



Module TSS4: Annotation of Broad Transcription Start Sites

Q1. According to the 9-state models for the CG46466-RB isoform, how is the chromatin state of this region classified?

Q2. How many DHS Positions (statistically significant DNase I hypersensitive sites) are located in this region?

Q3. How many TSSs were predicted in the Celniker data in this region?

Q4. *D. melanogaster* promoters are classified based on the number of annotated TSS (Celniker) positions and the number of DHS positions within a 300 bp window (Table 1). Type "chr4:662,969-663,268" into the "chromosome range, or search terms" text box, and then click the "go" button. Based on the data, how should we classify the promoter of the *D. melanogaster* CG46466-RB gene?

Q5. Why change the display of both the CAGE and RAMPAGE data to hide the data from the plus strand?

Q6. Do the CAGE and RAMPAGE data match with the locations of the Celniker predicted TSSs?

Q7. Are the CAGE and RAMPAGE data consistent with a peaked, intermediate, or broad promoter shape?

Q8. Which is the stronger evidence for the promoter shape — Celniker TSSs and DHS positions, or CAGE and RAMPAGE? Explain.

Q9. In Figure 10, numbered lines are drawn to ten potential landmarks in the vicinity of the TSS of the *myo* gene on contig40 (GEP/Dot) in *D. biarmipes*. Indicate the following for each in the table below: Is it a good landmark for TSS annotation? If it is a good landmark, indicate the quality of the landmark (use it 1st, 2nd or 3rd). Finally, provide a short explanation of your decision.

#	Use as Landmark? (Y/N)	Use 1 st , 2 nd , 3 rd ?	Explanation
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Q10. Using the landmarks you identified above, define a narrow and a wide search region for the *myo* TSS or TSSs. Describe the landmarks you used to determine these search regions.

Q11. Next, identify a wide search region for the *myo* TSS using the landmarks described above.

Q12. Was *blastn* able to locate the CG46466-RB exon 1 in the *D. biarmipes* contig38 project sequence? Explain the evidence you used to make this determination.

Q13. At which coordinate does the coding DNA sequence (CDS) of all of the CG46466-RB isoform begin (ATG codon)?

Q14. According to the Spaln alignment CG46466-RB seems to have only a single exon. What are the coordinates of this exon?

Q15. Can you conclude that the 5' end of the first CG46466 exon is located near the Spaln alignment or is other evidence present that would lead you to conclude that the first exon is actually in a different position? Explain.

Q16. Based on the data available in this region, designate a narrow TSS search region for the *D. biarmipes* CG46466-RB isoform. Describe the landmarks you used to determine these search regions. (Use the browser to zoom in and out to identify the precise boundaries to the nucleotide.)

Q17. For the CG46466-RB isoform would it make sense to use the entire RNA Pol II peak to designate a wide TSS search region? Explain why or why not.

Q18. List all of the consensus promoter motifs found in the narrow TSS search region you identified in the previous question (include the start position of each motif).