Módulo 3: Transcripción Parte II: ¿Qué le sucede al transcrito inicial (pre-ARNm) sintetizado por la ARN pol II? 

*Catherine Silver Key & Chiyedza Small  
(traducido por* Paola A Acosta-Crespo, Bethany Paris y Juan Carlos Martinez-Cruzado)

**Objetivos**

* Explicar cómo, mediante el uso de las secuencias clave identificadas en el Módulo 2, el transcrito generado por la ARN polimerasa II (el pre-ARN mensajero) es procesado para convertirse en ARN mensajero maduro.
* Utilizar el navegador genómico para analizar las relaciones entre:
  + pre-ARN mensajero
  + “capping” del extremo 5’
  + poliadenilación del extremo 3’
  + empalme
  + ARN mensajero

**Prerrequisitos**

* Módulos 1-2 de “[Entender los genes eucariotas](https://thegep.org/ueg/)”
* Definir el pre-ARN mensajero como el ARN inmaduro que resulta del proceso de transcripción; este transcrito inicial incluye exones e intrones.

**Instrucción en el aula**

* Debata las preguntas: ¿Qué le sucede al transcrito inicial (pre-mARN) sintetizado por la ARN pol II? ¿Sale del núcleo tal y como fue transcrito o debe sufrir cambios? (Pista: intrones frente a exones) (los estudiantes conversan en parejas y luego se abre una conversación con el conjunto de la clase).
* Haga una breve exposición del procesamiento del pre-ARN mensajero explicando que incluye tres eventos:
  + “capping” del extremo 5’
  + poliadenilación del extremo 3’
  + eliminación de intrones y empalme de exones
* Progrese en el trabajo del navegador genómico realizando pausas para responder a las preguntas.
* Concluya subrayando los puntos más importantes:
  + El pre-ARN mensajero es procesado en tres fases:
    - “capping” del extremo 5’
    - poliadenilación del extremo 3’
    - eliminación de intrones mediante empalme (vía el empalmosoma)

**Índice de contenidos**

[Introducción 2](#_Toc126575400)

[Configuración del Navegador (repaso): 2](#_Toc126575401)

[Investigación: procesamiento del ARN mensajero 3](#_Toc126575402)

[Conclusiones 9](#_Toc126575403)

# Introducción

En el Módulo 2, identificaste el sitio de iniciación de la transcripción (TSS, por sus siglas en inglés) de la isoforma A del gen *tra* (tra-RA). En este módulo, exploraremos cada uno de los tres pasos del procesamiento del pre-ARN mensajero.

# Configuración del Navegador (repaso):

1. Abre una nueva ventana del navegador de la web y dirígete al sitio “UCSC Genome Browser Mirror” en <https://gander.wustl.edu/>. Sigue las instrucciones dadas en el Módulo 1 para llegar al proyecto “contig1” en el ensamblado de *D. melanogaster* "July 2014 (Gene)".
2. Como recordarás del Módulo 1, el “contig1” es un fragmento del chr3L del genoma de *D. melanogaster*. Este “contig” contiene tres genes diferentes (CG32165, *spd-2* y *tra*). Introduce **"contig1:9,500-11,000"** en la sección de texto “chromosome range, or search terms, see examples” y luego pulsa el botón de “go” para navegar a la región genómica que rodea al gen *tra*.
3. Dado que el navegador genómico recuerda las configuraciones previas, tendrás que pulsar el botón de "default tracks" para restablecer la configuración estándar. Cambia el modo de visualización de la pista (“Track”) de “Base Position” a “**full**” y verifica que la pista “FlyBase Genes” esté en “**pack**”. Haga clic en el botón de “refresh”.
4. Dirígete hacia abajo en la zona de control hasta llegar a la sección de “RNA Seq Tracks” y haz clic en el enlace “**RNA-Seq Coverage**”. Cambia la configuración de visualización de la pista como se indica a continuación y como ya hicimos en el Módulo 2:
5. Cambia el modo de visualización a “**full**”.
6. Selecciona “Data view scaling” y escoge "**use vertical viewing range setting**"
7. Cambia el “max” que se encuentra bajo "Vertical viewing range" a **37**.
8. Bajo la sección "List subtracks", selecciona "**Adult Females**" y "**Adult Males**". (Selecciona las casillas que se encuentran junto a cada “subtrack” para activarlo.)
9. Haz clic en el botón de “Submit”. Verifica que la franja de “RNA-Seq Coverage” en la página del navegador esté en “**full**”.

