

Análisis de Datos Ómicos

Introducción a la *Next Generation Sequencing*

Alex Sanchez

Máster universitario de Bioinformática y Bioestadística



UNIVERSITAT DE
BARCELONA



Universitat Oberta
de Catalunya

Presentación

- ▶ Esta actividad presenta la ultrasecuenciación, también conocida como “Next generation sequencing”.
- ▶ Se trata de una tecnología que extiende la secuenciación tradicional (“Sanger”) permitiendo, literalmente hacer más que ésta en menos tiempo y menor coste.
- ▶ De forma parecida a los microarrays su aparición ha representado un cambio revolucionario en el estudio de la biología y la salud que no ha hecho más que empezar.
- ▶ En esta actividad introduciremos los conceptos básicos y un nuevo entorno de análisis, Galaxy, que se ha consolidado, a veces como complemento, a veces como alternativa Bioconductor.

Antes de empezar: conocimientos previos

- ▶ Como parte de la asignatura de análisis de datos ómicos se asume familiaridad con R y Bioconductor, que ya debéis haber adquirido en la primera parte del curso.
- ▶ También es importante cierta base de Biología Molecular, quizás algo más que para el estudio de los microarrays.
- ▶ En entornos profesionales de bioinformática se trabaja con muchos archivos de muchos tipos y a menudo de gran tamaño
 - ▶ La capacidad de trabajar con sistemas linux y bash suele ser fundamental.
 - ▶ No lo será para los ejercicios que haremos pero es importante tenerlo en cuenta.

Objetivos y competencias

Los objetivos de esta actividad son:

- ▶ Familiarizarse con la tecnología de ultrasecuenciación y con los tipos de problemas biológicos que permite abordar.
- ▶ Conocer las principales similitudes y diferencias entre los estudios basados en microarrays y los basados en ultrasecuenciación.
- ▶ Conocer las herramientas bioinformáticas adecuadas para el análisis de datos de ultrasecuenciación, en particular Galaxy y un nuevo conjunto de paquetes de Bioconductor.

Contenidos de la actividad

- ▶ Tecnología de secuenciación de segunda y tercera generación.
- ▶ Los datos de secuenciación: Estructura y formato de archivos.
- ▶ Control de calidad de datos de secuenciación.
- ▶ Paquetes de Bioconductor para trabajar con datos de secuenciación.
- ▶ Introducción a Galaxy

Materiales de estudio

- ▶ Los materiales del curso no contienen una parte dedicada a la ultrasecuenciación, puesto que se trata de una tecnología muy moderna y en rápida evolución.
- ▶ Utilizaremos, en su lugar:
 - ▶ Presentaciones y Vídeos que os proporcionaremos
 - ▶ Notebooks y workflows con ejemplos de análisis
 - ▶ Materiales de formación libres que se actualizan constantemente.

Materials de estudio (2)

- ▶ Introduction to NGS: [*Pres./Vid.*]
- ▶ NGS data. Quality Control and Preprocessing: [*Pres./Vid.*]
- ▶ Bioconductor for high-throughput sequence analysis [*Pres./Lab.*]
- ▶ A short introduction to Galaxy [*Tutorial de Galaxy*]
- ▶ NGS Data Management with Galaxy)[*Tutorial de Galaxy*]

- ▶ En esta actividad se iniciará un debate **que finalizará en la siguiente** y os permitirá poner en práctica los conceptos adquiridos y discutir las dudas que os surjan.
- ▶ Como en otros debates se busca promover la participación, y, para motivarla, ésta se evaluará (la media de los cuatro debates será el 10% en total de la calificación final).
- ▶ Para facilitar la evaluación *debéis compilar vuestras aportaciones en un documento que subiréis al Registro de Evaluación.*
 - ▶ Por cuestiones de funcionamiento el buzón de entrega sólo estará disponible durante la siguiente actividad.