



Universitat Oberta  
de Catalunya

## Análisis de datos Ómicos (M0-157)

### Primera prueba de evaluación continua.

**Fecha de publicación del enunciado: 31/10/2022**

**Fecha límite para presentar la PEC: 16/11/2022<sup>1</sup>**

## Presentación y objetivos

---

Esta PEC completa la introducción a las ómicas mediante un ejercicio de repaso y ampliación que nos permite introducirnos a una de las herramientas que trabajaremos más en este curso, el conjunto de paquetes para análisis de datos ómicos conocido como Bioconductor.

No tenéis que pensar en la PEC como algo que haréis tras completar los tres primeros módulos sino como unos ejercicios que os servirán, sí, para repasar y para ser evaluados, pero también para aprender conceptos nuevos que se han introducido, pero no practicado antes.

Para poder entender y llevar a cabo esta primera parte, tenéis que familiarizaros con [Bioconductor](#) y con el paquete [Biobase](#), en el que se implementan los contenedores de datos ómicos denominados [expressionSets](#). Para ello, disponéis de los siguientes materiales: - Dos vídeos que he preparado sobre Bioconductor, disponibles en el aula (con la PEC) y en el [documento de google drive con enlace a los materiales](#) - Un caso de uso de introducción a Bioconductor **que he actualizado para que se complemente bien con el Caso de Introducción a la Exploración de microarrays, que he actualizado también**. Podéis descargarlos del mismo documento de google drive o directamente de github con los siguientes enlaces: - [Omics Data Analysis-Case Study 0-Introduction to BioC](#).

Mi consejo es que **hagáis esto antes de empezar la PEC**.

## Descripción de la PEC

---

El objetivo de esta PEC es que seleccionéis un estudio de microarrays de la lista disponible en la tabla del final del documento, correspondiente al archivo “GEOdatasetsUpdatedNov2022.xls” y, tras descargarlo usando el paquete [GEOquery](#) llevéis a cabo un estudio exploratorio *similar* al que se hacía en el caso de estudio del primer debate que, recordad, estaba disponible, **actualizado** en: [Análisis de datos omicos-Ejemplo 0-Microarrays](#)

En la práctica esto significa que debéis:

1. Familiarizaros con Bioconductor, los ExpressionSets y GEOquery, usando los materiales que os indico en la sección anterior.
2. Seleccionar el dataset con el que trabajaréis de la lista que os proporcionamos
  - *Han habido dos versiones de esta lista debido a un error en su preparación. Aunque la lista correcta es la última, si ya lo hubierais hecho con un dataset anterior indicadlo y no habrá ningún problema, porque en principio el proceso funcionaría con casi cualquier dataset de GEO.*
3. Leer los datos desde GEO
  - Para descargar los datos debéis utilizar el paquete **GEOquery** que os permitirá obtenerlos un **expressionSet** con los datos normalizados (los mismos que la matriz de expresión del ejemplo 0, pero ya en formato de un expressionSet)
4. Determinar su estructura y el diseño del estudio que los ha generado.
  - La información del experimento podéis descargarla también de GEO, bien con GEOquery si proporcionáis el identificador de dataset GDSxxxx o accediendo a la página del estudio.
5. Realizar un análisis exploratorio univariante y multivariante similar al del caso de estudio.
  - Aquí podéis ser creativos, pero si no deseáis serlo os bastará con seguir el código del ejemplo de análisis. **OJO**, debéis seguir únicamente **el punto 3 del estudio. “Exploratory Analysis**. La lectura la haréis diferente, es decir con GEOquery, y la parte 4 de Selección de Genes, **no tenéis que hacerla**
6. Elaborar un informe de vuestro trabajo usando Rmarkdown. Aquí tenéis que tener en cuenta el contenido y la construcción.
  - En cuanto al contenido el informe tiene que tener la estructura habitual de cualquier trabajo: (i) Tabla de contenidos, (ii) Introducción y Objetivos, (iii) Métodos, (iv) Resultado y (v) Apéndices. En el apéndice podéis poner el código R que habréis utilizado para vuestro trabajo y así será un único documento.
  - En cuanto a la construcción debéis preparar documento en Rmarkdown que genere el informe en HTML y que debéis imprimir a pdf para entregarlo.
7. Tenéis que entregar un único archivo en formato pdf con la estructura anterior y cuyo nombre sea la concatenación de vuestros apellidos, el nombre y la palabra PEC1, **sin acentos ni espacios en blanco**.
  - Por ejemplo, en mi caso el archivo que entregaría se denominaría “Sanchez\_Pla\_Alex-PEC1.pdf”

## Recursos

---

Los recursos para la solución de la PEC son los que se han proporcionado en el aula para las tres primeras unidades, es decir los materiales del curso y casos de estudio.

## Criterios de valoración

---

Tal como se indica en el plan docente, cada PEC vale el 30% de la nota

tal como se indica en el plan docente, cada PEC vale el 30% de la nota.

