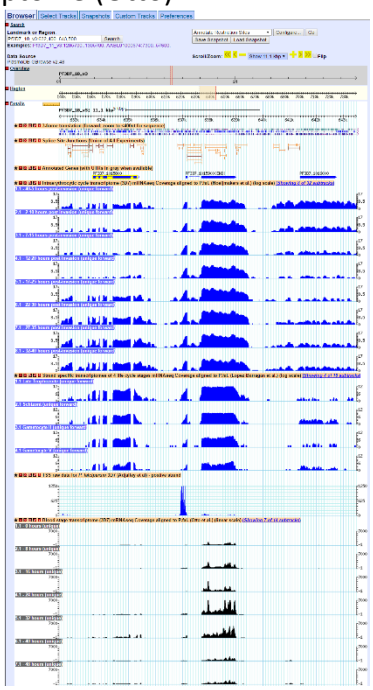
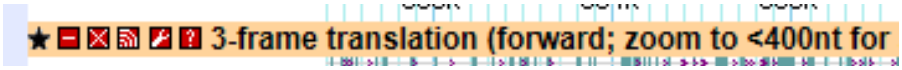
Title: Interpreting RNA Seq data: Gene models and stage specific expression

Gene models are hypotheses about the structure of transcripts produced by a gene and are either predicted *de novo* or inferred from closely related organisms. Gene models are often open to interpretation and can be improved using RNA sequencing data especially for transcript initiation/termination sites (5' and 3' untranslated regions, or UTRs) and alternative processing events. Even heavily curated genomes (*Plasmodium falciparum, Trypanosoma brucei, Saccharomyces cerevisiae*) may not fully reflect all available knowledge about stage-specific splicing, as new information is emerging all the time. In this exercise we will use RNA sequencing data to highlight how to estimate a gene's UTR length, stage-specific expression and possible alternative transcripts.

1. Examine the gene model for PF3D7_1015900 and surrounding genes.
   a. Use the URL to open GBrowse in a preconfigured view.
      - http://tinyurl.com/jp9dzs3
      - Landmark Region => Pf3D7_10_v3:632,400..643,500
      - Containing three genes PF3D7_1015800, PF3D7_1015900 (ENO), PF3D7_1016000
      - 7 data tracks should be turned on:
        - 3-frame translation (forward; zoom <400 for sequence)
        - Splice Site Junctions (Union of all experiments)
        - Annotated Genes (with UTRs in gray when available)
        - Intraerythrocytic cycle transcripme (3D7) (Hoeijmakers)
        - Strand specific transcriptomes of 4 life cycle stages (Lopez-Barragan)
        - TSS raw data from Pfal 3D7 (Adjalley)
        - Blood stage transcriptome (Otto)

