

Semana 1. 08/03/2024

Teoría:

- Presentación de los contenidos, objetivos y metodología de evaluación del Seminario.
- Historia de la Bioinformática. Introducción al Proyecto Genoma Humano.
- Evolución de las tecnologías de secuenciación de ADN:
 - Primera Generación (Método de Sanger).
 - Introducción a las Nuevas Tecnologías de Secuenciación en Paralelo (NGS): Segunda Generación (Pirosecuenciación – 454 Roche, Secuenciación por síntesis - Illumina, Secuenciación por Ligación de Oligonucleótido y Detección – SOLID, Secuenciación por semiconductores – Ion Torrent), Tercera Generación (Secuenciación de una única molécula a tiempo real – PacBio, Secuenciación electroquímica de molécula única – Nanopore). Utilización de videos para comprender cada metodología.
 - Explicación del funcionamiento de cada tecnología, diferencias y similitudes entre ellas, costos asociados, ejemplos de equipamientos, entre otras comparaciones.

Práctica:

- Lectura y análisis de artículos (*papers, reviews, y/o noticias*) relacionados con una aplicación puntual en la sociedad de las NGS (por ejemplo: secuenciación del virus del Ébola durante el mayor brote de la historia, secuenciación del genoma de SARS-CoV-2) y/o perspectivas futuras de su uso.
- Visualización de diferentes videos (ilustrativos y de tipo charlas TED) relacionados al uso, avance y aplicaciones de las NGS.
- Búsquedas web de diferentes *kits* comerciales “directo al consumidor” disponibles para el análisis de ADN, proveedores nacionales e internacionales. Preguntas disparadoras para reflexionar acerca de su uso, la privacidad de los datos, consentimiento informado, acceso universal a estos *kits*, entre otras.

Semana 2. 15/03/2024

Teoría:

- Preparación de bibliotecas para secuenciación de ADN y diferentes *workflows* de trabajo. Consideraciones en la preparación de bibliotecas para: *Whole Genome Sequencing, Whole Exome Sequencing, RNA-Sequencing, CHIP-Sequencing*.
- Control de calidad en diversas etapas del proceso de secuenciación, abarcando puntos claves durante la preparación de la muestra y durante el procesamiento de los datos generados.
- Conceptos generales NGS: qué son y de dónde provienen las lecturas o *reads*. *HiFi reads*. Ensamblado de lecturas.
- Comparación entre bibliotecas *Single-End, Paired-End* y *Mate-Paired*.
- Conceptos de Cobertura horizontal y vertical. Concepto y valores de Calidad.
- Introducción a formatos de archivo para NGS: formato FASTQ, GTF/GFF3, SAM/BAM/CRAM, BED, VCF, SRA.

- Introducción al uso de *software* para el control de calidad y pre-procesamiento de las secuencias.

Práctica:

- Exploración de repositorios públicos de datos de genómica funcional, por ejemplo GEO del NCBI.
- Descarga de datos públicos desde dichos repositorios. Uso del NCBI SRA-toolkit y SRA Run Selector.
- Introducción al análisis de datos producidos por NGS. Uso de los *software*: FASTQC para el control de calidad de lecturas originadas por diferentes plataformas de secuenciación, y Fastx-Toolkit y Trimmomatic para el procesamiento de datos reales producidos por tecnologías NGS. Uso de MultiQC para generar un reporte interactivo.
- Ejemplo procesamiento de secuencias a través del uso de un *cluster* de computadoras para ejecutar los trabajos mediante SLURM y Snakemake.
- Trabajo Práctico: los alumnos deberán elegir dos aplicaciones de las tecnologías NGS y explicar su fundamento, protocolo, análisis de los datos y dar ejemplos puntuales de su uso. Para ello usarán la bibliografía del Seminario, información confiable en la *web* que considere relevante, videos, *papers*, *reviews*, etc. Deberán armar una presentación para exponer ambas aplicaciones con sus ejemplos para la siguiente clase.

Semana 3. 22/03/2024

Teoría:

- Explicación, desarrollo y análisis de ejemplos de las aplicaciones más frecuentes de las NGS: Secuenciación *de novo*, Re-secuenciación, RNA-Seq (con secuencia de referencia, y sin secuencia de referencia), CHIP-Seq, *Amplicon Sequencing*, - *Target Sequencing*, *Whole Exome Sequencing*, *SNP Discovery*, Metagenómica.

