



Módulo 4: Empalme – Eliminación de intrones del ARN mensajero por empalme

Meg Laakso

(traducido por Mario Naranjo, Karla González-Cruz y Paula Croonquist)

Objetivos

- Identificar el sitio donante y el aceptor de empalme que están mejor justificados/validados por predicciones realizadas en base a la secuencia del ARN y TopHat
- Utilizar las secuencias del sitio canónico donante y aceptor para identificar los límites de los intrones y exones.

Prerrequisitos

- Comprender los Módulos 1-3 titulados Entendiendo Genes Eucaryotas: <https://thegep.org/ueg/>
- Definir el pre-ARN como el resultado del proceso de transcripción; este mensaje incluye exones e intrones

Instrucción en el Aula

- Examinar el procesamiento del pre-ARN usando las imágenes correctas del módulo 3 o libro de referencia.
- Investigación 1: Familiarizarse con la información provista por las secuencias del ARN (RNA-seq)
 - Examinar el consenso de las secuencias del sitio donante y aceptor de empalme.
- Investigación 2: Encontrar el sitio donante y aceptor de empalme del intrón 1.
 - Examinar el empalme de ARN del intrón 1 usando el gen tra-RA como ejemplo
 - Investigación 3: Encontrar el sitio donante y aceptor de empalme del intrón 2.
 - Discutir la medida del pre-ARN comparada con las del ARN maduro/procesado
 - Identificar las isoformas de los distintos sitios de comienzo de transcripción (TSS) o formas de empalme alternativo

Videos adjuntos

- Video de secuencia de ARN (RNA-Seq) y TopHat: <https://youtu.be/XD-egRcHYL4>
- Video de genes e isoformas: <https://youtu.be/ce6nVSiiV7I>

Índice de Contenidos

Investigación 1: Examinación de la secuencia de ARN (RNA-Seq)	2
Investigación 2: Identificación de sitios de empalme	7
Investigación 3: Identifique los sitios de donante de empalme de 5' y de aceptación de empalme 3' para el intrón 2	10
Asignación Parte 1: Identificando sitios de empalme para el intrón 1	13
Tarea Parte 2: Identificar los sitios donantes de empalme de 5' y de aceptación de empalme 3' para el intrón 2	14

Investigación 1: Examinación de la secuencia de ARN (RNA-Seq)

Continuaremos enfocándonos en la isoforma A del gen *transformer* (al cual nos referimos como tra-RA). Aquí nos enfocaremos en la información de los experimentos que investigan el ARN en células. Esta información se puede usar para ayudarnos a identificar exones e intrones del gen a estudiar.

Todos los ARNs en la célula se llaman colectivamente el transcriptoma, porque casi todo el ARN se produce por transcripción del molde de ADN. (En algunos casos, el ARN se puede producir con ARN como molde). El transcriptoma incluye ARN mensajero, ARN ribosómico, ARN de transferencia y otros ARNs que tienen funciones específicas en la célula.

El ARN puede ser purificado/extraído desde células u organismos enteros como *Drosophila* y convertido en ADN, y secuenciado. Primero, el ARN mensajero extraído que ha sido completamente empalmado/procesado es copiado a ADN con la enzima llamada transcriptasa inversa. Copias cortas del ADN complementario copiado son secuenciados y estos segmentos son mapeados/comparados en contra del genoma. Analizando esta información es posible saber cuáles y cuántos ARNs se han sintetizado.

Esta técnica poderosa nos permite ver cuando y donde los diferentes genes son expresados. Este tipo de información puede ayudar a los investigadores y clínicos a saber cuáles genes son expresados en diferentes tipos de cáncer, por ejemplo. Aquí y en los siguientes módulos vamos a usar datos de secuencia de ARN (RNA-Seq) para explorar como el gen llamado *transformador (tra)* está expresado en machos y hembras de *Drosophila*.

La información de secuencia de ARN puede indicar donde la transcripción ocurre con alta precisión. El número de copias de secuencia de RNA que mapean un sitio específico indica cuántas copias de ARN existen en la muestra. Recuerden que en el module 3 el número inicial de copias de ARN son procesadas rápidamente para eliminar los intrones. Por eso el total de ARN de una célula, las secuencias de exones serán mucho más abundantes que las secuencias de intrones. Usaremos información de la secuencia de ARN para buscar los límites entre exones e intrones del tra-RA, la isoforma del gen *tra* que se expresa en las moscas hembra.

1. Abre la ventana del navegador de internet y anda al sitio de GEP UCSC Genoma Browser Mirror <https://gander.wustl.edu/>.
2. Selecciona el enlace de Navegador Genómico o “Genome Browser” del panel del lado izquierdo. (Instrucciones detalladas de como navegar la página del Genome Browser Gateway son provistas en el [Módulo 1.](#))

Recordar: Usa el menú desplegable del navegador del Genome Browser para seleccionar lo siguiente.

1. Selecciona “D. melanogaster” debajo del sección de texto de “REPRESENTED SPECIES”.
2. Selecciona "July 2014 (Gene)" en el sección de texto "D. melanogaster Assembly".
3. Entra "contig1" en el sección de texto "Position/Search Term".
4. Después cliquea “GO” y la pantalla de abajo aparecerá (Figura 1). Como recordarás, esta sección de ADN tiene 11,000 pares de bases y es parte del brazo izquierdo del cromosoma 3, el cual tiene casi 28,100,000 pares de bases. Si la ventana de tu navegador está mostrando otras vías de evidencia, resetea cliqueando en “default tracks.”

