MODULE 3: TRANSCRIPTION PART II

S. CATHERINE SILVER KEY, CHIYEDZA SMALL

Lesson Plan:

Title	Transcription Part II: What happens to the initial (pre-mRNA) transcript made by RNA pol II?

Objectives	• Explain how the transcript generated by RNA polymerase II (the pre-mRNA) is processed to become mature mRNA, using the sequence signals identified in Module 2.
	• Use the genome browser to analyze the relationships among:
	  1. pre-mRNA
	  2. 5' capping
	  3. 3' polyadenylation
	  4. splicing
	  5. mRNA
	• Use the genome browser to identify the relationships among; pre-mRNA, 5' capping, 3' polyadenylation, splicing, mRNA (Investigation).
	• Describe how the transcript generated by RNA polymerase II (the pre-mRNA) is processed to become mature mRNA, using the sequence signals identified in Module 2 (Investigation).

Pre-requisites	• Modules 1 and 2 • Define pre-mRNA as the RNA that results from the process of transcription; this initial transcript includes exons and introns
Order	• Warm Up Discussion • Investigation
Homework	None

Class instruction	• Discuss the question: What happens to the initial (pre-mRNA) transcript made by RNA pol II? Does it leave the nucleus 'as is'? Or do changes have to occur? (Hint: introns vs. exons) (Discuss with a partner then as a class).
	• Mini-presentation illustrating that during pre-mRNA processing, three events occur:
	  1. 5' capping,
	  2. 3' polyadenylation
	  3. splicing out of introns

	- Work through the genome browser investigation, with pauses to discuss the answers to the questions. - Conclude with emphasis on main points: - Pre-mRNA is processed using 3 steps: 1. 5' capping, 2. 3' polyadenylation, 3. removal of introns through splicing (via spliceosome)

INTRODUCTION

In Module 2, you identified the transcription start site (TSS) for the A isoform of the *tra* gene (tra-RA). In this module, we will explore each of the three steps of pre-mRNA processing.

SETTING UP OUR BROWSER PAGE (REVIEW):

1. Open a new web browser window and go to the UCSC Genome Brower Mirror site at http://gander.wustl.edu/. Follow the instructions given in module 1 to navigate to the contig1 project in the *D. melanogaster* "July 2014 (Gene)" assembly.
2. As you may remember from Module 1, contig1 is derived from chr3L in the *D. melanogaster* genome. This contig contains three different genes (*CG32165*, *spd-2*, and *tra*). Enter "**contig1:9,500-11,000**" into the "enter position or search terms" textbox and then click on the "go" button to navigate to the genomic region surrounding the *tra* gene.
3. Because the Genome Browser remembers your previous display settings, you should click on the "default tracks" button to reset the display to the default settings. Change the display mode for the "Base Position" track to "**full**" and verify that the "FlyBase Genes" track is set to "**pack**". Click on the "refresh" button.
4. Scroll down to the "RNA Seq Tracks" section and then click on the "**RNA-Seq Coverage**" link. Change the track display settings to the following, as we did in Module 2:
   - Change the "Display mode" field to "**full**"
   - Select the "Data view scaling" field to "**use vertical viewing range setting**"
   - Change the "max" field under "Vertical viewing range" to **37**
   - Under the "List subtracks" section, select BOTH the "**Adult Females**" and the "**Adult Males**" (Select the check box next to subtrack to turn the subtrack on.)
   - Click on the "Submit" button. Verify that the RNA-Seq Coverage track on the browser page is set to "**full.**"

INVESTIGATION: MRNA PROCESSING

Learning Objectives:

- Use the genome browser to identify the relationships among; pre-mRNA, 5' capping, 3' polyadenylation, splicing, mRNA.
- Describe how the transcript generated by RNA polymerase II (the pre-mRNA) is processed to become mature mRNA, using the sequence signals identified in Module 2.