# Investigación: procesamiento del ARN mensajero

El procesamiento del pre-ARN mensajero a ARN mensajero comprende tres pasos claves (Figura 1):

* La adición de la estructura **caperuza** al extremo **5’** (“5’ cap”)
* La adición de la **cola poli(A)** al extremo **3’** (“3’ poly(A) tail”)
* La eliminación de intrones mediante **empalme**

La eliminación de intrones durante este proceso produce la unión de exones adyacentes en el ARN mensajero final.

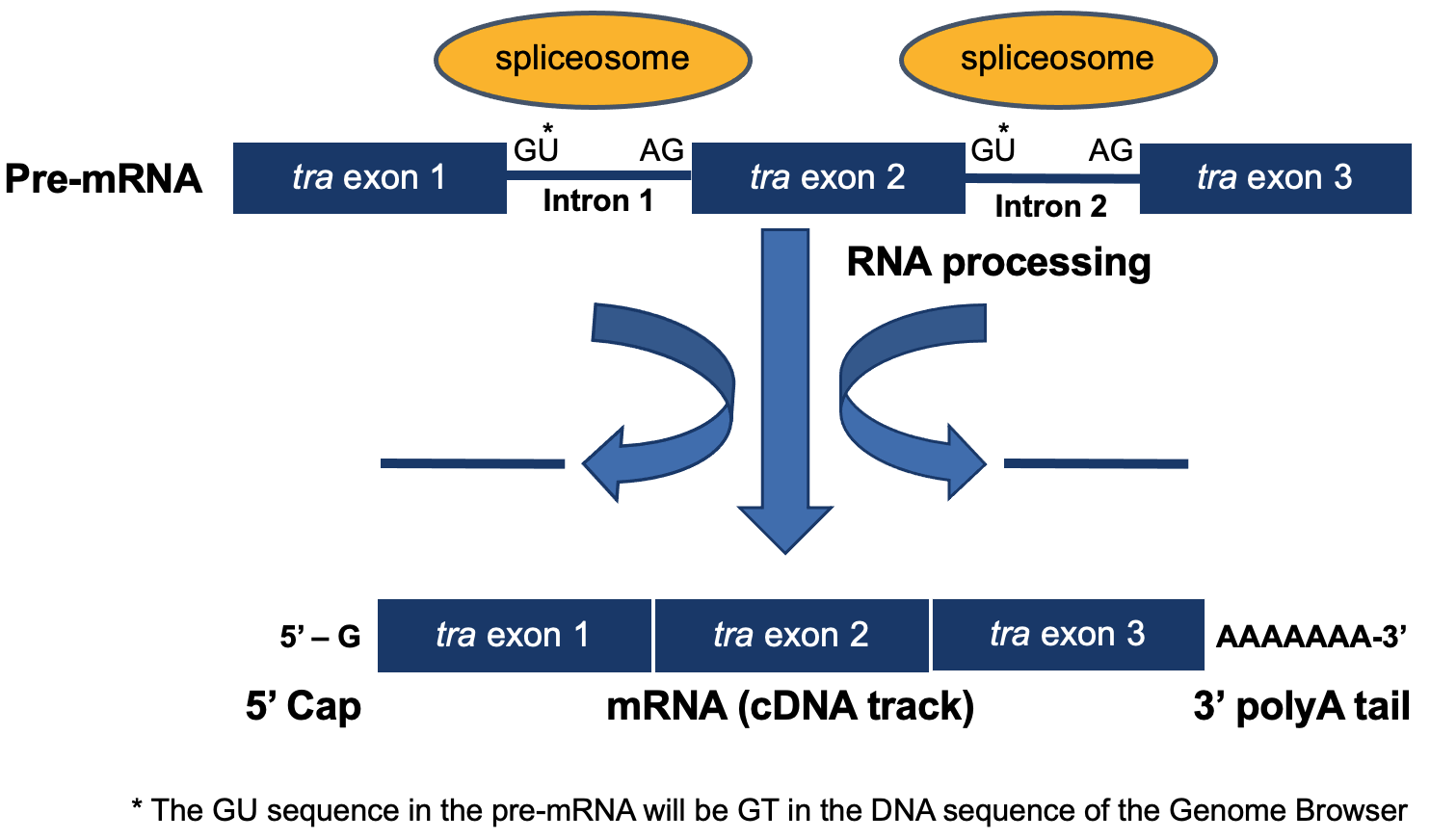


Figura Diagrama del procesamiento del ARNm mediante el cual el pre-ARNm es convertido en ARNm

