

# Análisis de Datos Ómicos

## El Proceso de Análisis de Datos Ómicos

Alex Sanchez

**Máster universitario de Bioinformática y Bioestadística**



UNIVERSITAT DE  
BARCELONA



Universitat Oberta  
de Catalunya

# Presentación

- ▶ Esta actividad enlaza con la anterior, en la que se introducen las tecnologías ómicas y los datos ómicos.
- ▶ Para ello se introduce lo que denominamos el “proceso de análisis” señalando las características que son únicas de los datos ómicos frente a las que son comunes a otros tipos de estudios.
- ▶ En este vídeo veremos una perspectiva general de la actividad
  - ▶ Los objetivos y competencias a adquirir
  - ▶ Los contenidos del tema
  - ▶ Qué debates o ejercicios prácticos se realizaran
  - ▶ De qué recursos disponéis para llevarla a cabo.

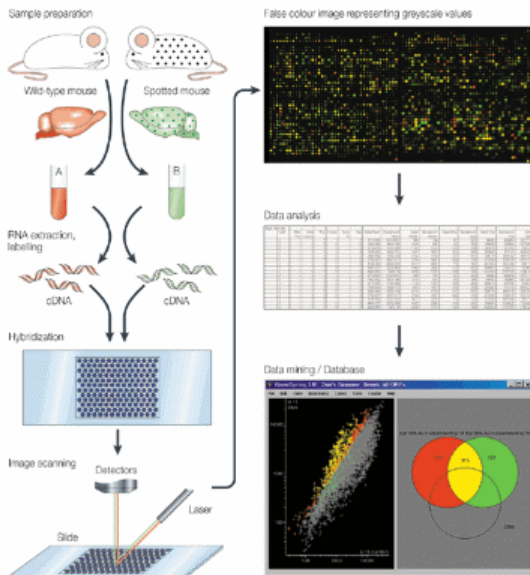
# Antes de empezar: conocimientos previos

- ▶ En esta unidad supondremos que ya os habéis familiarizado con las tecnologías ómicas, principalmente al análisis de microarrays con el que ilustramos los conceptos y métodos que se presentaran en esta primera parte.
- ▶ Naturalmente también se asume que tenéis un dominio razonable de R y de los fundamentos de inferencia estadística al nivel que se imparten en el máster.
- ▶ Si además os manejáis bien con Rmarkdown y habéis cursado la(as) asignaturas de Regresión y la de Análisis Multivariante, mejor que mejor.

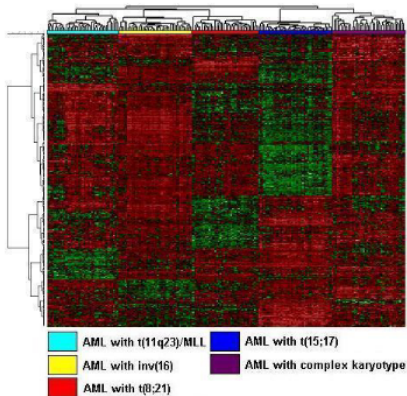
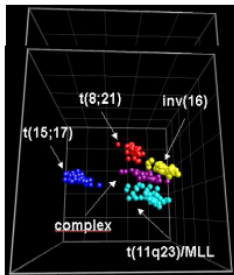
# Objetivos y competencias

- ▶ Adquirir una visión general de lo que denominamos “Proceso de Análisis de Datos Ómicos”, concretamente:
  - ▶ Conocer los tipos de estudios
  - ▶ Conocer las fuentes de datos y saber como obtenerlos
  - ▶ Conocer la herramientas básicas y saber como instalarlas.
- ▶ Adquirir un dominio básico de Bioconductor que permita obtener y explorar datos de microarrays desde el *Gene Expression Omnibus*
- ▶ Repasar conceptos de diseño de experimentos necesarios para una correcta planificación de estudios con datos ómicos