The processing of pre-mRNA into mRNA involves three key steps (Figure 1):

- The addition of a **5' cap**
- The addition of a **3' poly(A) tail**
- The removal of introns through **splicing**

Removal of the introns during this process results in adjacent exons being brought together in the final mRNA message.
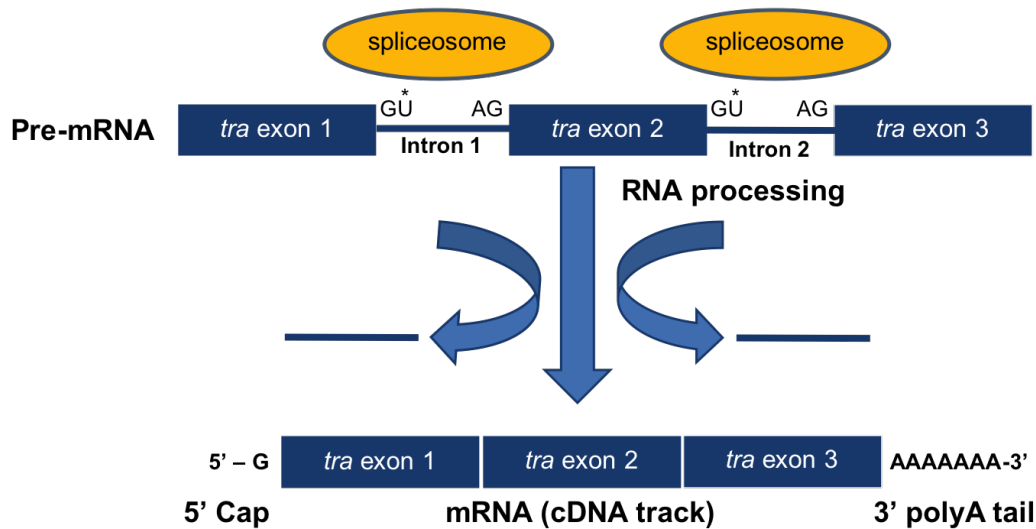
* The GU sequence in the pre-mRNA will be GT in the DNA sequence of the Genome Browser

Figure 1 Diagram of mRNA processing that converts a pre-mRNA to a processed mRNA.

The **first step in pre-mRNA** processing occurs at the 5' end of a messenger RNA. Recall that mRNA is synthesized in a 5' to 3' direction, so the 5' end of the mRNA was synthesized first. Let's examine the beginning of the *tra* gene. Type "**contig1:9,825-9,870**" into the "enter position or search terms " textbox and then click on the "go" button.

In Module 2, we identified the **transcription start site** (TSS) of the A isoform of *tra* at position **9,851**.

To show the TSS's that have been annotated by the modENCODE project, scroll down to the "Genes and Gene Prediction Tracks" and change the display mode for the "TSS Annotations" track to "**pack,**" and click refresh. The modENCODE project looked for TSSs by using a chemical method to tag the special structure that occurs at 5' ends of transcript, fishing out the RNA molecules that carried these tags, and mapping the sequence back to the genome, a method called "CAGE" (cap analysis of gene expression).

In addition, we will also display the "D. mel. cDNAs" track (also under the "Genes and Gene Prediction Tracks" section); change this to "**pack**". This track shows the alignment of *D. melanogaster* cDNAs (complementary DNAs, made by copying the mRNA) that have been sequenced by the Berkeley Drosophila Genome Project (BDGP). Click on the "refresh" button (Figure 2). These two tracks both provide an analysis based on the RNA population, and mapping the positions of these sequences indicates where the transcript started.