Ahora bien, y como cosa importante, recordad que la PEC en si misma es un ejercicio de síntesis y aprendizaje en la que intenta valorar vuestra capacidad para resolver un problema muy parecido a los que se encuentra un/a bioinformática/a en su día a día. Esto quiere decir que para más de uno de los pasos que debéis realizar no hay una solución única. Plantead vuestra propia solución y explicad porqué creéis que es la adecuada. Entre otras cosas valoraremos:

- Capacidad de definir correctamente los objetivos a alcanzar
- Capacidad de organizar el análisis, obtención de los datos, preparación de los archivos etc.
- Dominio adecuado de las herramientas propias del tema (R, Rmarkdown, BioC)
- Capacidad de explicar qué y porqué se hace en cada paso.
- Capacidad de interpretar los resultados obtenidos.
- Capacidad de discutir las posibles limitaciones del estudio.
- Presentación del trabajo en un documento legible y bien organizado.

## Código de honor

Cuando presentáis ejercicios individuales os adherís al código de honor de la UOC, con el que os comprometéis a no compartir vuestro trabajo con otros compañeros o a solicitar de su parte que ellos lo hagan. Asimismo, aceptáis que, de proceder así, es decir, en caso de copia probada, la calificación total de la PEC será de cero, independientemente

## Apéndice: Datasets

<b>GEO Dataset</b>	<b>GEO Serie</b>	<b>Especie</b>	<b>Título</b>
GDS1251	GSE2401	Rattus norvegicus	Acute hypotension effect on kidneys
GDS1736	GSE3737	Homo sapiens	Arachidonic acid effect on prostate cancer cells
GDS2107	GSE3311	Rattus norvegicus	Long-term ethanol consumption effect on pancreas
GDS2153	GSE5370	Homo sapiens	Dermatomyositis
GDS2294	GSE5583	Mus musculus	Histone deacetylase 1 deficient embryonic stem cells
GDS2300	GSE5668	Mus musculus	Germinal vesicle stage and metaphase II stage oocyte comparison

<b>GEO Dataset</b>	<b>GEO Serie</b>	<b>Especie</b>	<b>Título</b>
GDS2406	GSE6077	Mus musculus	Proto-oncogene Nmyc overexpression effect on the embryonic lung
GDS2629	GSE7381	Mus musculus	Epithelial transcription factor Get-1 deficiency effect on embryonic skin
GDS2637	GSE6299	Rattus norvegicus	Keratinocyte cell line response to a (56)Fe ion beam
GDS2646	GSE6868	Gallus gallus	Homocysteine effect on cardiac neural crest cells
GDS2648	GSE6766	Mus musculus	Palmitate effect on myoblast cell line
GDS2686	GSE6376	Mus musculus	MyD88 deficient macrophage response to zymosan
GDS2698	GSE6461	Mus musculus	Synovial sarcoma model
GDS2699	GSE6383	Mus musculus	Mesenchymal and epithelial compartments of the developing intestine
GDS2707	GSE4936	Mus musculus	Neuralized embryoid bodies response to Hedgehog agonist
GDS2727	GSE7196	Mus musculus	Estrogen-related receptor alpha deficiency effect on the heart
GDS2744	GSE7765	Homo sapiens	Dioxin effect on breast cancer cell line (HG-U133A)
GDS2766	GSE6850	Mus musculus	Dominant negative cJun effect on apolipoprotein E deficient livers
GDS2855	GSE3307	Homo sapiens	Various muscle diseases (HG-U133B)
GDS2917	GSE4734	Mus musculus	Various brain regions of several inbred strains
GDS2922	GSE5180	Homo sapiens	Ascending aortic aneurysms

<b>GEO Dataset</b>	<b>GEO Serie</b>	<b>Especie</b>	<b>Título</b>
GDS2935	GSE6281	Homo sapiens	Allergic contact dermatitis: time course
GDS2936	GSE8972	Mus musculus	Neural retina leucine zipper deficiency effect on retinas: time course
GDS4843	GSE42806	Homo sapiens	Skeletal muscle disuse atrophy
GDS3221	GSE6297	Mus musculus	Liver response to human and chimpanzee diets
GDS3223	GSE8853	Homo sapiens	Interleukin-13 effect on esophageal epithelial cells from eosinophilic esophagitis patients
GDS781	GSE1746	Homo sapiens	CD14 cells from granulocyte colony stimulating factor mobilized peripheral blood: expression profile
GDS885	GSE1417	Homo sapiens	Tumor cell response to topoisomerase poison camptothecin
GDS2613	GSE6955	Homo sapiens	Rett syndrome: brain frontal cortex

**Footnotes**

1. La fecha de entrega es la que se indica en el enunciado de la PEC. En caso de no coincidir con la indicada en el aula, ésta será la que predomine.↩