Práctica:

- Presentación por parte de los alumnos del Trabajo Práctico planteado en la Semana 2.
- Actividad titulada "Conociendo laboratorios de NGS y usando sus equipos", confeccionada en forma de consignas secuenciales en un formulario de Google. El objetivo de la misma es que los alumnos exploren las instalaciones de institutos dedicados a la secuenciación de ADN mediante NGS con fines clínicos y, además, "usar virtualmente" dos equipos de la plataforma de NGS Illumina y preparar las bibliotecas de ADN para su secuenciación. El propósito perseguido con esta actividad es que los estudiantes "salgan" de manera virtual del aula tradicional, alejándose por un momento de los datos y herramientas específicas de análisis, para que, en primer lugar, puedan ver y explorar lugares en los cuales ellos podrían trabajar en un futuro, que examinen cómo están formados estos espacios y que busquen similitudes entre un instituto internacional (Australia) y otro nacional. Con algunas preguntas se busca interpelar a los alumnos y conocer la opinión de los mismos en relación a la profesión.

Semana 4. 29/03/2024

FERIADO - Libre para turno de Examen Especial

Semana 5. 05/04/2024

Teoría:

- Conceptos vinculados al ensamblado de lecturas producidas por NGS. *Whole Genome Shotgun*, tipos de ensamblado (*de novo*, guiado por referencia).
- Ensamblado *de novo*: algoritmos *Greedy*, *Overlap-Layout-Consensus* y *de Bruijn graphs*.
- Evaluación de un ensamblado. Parámetros estadísticos importantes. El problema de las secuencias repetitivas.
- Estrategias para *Scaffolding* y *Finishing*. Anotación de genomas.
- Programas para el alineamiento para secuencias cortas (BWA, Bowtie), repaso de formatos SAM/BAM, entre otros.

Práctica:

- Parte 1:
 - Empleo y análisis de diferentes *software* para el ensamblado de secuencias como Velvet, SPAdes, Bowtie, entre otros.
 - Uso de Quast para la evaluación de los alineamientos producidos.
 - Alternativas exclusivas de acuerdo al sistema utilizado y a la estrategia de secuenciación (*de novo*, *re-sequencing*). Ensamblado contra referencia.
 - Anotación de un genoma ensamblado.
- Los alumnos deberán elegir un artículo científico sobre ensamblado y anotación de un genoma o pangenomas y presentarlo la siguiente clase.

Semana 6. 12/04/2024

Práctica Ensamblado (cont.):

- Parte 2:
 - Ensamblado de secuencias de ADN largas obtenidas mediante la tecnología PacBio. Pulido del ensamblado. Genómica comparativa.
 - Uso de programas como Canu, Circlator, BWA, Samtools, BLAST+, entre otros.
- Exposición del artículo científico elegido.

Semana 7. 19/04/2024

Teoría:

- Consideraciones para el envío de los datos genómicos producidos a bases de datos públicas. ¿Por qué compartir la información generada?
- Introducción a los visualizadores de genomas y bases de datos especializadas, por ejemplo: UCSC Genome Browser, Ensembl, NCBI Genome Data Viewer, Integrative

Genomics Viewer (IGV), Jbrowse, arabidopsis.org. Diversas actividades de exploración de datos a resolver en el foro del aula virtual.

- Introducción a la plataforma Galaxy.

Práctica:

- Uso de la plataforma Galaxy. Resolución de alguno de los siguientes tutoriales (u otro que se considere más conveniente/actualizado):
 - Ensamblado híbrido: “Hybrid genome assembly - Nanopore and Illumina”.
Fuente: https://www.melbournebioinformatics.org.au/tutorials/tutorials/hybrid_assembly/media/tutorial.pdf
 - Comparación de genomas: “From small to large-scale genome comparison”.
Fuente: <https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/hpc-for-lsgc/tutorial.html>

Semana 8. 26/04/2024

Teoría – Práctica:

- Introducción al lenguaje R. Tipos de datos, operadores, estructuras de datos, ciclos, estructuras de control, funciones.
- Uso de Rstudio. Instalación de paquetes (bioconductor).
- Importar y exportar archivos. Operaciones básicas con *Data Frames*.
- Introducción a tidyverse. Gráficos básicos con ggplot2.
- Ejemplos de análisis de datos producidos por tecnologías NGS.

Semana 9. 03/05/2024

Teoría:

- Secuenciación de transcriptomas: RNA-Seq. *Workflow* de trabajo, diseño experimental en relación a las muestras y estrategias de construcción de bibliotecas de ADNc.
- Análisis de transcriptomas: con referencia y *de novo*. Ejemplo de ciertos programas bioinformáticos para cada estrategia.
- Determinación de niveles de expresión y estadística comparativa. Principios de diseño experimental.
- Introducción al *Downstream analysis: Gene Ontology (GO) enrichment analysis, Pathway analysis* (o análisis de vías metabólicas) y *Network analysis*.