Remember from Module 2 that we also found a match to TCAKTY, a common initiation signal just upstream, at 9,834 (display this using the "Short Match" track). All of these pieces of evidence argue for a TSS in this region.
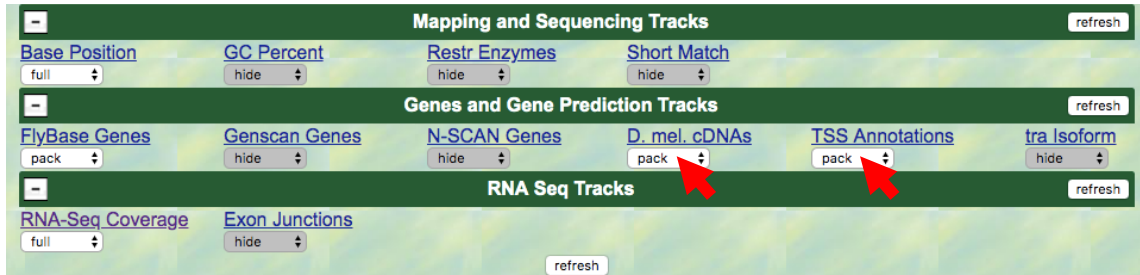

FIGURE 2 CHANGE THE DISPLAY MODES OF THE "D. MEL. CDNAS" AND "TSS ANNOTATIONS" TRACKS TO "PACK".

The new Genome Browser image (Figure 3) shows the 5' end of the pre-mRNA transcript (i.e. the start of transcription) based on the CAGE experiment (modENCODE track) with the additional lines of support. On this end of the pre-mRNA, a modified guanine nucleotide (7mG) is added to the nucleotide at position 9,851, forming the **5' cap**. Note that this additional nucleotide is **NOT** visible in the DNA track. It is added AFTER the transcript is made. This is the **first step** in **pre-mRNA processing**: capping.
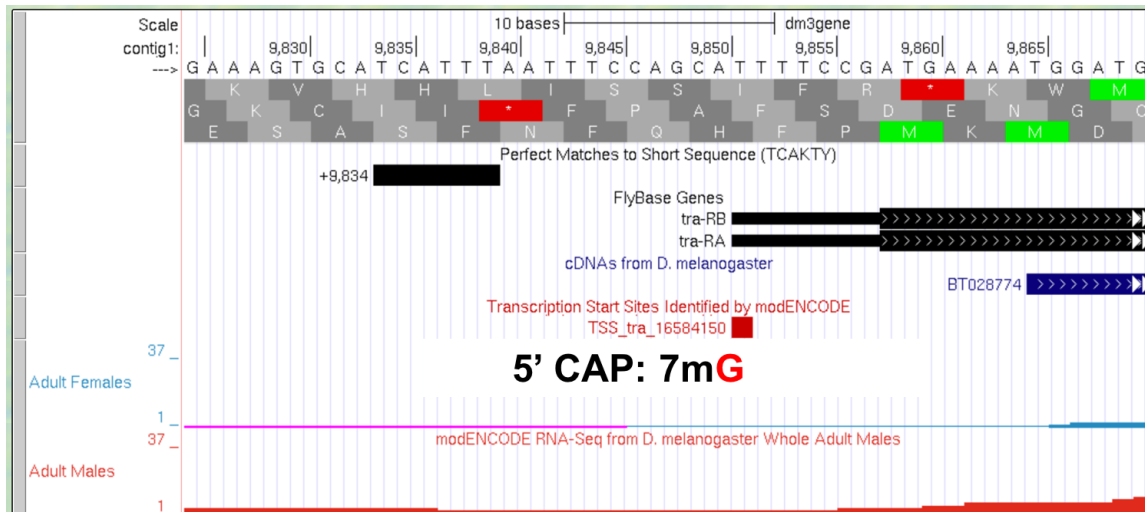

FIGURE 3 ADDITION OF A 5' CAP TO THE 5' END OF THE TRANSCRIPT.

Q1. What is the coordinate of the first nucleotide that is transcribed? In the DNA sequence, is it an A, C, T or G?

Q2. What are the coordinates for the start codon that codes for the first amino acid of the A isoform of the *tra* gene? (Assume reading frame +3.)