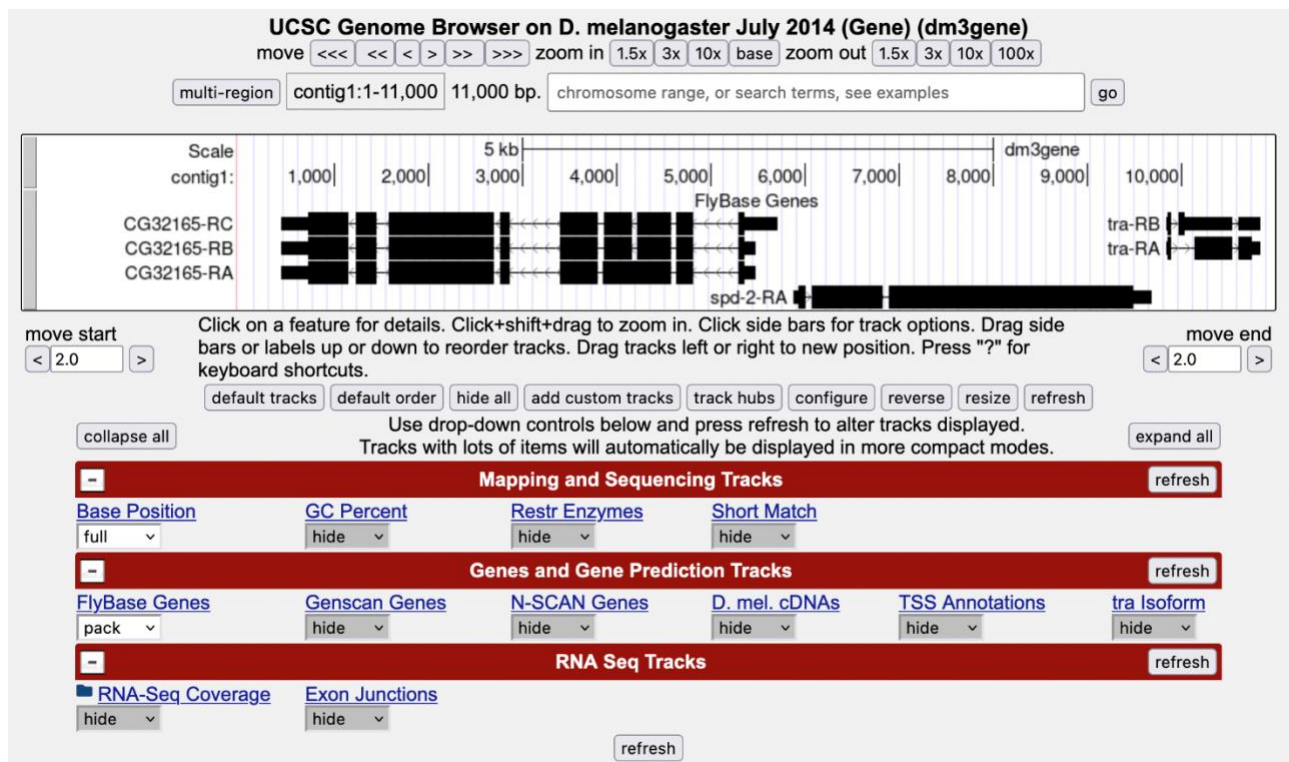


Figura 1 Una captura de pantalla del proyecto “contig1”.

- Comienza estableciendo las vías/pistas de evidencia que queremos ver. Presiona el botón "hide all". Entonces habrá solamente las pistas de evidencia que te proveerán la información necesaria para esta investigación. Fíjate que no vas a poder ver la secuencia de ADN hasta que la agrandes con el "zoom in."
- Cambia la cobertura de la secuencia de ARN (RNA-Seq) a "full," y cliquea "refresh."

Veras histogramas en azul y rojo representando la secuencia de ARN generada usando muestras de ARN de hembras y machos adultos respectivamente. Esto nos ayuda a inferir patrones de síntesis de ARN, y buscar similitudes y diferencias entre los sexos. ¡Ambas similitudes y diferencias son aparentes!

- Acércate para mirar solamente al gen *tra* gene tipeando **contig1:9,800-10,900** en la sección de texto "chromosome range, or search terms" (Figura 2).
- Ahora estamos mirando la región del cromosoma 3 donde se encuentra el gen *tra*. Compara los histogramas rojos y azules de las hembras y machos adultos. Nota que hay números en el eje vertical que muestra cuántas copias de ARN (secuencias de copias) se encuentran en esa posición.