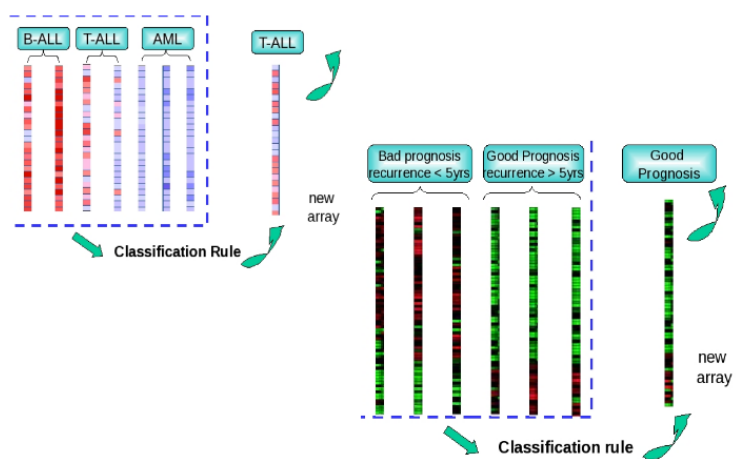
# Los contenidos en 5' (1a): Tipos de estudios



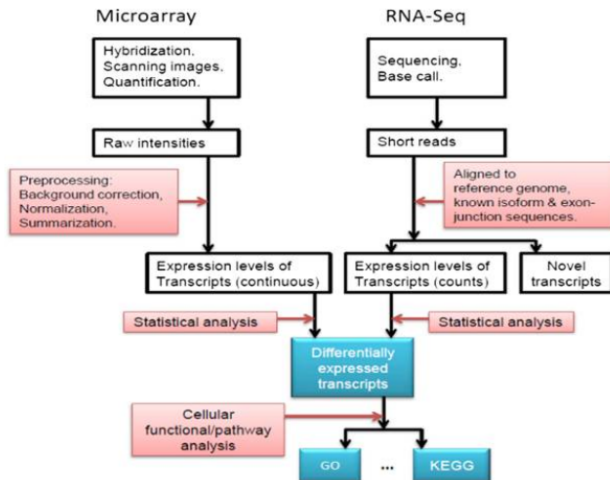
# Los contenidos en 5' (1b): Tipos de estudios



# Los contenidos en 5' (1c): Tipos de estudios



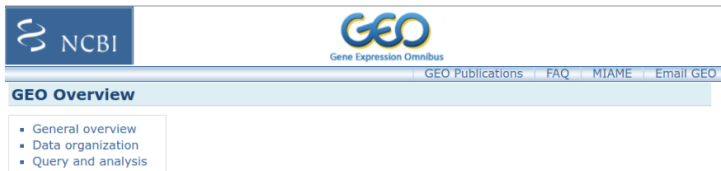
# Los contenidos en 5' (2): El proceso de Análisis



Fang et al. *Cell Biosci.* 2012; 2: 26.



# Los contenidos en 5' (3): GEO



The screenshot shows the top section of the GEO website. On the left is the NCBI logo. In the center is the GEO logo with the text 'Gene Expression Omnibus' below it. On the right is a navigation bar with links: 'GEO Publications', 'FAQ', 'MIAME', and 'Email GEO'. Below the navigation bar is a 'GEO Overview' section with a list of links: 'General overview', 'Data organization', and 'Query and analysis'.

## General overview

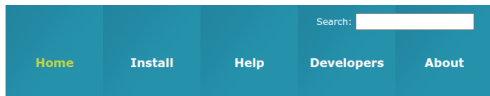
GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community.

The three main goals of GEO are to:

1. Provide a robust, versatile database in which to efficiently store high-throughput functional genomic data (see [Data organization](#))
2. Offer simple submission procedures and formats that support complete and well-annotated data deposits from the research community (see [Submission guide](#))
3. Provide user-friendly mechanisms that allow users to query, locate, review and download studies and gene expression profiles of interest (see [Query and analysis](#))

Please see the [GEO Documentation](#) listings to find more information about various aspects of GEO.

# Los contenidos en 5' (4): Bioconductor



## About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data.

Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and [Docker](#) Images.

## News

- Bioconductor [3.13](#) release schedule announced. Please view for important deadlines.
- Nominations for the Bioconductor 2021 Awards now Open! See [award page](#) for more details or use this [nomination form](#).
- See our [google calendar](#) for events, conferences, meetings, forums, etc. Add your event with email to events at bioconductor.org.
- Bioconductor [F1000 Research Channel](#) is available.

## BioC 2021

Visit the [BioC 2021](#) website for complete conference information! The virtual conference will be held August 4-6, 2021!

