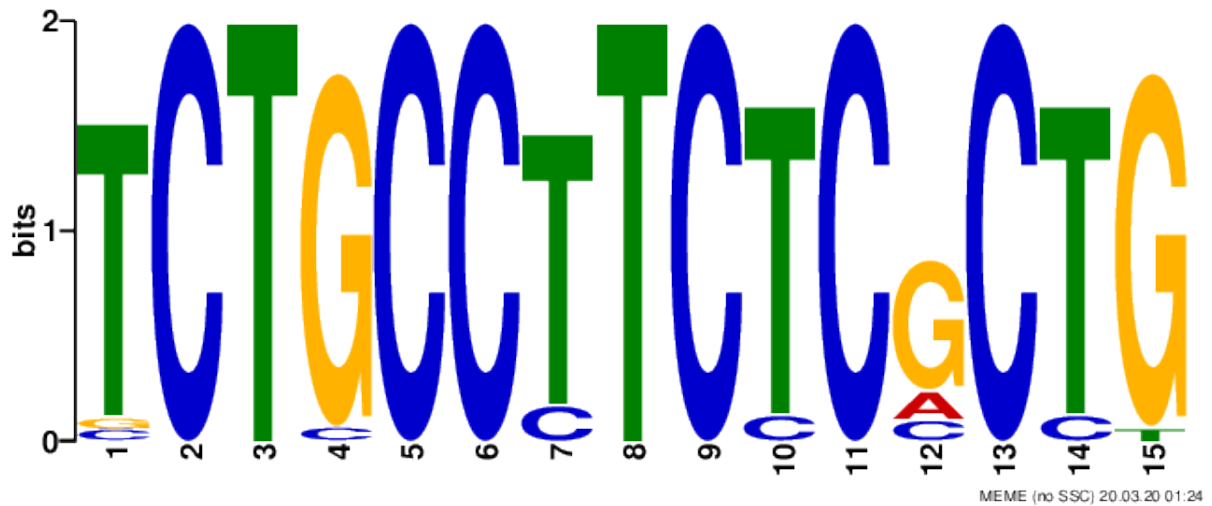


The peaks for *HSP90AA1* are on chromosome 14 and only indicate one binding site. The peaks for *HSP90AB1* are on chromosome 6 and have three peaks. Both genes are heat shock protein genes.

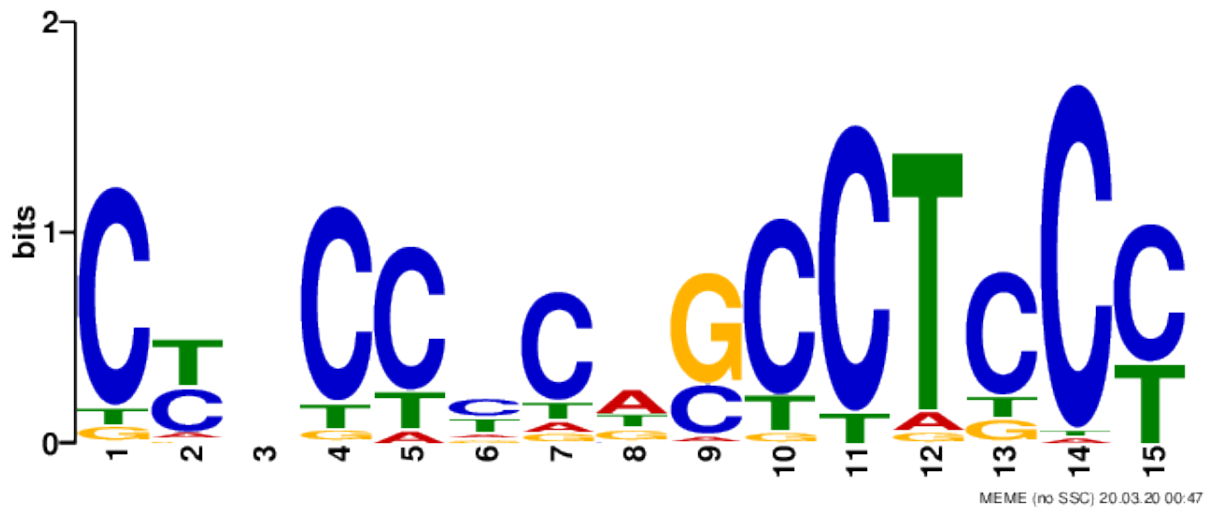
4. Top1k\_logol:



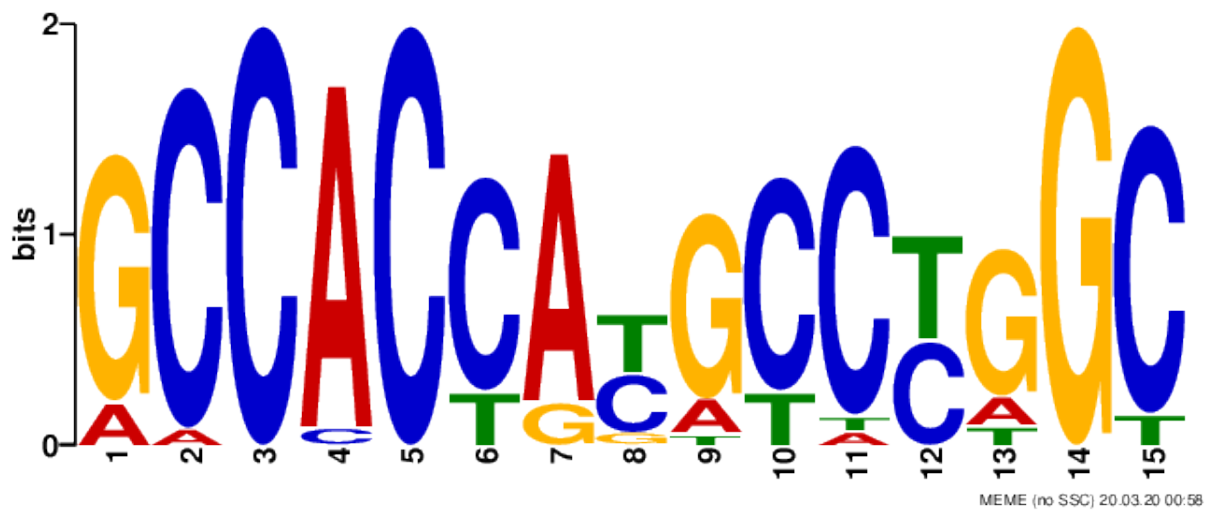
top1k\_logol2:



Ran\_logol:

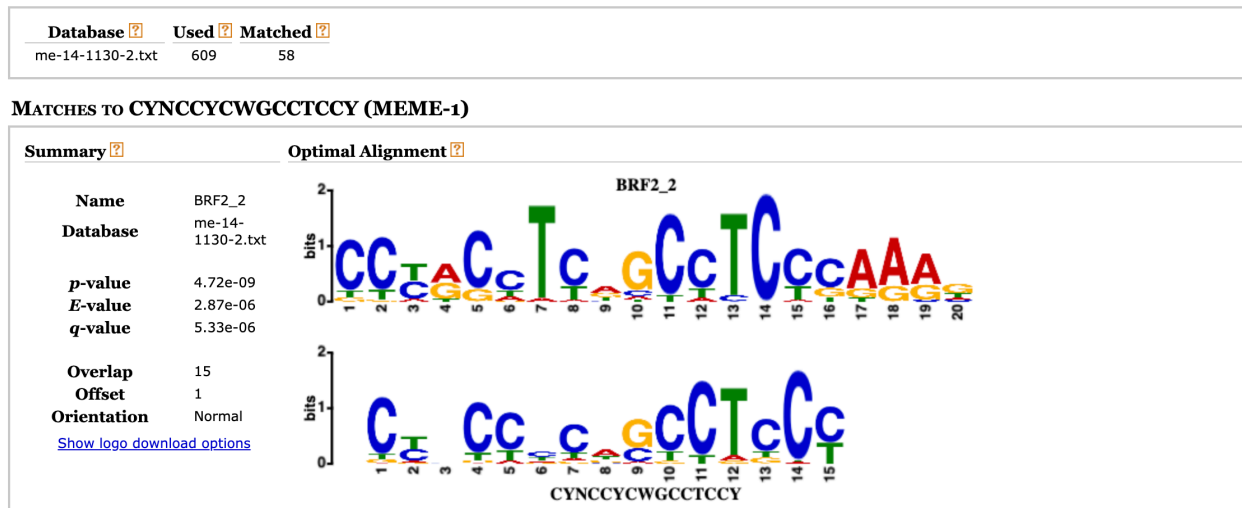


ran\_logo2:

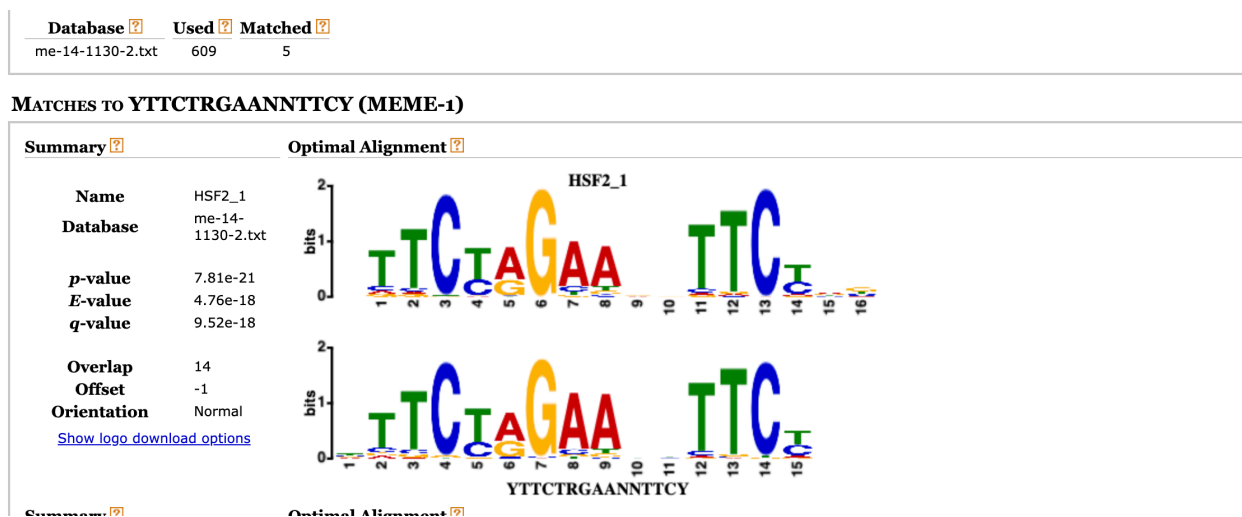


5. The logos are not similar to each other. Since MEME was ran on sequences containing the top 1,000 most intense CHIP-seq peaks as well as 1,000 random CHIP-seq peaks, it is expected that the results will not be the same. The random peaks will not align to the top 1,000 most intense peaks since they cannot be the exact same.

6. Tomtom for random sequences on one match:



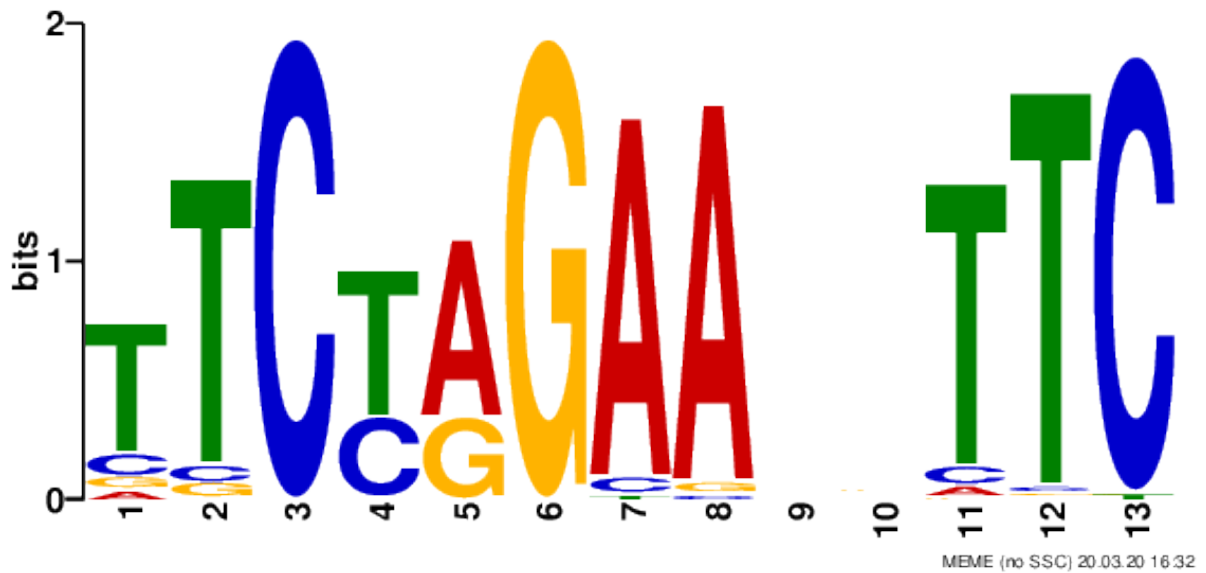
Tomtom for top1k on one match:



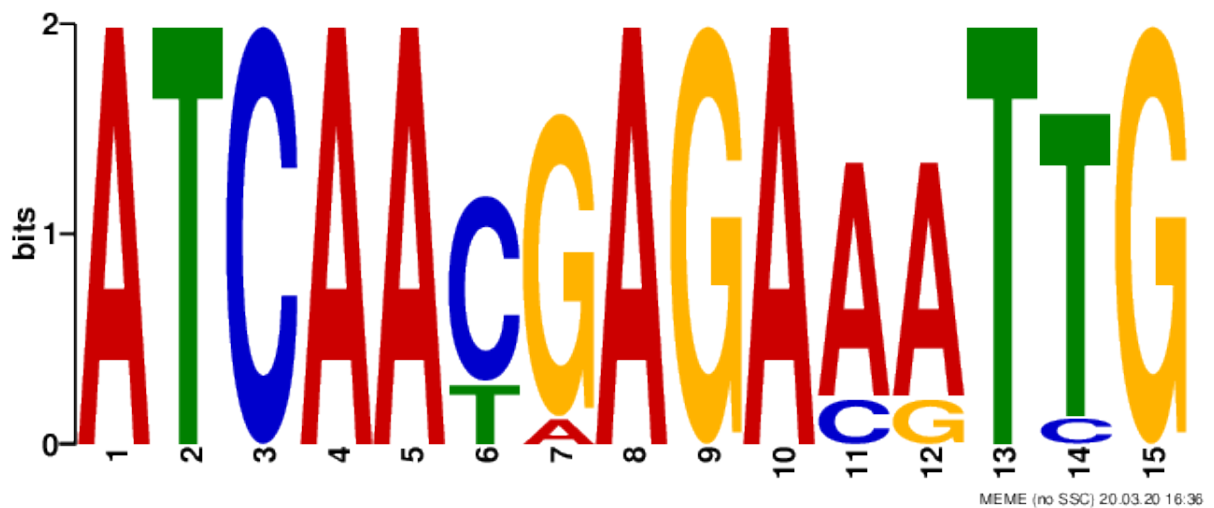
For this motif, there was no HSF motif match in the random sequences (there was a BRF match). For the top1k, there was a match to HSF2\_1. This is also due to the fact that the random sequences returned a different motif match compared to the top 1k. Since the sequences are random, the exact motif cannot be matched. HSF2 is on chromosome 6 while BRF2 is on chromosome 8, meaning that the sequences are on completely different chromosomes.

7.

Top1k.window50\_logo1:



Top1k.window50\_logo2:

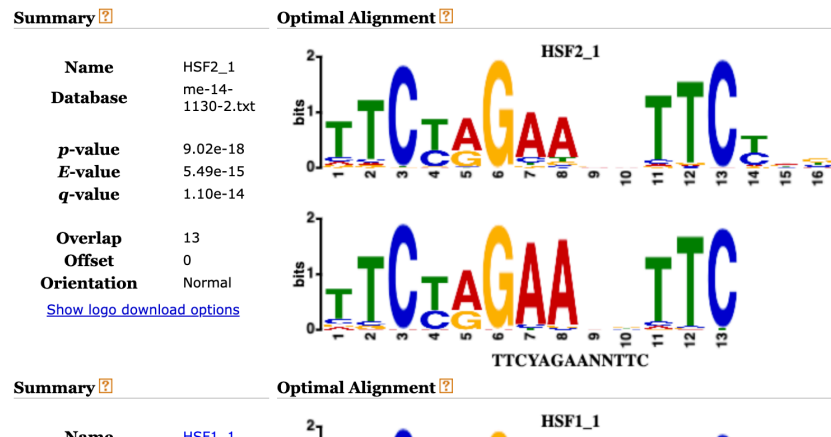


Logo1 motif in the 50bp windows is similar to logo1 in the 100bp windows. However, logo2 motif differs from the 100bp windows logo2.

8.

50bp window on one motif:

#### MATCHES TO TTCYAGAANN TTC (MEME-1)



This match has a p-value of 9.02e-18 and an E-value (expected number of false positives) of 5.49e-15, while the 100bp window has a p-value of 7.81e-21 and an E-value of 4.76e-18. The 50bp windows HSF motif matches really well to the HSF motif from the database, however the motif from the 100bp windows matches better to the motif from the database than the motif from the 50bp windows.