Q3. The region of the transcript from the 5' cap to the nucleotide just upstream of the start codon is called the 5' untranslated region (5'UTR) because it is part of the transcript that is not translated. How long (in ribonucleotides) is the 5'UTR?

The **second step** in pre-mRNA processing is **polyadenylation**.

5. To view the 3' end of the tra-RA gene, change the "enter position or search terms" field to "**contig1:10,633-11,000**" and then click on the "go" button.

Polyadenylation means that **many** (poly) **adenine nucleotides** (ribonucleotides) are added to the 3' end of the pre-mRNA **AFTER** transcription termination. This generates a poly-A tail (typically ~20 to ~250 As) that will be retained in the final mRNA but it is not present in the "Base Position" track of the Genome Browser. This is because the poly-A tail does not exist in the DNA template but is simply added to the RNA by a special polymerase as a long run of adenine nucleotides.

Our previous analysis in Module 1 has shown that the last coding exon of tra-RA is in frame +2 and the stop codon is located at 10,754-10,756. We can use the Genome Browser to determine the end of the tra-RA transcript indicated by the cDNA track (in blue). (Note that this aligns with the cDNA although there is some discrepancy between the two as to the exact end of the transcript.)

6

Q4. How long (in base pairs) is this 3' untranslated region (3'UTR) as indicated by the cDNA track (in blue)?

Q5. Zoom into the 3' end of the FlyBase Gene, near the termination site. What is the longest stretch of A nucleotides that you observe?

Q6. Do your findings support the conclusion that the poly(A) sequence observed in the mature mRNA transcript is not in the template DNA?

6. Perform a "Short Match" search for the poly-A signal (AATAAA) using the protocol you learned in Module 2. This search should place the poly-A signal at 10,818-10,823 (Figure 4). As mentioned in Module 2, the transcript is cleaved 11 to 30 nucleotides downstream of the poly(A) signal sequence, and then 150-200 adenines are added to the pre-mRNA. The nucleotides between the stop codon and the end of the poly-A tail comprise the 3' UTR.
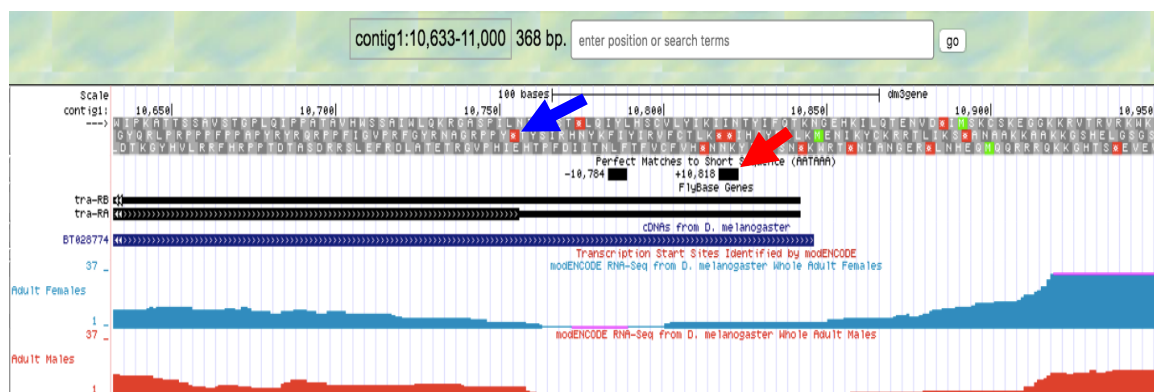

FIGURE 4 PREVIOUS ANALYSIS PLACED THE STOP CODON FOR THE A ISOFORM OF *TRA* IN FRAME +2 (BLUE ARROW) AND THE POLY-A SIGNAL AT 10,818-10,823 (RED ARROW).

We can see the polyadenylation sequence that is associated with the processed mRNA by examining the cDNA (BT028774) that has been aligned to this region.