News highlights:

- Registration is Open! [Register Here](#).
- Bioconductor 2021 Award nominations now open! See [award page](#) for more details
- See the list of confirmed speakers on the [website home page](#)

## Install »

- Discover [1974 software packages](#) available in Bioconductor release 3.12.

Get started with Bioconductor

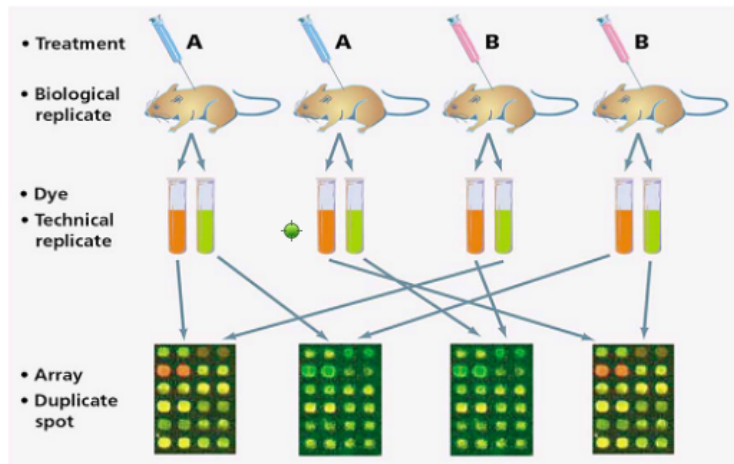
- [Install Bioconductor](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

## Learn »

Master Bioconductor tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

# Los contenidos en 5' (5): Diseño de experimentos



## Caso de uso. Presentación

- ▶ El caso se basa en un estudio para estudiar la función del gen ZBTB7B.
- ▶ Los autores utilizaron microarrays para investigar en qué forma este gen regula la expresión de los genes de la grasa marrón en dos condiciones experimentales, temperatura ambiente y después de la exposición al frío.
- ▶ Tenéis acceso a los datos, el código y las explicaciones en el siguiente repositorio de GitHub:  
[https://github.com/ASPteaching/Omics\\_Data\\_Analysis-Case\\_Study\\_1-Microarrays](https://github.com/ASPteaching/Omics_Data_Analysis-Case_Study_1-Microarrays)

# Materiales de estudio (1)

- ▶ Esta actividad utiliza los capítulos 1 (“El proceso de Análisis”) y 2 (“Análisis de Datos de Microarrays”) del documento “Análisis de Datos de Microarrays”.
- ▶ Además se os proporcionaran materiales complementarios sobre
  - ▶ github,
  - ▶ bioconductor y
  - ▶ diseño de experimentos.
- ▶ Naturalmente el caso de estudio, citado en la introducción, es un soporte más.
  - ▶ Recordad, sin embargo, que un ejemplo *no es una guía rígida* que seguir al pie de la letra.
  - ▶ Otros datasets u otros problemas requieran cambios u omisiones del código.

## Materiales de estudio (2)

- ▶ La bioinformática evoluciona rápidamente, a la vez que la tecnología y los programas de análisis. Por este motivo los materiales contienen información que o bien queda ya obsoleta (los “PM/MM” en arrays de Affymetrix ya no se utilizan) o desactualizada (algunos fragmentos de código de R que han evolucionado con Bioconductor.)
- ▶ El repositorio de github  
[https://github.com/ASPteaching/Analisis\\_de\\_datos\\_omicos-Materiales\\_para\\_un\\_curso](https://github.com/ASPteaching/Analisis_de_datos_omicos-Materiales_para_un_curso) contiene un documento de trabajo que **actualiza los materiales docentes**
  - ▶ En una primera versión se trata básicamente del mismo documento con cambios mínimos, corrección de erratas y el código de R actualizado.
  - ▶ Podéis clonar el repositorio con Rstudio o acceder al documento directamente desde este enlace.

- ▶ En esta actividad se realiza un debate que os permitirá poner en práctica los conceptos adquiridos y discutir las dudas que os surjan.
- ▶ Como en otros debates se busca promover la participación, y, para motivarla, ésta se evaluará (la media de los cuatro debates será el 10% en total de la calificación final).
- ▶ Para facilitar la evaluación *debéis compilar vuestras aportaciones en un documento que subiréis al Registro de Evaluación.*