



TRABAJO ENTREGABLE 2 DE ANÁLISIS DE DATOS ÓMICOS

Descripción

El trabajo consistirá en el análisis de datos NGS procedentes de repositorios públicos o proporcionados por grupos de investigación colaboradores. Queda abierta la posibilidad de utilizar datos de los organismos que se ajusten mejor a los intereses formativos de los estudiantes.

Características de los tipos de trabajos

1. Ensamblado de novo. Se podrá partir de un conjunto de datos FASTQ para obtener su genoma. Una vez obtenido el ensamblado, se habrá de intentar anotar el contenido.
2. Identificación de variantes. Se podrá partir de datos DNA-Seq y se aplicarán herramientas de identificación de variantes. Se deberán analizar y describir las variantes más significativas a partir del contenido de bases de datos existentes.
3. Análisis de expresión diferencial. Se deberá realizar un análisis de expresión diferencial escogiendo para ello la comparación de dos condiciones diferentes. Se procurará utilizar un mínimo de 3 muestras por condición. Se deberá realizar una caracterización funcional de los genes expresados diferencialmente.
4. Análisis metagenómico.
5. Otro tipo de trabajo que los estudiantes planteen.

Los trabajos podrán combinar características de varios tipos de trabajos. Los estudiantes deberán justificar la selección de los datos a partir del objetivo específico de su trabajo. Los trabajos deberán arrancar de los ficheros FASTQ disponibles, aunque en los repositorios de datos NGS puedan estar disponibles ficheros BAM con los alineamientos BAM ya obtenidos. No se podrán utilizar datos que posteriormente sean utilizados por los estudiantes para su TFM.

Herramientas a emplear

El análisis se podrá realizar aplicando los pipeline de análisis vistos en clase, pero los estudiantes tendrán libertad para aplicar el que consideren oportuno. A la hora de hacer los alineamientos podéis optar por usar nuestro servidor o usar servidores públicos como <http://usegalaxy.org>, que os dan cuotas de 250GB. En nuestro servidor tenéis cuota de 50GB, aunque estas cuotas no aplican a los usuarios de Galaxy.

Algunas posibles fuentes de datos

- ENCODE RNA Dashboard: http://genome.crg.es/encode_RNA_dashboard/hg19
- ArrayExpress: <http://www.ebi.ac.uk/arrayexpress/experiments/browse.html>
- The Cancer Genome Atlas: <https://portal.gdc.cancer.gov>
- NCBI SRA: <https://www.ncbi.nlm.nih.gov/sra>. Se recuerda que para descargar los datos se debe hacer uso de SRA Toolkit desde terminal de comandos o via wrappers Galaxy.
- EBI SRA: <http://www.ebi.ac.uk/ena> . <https://www.ebi.ac.uk/ena/pathogens/covid-19>
- GEO: <https://www.ncbi.nlm.nih.gov/geo>
- International Cancer Genome Consortium: <https://dcc.icgc.org>



Fases del trabajo

Los alumnos deberán realizar el trabajo en grupos de dos estudiantes. Se admitirán grupos de tres estudiantes si los objetivos y dificultad del trabajo lo justifica.

El trabajo se realizará en dos fases:

- **La primera fase** consiste en la confección del grupo y la selección del trabajo a realizar. La selección del trabajo incluye la selección de los datos a analizar para comprobar que están disponibles y que el trabajo es viable.
 - El resultado de esta fase es un documento PDF de máximo 1 página donde se expliquen los objetivos del trabajo propuesto, los datos a emplear y los miembros del grupo. Se entregará a través del Aula Virtual (Tarea Fase 1 de la Entrega 2) **hasta el 24 de marzo**. Solo será necesario que un miembro del grupo suba el informe al Aula Virtual.
 - Esta fase no es obligatoria, aunque es puntuable. Los profesores de la asignatura revisarán la propuesta para considerar su adecuación a los objetivos de la asignatura.
- **La segunda fase** es la ejecución del trabajo. Se entregará a través del Aula Virtual (Tarea Fase 2 de la Entrega 2) **hasta el 18 de abril**. La entrega deberá contener:
 - Informe de resolución, que explique y muestre la realización de cada paso. Solo será necesario que un miembro del grupo suba el informe al Aula Virtual.
 - En la medida de lo posible, los archivos deberán estar disponibles en la cuenta del máster, en un servidor Galaxy, Github o similar. En caso de hacerlo en Galaxy, el historial deberá ser publicado y la URL incluida en el informe para que se pueda comprobar la realización del trabajo en el servidor correspondiente, ya sea biomaster, o un servidor público en Internet. Por motivos de espacio de disco no será necesario guardar en dicho directorio los archivos FASTQ.

Criterios generales de valoración

	Resultado	Peso	Criterios de evaluación
Fase 1	Planteamiento del problema	10%	<ul style="list-style-type: none">• Viabilidad y adecuación
Fase 2	Resolución del trabajo	70%	<ul style="list-style-type: none">• Corrección del trabajo• Diseño y calidad metodológica de la solución• Dificultad del trabajo planteado• Desarrollo de scripts de programación y análisis de datos
	Informe de resolución	20%	<ul style="list-style-type: none">• Estilo de la redacción y organización del informe• Extensión máxima: 15 páginas (desde portada hasta referencias, no incluye anexos)
	Redacción en inglés	10%	<ul style="list-style-type: none">• Puntuación adicional en función de la calidad lingüística del documento