****

**Listado de proyectos ofrecidos para Trabajo Fin de Máster en Bioinformática o Ciencia de datos en el ConesaLab, I2SysBio, Paterna, Valencia.**

Ultima actualización Octubre 2024.

Lab url: [conesalab.org](http://conesalab.org/)

Contacto: [conesalab@gmail.com](mailto:conesalab@gmail.com)

1. **Meta-analysis of transcirptome inestability in human disease**

Consiste en obtener colecciones de datos de la base de datos Gene Expression Omnibus de diferentes enfermedades para evaluar la fracción de genes que tienen evidencia de fallos en la maquinaria de transcripción y estudiar su posible asociación con la enfermedad. Las competencias que se desarrollan son

1. Conocimiento de bases de datos de expresión génica
2. Utilización de clústers de cálculo
3. Desarrollo de pipeline análisis en Nextflow
4. Programación en R
5. Análisis de expresión diferencial y enriquecimiento
6. **Alphafold modelling of novel Long-read sequencing transcripts**

Modelado con el programa de IA Alphafold del potencial para formar proteínas funcionales a partir de tránscritos noveles asociados a enfermedades. El TFM se desarrolla en colaboración con Christine Orengo de la University College London. Las competencias que se desarrollan son:

* 1. Funcionamiento del algoritmo de IA Alphafold (premio Nobel 2024)
  2. Transcriptómica con Long-read sequencing
  3. Desarrollo de pipeline análisis en Nextflow
  4. Programación en R y Python
  5. Optimización y paralelización del cómputo

1. **SQANTI for non-model organisms**

Consiste en desarrollar un método de control de calidad para los tránscritos obtenidos a partir de secuenciación de Pacbio para aquellos casos donde no existe un genoma de referencia. Se aplicará el método a transcriptomas de especies de la biodiversidad. Las competencias que se desarrollan son:

* 1. Transcriptómica con Long-read sequencing
  2. Desarrollo de algoritmos, incluyendo benchmarking
  3. Programación en R y Python
  4. Análisis de splicing alternativo

1. **Upscaling SQANTI3 for large datasets**

Consiste en adaptar la herramienta de control de calidad de long-read transcriptomics SQANTI3 al procesamiento de experimentos con muchas muestras. El reto es la gestión de los datos, paralelización e integración. El TFM es en colaboración con Mark Diekans, del UC Santa Cruz Genome Browser. Las competencias que se desarrollan son:

* 1. Transcriptómica con Long-read sequencing
  2. Integración de datos
  3. Desarrollo de pipelines en Nexflow
  4. Paralelización y optimización de código.

1. **SQANTI3 for single-cell LRS and transcriptomics datasets**

Consiste en adaptar la herramienta de control de calidad de long-read transcriptomics SQANTI3 al procesamiento de experimentos de single cell y spatial transcriptomics usando long reads. El reto es la gestión de los datos, la representación efectiva de los resultados y paralelización e integración. Las competencias que se desarrollan son:

* 1. Transcriptómica con Long-read sequencing
  2. Transcriptómica de single-cell y spatial
  3. Integración de datos
  4. Desarrollo de pipelines en Nexflow
  5. Desarrollo de algoritmos y benchmarking.

1. **SQANTI3 analysis of Ribo-STAMP and LRS Nascent-seq data**

Consiste en usar la herramienta de control de calidad de lrRNA-seq SQANTI3 para evaluar nuevas aplicaciones al estudio del transcriptoma usando long reads en concreto la transcripción naciente y la traducción en ribosomas. Es un proyecto básicamente de análisis de datos y de obtener conclusiones sobre la calidad de los métodos para caracterizar la dinámica del metabolismo del RNA. Las competencias que se desarrollan son:

* 1. Transcriptómica con Long-read sequencing
  2. Nuevas aplicaciones de LRS.
  3. Análisis de datos con SQANTI3
  4. Desarrollo y documentación de código en R y Github
  5. Trabajo con bases de datos públicas