

CURSO DE BIOINFORMÁTICA Y MINERÍA DE DATOS APLICADOS A ESTUDIOS DE ONCOGENÓMICA FUNCIONAL

Docente: Dr. Martín C. Abba, Facultad de Ciencias Médicas – Universidad Nacional de La Plata, Argentina

Objetivos: Introducción metodológica al manejo de bases de datos, análisis estadístico y a la minería de datos de perfiles de expresión genómica derivados de microarreglos de ADN y SAGE.

Destinatarios: Curso dirigido al persona involucrado en la realización de estudios de genómica funcional estrechamente relacionados con investigación básica o aplicada del cáncer.

PROGRAMA:

Clase N°1: Introducción a la oncogenómica funcional .

Objetivos y alcances de la oncogenómica. Bases de la bioinformática aplicada al análisis de datos en transcriptómica. Descripción y caracterización del diseño tecnológico y de los datos derivados de microarreglos de ADN (ADNc / oligoarrrays) y SAGE (*Serial Analysis of Gene Expression*). Ventajas / desventajas de cada plataforma en función de los diseños experimentales. Características generales del entorno bioinformático para la administración, procesamiento y análisis de datos.

Clase N°2: Análisis estadísticos de perfiles de expresión génica y minería de datos.

Introducción a los principales métodos estadísticos con aplicación en genómica funcional. Descripción de métodos estadísticos supervisados (Chi², Ttest, PTM, ANOVA, etc.) y métodos estadísticos no supervisados (Agrupamiento Jerarquizado de Clusters, Análisis de Componente Principal, etc). Descripción de los conceptos de Gene Ontology (GO), análisis de enriquecimientos de grupos funcionales (ejemplo: genes relacionados con ciclo celular, apoptosis, etc). Instalación y uso de programas estadístico y de minería de datos para el análisis de expresión genómica: TIGR MultiExperiment Viewer (*MeV, The Institute For Genomic Research*), Discovery Space

(Canada's Michael Smith Genome Science Centre), GSEA (*Gene set Enrichment Analysis, Broad Institute - MIT*), STEM (*Short Time-series Expression Miner, Carnegie Mellon University*), Subio (Subio Inc.), STRING (*Search Tool for Retrieval of Interacting Genes/Proteins, EMBL*). Comparación de perfiles de expresión génica entre diferentes plataformas. Meta-análisis de datos.

Clase N°3: Manejo de los principales repositorios de perfiles de expresión genómica.

Descripción y uso de bases de datos públicas: GEO (*Gene Expression Omnibus, NCBI*), expO (*Internacional Genomic Consortium*), Array Express (EMBL-EBI), Oncomine (Compendia Biosciences), ITTACA (*Integrated Tumor Transcriptome Array and Clinical data Analysis, InstitutCurie*), CGAP (*The Cancer Genome Anatomy Project, NCBI*).

Clase N°4: Creación y manejo de bases de datos privadas.

Introducción al desarrollo de aplicaciones web en servidores con entornos WAMP (Windows – Apache – MySQL – PHP). Instalación y manejo de WAMP server 2.0. Concepto y estructura de una base de datos relacional. Programación básica de scripts en PHP y MySQL para la búsqueda de perfiles de expresión génica en base de datos locales. Descripción de un ejemplo práctico vPEG 1.0 (visualizador de Perfiles de Expresión Génica) hospedado en sitio URL.

Clase N°5: Aplicación de las herramientas bioinformáticas a la resolución de problemáticas específicas del alumnado en oncogenómica.

Búsqueda de genes diferencialmente expresados en distintas localizaciones tumorales. Empleo de bases de datos públicas para la búsqueda de genes de interés personal / grupal, para la generación de nuevas hipótesis y para sustentar datos experimentales en publicaciones científicas. **Examen Final.**