Práctica:

- Análisis de datos reales producidos por RNA-Seq.
 - Control de calidad de las secuencias y post-procesamiento.
 - Alineamiento de los datos contra el genoma de referencia. Reconstrucción del transcriptoma.
 - Comparación de la expresión génica entre dos condiciones diferentes con el fin de identificar genes diferencialmente expresados.

- Visualización de los datos producidos. Anotación funcional de genes diferencialmente expresados. Uso de los *software* y paquetes de R: Tophat, HISAT, Cufflinks, STAR, Samtools, Bedtools, IGV genome browser, HTSeq-count, DESeq2, Salmon, Kallisto, entre otros.
- Entrega de uno o varios artículos científicos para analizar su estructura en la siguiente clase.

Semana 10. 10/05/2024

Teoría – Práctica:

- Análisis del artículo científico entregado la Semana 9 con el objetivo de comprender como se escribe un *paper* vinculado con la temática, y como se presenta la información (estructura del artículo, título, orden de los autores y filiaciones, vocabulario, encabezados, figuras, leyendas, referencias bibliográficas, etc.). Esta actividad servirá como base para la elaboración del TP Final.
- Explicación del Trabajo Práctico Final. El mismo tendrá consignas semi-estructuradas y los alumnos (en grupos diferentes) deberán analizar datos reales obtenidos mediante NGS para la resolución de un problema biológico, aplicando las diversas estrategias de análisis vistas en el cursado del Seminario. Además elaborarán un artículo científico en español (opcional en inglés) con los datos generados (se les brindará un documento modelo para tal fin). Se comenzará a trabajar en el mismo.

Semana 11. 17/05/2024

Teoría:

- Modificaciones epigenéticas: metilación del ADN y modificaciones de histonas.
- Técnicas para la investigación de modificaciones epigenéticas a escala genómica.
 - Secuenciación de ADN por bisulfito.
 - ChIP-sequencing: técnica para mapear sitios de unión de factores de transcripción de todo el genoma y regiones enriquecidas con modificación de histonas.
 - Consideraciones para la preparación de las muestras y análisis de los datos.
- Ensayos de accesibilidad de la cromatina (ATAC-sequencing).
- Organización nuclear. Técnicas de captura conformacional de la cromatina (3C, 4C, HiC). Formación de TADs y *loops* de la cromatina. Interpretación de los datos producidos por HiC.

Práctica:

- Análisis de datos producidos por inmunoprecipitación de cromatina seguida de secuenciación de ADN de alto rendimiento (ChIP-seq).
 - Procesamiento de los datos y alineamiento de las lecturas con Bowtie.
 - *Peak calling* mediante el uso del programa MACS. Empleo de Bedtools para comparar los sets de picos obtenidos.
 - Visualización de los datos obtenidos en un visualizador de genomas.
 - Gráficos varios con los datos empleando deepTools.

- Descubrimiento de motivos.
- Avances y consultas en relación al TP Final.

Semana 12. 24/05/2024

Teoría:

- Introducción a la Medicina de Precisión. Desarrollo del Proyecto Genoma Humano, hitos históricos, principales logros y alcances.
- Aplicaciones de la Genómica clínica: oncología de precisión, enfermedades mendelianas raras, análisis de asociación.
- Farmacogenómica. Miedos en la era genómica.
- Ejemplos de casos de aplicación en la clínica y cómo trabajar en conjunto con médicos y profesionales de la salud. Consentimiento informado. Consideraciones éticas y de confidencialidad.
- Clase opcional complementaria titulada “(R)evolución genómica. Leyendo y modificando genomas”:
 - Tecnologías para la edición de genomas. ZFN, TALENs, CRISPR/Cas9.
 - CRISPR/Cas en la naturaleza. Historia de su descubrimiento.
 - Disputa por la patente de la tecnología CRISPR. Premio Nobel de Química.
 - Funcionamiento del sistema CRISPR/Cas. CRISPR *Toolbox*.
 - Consideraciones para el diseño de los ARN guía. Herramientas bioinformáticas para su diseño.
 - Ejemplos varios de aplicaciones de la tecnología CRISPR. Ensayos clínicos. Desarrollo de *kits* diagnósticos. Consideraciones bioéticas de su uso.

