

Máster interuniversitario de Bioestadística y Bioinformática

Análisis de datos Ómicos (M0-157)

Primera prueba de evaluación continua.

Fecha publicación del enunciado: 8-04-2021

Fecha límite de entrega de la solución: 29-04-2021

Presentación

Esta PEC consta de ejercicios similares a los discutidos en los debates con los que podréis contrastar vuestra asimilación de los conceptos y métodos presentados en la primera parte del curso.

Objetivos

El objetivo de esta PEC es ilustrar el proceso de análisis de microarrays mediante la realización de un estudio, de principio a fin, tal como se llevará a cabo en una situación real.

Descripción de la PEC

La PEC se basará en los datos de un estudio depositado en GEO con identificador “GSE46687” derivado de un estudio sobre identificación de genes que intervienen en la patogénesis del síndrome de Turner. Utilizando los archivos .CEL descargados del sitio y una función ad-hoc que os proporcionaremos deberéis crear un subconjunto de muestras que será el que analizaréis

Deberéis: (i) Plantear las cuestiones que deseáis responder (ii) Realizar los análisis necesarios y (iii) Elaborar un informe explicando problemas, métodos, resultados y discusión. Recordad que tan importante como el resultado es el razonamiento y el proceso que os lleva a ello, es decir el consultor debe poder ver no tan sólo donde habéis llegado sino también como y porque habéis llegado hasta allí.

Recursos

Los recursos para la solución de la PEC son los que se han proporcionado en el aula para las primeras unidades, es decir los materiales del curso y casos de estudio.

Criterios de valoración

Tal como se indica en el plan docente, la PEC vale el 40% de la nota.

Código de honor

Cuando presentáis ejercicios individuales os adherís al código de honor de la UOC, con el que os comprometéis a no compartir vuestro trabajo con otros compañeros o a solicitar de su parte que ellos lo hagan. Asimismo aceptáis que, de proceder así, es decir, en caso de copia probada, la calificación total de la PEC será de cero, independientemente del papel (copiado o copiador) o la cantidad (un ejercicio o todos) de copia detectada.

Formato

Para hacer la entrega se tiene que enviar un mensaje al buzón de entregas del aula. En este mensaje debéis adjuntar **únicamente** un fichero .pdf o .html (obtenido a partir de vuestro archivo Rmarkdown). El nombre del fichero debe ser la composición de vuestro apellido y vuestro nombre seguido de “_ADO_PEC1.doc” (por ejemplo: si vuestro nombre es “Hadly Wickam”, el fichero debe llamarse “wickam_hadly_ADO_PEC1.pdf”).

No olvidéis de poner vuestro nombre y apellidos en el informe!!!

Selección de las muestras para el análisis

Lo primero que debéis hacer es escoger las muestras que incluiréis en el estudio. Para ello debéis ejecutar el fragmento de código siguiente, pasando como argumento a la función vuestro número de identificación de la UOC (sin letras).

```
selectSamples<- function (myID){  
  set.seed(myID)  
  selected <- c(sample(1:10, 6), 11, sample(12:26, 5), sample(27:36, 6))  
  selected <- sort(selected)  
}
```

Esta función os retornará 18 valores que utilizaréis para subsetting el objeto “targetsAll” creando vuestro targets propio. En el ejemplo ID es 1234567 y llamaos a nuestro objeto targets “myTargets”

```
mySelected <- selectSamples(1234567)  
targetsAll <- read.csv(file="targetsAll.csv", row.names = 1, head=TRUE)  
myTargets <- targetsAll[mySelected,]
```

Este objeto “myTargets” os permitirá leer únicamente un subconjunto de 18 muestras (distinto para cada uno de vosotros) con el que realizaréis los análisis.

“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:

1. Identificar que grupos hay y a qué grupo pertenece cada muestra.
2. Exploración y 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

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 microarrays utilizados,...
 2. Métodos que habéis utilizado en el análisis:
 1. Procedimiento general de análisis (pasos, “workflow” o “pipeline” que habéis seguido)

3. Que habéis hecho en cada paso (NO ES PRECISO entrar en el detalle de los métodos, más bien hacer una descripción cualitativa indicando porque se ha llevado a cabo cada paso, y cual ha sido el “input” suministrado al procedimiento y el “output” obtenido.
4. **Resultados**
 1. Que se obtiene como resultado del análisis
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.
7. **Apéndice:** Podéis poner el código de R que hayáis utilizado en un apéndice con comentarios.

Algunos comentarios sobre el formato de entrega

- La estructura indicada no es más que una propuesta. Podéis modificarla o adaptarla según vuestro propio criterio.
- Procurad facilitar la revisión
 - Tabla de contenidos
 - Secciones y subsecciones bien organizadas.
 - Gráficos bien centrados, preferiblemente con número y pie
 - Código o salida en formato courier y bien justificado
 - Páginas numeradas
 - Referencia bibliográficas completas.

Una cosa importante: El informe NO DEBE SER una colección de salidas de R como en algunos de los scripts y markdown de ejemplo que os he ido facilitando. Podéis intercalar fragmentos de R en el texto si lo consideráis interesante pero tenéis que separar el informe del código (poniendo el código en un apéndice sencillo de reproducir).

Si, como es de esperar, trabajáis con Rmarkdown os será muy sencillo ocultar las salidas de código que no deseáis mostrar utilizando las opciones del paquete knitr.

Observad especialmente que el objetivo de la práctica no es que generéis un “tocho” con un montón de información cogida de todas partes (que luego yo deberé leer) sino que realicéis un trabajo de síntesis que ilustre, de forma general, el proceso que va desde que el investigador se presenta delante vuestro diciendo “tengo unos datos que me gustaría que analicéis” hasta que le presentáis un informe con un “esto es lo que ha salido”.