due in 2hr 5m

Submission Phase

1. Do assignment ☐ (/hiddenmessages-004/human_grading/view/courses/976025/assessments/3/submissions)

Evaluation Phase

2. Evaluate peers 🔒 (/hiddenmessages-004/human_grading/view/courses/976025/assessments/3/peerGradingSets)

Results Phase

3. See results 🔒 (/hiddenmessages-004/human_grading/view/courses/976025/assessments/3/results/mine)

☐ In accordance with the Honor Code, I certify that my answers here are my own work, and that I have appropriately acknowledged all external sources (if any) that were used in this work.

Save draft   Submit for grading

Note: you should attempt this Bioinformatics Application Challenge only after reading both chapters of the interactive text and completing the quizzes.

Mycobacterium tuberculosis (MTB) can persist in a latent state in humans for many years before causing disease. Latency has been found to be linked to **hypoxia** (lack of oxygen) in the host. You suspect that genes that are activated in hypoxia are regulated by a common transcription factor, so you collect the upstream sequences for all of the MTB genes that are upregulated in hypoxia, looking for the motif that corresponds to the binding site for the transcription factor regulating these genes. Your biologist colleague tells you that you should look at the 250 bp upstream region of each gene (which have been conveniently compiled for you in a FASTA (https://en.wikipedia.org/wiki/FASTA_format) file named upstream250.txt (http://bioinformaticsalgorithms.com/software_challenges/motifs/upstream250.txt) -- right click and download this file). Your colleague also tells you that the motif is probably about 20 bp long.

We will begin by running three different software tools on the dataset provided. For each tool, use default parameters except for motif length (specify that the motif length should be 20), and make a decision about which motif you feel is best. Then:

- Record all parameters that the program uses.
- Record all results reported by the program.
- Return a motif logo for the motif chosen.

Begin by performing these tasks with Consensus (http://stormo.wustl.edu/consensus/html/Html/main.html) (Hertz and Stormo, 1999). Note: you will need to use WebLogo (http://weblogo.berkeley.edu/logo.cgi (http://weblogo.berkeley.edu/logo.cgi)) to generate the motif logo.

| **B** | *I* | ☰ | ☷ | % Link | `<code>` | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Perform these tasks with MEME (http://meme-suite.org/tools/meme) (Bailey and Elkan, 1994).

| **B** | *I* | ☰ | ☷ | % Link | `<code>` | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Perform these tasks with AlignACE (http://www1.spms.ntu.edu.sg/~chenxin/W-AlignACE/) (Chen and Jiang, 2006).

| **B** | *I* | ☰ | ☷ | % Link | `<code>` | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Did all three programs generate similar motifs? Provide a brief (1-2 sentence) explanation.

| **B** | *I* | ☰ | ☷ | % Link | `<code>` | Math | | Edit: Rich ▾ | Preview |

Attach a file (supports: txt, png, jpg, gif, pdf)

If you had no information about the length of the motif, which program would you use?

Rerun the program you select without specifying the motif length, upload the motif you find as a motif logo, and compare the result to your previous one.

B	I	☰	☷	% Link	<code>	Math		Edit: Rich ▾	Preview

Attach a file (supports: txt, png, jpg, gif, pdf)

Can you find very short motifs in your dataset? What is the minimum motif length for which you would expect to get meaningful results for 250 bp long regions? Give a short (1-2 sentence) explanation for your answer.

B	I	☰	☷	% Link	<code>	Math		Edit: Rich ▾	Preview

Attach a file (supports: txt, png, jpg, gif, pdf)

To evaluate the statistical significance of an identified motif, we need to ensure that a motif with the same or even larger score is unlikely to occur in a collection of "typical" DNA strings (of the same length). How would you generate these strings? Justify your answer.

B	I	☰	☷	% Link	<code>	Math		Edit: Rich ▾	Preview

Attach a file (supports: txt, png, jpg, gif, pdf)

Thus far, we have only analyzed the 250 bp regions upstream of each gene. This makes us wonder whether we will identify the same motif for upstream regions of different lengths.

Give the motif logo obtained by running MEME on upstream regions of length 25 bp (download upstream25.txt (http://bioinformaticsalgorithms.com/software_challenges/motifs/upstream25.txt)).

| **B** | *I* | ☰ | ☰ | % Link | <code> | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Report the sequence logo obtained by running MEME on upstream regions of length 100 bp (download upstream100.txt (http://bioinformaticsalgorithms.com/software_challenges/motifs/upstream100.txt)).

| **B** | *I* | ☰ | ☰ | % Link | <code> | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Report the sequence logo obtained by running MEME on upstream regions of length 500 bp (download upstream500.txt (http://bioinformaticsalgorithms.com/software_challenges/motifs/upstream500.txt)).

| **B** | *I* | ☰ | ☰ | % Link | <code> | Math | | Edit: Rich ▾ | Preview |

Attach a file *(supports: txt, png, jpg, gif, pdf)*

Report the sequence logo obtained by running MEME (http://meme-suite.org/tools/meme) on upstream regions of length 1000 bp (download upstream1000.txt). (http://bioinformaticsalgorithms.com/software_challenges/motifs/upstream1000.txt)

B I ≔ ≔ % Link <code> Math   Edit: Rich ▾   Preview

Attach a file   (supports: txt, png, jpg, gif, pdf)

Which of the motif logos that you created are similar to the motif logo generated from upstream250.txt?

B I ≔ ≔ % Link <code> Math   Edit: Rich ▾   Preview

Attach a file   (supports: txt, png, jpg, gif, pdf)

☐ In accordance with the Honor Code, I certify that my answers here are my own work, and that I have appropriately acknowledged all external sources (if any) that were used in this work.

Save draft   Submit for grading