El **primer paso** en el procesamiento del **pre-ARN mensajero** ocurre en el extremo 5’ del ARN mensajero. Recuerda que el ARN mensajero es sintetizado en dirección de 5’ a 3’, lo que significa que el extremo 5’ del ARN mensajero se sintetizó en primer lugar. Examinemos el inicio del gen *tra*. Introduce "**contig1:9,825-9,870**" en la sección de texto “chromosome range, or search terms, see examples” y luego pulsa el botón de “go”.

En el Módulo 2, determinamos que el **sitio de iniciación de la transcripción** (TSS) de la isoforma A de *tra* está en la posición **9,851**.

Para ver los TSSs que han sido anotados por el proyecto modENCODE, dirígete hacia abajo hasta llegar a la sección "Genes and Gene Prediction Tracks" y cambia el modo de visualización de la pista "TSS Annotations" a “**pack**”, después haz clic en “refresh”. Para localizar los TSSs, el proyecto [modENCODE](https://www.genome.gov/26524507/the-modencode-project-model-organism-encyclopedia-of-dna-elements-modencode) utiliza un método químico que primero detecta y etiqueta una estructura particular que se encuentra en el extremo 5’ de los transcritos (cap), luego aísla los transcritos que llevan la etiqueta, y por último localiza las secuencias correspondientes a esos ARNs en el genoma, un método llamado “CAGE” (Cap Analysis of Gene Expression).

Además, visualizaremos los datos de la pista "D. mel. cDNAs" (también bajo la sección de "Genes and Gene Prediction Tracks") escogiendo la opción de “**pack**”. Esta pista muestra la localización en el genoma *D. melanogaster* de las secuencias complementarias a los ADNcs (ADNs complementarios, producidos al copiar el ARN mensajero), los cuales fueron secuenciados por el “Berkeley Drosophila Genome Project” (BDGP). Haz clic en el botón de “refresh” (Figura 2). Ambas pistas nos proporcionan un análisis que se basa en la población de ARN, y si determinamos las posiciones de las secuencias de ARN podremos saber dónde empezó el transcrito.

Recuerda que en el Módulo 2 también vimos que el motivo TCAKTY, una señal de iniciación muy común, se encuentra justo río arriba, en la posición 9,834 (visualízalo utilizando la pista “Short Match”). Todas estas pruebas apuntan a la presencia de un TSS en esta región.