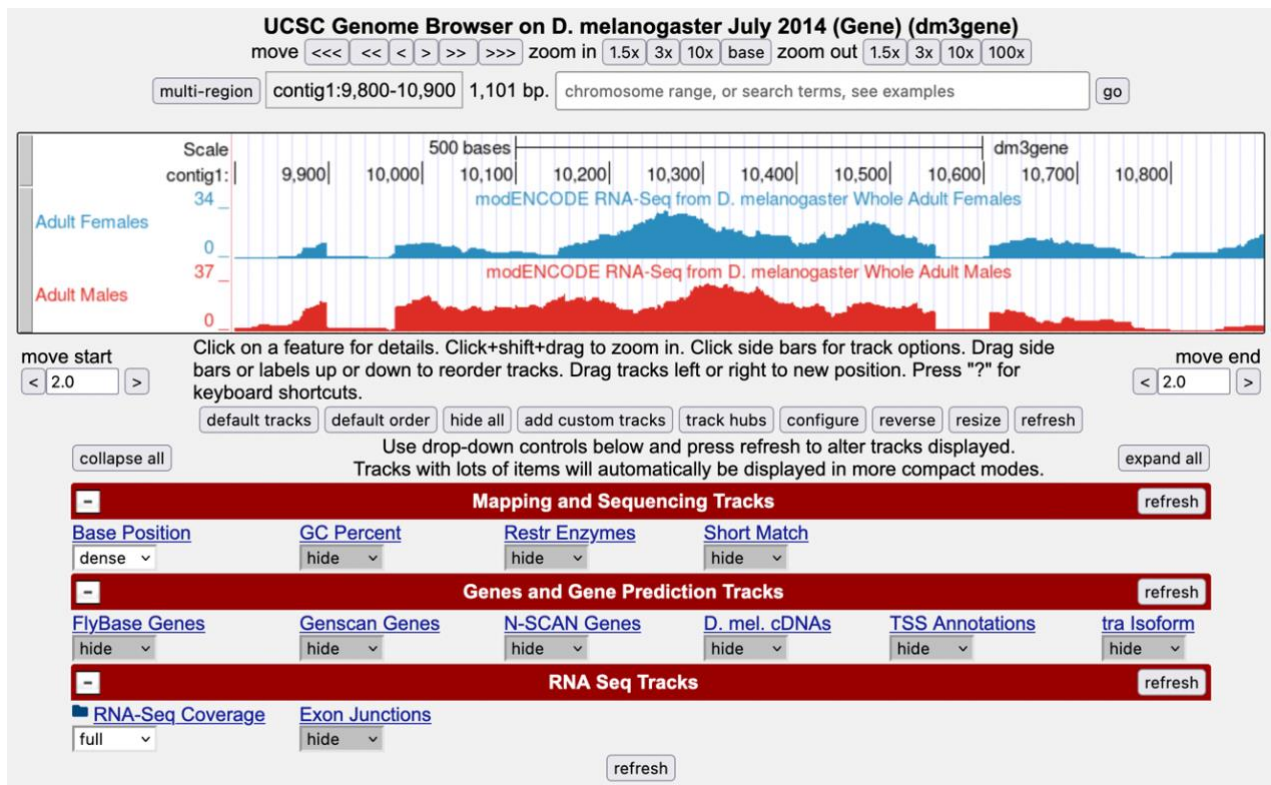


Figura 2 Histogramas representando RNA-Seq data en transformador (*tra*).

Pregunta 1. Enumera dos formas en las que los histogramas son similares.

a. _____

b. _____

Pregunta 2. Enumera una forma en que los histogramas difieren (aparte del color).

a. _____

En resumen: el histograma azul representa el ARN mensajero secuenciado de las hembras de la mosca de fruta y representa la forma del gen tra denominado isoforma A (tra-RA). El histograma rojo representa los datos de RNA-Seq de moscas de fruta masculinas. El ARN mensajero en los machos es diferente al de las hembras, y esta segunda isoforma del gen tra se llama isoforma B (tra-RB).

Pregunta 3. Recuerda que nuestra otra evidencia (ve el Módulo 3) indica que el tra-RA se extiende desde 9,851 a 10,846. ¿Cuántos espacios vez en los histogramas en este intervalo? _____

Cada espacio corresponde a un intrón; hay muy poca señal de **(RNA-Seq)** en esa parte del gen porque el ARN intrónico se empalmó antes de que se recogiera y se secuenció el ARN mensajero. [**Clue: Nota que en el primer intrón, el histograma de RNA-seq para tra-RA (azul, específico para las hembras) tiene un tramo largo de barras bajas. (Las barras azules son mucho más cortas que las barras para los exones expresados, y mucho más cortas que la región correspondiente en el histograma de ARNs de las moscas macho.) Esta región es parte del intrón en moscas hembra, y refleja el evento de empalme alternativo que crea la isoforma A que es específicamente de hembras. Sin embargo, porque una fracción pequeña de los RNA-Seq copias es derivada de pre-ARNs mensajeros, podemos encontrar secuencias intrónicas los datos de RNA-seq. Podemos usar un subconjunto de copias de RNA-Seq que abarcan exones adyacentes para verificar que esta región es parte del intrón en los ARNs mensajeros producidos en moscas hembra (e.g., predicciones del empalme de exones derivadas de las copias de RNA-seq empalmados, explicado en el Módulo 4, Investigación 2 siguiente).*]

Pregunta 4. ¿Cuántos exones vez? _____

Recuerda que el exón es la parte "expresada" del gen y habrá un pico agudo o ancho en el histograma de RNA-Seq.

Pregunta 5. ¿Las hembras o los machos producen más ARN mensajero transformador o lo expresan aproximadamente al mismo nivel? _____

Ahora que has examinado la evidencia tú mismo, regresemos y revisemos.

- **Brechas (Gaps) (intrones)** - La Figura 3 es una captura de pantalla del navegador del genoma con las brechas identificadas con un círculo. Tenga en cuenta que hay 2 brechas en el femenino y 2 en el masculino. La primera brecha en el femenino parece un poco extraña porque no tiene límites limpios, lo que sugiere una población mixta de transcripciones procesadas.
- **Exones:** Se han dibujado corchetes debajo de los datos de RNA-Seq para moscas hembra adultas, correspondientes a los tres exones que se expresan.
- **Isoformas:** Tenga en cuenta que la isoforma A y la isoforma B del gen tra son el resultado de un empalme alternativo. Para otros genes, las isoformas pueden tener diferentes sitios de inicio de la transcripción. Las isoformas de un gen siempre tienen diferentes secuencias de ARN mensajero, pero pueden tener la misma secuencia de proteínas. Para obtener más información sobre las isoformas y los genes, vea el video [Genes and Isoforms video](#).

