

Planificación de la Asignatura: Tópicos Especiales en Bioinformática: Análisis de Secuencias de Nuevas Tecnologías de Secuenciación en Paralelo

Fecha: 23/10/2024 13:02

Código: OP002-1

Carrera: Licenciatura en Bioinformática

Departamento Académico: Biología

Docente a cargo:

Correo del docente a cargo: ileana.tossolini@uner.edu.ar

Régimen de Dictado: Cuatrimestral 1º Cuatrimestre

Carga Horaria Semanal: 6 horas semanales

Carga Horaria Total: 84 horas

Contenidos Mínimos:

Principios de Secuenciación en Paralelo, Diferentes tecnologías disponibles, Aplicaciones de la tecnología, Nuevos algoritmos para su utilización, Práctica de análisis de diferentes aplicaciones.

Correlativas Regulares para cursar:

No posee

Correlativas Aprobadas para cursar:

Tercer año

Correlativas Aprobadas para promocionar o rendir el examen final:

Tercer año

Objetivo General:

Contribuir al desarrollo de habilidades y competencias en el estudiante próximo a graduarse como Licenciado en Bioinformática, vinculadas al análisis de datos producidos por las nuevas tecnologías de secuenciación de ADN (NGS) y los métodos computacionales de análisis y comparación de secuencias obtenidas mediante dichas tecnologías. Se analizarán en forma comparativa y complementaria con los métodos convencionales dictados en la asignatura Análisis y Alineamiento de Secuencias. Se analizarán las estrategias utilizadas en las diferentes aplicaciones disponibles gracias al surgimiento de estas nuevas tecnologías, y los nuevos desafíos computacionales y de almacenamiento que conllevan la gran acumulación de información mediante estas técnicas.

Se propone además el objetivo general de promover el desarrollo de los estudiantes en el uso del lenguaje propio de ámbitos institucionales como la academia, la ciencia y los espacios de actuación profesional. A la vez, promover la articulación entre materias de distintos años de la carrera y propiciar el uso del lenguaje científico propio de la academia y los espacios profesionales. Además, otro objetivo es enriquecer la experiencia de cursado de la asignatura mediante la realización de una salida educativa para visitar empresas e institutos de investigación vinculados con el uso de tecnologías de secuenciación y la Bioinformática, con el fin de reducir la brecha de contacto entre los estudiantes y el mundo laboral.

Objetivos Particulares:

Que los alumnos:

1. Conozcan las distintas plataformas de secuenciación en paralelo.
2. Comprendan los diferentes formatos de secuencia y los formatos de salida de cada una de las plataformas, su manejo y transformación.
3. Comprendan las limitaciones y errores de las tecnologías, y los métodos y herramientas de análisis de calidad de las secuencias obtenidas por las diferentes plataformas.
4. Potencien su espíritu crítico para que puedan valorar adecuadamente los alcances y limitaciones de las técnicas de análisis de datos.
5. Comprendan las similitudes y diferencias entre los distintos tipos de bibliotecas utilizadas (unpaired, paired-end, mate-pair, entre otras).
6. Conozcan las diferentes aplicaciones de las nuevas tecnologías de secuenciación (entre las más generales: secuenciación de novo, resecuenciación, metagenómica, RNA-Seq, CHIP-Seq, análisis y anotación de variantes)
7. Comprendan los algoritmos para el ensamblado de secuencias de novo y guiado por una referencia. Comprendan las limitaciones de cada tipo de ensamblado (Overlap-Layout-Consensus, gráficos de De

Bruijn).

8. Analicen y comprendan los resultados de los diferentes programas utilizados para el ensamblado de novo, guiado por referencia e híbrido, de programas para el ensamblado de RNA-Seq con o sin referencia para su ensamblado, entre otros análisis de resultados en diferentes aplicaciones.
9. Interpreten el funcionamiento y aplicación de diferentes herramientas para visualización de datos de NGS, de tipo independiente o como complemento de navegadores.
10. Realicen el análisis de datos reales, y obtengan conclusiones acerca de las mejores prácticas para los análisis así como conclusiones biológicas a partir de los mismos.
11. Analicen trabajos científicos, tanto de software desarrollado para el análisis de datos de NGS, como trabajos que utilizan las nuevas tecnologías para abordar problemáticas biológicas actuales.
12. Conozcan las diferentes aplicaciones de la Genómica clínica (oncología de precisión, enfermedades mendelianas, enfermedades poco frecuentes, análisis de asociación - GWAS) y sus consideraciones éticas y de confidencialidad.
13. Comprendan y se apropien del proceso de investigación científica de un tema afín a la temática, desde la obtención y análisis de los datos, pasando por la redacción, publicación y revisión de un artículo científico, hasta la comunicación oral en eventos científicos de los resultados obtenidos.
14. Conozcan y exploren diferentes lugares en los cuales podrían trabajar en un futuro, que examinen cómo están formados estos espacios y que comparen las características de diversas instituciones nacionales e internacionales.
15. Comprendan el rol de un/a Licenciado/a en Bioinformática en la interpretación de los resultados a partir de la información que se genera principalmente en el campo de las “-ómicas” y se conciben como futuros profesionales dedicados a la bioinformática, tanto en el ámbito privado como académico.
16. Desarrollen fundamentos para evaluar y actuar en relación con el impacto social de su actividad profesional en el contexto global y local.
17. Exploten sus aptitudes para el desempeño en equipos de trabajo, creatividad y desarrollo de soluciones para problemas biológicos concretos.
18. Potencien su aprendizaje de manera continua y autónoma, y puedan comunicar contenidos técnicos vinculados a la profesión de manera efectiva.
19. Desarrollen competencias comunicativas para la escritura de artículos científicos y para la presentación oral de los resultados de su trabajo.
20. Apliquen la autoevaluación y la evaluación entre pares, para identificar fortalezas, debilidades y potencialidades de las diversas producciones de los demás estudiantes (informes escritos, exposiciones orales, trabajo final).

