



1) Título del Curso

“Bioinformática aplicada al análisis de datos de secuenciación de última generación (NGS)”

2) Unidades de Créditos Académicos (UCAs) que otorga: 3 UCAs
(45 horas)

3) Número de inscriptos admisibles o cupo: 35

4) Docente responsable: Dra. Marcela Dotto

5) Docentes del curso:

Dr. Hernán Rosli
Dra. Marina Pombo
Dr. Diego Sánchez

6) Destinatarios: estudiantes de doctorados en ciencias de la vida, jóvenes científicos y profesionales que no posean conocimiento previo de bioinformática y quieran aprender sobre las metodologías de análisis de datos de NGS.

7) Justificación

El uso de secuenciación de última generación (Next Generation Sequencing) es muy utilizado a nivel mundial y ha demostrado su utilidad para responder la más amplia variedad de preguntas de la comunidad científica actual. En el país se ejecutan una gran cantidad de proyectos de investigación científica que utilizan esta tecnología, por lo que la demanda de capacitación para estudiantes de doctorado es muy alta.

8) Objetivos

El presente curso pretende introducir al alumno a las aplicaciones más comunes de la secuenciación de última generación (DNAseq, RNAseq, smRNAseq, entre otros) y aportar los conocimientos necesarios para que los alumnos sean capaces de llevar a cabo el análisis bioinformático de los datos obtenidos usando esta tecnología.



9) Programa

Módulo 1: Tipos de NGS y Preparación de bibliotecas

Desde la secuenciación Sanger hasta los métodos modernos. Funcionamiento de los distintos tipos de NGS: Illumina HiSeq, Roche 454, ABI/SOLiD, Gene Reader, BGI. Comparación de las mismas. Tercera generación de secuenciación: Pacific Biosciences, Oxford Nanopore. Secuenciadores de mesada: Ion Torrent PGM, Illumina MiSeq, 454 GS Junior. Nuevas opciones para secuenciación de fragmentos largos y single cell sequencing. Plataformas de secuenciación disponibles desde Argentina.

Detalle de la secuenciación por Illumina. Diseño experimental. Bibliotecas de RNA-Seq y DNA-Seq (kit comercial y protocolos alternativos). Ventaja de las bibliotecas hebra específicas. Multiplexing. Elección del tipo de biblioteca.

Módulo 2: Linux

Introducción al sistema operativo Linux. La terminal y la línea de comandos. Directorio home y directorio de trabajo. Moverse dentro del sistema de archivos. Comandos básicos para trabajar sobre archivos y directorios. Redireccionamiento y pipes. Comandos útiles para el manejo de archivos fasta y fastq.

Módulo 3: Archivos FASTQ y Controles de calidad

Características del formato fastq. Uso de los paquetes FastQC y fastx_toolkit para control de calidad de los datos de secuenciación. Pre-procesado de secuencias.

Módulo 4: RNAseq

Descripción del tipo de archivos manipulados en el análisis y unidades habitualmente empleadas para la cuantificación de la abundancia de transcritos. Filtrado de la contaminación por ARN ribosomal y mapeo al genoma de referencia. Ensamblado y cuantificación de transcritos y análisis de expresión diferencial. Estrategias para realizar un análisis global de la información.

Módulo 5: DNAseq

Mapeo a genoma de referencia usando Bowtie2. Tipos y características de los archivos obtenidos. Visualización de los datos de mapeo. Herramientas para el manejo y procesamiento como SAMtools y VcfTools. Llamado e



identificación de SNPs. Determinación del efecto asociado a cada SNP utilizando snpEff.

Módulo 6: smRNAseq

Mapeo usando Bowtie. Clasificación de acuerdo al tamaño, normalización. Ejemplo: pipeline de análisis de expresión diferencial. Ejemplo de programas para predicción de genes blanco.

Módulo 7: Ensamblado de genomas

Consideraciones generales. Ensambladores más utilizados. Criterios para la evaluación del ensamblado. Problemas y soluciones.

10) Actividades Prácticas

TP1: Introducción a Linux. Ejercicios en la PC

TP2: Control de calidad de archivos fastq y pre-procesado. Ejercicios en la PC

TP3: Mapeo a genoma de referencia y DNAseq. Ejercicios en la PC

TP4: RNAseq. Ejercicios en la PC

11) Cronograma de dictado y duración del curso

Día 1:

Introducción general y Módulo 1

Docentes: Dras. Marcela Dotto y Marina Pombo

Día 2:

Módulos 2 y 3

Docente: Dra. Marcela Dotto

Día 3:

Módulo 4

Docente: Dr. Hernán Rosli

Día 4:

Módulo 5

Docente: Dr. Diego Sánchez

Día 5:

Módulos 6 y 7.

Docente: Dra. Marcela Dotto



12) Número de horas teóricas: 20 h

13) Número de horas prácticas y seminarios: 25 h

14) Sistema de Evaluación

80% asistencia y evaluación final

15) Referencias Bibliográficas

- Ansorge WJ (2009). Next generation DNA sequencing techniques. *New Biotechnology* 25(4): 195-203
- Darst RP, Pardo CE, Ai L, Brown KD, Kladde MP (2010). Bisulfite sequencing of DNA. *Curr Protoc Mol Biol.*; Chapter 7: Unit–7.9.17.
- Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, and Sunyaev S (2013). Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet.* 2013 Jul; 14(7): 460–470.
- Hawkins RD, Hon GC, and Ren B (2010). Next-Generation Genomics: an Integrative Approach. *Nat Rev Genet.* Jul; 11(7): 476–486.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-9
- Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.
- Mehta JP (2014). Sequencing small RNA: introduction and data analysis fundamentals. *Methods Mol Biol.*; 1182:93-103.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protocols* 7(3): 562-578
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews* 10(1): 57-63
- Zhong S, Joung JG, Zheng Y, Chen YR, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ (2011) High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* 2011(8): 940-949