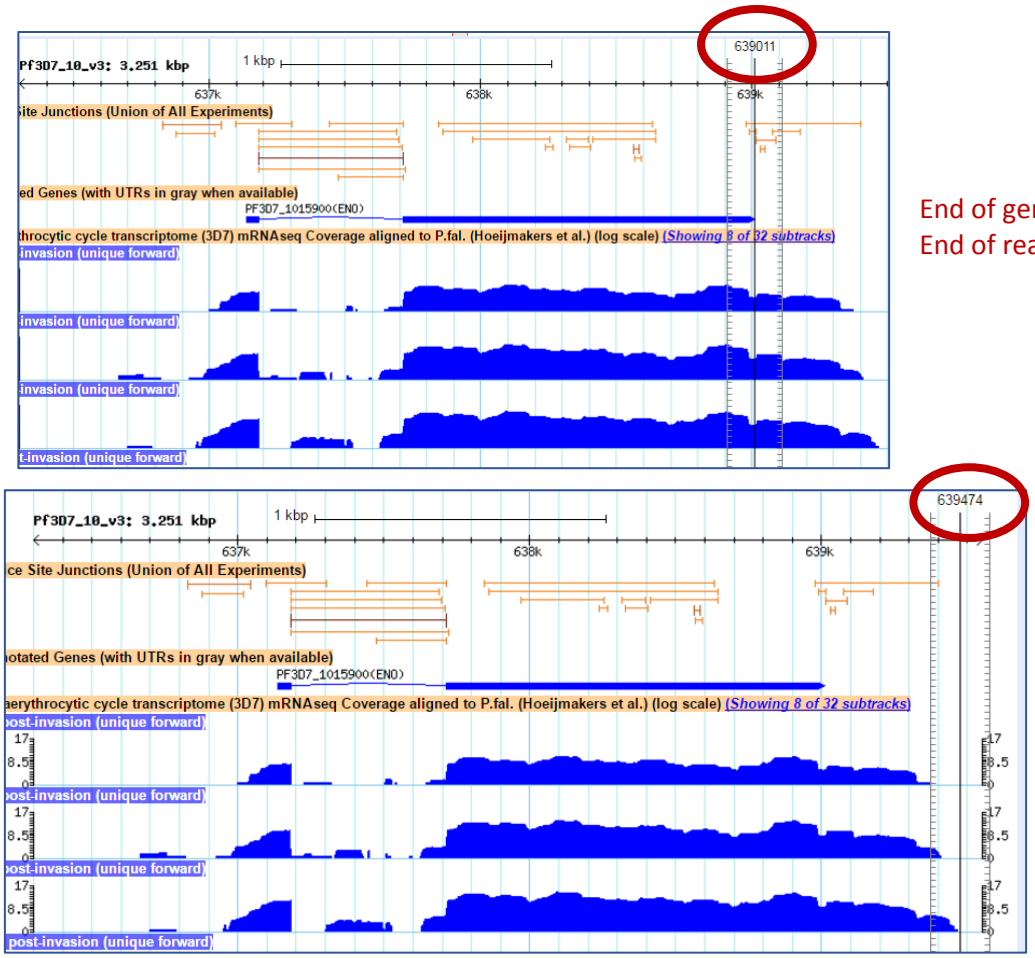


b. Look at each of the tracks and consider the information contained within.
   - Click on the question mark icon that precedes the track name to bring up information about the track.



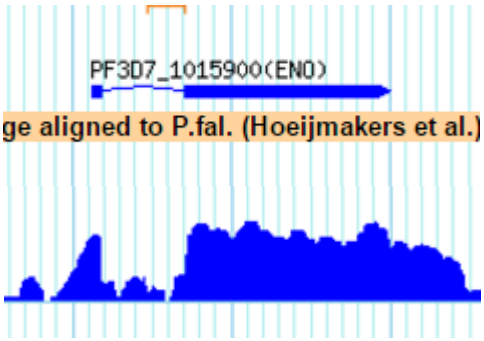
   - What track would you use to decide if a translation start site exists at a specific location? (methionine)
   - What type of data and what track will help confirm alternative transcripts?
   - Hover over the glyphs in the splice junction track. What information is here to help you confirm or discredit a splice junction?
   - Examine the RNA sequencing tracks and zoom in and out. Drag the annotation track up and down to examine the gene model overlaid with the RNA sequencing data. Can you find evidence to support or refute the gene model?

c. Consider the legitimacy of a 3' UTR length for PF3D7_1015900.
   - Use the landmark region to zoom in on PF3D7_1015900. (Pf3D7_10_v3:636,300..639,550)
   - Look at the splice junction track and concentrate on the sites predicted in the downstream region. Do you see evidence to support the existence of an extra exon? Do the splice sites correlate with a possible stop codon in the 3-frame translation track?
   - Look at the reads in Intraerythrocytic cycle transcripme (3D7) (Hoeijmakers) and compare them to the gene model. Could this indicate an untranslated 3' region?
   - Estimate the size of a putative 3' UTR. Use the ruler to determine the difference between the coordinates of the end of the gene and the end of the sequencing reads.



End of gene = 639011
End of reads = 639474

d. Consider the possibility of an alternative transcript for PF3D7_1015900.
   - Look at the reads in Intraerythrocytic cycle transcripme (3D7) (Hoeijmakers) and compare them to the gene model in the area of the intron
   - Do you see evidence for intron read-through?
   - What alternative gene models would you suggest based on the mapped reads?



e. Consider the possibility of stage specific expression for PF3D7_1015900.
   - Scroll through the page paying attention to the change in y-axis intensity for the reads.
   - Are there stages or time points that appear to have increased or decreased expression? At what time point or life cycle stage is the gene's maximal expression.

2. Consider the possibility of alternative transcripts, stage specific expression and possible untranslated regions for the other genes in the region: genes PF3D7_1015800 and PF3D7_1016000.
   - Change the Landmark Region => Pf3D7_10_v3:632,400..643,500 and choose a gene to investigate.
   - Work with your partner ☺
   -