



Especialización en Bioinformática

Aproximaciones ómicas al estudio de los sistemas biológicos. Tecnologías asociadas

Año 2022

Duración total: 55hs.

Responsable: Facultad de Ciencias Exactas

Docente Responsable: Dr. Antonio Lagares

Docentes: Dr. Bellora Nicolás
Dr. Erijman Leonardo
Dr. Guerrero Leandro
Dr. López José Luis
Dra. Moreno De Colonna Silvia
Dra. Valacco María Pía

OBJETIVO GENERAL

El objetivo general del curso es introducir a los alumnos al enfoque ómico del estudio de sistemas biológicos, con énfasis en las tecnologías más utilizadas, en el tipo de datos obtenidos, y en el procesamiento e interpretación de la información asociada

COMPETENCIAS A DESARROLLAR EN RELACION CON EL OBJETIVO DE LA CARRERA

C. 1 - Poder analizar problemas de Bioinformática con conocimiento de los fundamentos biológicos e informáticos, para luego resolverlos seleccionando los métodos y herramientas más adecuadas/eficientes para cada caso.

PROGRAMA

MÓDULO A. Plataformas de secuenciación de segunda y tercera generación (NGS).

Principios de funcionamiento y principales características de las plataformas actuales de secuenciación de segunda y tercera generación. El foco estará puesto en analizar los fundamentos físicos/químicos asociados a cada una de las tecnologías de secuenciación de DNA. Se analizarán las ventajas y limitaciones de cada plataforma, su complementación y sus usos más frecuentes (genomas, transcriptomas, metilomas). Capacidades y costos relativos.

- Contexto histórico. De los métodos de Sanger y Maxam y Gilbert a las plataformas de NGS.
- Secuenciadores de segunda generación: Ion Torrent PGM. Tecnología Illumina con sus distintas plataformas. Secuenciamiento *paired-end*, secuenciamiento *mate pair*. Capacidades, largo de las secuencias, tasas de error, limitaciones, costos.
- Secuenciadores de segunda generación (*single molecule*): Tecnología PACBIO. Tecnología Oxford Nanopore. Capacidades, tasas de error, largo de las secuencias, limitaciones, costos.



MÓDULO B. Ensamblado de genomas y metagenomas.

Uso de herramientas bioinformáticas para el análisis metagenómicos obtenidos por secuenciación al azar (shotgun). El foco estará puesto en el ensamblado de genomas a partir de metagenomas. Estos métodos permiten ir más allá de la descripción de la composición taxonómica en ecosistemas microbianos, brindando una visión detallada de flujos metabólicos e interacciones microbianas, además de hacer posible el descubrimiento de nuevos linajes y vías metabólicas.

- Archivos de datos. Control de calidad y procesamiento
- Ensamblado de metagenomas
- *Binning*
- Curado de genomas
- Clasificación taxonómica y anotación funcional.
- Estimación de dinámica de replicación bacteriana a partir de datos metagenómicos

MÓDULO C.

Abordaje al uso de herramientas de NGS en estudios de regulación de la expresión génica.

- Niveles de regulación de la expresión génica en eucariotas. Epigenética: estructura cromatina, dominios cromosómicos, código histonas y metilación DNA. Transcripción: Factores de transcripción, *enhancers*, Pioneer TFs, clases de promotores, PolyA alternativos. *Splicing* alternativo. Regulación sobre RNAs.
- (Repaso) Introducción a las secuenciaciones de alto rendimiento (*NGS, Next Generation Sequencing*) y sus aplicaciones en biología molecular. Historia reciente, tipos de secuenciaciones, *long reads*. RNAseq vs *microarrays*.
- Transcriptómica (RNAseq). Preproceso de lecturas (*reads*), filtrado por contaminaciones, calidad y *trimming*. Transcriptoma *de novo* y mapeo a referencia. Estrategias, algoritmos (básico histórico). Limitaciones y requerimientos de secuenciación según objetivo estudio. Cuantificación: *mapping* y *pseudo mapping*, algoritmos (básico), genes/isoformas. Métricas, normalizaciones. Análisis replicas. Análisis expresión diferencial. Gráficos: MAplots, volcano *plots* y *heatmaps*. Estrategias para relacionar genes de-regulados con procesos biológicos. Análisis del *splicing* alternativo, estrategias y limitaciones.
- Estudios de regulación génica por NGS. Técnicas: ChIP-seq, CLIP-seq, ATAC-seq y Hi-C. Fundamentos básicos de comunes a las diferentes técnicas. ChIPseq: wet-lab, mapeo, *input-control*, limitaciones, *peak-calling*, tipos de picos, análisis cuantitativo y cualitativo, *motif discovery*. Estrategias para la integración con transcriptómica. Para las restantes técnicas se dará una visión general: etapa wet-lab, secuenciación requerida, ventajas con técnicas precedentes, procesamiento *in silico*, resultados, visualización e interpretación.
- Visualización en entorno genómico (*Genome Browsers*). Local (IGV) y remoto (UCSC). Información disponible, tipos de pistas (bed, wig, bam, etc), carga de datos propios, compartir resultados y vistas. ENCODE *project*. Ejercicio sobre el *browser*: genes, expresión en tejidos, estructura, sintonía, conservación, estructura secundaria y *splicing* alternativo.