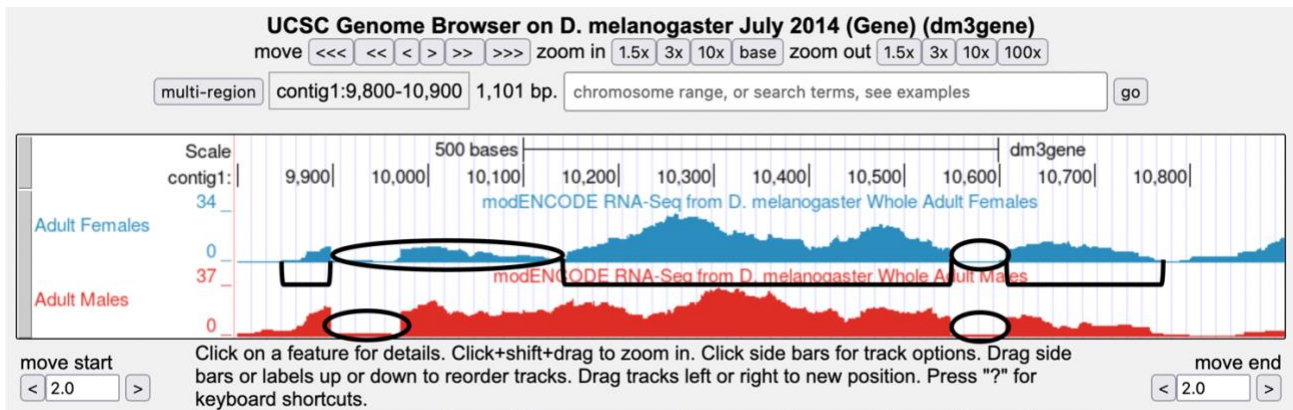


Figura 3 En esta captura de pantalla del gen tra, los intrones se han identificado con un círculo y los exones tienen corchetes debajo.

Pregunta 6. Con la información que has obtenido hasta ahora, haz un diagrama de la isoforma tra-RA (específica de hembras) con 3 exones y 2 intrones. Representa los exones como cajas rectangulares y los intrones como líneas que conectan las cajas. Numera cada exón e intrón (comienza desde la izquierda con el "exón 1").

Pregunta 7. ¿Dónde está el promotor en relación con los exones y los intrones?
 _____ . En tu diagrama, marca el sitio de inicio de la transcripción putativo en su diagrama con una flecha apuntando en la dirección de la transcripción.

Investigación 2: Identificación de sitios de empalme

Nuevamente nos centraremos en tra-RA para identificar los sitios de empalme, es decir, los nucleótidos exactos donde se produce el empalme para eliminar los intrones del pre-ARN mensajero.

Antecedentes:

Dos programas de computación llamados TopHat y Bowtie, utilizan los datos de RNA-Seq para representar gráficamente las uniones de exones. El gráfico resultante en el navegador del genoma parece casualmente algo así como una pequeña pajarita (dos pequeñas cajas conectadas por una línea delgada) (Figura 4). Los recuadros representan el ARN mensajero secuenciado (los exones) y la línea representa un espacio (el intrón). La unión del exón se puede inferir cuando la primera parte de un fragmento secuenciado de la población de ARN coincide (por ejemplo) con las posiciones de ADN 50-100 y la segunda parte del mismo fragmento coincide con las posiciones de ADN 200-250; ¡el ARN de las posiciones 101-199 debe haberse tomado del medio!

Las secuencias cortas están presentes al principio y al final de cada intrón que permiten que el espliceosoma, la maquinaria molecular que corta los intrones elimine con precisión el intrón, dejando solo las secuencias del exón en el ARN mensajero maduro. Los dos primeros nucleótidos del intrón son el sitio donante de empalme y casi siempre los nucleótidos "GT". Los dos últimos nucleótidos del intrón son el sitio aceptor de empalme y casi siempre los nucleótidos "AG". (Recuerde sus observaciones en el Módulo 3.) Para obtener más información sobre RNA-Seq y la búsqueda de uniones de empalme, vea el [video RNA-Seq y TopHat](#).

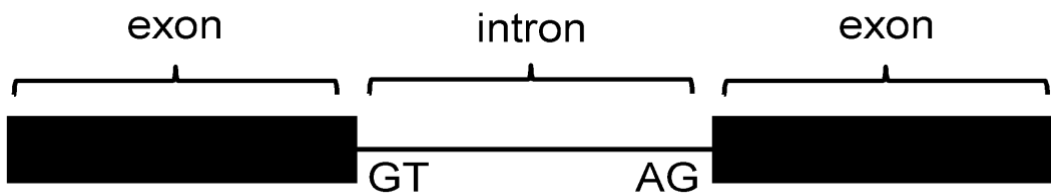


Figura 4 Un diagrama de uniones intrón-exón.

9. Usando la misma página del Navegador Genómico, reinicia el Navegador haciendo clic en "**Hide all**". Luego, abre las pistas que proporcionarán la información que queremos para la Investigación 2. (Nota: Si te detuviste después del Módulo 4: Investigación 1, entonces puedes volver a la página 2 para recordar cómo llegar a la página del Navegador Genómico.

10. Cambia la posición de base a "**dense**."

Ten en cuenta que no podrás ver la secuencia de ADN hasta que se "acerque". "**Zoom in**"

11. Cambia la cobertura de RNA-Seq a "**Full**", luego haga clic en "**Refresh**".

Verás nuevamente los histogramas azul y rojo que representan los datos de RNA-Seq (que indican la cantidad de ARN mensajero sintetizado) en mujeres y hombres, respectivamente. Nos centraremos en el histograma azul (mujeres adultas) nuevamente.

- Como hicimos en el Módulo 3, personalizemos la pista RNA-Seq configurando el sección de texto "Escala de vista de datos" ("Data view scaling") para "usar la configuración del rango de visualización vertical" ("use vertical viewing range setting") y el sección de texto "máximo" ("max") debajo de "Rango de visualización vertical" ("Vertical Viewing range") en 37. Recuerda que accedes a estas configuraciones haciendo clic en el enlace "RNA-Seq Coverage" debajo de la barra roja de RNA-Seq Tracks.

- Cambia "Exon Junctions" a "Full", luego haz clic en "Refresh".

Las cajas rectangulares unidas por una delgada línea negra te ayudarán a identificar los límites exón-intrón. Nuestro visor de salida gráfica se verá como la captura de pantalla a continuación (Figura 5).