Programa Analítico:

Unidades

1. Introducción a Nuevas Tecnologías de Secuenciación en Paralelo

- 454 - Roche
- Illumina (HiSeq, MiSeq, HiSeq X)
- IonTorrent - Applied Biosystems
- PacBio - Pacific Biosciences
- Oxford Nanopore

Explicación del funcionamiento de cada tecnología, diferencias y similitudes entre ellas, costos asociados, ejemplos de equipamientos, entre otras comparaciones.

2. Preparación de bibliotecas y diferentes workflows de trabajo.

Control de calidad en diferentes etapas del proceso de secuenciación (preparación de la muestra y de los datos generados). Bibliotecas single-end, paired-end, mate-paired. Conceptos generales del análisis de datos producidos por las NGS. Introducción a formatos de archivo para NGS: formato FASTQ, GTF/GFF3, SAM/BAM/CRAM, BED, VCF, SRA.

3. Aplicaciones varias de las NGS

- Whole Genome Sequencing (Secuenciación de novo, Resecuenciación)
- Amplicon Sequencing
- Target sequencing
- RNA-Seq (con y sin secuencia de referencia)
- Epigenómica (CHIP-Seq, ATAC-Seq, Methyl-Seq)
- SNP Discovery
- Metagenómica
- Medicina de precisión
- Single cell sequencing

4. Ensamblado

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. Alternativas de software: Velvet, SPAdes, MIRA, SOAPdenovo, etc. Programas para el alineamiento para secuencias cortas (BWA, Bowtie), repaso de formatos SAM/BAM, entre otros. Estrategias exclusivas de acuerdo al sistema utilizado y a la estrategia de secuenciación (denovo, reseq). Alternativas de ensamblado híbrido y de lecturas largas

5. Repositorios públicos de datos de genómica funcional, visualizadores de genomas, bases de datos especializadas y uso de la plataforma Galaxy.

Descarga de datos públicos desde repositorios públicos de datos de genómica funcional, por ejemplo GEO-NCBI. 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. Introducción a la plataforma Galaxy.

6. Introducción a la programación en R aplicada a Bioinformática

R es un lenguaje de programación fundamental en la bioinformática, permitiendo manipular y analizar grandes cantidades de datos. Además, ayuda a automatizar los complejos cálculos estadísticos para el análisis de datos. En esta unidad se verá una introducción a dicho lenguaje:

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

7. Transcriptómica

Principios generales de la técnica RNA-sequencing (secuenciación de ARN mensajero y ARNs pequeños). Estrategias para la construcción de bibliotecas. Análisis de transcriptomas, con referencia y de novo. Determinación de niveles de expresión y estadística comparativa. Principios de diseño experimental.

8. Epigenómica

Modificaciones epigenéticas: metilación del ADN y modificaciones de histonas. Técnicas para la investigación de modificaciones epigenéticas a escala genómica. 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. Análisis de datos producidos por inmunoprecipitación de cromatina seguida de secuenciación de ADN de alto rendimiento (ChIP-seq).

9. Genómica clínica

Introducción a la Medicina de Precisión. 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. Workflow para el análisis de los datos, llamado y anotación de variantes, plataformas para la visualización de los datos e interpretación biológica.