7. Click on the "BT028774" feature under the "cDNAs from *D. melanogaster*" track and then click on the "View details of parts of alignment within browser window" link (Figure 5).
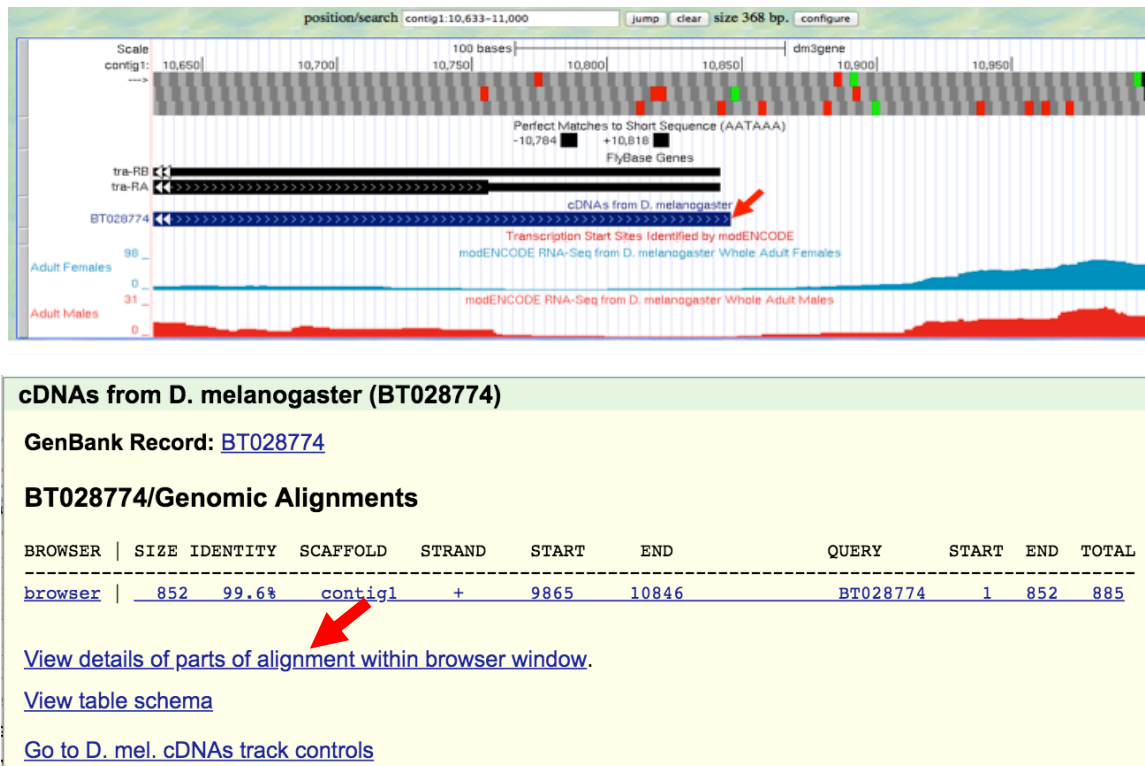

FIGURE 5 EXAMINE THE ALIGNMENT OF D. MELANOGASTER CDNA BT028774 AGAINST CONTIG1.

The next figure shows the actual alignment between the *D. melanogaster* cDNA BT028774 and the genomic sequence in contig1 (Figure 6). Nucleotides that are identical between the two sequences are shown in blue capital letters while nucleotides that differ are shown as black lowercase letters. The light blue bases denote the start and the end of the gap in the alignment. The side-by-side alignment shows the pairwise alignment between the cDNA (top) and the contig1 sequence (bottom) within the viewing region (i.e. contig1:10,633-11,000).