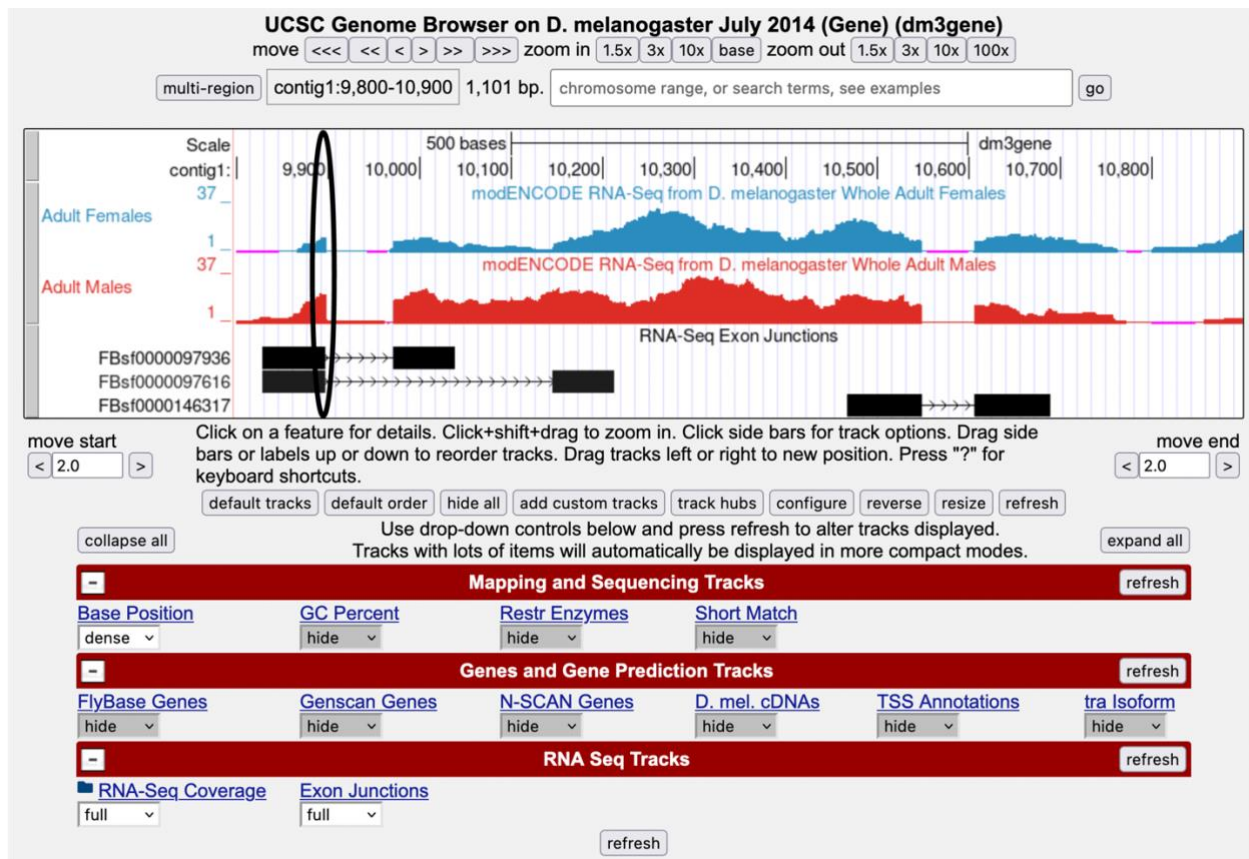


Figura 5 Vista de los datos de tra RNA-Seq con el sitio donante de empalme en el intrón 1 identificado con un círculo.

- Amplía (**Zoom in**) el área que está encerrada en un círculo: haz clic y arrastre el cursor justo encima de los números, o usa los botones de zoom.

15. Configura la pantalla de modo que pueda ver entre 15 y 20 nucleótidos, como se muestra en el siguiente ejemplo (Figura 6).

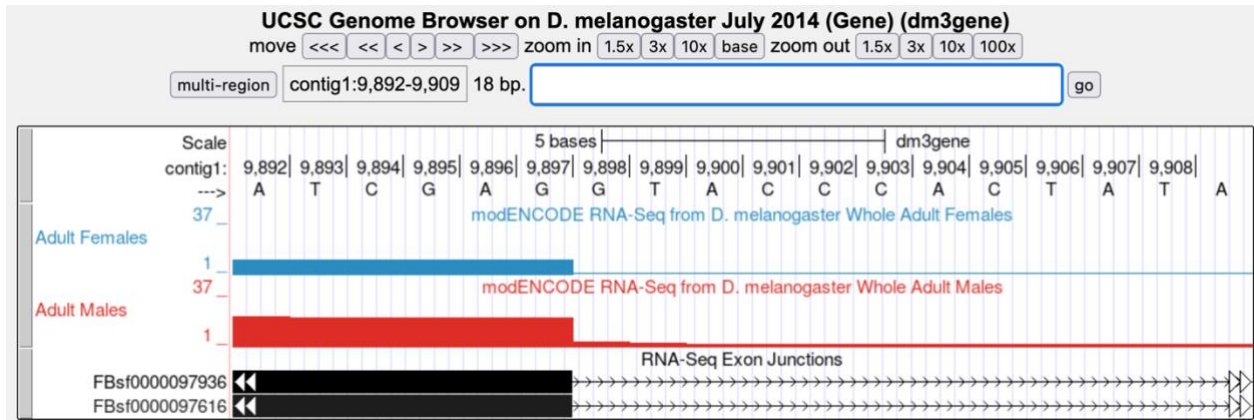


Figura 6 Datos de TopHat (en negro) para machos y hembras adultos que indican una unión de exón.