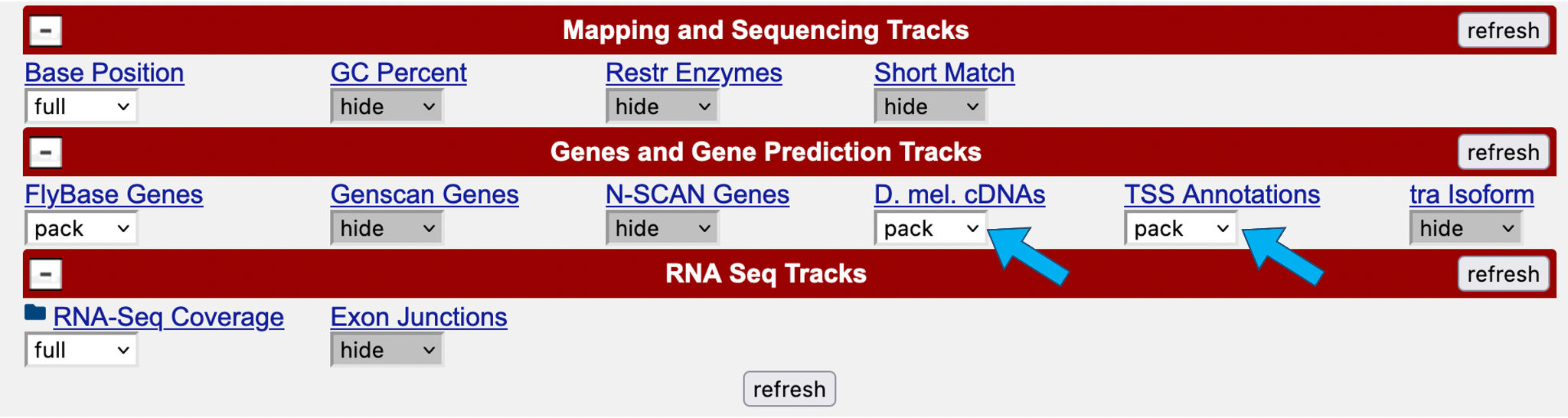


Figura Selecciona el modo de visualización “pack” en las pistas "D. mel. cDNAs" y "TSS Annotations".

La imagen actualizada del navegador genómico (Figura 3) nos muestra el extremo 5’ del transcrito del pre-ARN mensajero (en otras palabras, el punto de la iniciación de la transcripción) basado en el experimento CAGE (franja “modENCODE”), en el contexto de datos adicionales que refuerzan dicha selección. En esta punta del pre-ARN mensajero, un nucleótido con guanina metilada (7mG) ha sido añadido al nucleótido de la posición 9,851, formando así el “**5’ cap**”.Observa que este nucleótido adicional **NO** aparece en la secuencia del ADN. Es añadido DESPUÉS de que el transcrito ha sido sintetizado. Este es el **primer paso** en el **procesamiento del pre-ARN mensajero**: “capping”.

Timeline

Description automatically generated

Figura Adición del “5' cap" al extremo 5' del transcrito.

**Pregunta 1.** ¿Cuál es la coordenada del primer nucleótido transcrito?En la secuencia de ADN, ¿es una A, una C, una T o una G?

**Pregunta 2**. ¿Cuáles son las coordenadas del codón de iniciación que codifica para el primer aminoácido de la isoforma A del gen *tra*? (Supón que el marco de lectura es el +3)

**Pregunta 3.** La región del transcrito que va desde el punto de la iniciación de la transcripción hasta el nucleótido que está justo antes del codón de inicio se conoce como la región 5’ no traducida (5’UTR, por sus siglas en inglés) porque es parte del transcrito que no se traduce. ¿Qué longitud (en ribonucleótidos) tiene el 5’UTR?

El **segundo paso** en el procesamiento del pre-ARN mensajero es la **poliadenilación**.

1. Para ver el extremo 3’ del gen tra-RA, cambia la información en la sección de texto “chromosome range, or search terms, see examples” a "**contig1:10,633-11,000**" y luego pulsa el botón de “go”.

La poliadenilación significa que **muchos** (poli) **nucleótidos de adenina** (ribonucleótidos) son añadidos al extremo 3’ del pre-ARN mensajero **DESPUÉS** de que la transcripción ha terminado. Esto genera una cola de poli-A (normalmente de ~20 a ~250 As) en el ARN mensajero final que no está presente en la pista "Base Position" del Navegador Genómico. Esto se debe a que la cola de poli-A no existe en la hebra molde de ADN que sirve de plantilla para producir ARNm. El poli-A simplemente es añadido al ARN por una polimerasa especial que sintetiza una larga cadena de nucleótidos de adenina.

Nuestro análisis previo en el Módulo 1 mostró que el último exón codificante de tra-RA se encuentra en el marco de lectura +2 y el codón de terminación está en las coordenadas 10,754-10,756. Podemos utilizar el Navegador Genómico para determinar el punto donde termina el transcrito de tra-RA según indica la pista de ADNc (en azul). (Observa que la pista de ADNc concuerda bastante bien con la de “FlyBase Genes” aunque discrepan en señalar el punto exacto donde termina el transcrito.)

**Pregunta 4.** Según la pista de ADNc (en azul), ¿Qué longitud (en pares de bases) tiene esta región 3’ no traducida (3’UTR, por sus siglas en inglés)?

