



## ANÁLISIS E INTERPRETACIÓN DE DATOS DE ALTO RENDIMIENTO

### PRESENTACIÓN

- **Titulación:** MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS PARA CIENCIAS EXPERIMENTALES
- **Módulo y materia:** Módulo III Optativo. Materia 3.1. Optativas
- **Carácter:** Optativa
- **ECTS:** 3
- **Curso y semestre:** Curso 1º y semestre 2º
- **Idioma:** Español
- **Profesor responsable de la asignatura:** Igor Ruiz de los Mozos
- **Profesores:** Igor Ruiz de los Mozos
- **Horario y aula:** consultar calendario del máster

### RESULTADOS DE APRENDIZAJE

RAO5 Diseñar y analizar experimentos en el ámbito de las tecnologías ómicas (genómica y transcriptómica) para la búsqueda de nuevos biomarcadores, dianas terapéuticas o estudiar mecanismos moleculares. Al finalizar este módulo, los alumnos estarán capacitados para trabajar eficientemente en entornos HPC-UNIX y preparar algoritmos para llevar a cabo todos los procesos necesarios en el análisis de datos masivos. Esto les permitirá contribuir significativamente en la investigación biomédica y la medicina de precisión, generando conocimiento y respuestas a importantes preguntas en el campo de la biología molecular.

### PROGRAMA

#### Tema 1: Teoría de la secuenciación

- Evolución de las plataformas de secuenciación
- Desde las muestras hasta las librerías de secuenciación
- Desde la librería hasta las lecturas (*reads*)
- Estructura de los NGS (*next-generation sequencing*), control de calidad del preprocesado
- Visión general de los flujos de trabajo bioinformáticos

#### Tema 2: Secuenciación práctica

- Visita al centro de secuenciación masiva de Nasertic
- Secuenciación del exoma y genoma completo (Whole Exome/Genome)
- Experimentos RNA-Seq
- Secuenciación Single-cell
- *Spatial Transcriptomics*

#### Tema 3: Otras tecnologías Ómicas

- Introducción al estudio de la proteómica, metabolómica, epigenética, epigenética y microbioma
- Técnicas de separación de proteínas (ej. electroforesis en gel bidimensional)
- Análisis de interacción proteína-proteína
- Análisis de datos metabolómicos
- Análisis de datos epigenéticos y epitranscriptómicos
- Estudio de patrones de metilación del ADN y modificaciones de histonas



- Aplicaciones de la epigenómica en la investigación de enfermedades
- Análisis de la diversidad microbiana y funciones metabólicas
- Integración de datos ómicos (multi-omics)
- Métodos bioinformáticos para la integración de datos transcriptómicos, proteómicos y metabolómicos
- Análisis de redes biológicas

**Tema 4: Unix y clústeres de alto rendimiento**

- Visita al centro de Super Computación Urederra de Nasertic
- Acceso a recursos computacionales en la nube emulando a un HPC
- Curso intensivo de UNIX/bash
- Curso intensivo en Ordenadores de Alto Rendimiento (*High-Performance Computers*)

**Tema 5: Alineamiento NGS con computación avanzada**

- Evaluación de calidad de datos de secuenciación
- Limpieza, descontaminación y *demultiplex*
- Teoría del alineamiento NGS, desde datos crudos multiplexados a lecturas alineadas.
- Mejorar los parámetros de alineamiento y estado del arte de los análisis NGS.
- Alineamiento contra referencia y *de novo*

**Tema 6: Detección de diferencias NGS (*variant calling*) y secuenciación genómica**

- Traducción de las lecturas de secuenciación para obtener el genoma completo de interés
- Identificar mutaciones de nucleótidos puntuales que pueden convertirse en diferencias fenotípicas o virulentas

**Tema 7: Comparación computacional del genoma**

- Desde la secuencia consenso del DNA hasta la exploración genómica para inferir ganancia o pérdida de funciones.
- Especialización y la evolución genómicas.
- Análisis genómico en un caso de pandemia.
- Diseño y selección de antibióticos susceptibles
- Evaluación de mecanismo de virulencia microbiológica

**Tema 8: Análisis de RNA-seq**

- Tecnologías transcriptómicas de secuenciación
- Cuantificación de la expresión génica
- Clusterización y caracterización de muestras, hierarchical y k-means
- Reducción de dimensiones PCA
- Análisis de expresión diferencial: teoría y práctica
- Integración y anotación de datos
- Análisis funcional: Gene Set Enrichment Analysis
- Corrección de desviación por lotes
- Cuantificación a nivel de transcrito, nuevos algoritmos sin mapeo genómico
- Predicción de efectos sobre el transcriptoma

**Tema 9: Computación de alto rendimiento con *workflows* en Snakemake**

- Iniciación a lenguaje de *workflows* en Snakemake.
- Análisis NGS en Snakemake.
- Análisis de reproducibilidad, trazabilidad y aprovechamiento de recursos en entornos HPC.
- Descripción de Nextflow y nf-core para la creación de pipelines bioinformáticas

**Tema 10: Aplicación de técnicas de Inteligencia Artificial para la creación de Código Bioinformático**



- Iniciación IA para creación de código computacional
- *Prompt engineering* en bioinformática
- Auto digestión y auto-comprension de datos biológicos ayudado por AI
- Programación en Bash, Python y R guiada por AI

### ACTIVIDADES FORMATIVAS

ACTIVIDADES FORMATIVAS				
ACTIVIDADES FORMATIVAS	Horas totales	Horas lectivas*	% Presencialidad**	% Interacción virtual***
AF1 - Clase expositiva	10	10	100%	0%
AF2 - Clases prácticas, seminarios y/o talleres	20	20	100%	0%
AF3- Trabajos dirigidos, individuales o en equipo	40	0	0%	0%
AF4 - Estudio personal	20	0	0%	0%
AF6 - Evaluación	3	3	100%	0%
AF7 - Tutorías	2	2	100%	0%

### SISTEMAS DE EVALUACIÓN

SISTEMA DE EVALUACIÓN	PORCENTAJE EVALUACIÓN
SE2 - Resolución de problemas, casos prácticos y trabajos	50%
SE3 - Evaluaciones parciales y/o evaluación final	40%
SE13 - Evaluación la intervención y participación.	10%