El histograma azul se detiene al final del exón 1. Los últimos tres nucleótidos del exón 1 son "G-A-G".

Pregunta 8. ¿Cuál es la coordenada del último nucleótido del exón 1 de tra-RA? _____

Pregunta 9. ¿Cuáles son los dos primeros nucleótidos del intrón 1 tra-RA? _____

Esto se denomina sitio de empalme 5' o "sitio donante de empalme" "**splice donor site.**"

16. Aléjate para que puedas ver todo el gen tra. Usemos TopHat para ayudarnos a encontrar el extremo 3' del intrón. Examina la pista **Exón Junctions**. La primera unión intrón-exón predicha por TopHat (negro) parece alinearse con los datos del histograma rojo de los machos; la segunda unión se alinea mejor con los datos del histograma azul de las mujeres (Figura 7).

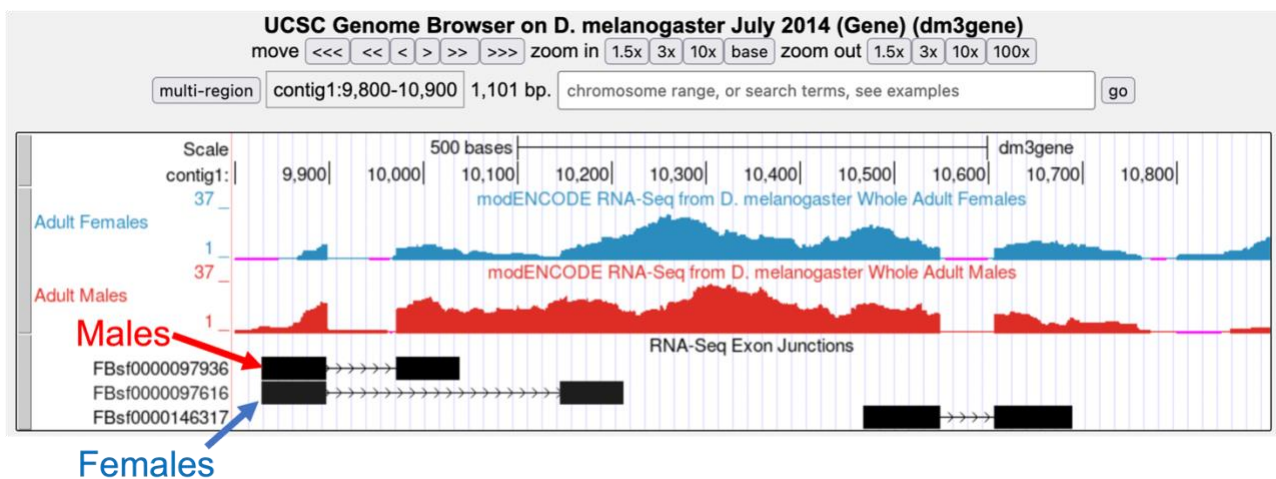


Figura 7 Datos de TopHat y RNA-Seq para hombres y mujeres.

17. Examinemos el extremo 3' del intrón 1 más cerca. Cambie el seccion de texto "ingresar posición o términos de búsqueda" ("chromosome range, or search terms") a **contig1: 10,140** y luego haga clic en "ir" ("Go"). Reduzca el zoom 10x y luego 3x (Figura 8).

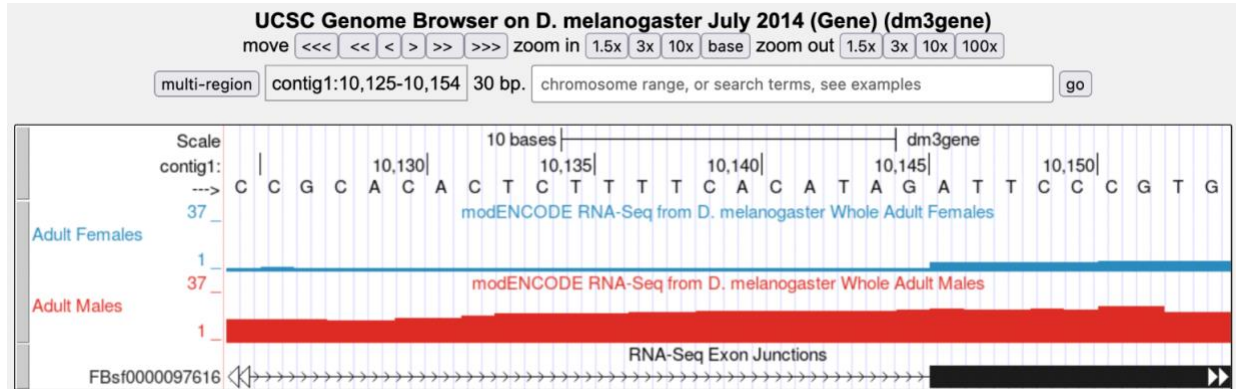


Figura 8 Pantalla centrada en la unión entre el intrón 1 y el exón 2 de la isoforma tra-RA (específico de hembras).

Pregunta 10. ¿Cuáles son los últimos dos nucleótidos del intrón 1 de tra-RA? _____

Esto se denomina el sitio de empalme 3' o "sitio de aceptación de empalme".

Pregunta 11. ¿Cuál es la coordenada del primer nucleótido del exón 2 de tra-RA? _____

Investigación 3: Identifique los sitios de donante de empalme de 5' y de aceptación de empalme 3' para el intrón 2

Podemos utilizar el mismo acercamiento para mapear los límites exón-intrón del segundo intrón.

Repasa los pasos #2-5 en la Investigación 2. Luego, repite el proceso para contestar las siguientes preguntas sobre el sitio de donantes de empalme 5' y el sitio de aceptación de empalme 3' para el tra-RA específico femenino del intrón 2.