**Pregunta 5**. Amplía el campo de visión en el extremo 3’ del “FlyBase Genes” de forma que puedas leer la secuencia de ADN, cerca del final del transcrito. ¿Cuál es la cadena de adeninas más larga que observas en la secuencia de ADN?

**Pregunta 6.** ¿Los datos que has recogido hasta ahora reafirman la conclusión de que la secuencia de poli(A) observada en el ARN mensajero maduro no es codificada por el ADN?

1. Realiza una búsqueda “Short Match” de la señal de poli-A (AATAAA) siguiendo el protocolo que ya usaste en el Módulo 2. El resultado de esta búsqueda debería ubicar la señal poli-A en 10,818-10,823 (Figura 4). Como ya se mencionó en el Módulo 2, el transcrito es cortado entre 11 y 30 nucleótidos más allá de la secuencia de la señal de poli(A), y luego unas 150-200 adeninas son añadidas al pre-ARN mensajero. Los nucleótidos que están entre el codón de terminación y el final de la cola de poli-A componen el 3’UTR.

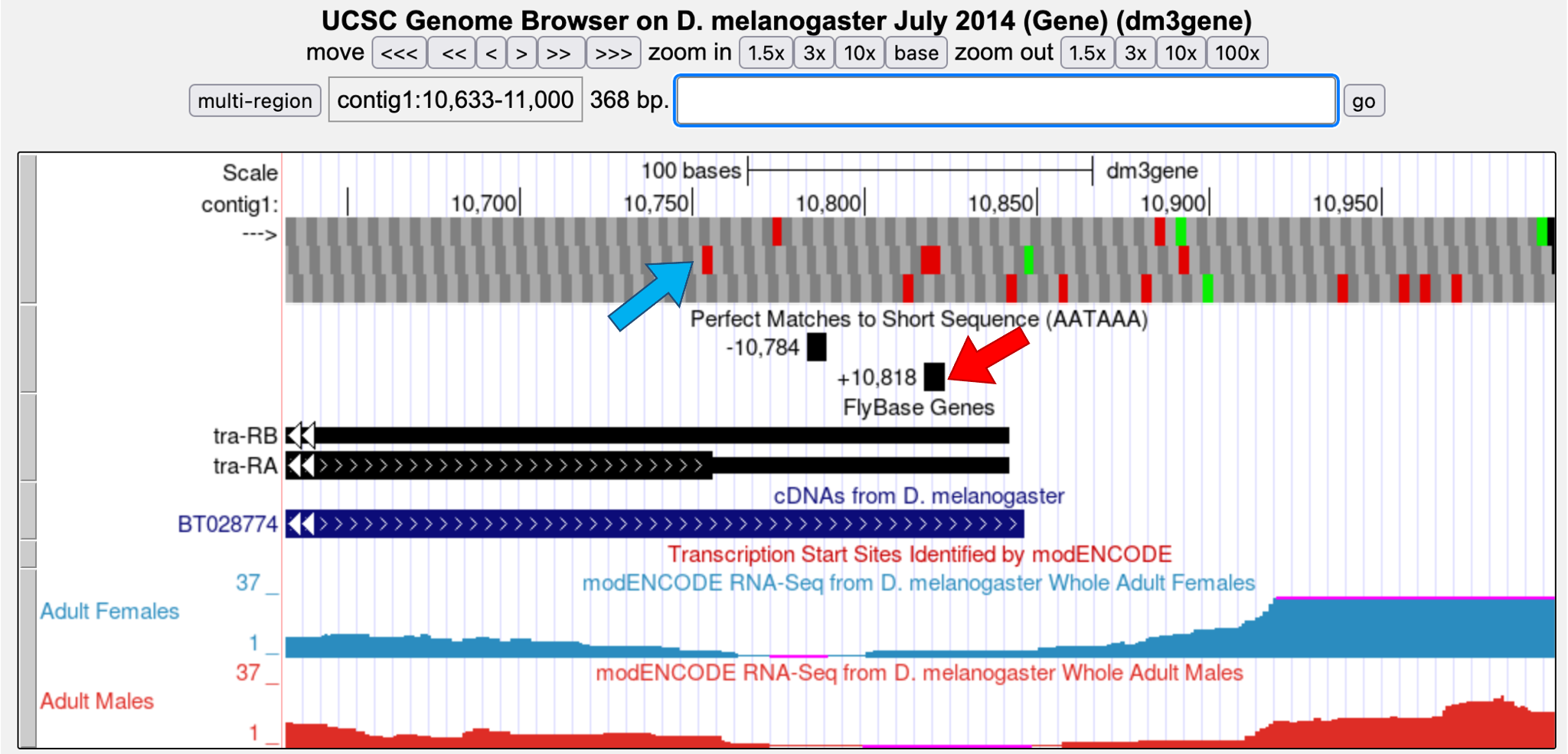


Figura El codón de terminación de la isoforma A de *tra* se encuentra en el cuadro de lectura +2 (flecha azul) y la señal de poli-A en 10,818-10,823 (flecha roja).

Podemos ver la secuencia de poliadenilación que normalmente se asocia con el ARN mensajero procesado examinando el ADNc (BT028774) previamente alineado con esta región.

1. Haz clic en la barra azul de "BT028774” bajo la pista "cDNAs from D. melanogaster" y haz clic en el enlace bajo la columna “SIZE” de la sección "BT028774/Genomic Alignments" (Figura 5).