MÓDULO D. Abordaje a estudios proteómicos.

Uso de estrategias de espectrometría de masas para la identificación y cuantificación de proteínas

Proteómica. Conceptos Generales.

- Introducción a la espectrometría de masa y proteómica. Introducción a la técnica - Equipamiento - Ejemplo de espectros. Concepto de masa monoisotópica y promedio. Ionización por MALDI (iones +1); Ionización ESI (iones multicarga). Identificación de proteínas por MS - Fragmentación - MS/MS- Breve introducción al abordaje experimental.
- Bases teóricas de la identificación de proteínas por espectrometría de masa: identificación por MS y MSMS. Proteómica global y dirigida. Conceptos generales: Shotgun vs targeted. Identificación de proteínas puras, mezclas simples y complejas. Introducción a la proteómica cualitativa y cuantitativa emPAI.
- Ejemplos de Usos y aplicaciones: Estrategias de *pull down* y coimmunoprecipitación, Interactómica, PTMs, proteómica diferencial por geles 2D MALDI, Biotyping, biosimilares.
- Programas de análisis de datos - MASCOT y *Proteome Discoverer*. Criterios de búsqueda para identificación - Mostración con datos concretos utilizando Mascot y visualización de datos en PD.

Introducción a la proteómica cuantitativa

- Perspectiva histórica de la proteómica cuantitativa: evolución de metodologías.
- Generalidades de la aproximación actual al estudio cuantitativo del proteoma. Estrategias de separación, identificación y cuantificación de péptidos y proteínas. Abordaje *bottom-up*, separación de péptidos por MD-LC acoplado a su análisis por MSn, adquisición de datos del tipo *shotgun*.
- Origen de los datos cuantitativos. Cuantificación basada en identificación o en señal en MSn. Conceptos de XIC (extracted ion chromatogram) y Spectral Counting.
- Dimensiones de cuantificación: relativa y absoluta. Estrategias de cuantificación relativa con marcaje isotópico o libre de marca. Métodos de marcaje metabólicos y químicos.

Procesamiento informático de los datos crudos de ensayos de proteómica cuantitativa

- Descripción de la arquitectura e interfaz de usuario del software. Descripción del motor de búsqueda integrado Andromeda para la identificación de espectros. Descripción de los algoritmos para la cuantificación relativa y absoluta de péptidos y proteínas identificadas. Interpretación de tablas *output*.
- Procesamiento de datos espectrales crudos. Flujo de trabajo en software MaxQuant para recuperar datos cuantitativos a partir de un dataset de espectros obtenido en un abordaje de proteómica cuantitativa del tipo *shotgun* libre de marca.

Análisis estadístico de datos cuantitativos de un ensayo proteómico

- Introducción al software Perseus y flujo de trabajo. Descripción del software. Carga de datos y anotaciones funcionales. *Workflow* para curado de los datos, imputación de valores nulos, y análisis estadístico de los resultados. Generación de gráficos de correlación, análisis multivariado, diagramas de Venn. Elaboración de volcano-plots.



ACTIVIDADES EXPERIMENTALES y DE INVESTIGACION

Los trabajos experimentales se desarrollarán en una o varias clases en las salas de informática, y mediante conexión remota a los servidores utilizando los lenguajes Bash, R y/o Python. Finalizada cada actividad, habrá una sesión de discusión conjunta donde cada participante comunicará los resultados obtenidos. La evaluación se realizará mediante la entrega de un trabajo final.

METODOLOGIA Y MODALIDAD DE EVALUACION

Para aprobar el curso se requiere un 80% de asistencia a las clases presenciales (teóricas y prácticas).

La evaluación se realizará mediante un examen escrito al final de las clases del curso para evaluar el grado de conocimientos del alumno (20%), el proyecto/desarrollo experimental que deberá entregar el alumno al final de las horas programadas (70%) y la participación y aportaciones de calidad/excelencia a las soluciones propuestas (10%). El proyecto/desarrollo experimental consistirá en un reporte técnico (símil artículo científico) producto de un estudio teórico o experimental específico.

BIBLIOGRAFÍA BASICA

- Data Analysis and Visualization in Genomics and ProteomicsK (2005). Francisco Azuaje, Joaquin Dopazo (eds). Wiley.
- Cell biology by the numbers (2015). Ron Milos and Rob Phyllips. Garland Science.
- Bioinformatics and Functional Genomics (2015). Pevsner, Jonathan. Tercera edición. Wiley-Blackwell.
- The dictionary of genomics, transcriptomics and proteomics (2015). Guenter Kahl. Wiley-Blackwell.
- Transcriptomics and gene regulation (2016). Jiaqian Wu (eds.). Springer Netherlands.
- Protein bioinformatics: from protein modifications and networks to proteomics (2017). Cathy H. Wu, Cecilia N. Arighi, Karen E. Ross (eds.). Humana Press.
- Transcriptome data analysis: Methods and protocols (2018). Yejun Wang, Ming-an Sun (eds.).
- Genome data analysis (2019). Ju Han Kim. Springer Singapore.