18. Aumenta hacia la secuencia que rodea el sitio de donantes de empalme 5' para el intrón 2 del tra-RA (**específica de hembras**).

Pregunta 12. ¿Cuáles son los últimos tres nucleótidos del exón 2 de tra-RA? _____

Pregunta 13. ¿Cuál es la coordenada del último nucleótido del exón 2 de tra-RA? _____

Pregunta 14. ¿Cuáles son los primeros dos nucleótidos del intrón 2 de tra-RA? _____

19. Aléjate según sea necesario para que puedas ver todo el intrón 2. Utiliza TopHat para encontrar el extremo 3' del intrón 2.

20. Presiona y arrastra para que el final del intrón 2 este centralizado en la pantalla. Luego, aumenta la pantalla para que puedas ver la secuencia de los nucleótidos.

Pregunta 15. ¿Cuáles son los últimos dos nucleótidos del intrón 2 de tra-RA? _____

Pregunta 16. ¿Cuál es la coordenada del primer nucleótido del exón 3 de tra-RA? _____

Pregunta 17. Utilizando la información adquirida, realiza una imagen gráfica de la isoforma de tra-RA (**especifica de hembras**) con tres exones y dos intrones. Numera cada exón e intrón a las coordenadas de ADN correspondientes. Agrega las coordenadas para el primer y último nucleótido de los exones que has encontrado hasta ahora. Añade las secuencias de los sitios donantes y aceptación de empalme en las ubicaciones apropiadas.

Pregunta 18. ¿Dónde crees que el promotor está localizado en relación con su modelo genético? ¿Qué evidencia tienes que respalda su idea, utilizando las pistas de evidencia que hemos mostrado (la posición de la base, la cobertura de RNA-Seq, los cruces de exones)?

Pregunta 19. ¡Pregunta de bono! Respalda tu hipótesis mediante la recopilación de datos adicionales. Recuerda nuestras exploraciones en los Módulos 2 y 3. Es posible que desees abrir las pistas "**D. mel. cDNAs**," y "**TSS Annotations**," ambos en su **totalidad**. ¿Qué tipo de evidencia es mostrada por cada una de estas pistas (consulte el Módulo 2)? Finalmente, para ver tra-RA como actualmente se encuentra anotado en FlyBase, abre la pista "**tra Isoform**" en su **totalidad**, o para ver ambas isoformas abre la pista "**FlyBase Genes**" en **paquete**. ¿Estos resultados avalan tu modelo? ¿Queda alguna ambigüedad?



Asignación Módulo 4: Determinando sitios de empalme para el gen *spd-2*

Meg Laakso

- Abre una pestaña nueva en su navegador y ve al sitio de GEP Navegador Genómico Espejo UCSC (Genome Browser Mirror) en <https://gander.wustl.edu/>.
- Sigue las instrucciones dadas en el Módulo 1 para poder navegar hacia el proyecto del contig1 en el ensamblaje de *D. melanogaster* "July 2014 (Gene)".

Recordar: utiliza los menús desplegables en el Navegador Genómico para seleccionar lo siguiente:

- Debajo de la sección de texto "REPRESENTED SPECIES" = "D. melanogaster"
- Debajo de la sección de texto "D. melanogaster assembly" = "July 2014 (Gene)"
- En la sección de texto "Position/Search Term" = "contig1"

Posteriormente, presione "GO" y aparecerá la pantalla de abajo (Figura 9). Como recordarás, esta sección de ADN tiene 11,000 pares de bases de largo y es parte del brazo izquierdo del cromosoma 3, que mide unos 28,100,000 pares de bases de largo. Si la ventana de tu navegador muestra otras pistas de evidencia, restablece presionado en "default tracks".

