Module 3: 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:
  - pre-mRNA
  - 5’ capping
  - 3’ polyadenylation
  - splicing
  - mRNA

Prerequisites

- Understanding Eukaryotic Genes Modules 1-2
- Define pre-mRNA as the RNA that results from the process of transcription; this initial transcript includes exons and introns

Class Instruction

- Discuss the questions: 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:
  - 5’ capping,
  - 3’ polyadenylation,
  - 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:
    - 5’ capping,
    - 3’ polyadenylation,
    - removal of introns through splicing (via spliceosome)
Table of Contents

Introduction ................................................................................................................................................. 2
Setting up our Browser page (review): ........................................................................................................ 2
Investigation: mRNA processing ..................................................................................................................... 3
Conclusions ..................................................................................................................................................... 9

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 Browser Mirror site at
   https://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 "chromosome range, or search terms, see examples" 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:
   5. Change the "Display mode" field to "full"
   6. Select the "Data view scaling" field to "use vertical viewing range setting"
   7. Change the "max" field under "Vertical viewing range" to 37.
   8. 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.)
   9. Click on the "Submit" button. Verify that the RNA-Seq Coverage track on the browser page is
      set to “full.”
**Investigation: mRNA processing**

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

![Diagram of mRNA processing that converts a pre-mRNA to a processed mRNA.](image)

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 "chromosome range, 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 TSSs 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.

10. To view the 3’ end of the tra-RA gene, change the "chromosome range, 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.)

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?

11. 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.

12. Click on the "BT028774" feature under the "cDNAs from D. melanogaster" track and then click on the link under the "SIZE" column of the "BT028774/Genomic Alignments" section (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).

![Alignment of BT028774 and contig1: 9865-10846](Alignment of BT028774 and contig1: 9865-10846)

**Figure 6**  **Alignment of the *D. melanogaster* cDNA BT028774 with the end of contig1.**

**Q7.** Scroll up 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.

13. Change the "chromosome range, 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).

![Image](image_url)

Figure 7  The genomic region surrounding the first intron (red arrows) of tra-RA.

14. 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? ______________

15. 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? ______

16. 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? ____________

Conclusions

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

- The addition of a 5' cap
- The addition of a 3' poly(A) tail
- 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.