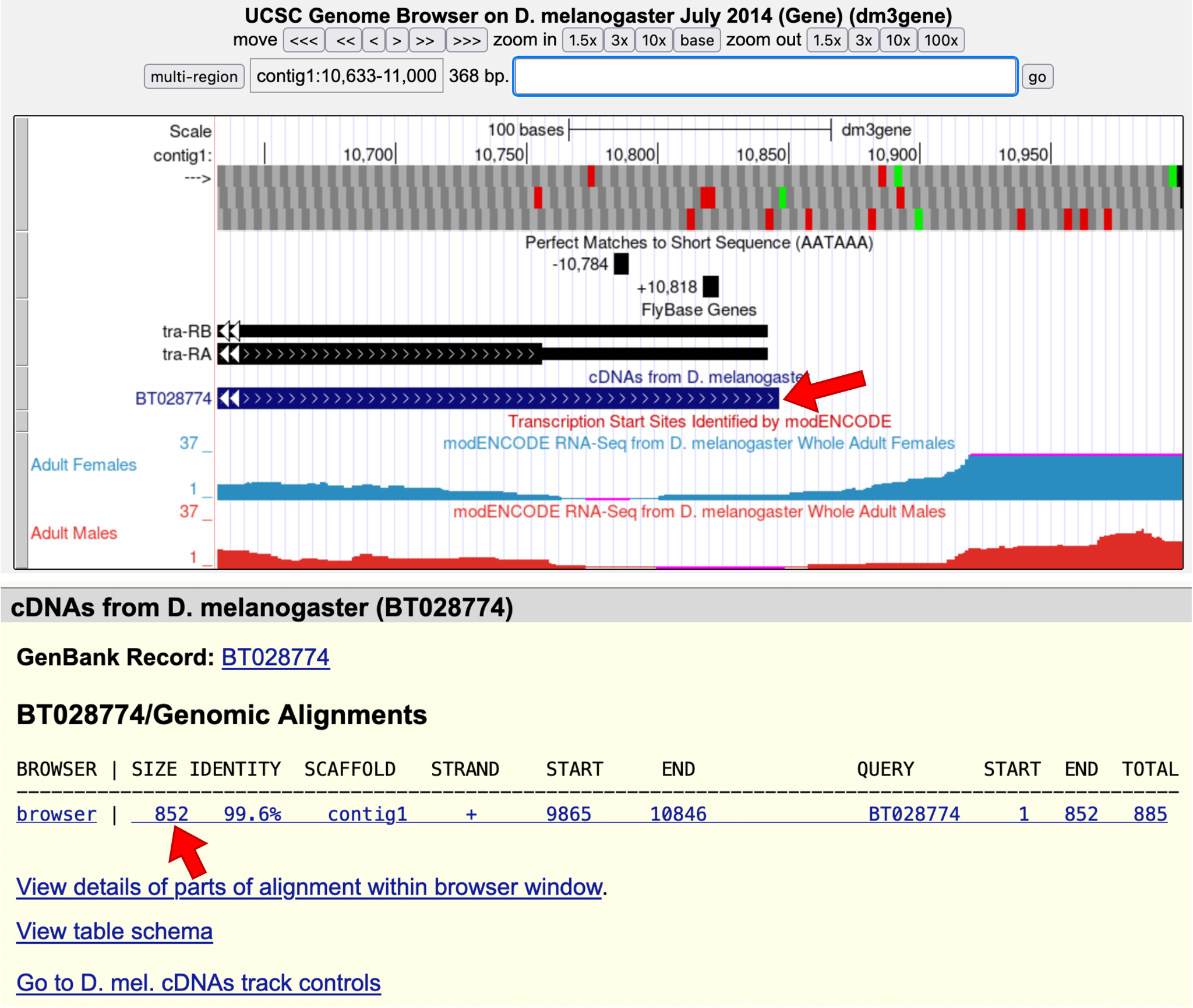


Figura Examen de la alineación del ADNc BT028774 de *D. melanogaster* con el contig1.

La siguiente figura muestra la alineación entre el ADNc BT028774 de *D. melanogaster* y su secuencia genómica en el “contig1” (Figura 6). Los nucleótidos que las dos secuencias tienen en común se muestran con letras mayúsculas de color azul, mientras que los nucleótidos que son diferentes se muestran con letras minúsculas de color negro. Las bases en color azul claro denotan el principio y el final de las regiones alineadas. La alineación “side-by-side” en la parte inferior de la imagen muestra los tres bloques de apareamiento entre el ADNc (arriba) y la secuencia del “contig1” (abajo) comprendidos en la región contig1:9,865-10846 y que se representaban con letras azules en la parte superior.



Figura Alineación del ADNc BT028774 de *D. melanogaster* con el final del contig1.

**Pregunta 7.** Dirígete hacia arriba hasta llegar a la secuencia “cDNA BT028774”. ¿A partir de qué coordenada (número en el ADNc) comienza la cadena de adeninas (en letras minúsculas de color negro)?

**Pregunta 8**. ¿Cuántos ribonucleótidos ‘A’ se añadieron al ARN mensajero de *tra* (representados en el ADNc)?

**Pregunta 9.** Localiza la señal de poliadenilación **AATAAA** en la secuencia del ADNc. ¿Cuántos nucleótidos más allá de la última A de la señal de poliadenilación ves que comienza la cadena de poli-A? (Deberán ser entre 11-30 nucleótidos.)

El último paso en el procesamiento del pre-ARN mensajero es la eliminación de intrones y el empalme de exones adyacentes para formar un marco de lectura abierto y continuo de manera que el ARN mensajero esté listo para ser traducido a proteína.

1. Introduce "**contig1:9,870-10,170**" en la sección de texto “chromosome range, or search terms, see examples” y pulsa el botón de “go” para navegar al primer intrón del transcrito de tra-RA (Figura 7).

