



Programa escuela de Bioinformática de verano 2018 A2B2C Fundación Instituto Leloir.

Duración:

28 horas.

Objetivos:

El curso introduce conceptos relacionados con el análisis de datos, particularmente de RNAseq, haciendo uso de herramientas y enfoques varios. Incluye una introducción al entorno de análisis estadístico de datos R, a conceptos de análisis exploratorio de datos y a la construcción y análisis de redes complejas a partir de perfiles transcripcionales.

Audiencia:

El curso está pensado para investigadores y estudiantes de grado, que tengan interés en el análisis de datos en general y de datos transcripcionales en particular.

Docentes:

Lic. Andres Rabinovich (arabinovich@leloir.org.ar)

Lic. Maximiliano Beckel (mbeckel@leloir.org.ar)

Mag. Bruno Kaufman (bkaufman@leloir.org.ar)

Lic. Fernando Orti (ferorti@gmail.com)

Dr. Javier Iserte (jiserte@leloir.org.ar)

Organización:

Dra. Cristina Marino-Buslje

Programa:

Clase 0) (ANTES DE COMENZAR EL CURSO)

Preparación del entorno de trabajo antes de venir al curso:

- Instalación de R, RStudio y paquetes que serán indicados.

Clase 1)

- Introducción a R y RStudio.
- Variables y tipos de datos (Capítulo 3 Statistics Using R with Biological Examples), estructuras de control, funciones, subsetting y vectorización(Capítulo 4 Statistics Using R with Biological Examples).
- Paquetes.
- Importación, exportación y limpieza de datos.
- Ayudas.

Clase 2)

- Introducción a RNASeq
- Medidas de similitud y distancia. Escala de los datos.
- Agrupamientos (clustering). Tipos de agrupamientos.
- Algoritmos de agrupamiento. Implementación en R.
- Ejemplos biológicos.

Clase 3)

- Sistemas complejos. Ejemplos biológicos y biología de sistemas.
- Conceptos básicos de redes complejas. Definición, topología y representación.
- Tipos de redes y sus características. Propiedades topológicas.
- Ejemplos biológicos.
- El paquete Igraph, uso de redes con R.

Clase 4)

- Detección de comunidades en redes. Tipos de algoritmos, modularidad.
- Similitud en espacios topológicos. Similitud en redes, equivalencias.
- Transcriptómica. Introducción al análisis de datos transcriptómicos con weighted gene co-expression network analysis.
- Introducción al paquete WGCNA en R.

Clase 5)

- WGCNA.
- Redes pesadas.
- Similaridad topológica (TOM).
- Clustering y detección de genes “importantes”.
- Enriquecimiento funcional de temas biológicos. GO enrichment.
- Similaridad semántica.
- Pipeline completo de trabajo de WGCNA en R.

Bibliografía recomendada (manuales, tutoriales y libros):

- An introduction to R. [\[online\]](#)
- Peng R. (2015), R Programming for Data Science, Lean Publishing
- Everitt B. (2010), A Handbook of Statistical Analyses Using R, CRC Press.
- Verzani J. (2002), simpleR: Using R for Introductory Statistics. [\[pdf\]](#)
- Dalgaard P. (2008), Introductory Statistics with R, Springer
- Mangiafico S. (2015), AN R COMPANION FOR THE HANDBOOK OF BIOLOGICAL STATISTICS, New Brunswick, Rutgers University [\[pdf\]](#)
- McDonald J. (2014). HANDBOOK OF BIOLOGICAL STATISTICS, SPARKY HOUSE PUBLISHING. [\[pdf\]](#)
- Zeileis A. (2013), Extended Model Formulas in R: Multiple Parts and Multiple Responses. [\[pdf\]](#)
- Seefeld K. (2007), *Statistics Using R with Biological Examples*, Durham, NH, University of New Hampshire, Department of Mathematics & Statistics. [\[pdf\]](#)
- Mangiafico, S. (2016). *Summary and analysis of extension program evaluation in R*. New Brunswick, Rutgers University [\[pdf\]](#)
- Wickham h. (2009), *ggplot2: Elegant graphics for data analysis*, Springer
- Barabasi A. Network Science. [\[online\]](#)
- Tan, Steinbach & Kumar "Introduction to Data Mining". [\[online\]](#)
- Hastie, Tibshirani & Friedman, "The Elements of Statistical Learning", 2nd ed, Springer, 2009. [\[online\]](#)
- James, Witten, Hastie & Tibshirani, "An Introduction to Statistical Learning with Applications in R", 6th ed, Springer, 2015 [\[online\]](#)