

Primera Prueba de Evaluación Continua(PEC1) Análisis de datos Ómicos, (2-26 Noviembre 2020)

Enunciado

El objetivo de esta práctica es doble:

- Partiendo de un problema y unos datos preparados deberéis analizarlos siguiendo las pautas discutidas en el segundo debate.
- Una vez obtenidos los resultados deberéis redactar un informe con la estructura tradicional de un informe técnico-científico (ver sección “Informe del análisis”).

Los datos para el análisis

Lo primero que debéis hacer es escoger un estudio para re-analizar. *Si disponéis de uno podéis pasar a la sección siguiente.* Si no, seguid leyendo.

El archivo adjunto “GEO DATASETS.XLS” contiene una lista de identificadores de estudios depositados en el *Gene Expression Omnibus* (<http://www.ncbi.nlm.nih.gov/geo/>) que pueden descargarse para ser re-analizados en este ejercicio (*sin la necesidad de obtener los mismos resultados que los autores*).

Hojeadlos un poco hasta encontrar uno que os parezca interesante. El estudio ideal debería tener pocos muestras (5-15) para que no tengáis problemas de memoria y preferiblemente 2-3 comparaciones aunque si tiene sólo una también servirá.

Supongamos **por ejemplo** que queráis utilizar el primer estudio “GDS1251” para hacer la PEC. Deberíais proceder de la forma siguiente

1. Acceder a: <http://www.ncbi.nlm.nih.gov/sites/GDSbrowser/>
2. Realizar una búsqueda con el código proporcionado (GDS1251). Ésta os llevara al registro de estudio:
3. Si los archivos .CEL están disponibles podéis descargarlos desde esta misma página. (Haciendo clic donde dice “CEL” o en la flecha que aparece a su lado).
4. Alternativamente haciendo clic en el enlace correspondiente al “Accession”(Reference series) (GSE2401) accederéis a la página del estudio.
 - En la parte inferior encontrareis el enlace a los archivos suplementarios desde donde podéis descargar los archivos CEL, *siempre que estén disponibles. OJO, porque podrían no estarlo. Si es así cambiad de estudio.*
 - Para saber a que grupo pertenece cada muestra (En este caso “Case” o “Control”) no tenéis más remedio que hacer clic en el enlace de la muestra y mirar a que grupo pertenece.
 - Esto lo encontrareis en el campo “Sample Subsets” de la ficha de la muestra. Por ejemplo la muestra GSM45184 pertenece al grupo “Control”
 - Una vez sepáis a que grupo pertenece cada muestra podéis preparar el archivo “targets” y utilizarlo para crear la matriz del diseño y analizar los datos del ejemplo.
 - Si el estudio se ha publicado podéis acceder a él desde el apartado “citations”. El abstract debería bastar para haceros una idea sobre de que trata el estudio y probablemente que comparaciones realizar.

Unos últimos comentarios sobre el formato de los datos:

- Recordad que lo más sencillo es que analicéis datos de affymetrix y que empecéis con los archivos .CEL
- Para realizar las anotaciones debéis saber con que modelo de array trabajáis. Esta información la encontrareis en el apartado “platform” en la ficha del estudio Gene Expression Omnibus.

“Pipeline” de análisis

Tal como habéis aprendido en esta unidad un análisis de microarrays suele seguir una serie de pasos ordenados. El proceso estándar consistirá en:

Primera Prueba de Evaluación Continua(PEC1) Análisis de datos Ómicos, (2-26 Noviembre 2020)

1. Identificar que grupos hay y a qué grupo pertenece cada muestra.
2. Control de calidad de los datos crudos
3. Normalización
4. [Control de calidad de los datos normalizados] (opcional)
5. Filtrado no específico [opcional]
6. Identificación de genes diferencialmente expresados
7. Anotación de los resultados
8. Comparación entre distintas comparaciones (si hay más de una comparación, ver que genes han sido seleccionados en más de una comparación)
9. Análisis de significación biológica ("Gene Enrichment Analysis")

Los capítulos 4 a 7 de los materiales muestran como llevar a cabo cada paso utilizando bioconductor.

Informe del análisis (rmarkdown: formato pdf o html)

Una vez realizado el análisis debéis redactar un informe exponiendo qué habéis hecho, como lo habéis hecho y qué resultados habéis obtenido.

Como cualquier informe científico-técnico vuestro informe tiene que tener las partes siguientes:

1. **Abstract**, con un resumen breve de no más de cinco líneas.
2. **Objetivos**: Que se pretende con este estudio
3. **Materiales y Métodos**
 1. Naturaleza de los datos, tipo de experimento, diseño experimental, tipo de micro-arrays utilizados,...
 2. Métodos que habéis utilizado en el análisis:
 1. Procedimiento general de análisis (pasos, "workflow" o "pipeline" que habéis seguido)
4. **Resultados**
 1. Incluir **los resultados que habéis obtenido en cada paso** (anteriormente especificados: Control de calidad, normalización, etc, aunque NO ES PRECISO entrar en el detalle de los métodos, más bien hacer una descripción cualitativa indicando por qué se ha llevado a cabo cada paso, y cual ha sido el "input" suministrado al procedimiento y el "output" obtenido.
5. **Discusión**
 1. Que limitaciones consideramos que pueden haber en el estudio (si consideramos que hay alguna...)
6. **Conclusión**: NO HACE FALTA. Vuestro "rol" aquí es técnico. Como bioinformáticos se os presupondrá la capacidad de manejar la información biológica mediante los programas adecuados, pero ello no implica que debáis tener los conocimientos específicos que puede requerir la interpretación biológica de los resultados.

ALGUNOS COMENTARIOS SOBRE EL FORMATO

- Procurad facilitar la revisión
 - Tabla de contenidos
 - Documento preferiblemente en pdf
 - Gráficos bien centrados, preferiblemente con número y pie
 - Código o salida en formato courier y bien justificado
 - páginas numeradas...

De cara a evaluar el trabajo deberéis entregar

Idealmente un rmarkdown (fuente: Rmd o Rnw) y su output en pdf o html

Los archivos deben contener vuestro nombre y apellidos (CivitVivesSergi.Rnw, CivitVivesSergi.pdf),

Primera Prueba de Evaluación Continua(PEC1)
Análisis de datos Ómicos, (2-26 Noviembre 2020)

Debéis comprimir los archivos y nombrarlos con vuestros apellidos y nombre
(p.ej. *CivitVivesSergi-informePEC1.zip*).

Observad especialmente que el objetivo es que realicéis un trabajo de síntesis que os sirva para tener una visión de conjunto del proceso que va desde que el investigador os hace la pregunta hasta que vosotros la procesáis y generáis el informe. **En ningún momento se trata de reproducir los resultados del paper al que van asociado los datos**