Alignment of BT028774

BT028774
BT028774 in browser window
D. melanogaster.contig1
together

cDNA BT028774

```
TGGATGCCGA CAGCAGTGGA ACCCAGCATC GAGTGTCATA TTGTGTGAAA  50
TGTGAAATGG ATGAGAATGA ACGCTGGACG ACCAGACAGA GAAAGAGAAG  100
CTAGGACAAT AGGACTCTCA ACTGCGCATT ACGTGGATTC CGTCTCCGAC  150
GATGCGCCAA ACACTATGCG TTAGATGCAT TGACTAACCG CACACTCTTT  200
TCACATAGAT TCCCGTGGCT CAAGGTCTCG ATCCCGGCGA GAAAGAGAAT  250
ACCATGGGCG ATCAAGCGAG AGGGACAGCA GAAAGAAGGA GCACAAGATT  300
CCGTACTTTG CAGACGAGGT TCGAGAACAG GATCGGTTGA GAAGATTGCG  350
CCAAAGAGCG CACCAATCCA CCAGACGCAC TCGCTCCAGA TCCAGATCAC  400
AGTCGTCCAT CAGAGAGAGC AGGCACAGAA GGCATCGCCA GCGCTCTAGG  450
AGCCGCAATC GCAgCCGAAG TCGCAGCAGT GAACGAAAAC GCCGTCAACg  500
gAGCCGAAGT CGCAGCAGTG AACGAAGACG CCGTCAACGG AGCCCGCATC  550
GGTATAATCC TCCGCCAAAG ATCATCAACT ACTATGTGCA AGTGCCACCA  600
CAGGATTTCT ACGGGATGTC TGGcATGCAG CAAAGTTTTG GATACCAAAG  650
GCTACCACGT CCTCCGCCGT TTCCACCGGC CCCCTACAGA TACCGCCAGC  700
GACCGCCGTT CATTGGAGTT CCGCGATTTG GCTACAGAAA CGCGGGGCGT  750
CCCCCATATT GAACATACTC CATTCGACAT AATTACAAAT TTATTTACAT  800
TCGTGTGTTT TGTACATTAA AATAATAAAT ACATATATAT TTCAAACTAA  850
AAcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa
```

Genomic contig1 :

```
CACCACAGGA TTTCTACgtt agtagttatt tgtattgaaa caaataacaa  10582
atctaactta aatcgatttt gcagGGGATG TCTGGtATGC AGCAAAGTTT  10632
TGGATACCAA AGGCTACCAC GTCCTCCGCC GTTTCCACCG GCCCCCTACA  10682
GATACCGCCA GCGACCGCCG TTCATTGGAG TTCCGCGATT TGGCTACAGA  10732
AACGCGGGGC GTCCCCCATA TTGAACATAC TCCATTCGAC ATAATTACAA  10782
ATTTATTTAC ATTCGTGTGT TTTGTACATT AAAATAATAA ATACATATAT  10832
ATTTCAAACT AAAAatggag aacataaaat attgcaaacg gagaacgttg  10882
attaaatcat gagcaaatgc agcaaagaag gcggcaaaaa aagggtcaca  10932
cgagttagga agtg
```

Side by Side Alignment

```
00639 tggataccaaaggctaccacgtcctccgccgtttccaccggccccctaca 00688
>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>
10633 tggataccaaaggctaccacgtcctccgccgtttccaccggccccctaca 10682

00689 gataccgccagcgaccgccgttcattggagttccgcgatttggctacaga 00738
>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>
10683 gataccgccagcgaccgccgttcattggagttccgcgatttggctacaga 10732

00739 aacgcggggcgtcccccatattgaacatactccattcgacataattacaa 00788
>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>
10733 aacgcggggcgtcccccatattgaacatactccattcgacataattacaa 10782

00789 atttatttacattcgtgtgttttgtacattaaaataataaatacatatat 00838
>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>
10783 atttatttacattcgtgtgttttgtacattaaaataataaatacatatat 10832

00839 atttcaaactaaaa 00852
>>>>> |||||||||||||| >>>>>
10833 atttcaaactaaaa 10846
```

FIGURE 6 ALIGNMENT OF THE D. MELANOGASTER CDNA BT028774 WITH THE END OF CONTIG1.