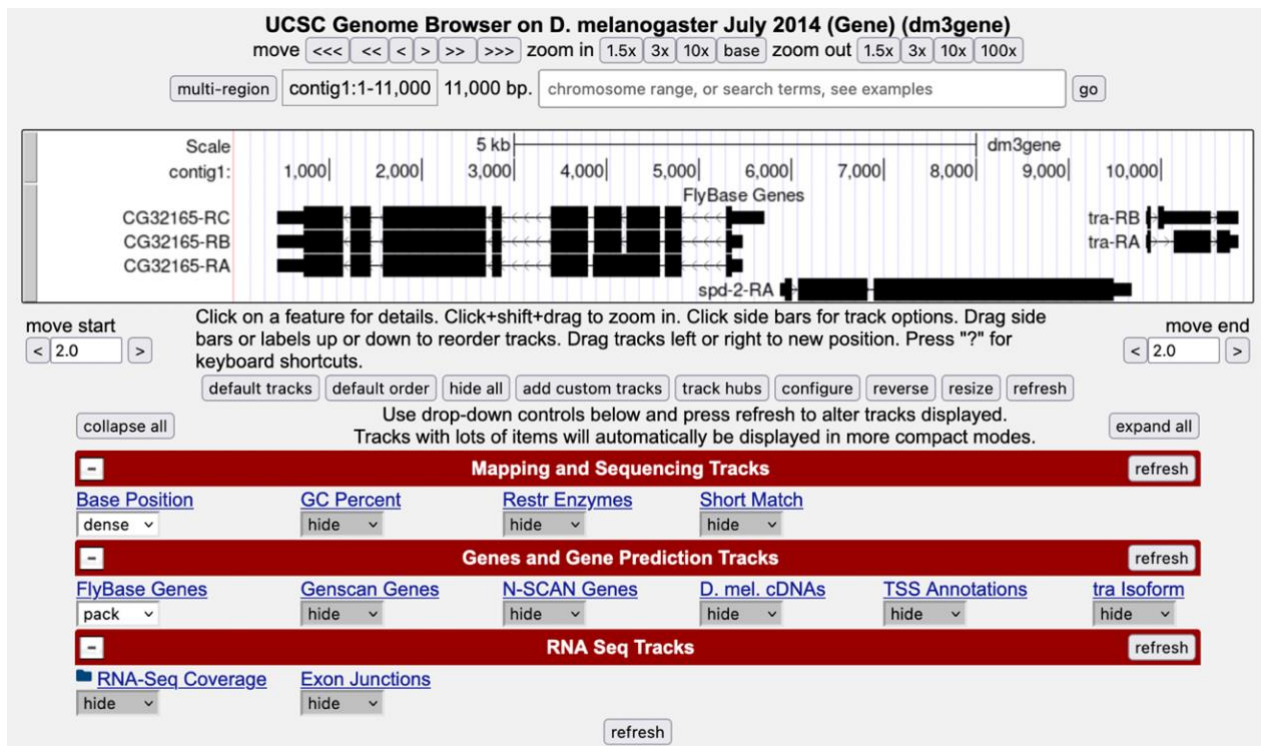


Figura 9 Captura de pantalla del proyecto "contig1".

23. Comencemos por establecer las pistas de evidencia que deseamos ver. Presiona el botón **“hide all”**. Seguidamente, abre solo las pistas que te proveerán datos para esta investigación.

- **Base position "dense"** - Debes tener en cuenta que no podrás ver la secuencia de ADN hasta que te acerques. Si la posición de la base se cambia a **"full"**, también puedes ver las pistas de aminoácidos.
- **RNA-Seq coverage "full"** - Verás histogramas en azul y rojo representando la secuencia de ARN generada usando muestras de ARN de hembras y machos adultos, respectivamente.
- **Exon Junctions "full"** - Observar los cruces de exones te ayudará a encontrar los sitios de donante y aceptación de empalme.
- **D. mel. cDNAs "full"** - Esta pista fue utilizada extensivamente en módulos anteriores y es útil para confirmar la secuencia del ARN mensajero maduro. Recuerda que un ADN complementario es una copia de ADN de un ARN mensajero.

Acércate para ver sólo el gen *spd-2* entrando "contig1:5,750-9,800" en la sección de texto "chromosome range, or search terms".

Estaremos mirando la región del cromosoma 3 donde el gen *spd-2* gen está localizado.

Pregunta 1. ¿Cuántos exones tiene *spd-2*? _____

Pregunta 2. ¿Cuántos intrones tiene *spd-2*? _____

Asignación Parte 1: Identificando sitios de empalme para el intrón 1

Recuerda de la clase que las secuencias cortas están presentes al principio y al final de cada intrón que permiten que el espliceosoma quite exactamente cada intrón, dejando solamente las secuencias del exón en el ARN mensajero maduro. Los dos primeros nucleótidos del intrón son "GT", y los dos últimos nucleótidos son "AG" (Figura 10).

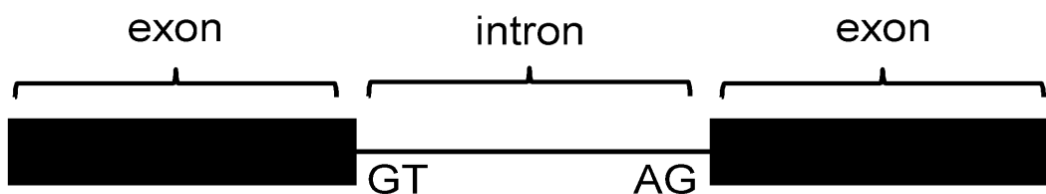


Figura 10 Un diagrama "bowtie".

24. Acércate al final del exón 1. Configura la pantalla para que pueda ver entre 15 y 20 nucleótidos.

Pregunta 3. ¿Cuál es la coordenada del último nucleótido en el exón 1? _____

Pregunta 4. ¿Cuáles son los primeros dos nucleótidos del intrón 1? _____

25. Aléjate para que puedas ver todo el intrón 1 y usa TopHat para ayudarte a encontrar el final del intrón. Examina la pista “**Exon Junctions**”.

26. Luego, acércate para que puedas ver la secuencia de nucleótidos.

Pregunta 5. ¿Cuáles son los últimos dos nucleótidos del intrón 1? _____

Pregunta 6. ¿Cuál es la coordenada del primer nucleótido en el exón 2? _____

Tarea Parte 2: Identificar los sitios donantes de empalme de 5' y de aceptación de empalme 3' para el intrón 2

Usemos el mismo enfoque para definir los límites exón-intrón para el intrón 2.

27. Revisa los pasos que utilizaste en la parte 1, luego repite el proceso para responder las siguientes preguntas sobre los sitios donante de empalme de 5' y de aceptación de empalme 3' para el intrón 2.

28. Amplía la secuencia que rodea al donante de empalme 5' para el intrón 2.

Pregunta 7. ¿Cuál es la coordenada del último nucleótido del exón 2? _____

Pregunta 8. ¿Cuáles son los primeros dos nucleótidos del intrón 2? _____

29. Reduce la imagen lo necesario para que se puedas ver todo el intrón 2. Utiliza TopHat para encontrar el final del intrón 2.

30. Presiona y arrastra para que el final del intrón 2 esté centrado en el visor. Luego, acércate para que puedas ver la secuencia de nucleótidos.

Pregunta 9. ¿Cuáles son los últimos dos nucleótidos del intrón 2? _____

Pregunta 10. ¿Cuál es la coordenada del primer nucleótido en el exón 3? _____

Pregunta 11. Con la información que has adquirido, haz una imagen gráfica del gen *spd-2* con 3 exones y 2 intrones. Numera cada exón e intrón. Suma las coordenadas del primer y último nucleótido de los exones que has encontrado hasta ahora. Agrega las secuencias de los sitios donantes de empalme y de aceptación de empalme en las ubicaciones apropiadas. Finalmente, agrega una flecha doblada para el sitio de inicio de la transcripción.