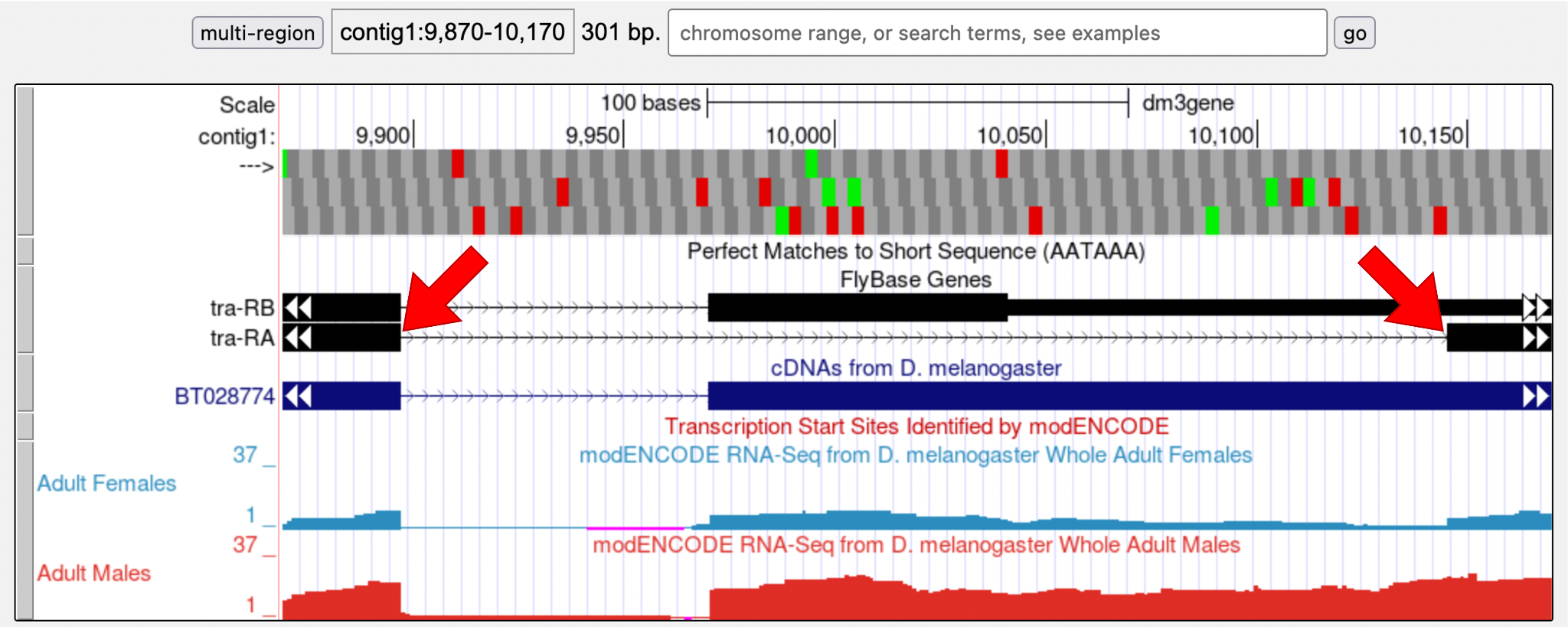


Figura Región genómica que rodea el primer intrón (flechas rojas) de tra-RA.

1. Ajusta el campo de visión para acercarte al final del primer exón de tra-RA.

**Pregunta 10.** ¿Cuáles son los dos nucleótidos que se encuentran justo después del final del primer exón de tra-RA? \_\_\_\_\_\_

Repite este análisis e identifica los dos nucleótidos del principio del segundo intrón de tra-RA. \_\_\_\_\_

Estos dos nucleótidos son la señal del **empalme** que se encuentra en el extremo 5’ de cada intrón; representan las primeras dos bases del intrón, y son conocidos comúnmente como el sitio donante (o sitio de empalme 5’).

**Pregunta 11.** ¿Con qué base termina el exón 1? \_\_\_\_\_\_\_\_\_\_\_\_\_

1. Aléjate y luego acércate al inicio del segundo exón de tra-RA usando el botón de “zoom”.

**Pregunta 12.** ¿Cuáles son los dos nucleótidos que se encuentran justo antes del principio del exón 2 de tra-RA? \_\_\_\_\_

1. Aléjate y luego acércate al inicio del tercer exón de tra-RA usando el botón de “zoom”.

**Pregunta 13.** ¿Cuáles son los dos nucleótidos que se encuentran justo antes del principio del exón 3 de tra-RA? \_\_\_\_\_\_

Estos dos nucleótidos son la señal del **empalme** que se encuentra en el extremo 3’ de cada intrón; representan las últimas dos bases del intrón, y son conocidos comúnmente como el sitio aceptor (o sitio de empalme 3’).

**Pregunta 14.** ¿Con qué base comienza el exón 2 de tra-RA? ¿Cuál es la coordenada? \_\_\_\_\_\_\_\_\_\_\_\_

# Conclusiones

En este módulo hemos estudiado los tres pasos claves que comprenden la conversión del pre-ARN mensajero a ARN mensajero maduro:

* La adición de la estructura **caperuza** al extremo **5’** (“5’ cap”)
* La adición de la **cola poli(A)** al extremo **3’** (“3’ poly(A) tail”)
* La eliminación de intrones mediante **empalme**
  + Observa que los intrones son eliminados en este proceso y los exones adyacentes se unen para conformar el ARN mensajero maduro.

Después del procesamiento del ARN mensajero, el ARN mensajero maduro (tra-RA) puede salir del núcleo para ser traducido a proteína (tra-PA) por los ribosomas del citoplasma.