

Seminario Análisis de secuencias de nuevas tecnologías de secuenciación en paralelo

Trabajo Práctico Final

El presente Trabajo Práctico corresponde a la instancia evaluativa final del Seminario. Para la realización del mismo, se brindarán datos reales y se requerirá la selección de herramientas para su análisis apropiado. A partir de los resultados obtenidos se elaborará un artículo científico (para ello se le brindará un documento modelo) y se realizará la presentación oral de los mismos en el congreso "Bioinformática" a realizarse el día 6 de junio de 2025.

Título: "Análisis transcriptómico de pupas de *Drosophila melanogaster* crecidas en hipergravedad"

Problema biológico:

El objetivo de este trabajo es estudiar el transcriptoma e investigar los cambios en los niveles de expresión génica de la mosca de la fruta, particularmente de pupas de *Drosophila melanogaster* crecidas en condiciones de hipergravedad. La gravedad alterada puede perturbar el desarrollo normal e inducir cambios correspondientes en la expresión génica. Comprender esta relación entre el entorno físico y la respuesta biológica es importante para los objetivos de viajes espaciales de la NASA.

Durante la etapa pupal, la mayoría de los insectos holometábolos experimentan una metamorfosis extensa de larva a adulto alado. Esta metamorfosis está rigurosamente regulada a nivel transcripcional y se conserva evolutivamente. También depende de la gravedad terrestre. Por ejemplo, la orientación de la pupa se basa en señales gravitatorias para una alineación adecuada, y la exposición a entornos de microgravedad o hipergravedad puede influir de manera profunda en la metamorfosis y alterar la expresión génica. El experimento consta de dos condiciones: g3, en la que las pupas se desarrollaron a tres veces la gravedad terrestre (3 g), y g1, el control, en la que se desarrollaron a la gravedad estándar de la superficie terrestre (1 g).

El set de datos está compuesto por 12 muestras, dos tratamientos con tres réplicas biológicas cada uno. Los datos son de tipo *paired-end*. Debido a que analizar los archivos originales llevaría mucho tiempo de procesamiento, se le brinda un set de datos reducido que incluye sólo las lecturas que alinean contra el cromosoma 4 de la mosca de la fruta (*Drosophila melanogaster*). Los datos de RNA-Seq completos están disponibles en la base de datos GEO de NCBI, bajo el número de acceso GSE80323.

Completar la siguiente tabla con información (metadatos) de las muestras.

Muestra	Condición	Experimento	Réplica	SRA Accession	Protocolo preparación biblioteca	Total Read length	Equipo de secuenciación	Stranded?
g1_01	Control	RNA-seq	1	SRX1707053				No
g1_02	Control	RNA-seq		SRX1707048				No
g1_03	Control	RNA-seq		SRX1707049				No
g3_01	Tratamiento	RNA-seq	1	SRX1707051				No
g3_02	Tratamiento	RNA-seq		SRX1707052				No
g3_03	Tratamiento	RNA-seq		SRX1707047				No

Los pasos a seguir sugeridos son los siguientes:

- Control de calidad y estadísticas de las secuencias.
- Pre-procesamiento de las secuencias.
 - Las secuencias ya están pre-procesadas. Pero si lo considera necesario puede hacerles un procesamiento extra.
- Alineamiento de las lecturas contra el genoma de referencia.
 - Se brinda el genoma de referencia que corresponde solo al cromosoma 4 de *D. melanogaster* (ensamblado dm3) y el archivo de anotación (gtf) de dicho cromosoma, por si lo necesita emplear.
 - Se recomienda usar alguno de los programas vistos en clase del tipo “splice-aware”.
 - Tener en cuenta el tipo de biblioteca de cada muestra.
 - Reportar los resultados/estadísticas de los alineamientos.
 - Puede ser necesario aplicar algunas operaciones con [samtools](#) a los archivos generados para usarlos en pasos posteriores.
- Control de calidad de los mapeos.
 - Utilizar algún *software* específico para RNA-seq QC.
 - Puede utilizar la herramienta [bamCoverage](#) de deepTools para generar archivos de cobertura (dada a características de los datos reducidos, aplicar un valor de *bin size* igual a 5).

→ Visualizar los alineamientos y/o archivos de cobertura en un *genome browser*. Por ejemplo, en [Jbrowse2](#) (local) puede crear tracks del tipo “MultiWiggle track” para visualizar archivos .bw bajo una misma escala.

5. Recuento del número de lecturas por gen (*reads counting*).

→ Si utilizó STAR en el paso 2 puede incorporar la opción del conteo de reads durante el alineamiento. En caso contrario, puede emplear programas como [featureCounts](#) o [HTSeq-count](#).

→ Tener en cuenta formatear de manera apropiada las salidas de este paso (matriz de conteos *-count matrix-*) para usar esta información en el siguiente, según el programa que utilice.

6. Análisis de la expresión diferencial entre condiciones.

→ Emplear los paquetes de R DESeq2 o edgeR.

→ Preparar adecuadamente la matriz de conteos y la de información de las muestras, y setear correctamente el control y el tratamiento.

→ Establecer valores de filtrado para obtener una lista reducida de los genes diferencialmente expresados de manera significativa.

→ Realizar un análisis exploratorio de los resultados mediante diferentes visualizaciones de los datos.

7. Análisis funcionales de un conjunto de genes de interés.

→ A partir del paso previo, definir un *set* de genes que muestren cambios significativos entre las condiciones analizadas y realizar un análisis funcional.

→ Puede utilizar por ejemplo [Enrichr](#) específico para mosca, [Panther](#) o algún paquete de R desarrollado para este fin.

→ En este punto puede ser necesario convertir los ID de los genes si no lo hizo durante el análisis del paso anterior.

Material proporcionado:

- Archivo zip con los datos.
- Modelo para la escritura del artículo científico.
- Rúbrica para realizar la revisión por pares (*peer review*) de un artículo científico.

Entregables:

1. Documento Resumen con los programas y comandos ejecutados en cada una de las actividades realizadas.
2. Todos los datos procesados en archivo zip.
3. Artículo científico con un resumen de los resultados, los principales hallazgos y la metodología aplicada (se le proporciona un modelo para tal fin).
4. Presentación oral en el congreso “Bioinformática”. Dispone de un tiempo máx. de 15 min + 5 min de espacio para preguntas. No extenderse de este límite.

5. Revisión del artículo científico de un compañero/a (se le brinda una rúbrica a completar).

Fechas importantes:

- Lo solicitado en los puntos 1, 2 y 3 se deberá enviar como fecha límite el día **05/06** en el recurso "Tarea" dentro de la pestaña "Trabajo Final" del aula virtual.
- La presentación en el congreso (punto 4) se realizará el día **06/06**.
- La revisión de otro artículo científico (punto 5) deberá completarse como fecha límite el **13/06**.
- En caso de recuperar este TP Final, la fecha de entrega del artículo científico revisado con los cambios sugeridos por los pares (ya sean cuestiones de redacción o del análisis de los datos) es el **27/06** (semana de recuperatorios).