Listado de Actividades de Formación Práctica:

Son las actividades prácticas planificadas para el cuatrimestre (detalladas por semana en el cronograma extendido). Estas incluyen:

- * Visualización de videos sobre las nuevas tecnologías de secuenciación y sus aplicaciones actuales (charlas TED, conferencias) que inviten a la reflexión del presente de estas tecnologías y las aplicaciones futuras, incentivando a los estudiantes que asuman el rol de profesionales en este campo. Cuestionario con preguntas que fomenten el debate de cuestiones técnicas y aspectos éticos entre los estudiantes.
- * Realización de un tour virtual en un centro de secuenciación. Recorrido de las instalaciones. Preguntas que inviten a la reflexión como futuros profesionales en el área de la bioinformática en un lugar similar.
- * Realización de un Laboratorio virtual para la preparación de bibliotecas de ADN y la utilización de un secuenciador Illumina. Resolución de un cuestionario sobre la actividad.
- * Prácticas en laboratorio de computación:
 - Análisis de estadísticas correspondientes a las corridas
 - Análisis de nuevos formatos de secuencias (Fastq, BAM, SAM, etc)
 - Análisis de calidad de datos reales.
 - Descarga de datos públicos desde bases de datos y uso de diferentes herramientas para tal fin.
 - Exploración de diferentes visualizadores de genomas.
 - Ejemplo procesamiento de secuencias a través del uso de un cluster de computadoras para ejecutar los trabajos mediante SLURM y Snakemake.
 - Ensamblado de datos simulados y reales.
 - Análisis de resultados e interpretación de parámetros críticos de control de calidad, ensamblado, etc.
 - Análisis de datos de RNA-Seq, determinación de genes diferencialmente expresados. Anotación funcional.
 - Tutoriales para el uso de la plataforma Galaxy.

- Análisis de datos de proyectos de Metagenómica y/o Epigenómica.
- Análisis de datos de casos clínicos, diagnóstico de enfermedades raras, llamado y anotación de variantes, interpretación funcional.

* Trabajos prácticos

1. 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.

2. Análisis y discusión de artículos destacados recientes, vinculados a los temas abordados.

Análisis de un artículo científico específico, a fin 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 para la elaboración del TP Final.

3. 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.

Metodología de Evaluación Durante el cursado:

Esta propuesta de cátedra ha sido impulsada por la búsqueda de estrategias que promuevan un aprendizaje significativo, que desafíen y motiven a los estudiantes, y que a la vez los acerquen al mundo laboral del que formarán parte y les proporcionen herramientas para integrarse en él. Se optó por una metodología de evaluación estimulante que hiciera posible valorar conocimientos, habilidades, destrezas y competencias, promoviendo la necesidad investigativa de mantenerse actualizado en el campo de saber así como reconocer los saberes de los pares, y potenciando habilidades blandas típicas, como son las habilidades de comunicación, la capacidad de organización y el pensamiento crítico.

Se realizará una evaluación individual de un Trabajo Práctico Final que requiere de horas de trabajo extra áulico. 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 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 (instancia de evaluación por pares). Los docentes oficiarán como los editores de dicha revista, dejando sus revisiones también. Quienes necesiten hacer muchas correcciones y/o nuevos análisis de datos, entregarán el artículo re-elaborado para la semana de recuperatorio estipulada en el cronograma.

La Nota final será la derivada de la entrega del TP Final donde se incluye la evaluación del proceso de desarrollo (participación, asignación de roles, grado de compromiso con la tarea, autonomía), el producto final (plazo de entrega, impacto, calidad, estado del arte, coherencia, discusión, etc.), la defensa oral (claridad expositiva, síntesis, recursos usados, etc.) y la evaluación por pares.

Metodología de Evaluación en Exámenes Finales:

Se evaluará mediante examen escrito teórico y práctico. La evaluación teórica consistirá en preguntas conceptuales a desarrollar. La evaluación práctica consistirá en el análisis de datos reales utilizando software específico y la interpretación de los resultados obtenidos, o bien mediante el análisis de un trabajo científico que aborde una metodología/técnica de NGS y la exposición de dicho análisis. Los alumnos libres

deberán realizar ambas evaluaciones.

La calificación final será el promedio entre la evaluación teórica y práctica.

Condiciones de Regularidad :

La regularización de la materia, se logrará con una nota de al menos 60% en la evaluación integral del TP Final. La promoción de la materia se logrará si la nota supera el 80%, los alumnos que la logren no realizarán examen final. Los alumnos que regularicen pero no promocionen la materia deberán aprobar un examen final.

Bibliografía Principal:

Básica:

Datta S, Nettleton D. 2014. Statistical Analysis of Next Generation Sequencing Data. Springer. ISBN: 978-3-319-07211-1 (Print) 978-3-319-07212-8 (Online)

Brown, S. 2015. Next-Generation DNA Sequencing Informatics, Second Edition. Cold Spring Harbor Laboratory Press (Second Edition). ISBN 978-1-621821-23-6

Bibliografía Complementaria:

Específica:

Trabajos científicos de diversas revistas, con énfasis en la utilización de nuevas tecnologías de secuenciación en el análisis de información biológica.