Q7. Scroll up to the to the 'cDNA BT028774' area. After which coordinate (number in the cDNA) do you see the polyadenylation track (in lower case black letters)?

Q8. How many 'A' ribonucleotides have been added to the *tra* mRNA (represented in the cDNA)?

Q9. Locate the **AATAAA** termination signal in the cDNA sequence. How many nucleotides 3' of the final 'A' in the signal sequence does the poly(A) run start? (This number should be between 11-30 nucleotides.)

The final step in pre-mRNA processing is splicing out of introns and merging adjacent exons into one continuous open reading frame so that the mRNA is ready for translation into a protein.

8. Change the "enter position or search terms" field to "**contig1:9,870-10,170**" and then click on the "go" button to navigate to the first intron of the tra-RA transcript (Figure 7).
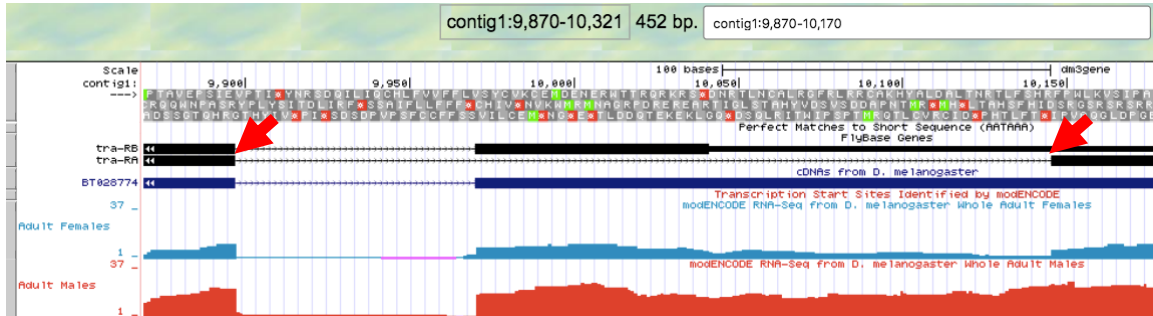

FIGURE 7 THE GENOMIC REGION SURROUNDING THE FIRST INTRON (RED ARROWS) OF TRA-RA.

9. Zoom in to the region near the end of the first exon of tra-RA.

Q10.   Which two nucleotides are found just after the end of the first exon of tra-RA? _____ Repeat this determination, identifying the two nucleotides at the start of intron 2 of tra-RA. _____

These two nucleotides are a signal for **splicing** to occur at the 5' end of an intron; these represent the first two bases of the intron, often called the donor site (or 5' splice site).

Q11.   At which base does exon 1 end? _____

10. Zoom out and then zoom in to the region near the beginning of the second exon of tra-RA.

Q12.   Which two nucleotides are found right before the start of tra-RA exon 2? _____

11. Zoom out and then zoom in to the region near the beginning of the third exon of tra-RA.

Q13. Which two nucleotides are found right before the start of tra-RA exon 3? _____

These two nucleotides are the signal for **splicing** out of the 3' end of the intron, often called the acceptor site (or 3' splice site). These represent the last two bases of the intron.

Q14. At which base does exon 2 of tra-RA begin? What is its coordinate?

_____

Summary

In this module, we learned about the three key steps that are involved in converting the pre-mRNA into a mature mRNA:

1. The addition of a **5' cap**
2. The addition of a **3' poly(A) tail**
3. The removal of introns through **splicing**
   - Note that introns are removed during this process and adjacent exons are brought together in the mRNA message

After mRNA processing, the mature mRNA (tra-RA) can now exit the nucleus so that it can be translated into a protein (tra-PA) by the cytoplasmic ribosomes.

Summary Question 1. Describe how the transcript generated by RNA polymerase II (the pre-mRNA) is processed to become mature mRNA, using the sequence signals identified in Module 2.