



Universitat Oberta de Catalunya

Análisis de datos Ómicos (M0-157) **Primera prueba de evaluación continua.**

Presentación y objetivos	1
Descripción de la PEC.....	1
Presentación del informe.....	3

Fecha de publicación del enunciado: 1/11/2023

Fecha límite para presentar la PEC: 15/11/2023¹

Presentación y objetivos

Esta PEC completa la introducción a las ómicas mediante un ejercicio de repaso y ampliación que nos permite consolidar nuestra introducción a la secuenciación y a las herramientas que trabajaremos este curso, el software Galaxy y el conjunto de paquetes para análisis de datos ómicos conocido como Bioconductor.

No tenéis que pensar en la PEC como algo que haréis tras completar los tres primeros módulos sino como unos ejercicios que os servirán, sí, para repasar y para ser evaluados, pero también para aprender conceptos nuevos que se han introducido, pero no practicado antes.

Descripción de la PEC

El objetivo de esta PEC es que os introduzcáis de forma práctica al uso de Bioconductor y R para analizar datos de NGS. Aprenderéis cómo descargar y

¹ La fecha de entrega es la que se indica en el enunciado de la PEC. En caso de no coincidir con la indicada en el aula, ésta será la que predomine.

manipular datos de NGS, cómo realizar un control de calidad básico y cómo explorar los datos para obtener una idea general de su calidad y estructura. Además, practicaréis la presentación de resultados en un informe claro y conciso.

Para ello vais a trabajar con datos de secuenciación que os proporcionaremos.

Tenéis que llevar a cabo los pasos siguientes:

1. Descargar a vuestro ordenador el archivo “fastq” con el que trabajaréis. Podéis descargarlos de [este enlace](#)².

Observaréis que hay 10 archivos distintos, lo que nos permite personalizar un poco el ejercicio: Básicamente tenéis que escoger la muestra que coincide con el último dígito numérico de vuestro DNI o pasaporte. Por ejemplo, si vuestro DNI es 23546798K, descargaréis la muestra “S08_reads.fastq”

2. Realizar una breve descripción de los datos.
 - Cuál es el formato del archivo.
 - Cuantas secuencias contiene.
 - Otra información adicional que consideréis relevante (tecnología, tipo de secuenciación utilizada, especie que se ha secuenciado etc.)
3. Llevar a cabo un control de calidad usando Galaxy.
 - Galaxy incorpora múltiples herramientas para el control de calidad de datos de secuenciación. Con que utilicéis la herramienta FastQC bastará para nuestros propósitos.
 - Tened en cuenta que no se os pide que resolváis eventuales problemas que podáis detectar (como eliminación de adaptadores o otros). **Tan sólo debéis determinar la calidad de las secuencias**, y elaborar un pequeño informe que lo explique.
4. Llevar a cabo un control de calidad usando el paquete Rqc de Bioconductor Bioconductor.
 - Este paquete automatiza completamente el control de calidad con lo que tan sólo debéis instalar el paquete y seguir las instrucciones de la viñeta.
 - También aquí recordad que tan sólo debéis correr el programa y comentar brevemente los resultados del control de calidad
5. Escribir un informe explicando vuestros resultados y comparando lo que obtenéis con Galaxy y con Bioconductor.

² Recordad que para acceder a cualquiera de los enlaces que os proporciono, tenéis que haber iniciado vuestra sesión en google con vuestro correo de la UOC.

Presentación del informe

Aunque podéis organizar el informe a vuestro aire nuestra recomendación es que consista de las siguientes partes.

1. **Introducción y Objetivos:** Explicad aquí cual es el objetivo del ejercicio, porqué es importante llevar a cabo el control de calidad y cual será el esquema general de la solución.
2. **Material y Métodos** Explicad con que datos vais a trabajar (aquí podría venir la descripción breve) y como llevaréis a cabo el control de calidad
3. **Resultados** Aquí vendrán los resultados del control de calidad. Podéis organizarlo como deseéis, pero básicamente tiene que contener los gráficos que el programa produzca y vuestra explicación de estos. No es preciso que os extendáis en exceso con las explicaciones. Con una o dos páginas para cada programa debería ser suficiente.
4. **Discusión y limitaciones** Aquí podéis evaluar lo que habéis obtenido, comparar ambos programas y comentar que creéis que faltaría por hacer, en caso de que penséis que falta algo.
5. **Conclusiones** A que conclusiones llegáis sobre la calidad de vuestros datos.

Entregad vuestro informe **como un único archivo** en formato PDF.