Práctica:

- *Workflow* para el análisis de los datos, llamado y anotación de variantes, plataformas para la visualización de los datos e interpretación biológica. Ejemplo de reportes de casos clínicos.
- Trabajo Práctico: Actividad de lectura de capítulos a definir del libro “Genoma y Salud” de M. Martí (2019). Deberán realizar una presentación resumiendo los principales puntos de dichos capítulos, y asociar estos conceptos con noticias/artículos que ejemplifiquen la temática abordada, incorporando de ser posible un enfoque bioético y brindando su opinión personal.

Semana 13. 31/05/2024

Teoría – Práctica:

- Entrega y presentación por parte de los alumnos del Trabajo Práctico planteado en la Semana 12.
- Metagenómica.
 - Microbiota, microbioma y metagenómica: Definiciones, Historia. Importancia. Proyecto Microbioma Humano. Proyecto Microbioma de la Tierra, etc.

- Concepto de Metagenómica y su desarrollo, relaciones, ventajas y desventajas con respecto a otros enfoques tradicionales. Concepto de Unidad Taxonómica Operativa (UTO).
- Protocolo experimental para el estudio del microbioma a través del 16S rRNA: Métodos de extracción de DNA. Construcción de bibliotecas de 16S rRNA para secuenciación por Illumina.
- Aplicaciones actuales y potenciales de la Metagenómica en estudios agrícolas, ecológicos y ambientales.
- Bases de datos en Metagenómica. Métodos y técnicas empleados en estudios metagenómicos agrícolas y ambientales.
- Avances y consultas en relación al TP Final.

Semana 14. 07/06/2024

- Entrega y evaluación del TP Final.
 - Envío del artículo científico elaborado por medio del aula virtual de la materia.
 - Se realizará un “congreso” para la presentación del TP Final por parte de los alumnos. Deberán hacer una presentación oral de 15-20 min en la cual contarán los resultados obtenidos y las conclusiones generales del trabajo realizado, emulando que están participando en un congreso de Bioinformática. Posterior a la presentación se destinarán aprox. 5 min para la realización de preguntas.
 - Durante los tres días siguientes, cada estudiante (o grupo de trabajo) evaluará el artículo científico del TP Final de otro alumno (o grupo), emulando la fase de revisión por parte de una revista científica. Los docentes oficiarán como los editores de dicha revista, dejando sus revisiones/comentarios también. Quienes necesiten hacer muchas correcciones y/o nuevos análisis de datos, entregarán el artículo re-elaborado para la siguiente instancia de recuperatorio.
- Se llevarán a cabo conferencias con profesionales (a definir) que trabajen en áreas vinculadas a las temáticas abordadas en el Seminario. Dichas charlas serán parte del “congreso”.
- Se invitará a los alumnos de primer y segundo año a participar como oyentes.

Semana 15. 14/06/2024

- Posible viaje de un día a Rosario con los alumnos para visitar empresas e institutos de investigación vinculados con el uso de tecnologías NGS y el análisis bioinformático de los datos. Visita a programar sujeta a aprobación de fondos/disponibilidad a:
 - *INDEAR* (www.indear.com): empresa de investigación y desarrollo en el sector de la biotecnología agropecuaria. En sus laboratorios alberga grupos de trabajo en las áreas de genómica y bioinformática, biología molecular, biología sintética y estudio de proteínas.
 - *Héritas* (<https://heritas.com.ar>): empresa con foco en medicina de precisión, ofrece servicios de diagnóstico genómico orientados a genómica clínica, oncología, microbioma humano y genómica de la reproducción.
 - *Instituto de Biología Molecular y Celular de Rosario (IBR)* (www.ibr-conicet.gov.ar): instituto de investigación dependiente de CONICET y de la Universidad Nacional de Rosario. Su objetivo principal es la generación, difusión y aplicación del

conocimiento científico en Ciencias Biológicas a través de la investigación, la docencia y la transferencia de este conocimiento al sector productivo (Biotecnología), de salud y distintos estamentos de la sociedad.

- *Centro Científico, Tecnológico y Educativo. Acuario del Río Paraná* (<http://www.acuariodelrioparana.gob.ar>): centro multifunción, único en su tipo en Argentina y Latinoamérica, donde se entrelazan de manera innovadora la educación, la ciencia, el ambiente y la biodiversidad, con la comunidad. El Área Científica se coordina a través del Laboratorio Mixto de Biotecnología Acuática dedicado a estudios de peces del Río Paraná y su ecosistema, principalmente en genética y genómica de peces.

Semana 16. 21/06/2024

FERIADO

Semana 17. 28/06/2024

Recuperatorio TP Final (entrega en “tarea” del aula virtual del artículo científico re-estructurado con las correcciones señaladas durante la exposición oral y la evaluación por pares).