# Introducción a las ómicas

María Elena Grela Gomes

**1. Leer un dataset del *Gene Expression Omnibus***

```r
if (!require(GEOquery)) {
  BiocManager::install("GEOquery")
}
```

Cargando paquete requerido: GEOquery

Cargando paquete requerido: Biobase

Cargando paquete requerido: BiocGenerics

Adjuntando el paquete: 'BiocGenerics'

The following objects are masked from 'package:stats':

    IQR, mad, sd, var, xtabs

The following objects are masked from 'package:base':

    anyDuplicated, aperm, append, as.data.frame, basename, cbind,
    colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
    get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
    match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
    Position, rank, rbind, Reduce, rownames, sapply, saveRDS, setdiff,
    table, tapply, union, unique, unsplit, which.max, which.min

```
Welcome to Bioconductor

    Vignettes contain introductory material; view with
    'browseVignettes()'. To cite Bioconductor, see
    'citation("Biobase")', and for packages 'citation("pkgname")'.


Setting options('download.file.method.GEOquery'='auto')


Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```r
require(GEOquery)
gse <- getGEO("GSE5668", GSEMatrix=TRUE, AnnotGPL=TRUE)
```

```
Found 1 file(s)


GSE5668_series_matrix.txt.gz
```

Como el formato requerido de GEO es una serie, la función **getGeo()** va a devolver una lista, de la cual cada elemento es un objeto de la clase *expressionSet*. Extraemos el primer elemento y mostramos las primeras líneas de los datos de expresión.

```r
eSetFromGEO <- gse[[1]]
head(exprs(eSetFromGEO))
```

```
             GSM132592 GSM132657 GSM132658 GSM132659 GSM132660 GSM132661
1415670_at    8.267416  8.306516  8.075975  6.837325  6.719677  7.155659
1415671_at    8.646356  8.511615  8.460421  7.034081  7.261820  6.796902
1415672_at   10.636901 10.874015 10.600622 10.268658 10.268658 10.348310
1415673_at    4.659725  4.692447  4.415703  4.928355  4.733324  4.923098
1415674_a_at  5.597327  5.796183  5.772778  5.007548  5.039642  5.322467
1415675_at    6.873229  7.052863  6.753307  5.536621  5.836841  5.324139
```

Exploramos la infomación existente de las covariables experimentales.

```r
colnames(pData(eSetFromGEO))
```

```
 [1] "title"                  "geo_accession"
 [3] "status"                 "submission_date"
 [5] "last_update_date"       "type"
 [7] "channel_count"          "source_name_ch1"
 [9] "organism_ch1"           "characteristics_ch1"
[11] "biomaterial_provider_ch1" "treatment_protocol_ch1"
[13] "growth_protocol_ch1"    "molecule_ch1"
[15] "extract_protocol_ch1"   "label_ch1"
[17] "label_protocol_ch1"     "taxid_ch1"
[19] "hyb_protocol"           "scan_protocol"
[21] "description"            "data_processing"
[23] "platform_id"            "contact_name"
[25] "contact_email"          "contact_institute"
[27] "contact_address"        "contact_city"
[29] "contact_state"          "contact_zip/postal_code"
[31] "contact_country"        "supplementary_file"
[33] "data_row_count"         "relation"
```

## 2. Determinar la estructura de los datos y el diseño del estudio

### Estructura de los datos

```
gds <- getGEO("GDS2300")
eSetFromGEO2 <- GDS2eSet(gds,do.log2=FALSE)
```

```
Using locally cached version of GPL1261 found here:
C:\Users\maria\AppData\Local\Temp\RtmpUPGhGw/GPL1261.annot.gz
```

```
dim(exprs(eSetFromGEO2))
```

```
[1] 45101      6
```

La matriz de los datos de expresión tiene 45.101 filas y 6 columnas.

### Diseño del estudio

Utilizando **experimentData()** sobre el *GEO DataSet* convertido en *expressionSet* podemos obtener información sobre el estudio.

3

`experimentData(eSetFromGEO2)`

```
Experiment data
  Experimenter name:
  Laboratory:
  Contact information:
  Title: Germinal vesicle stage and metaphase II stage oocyte comparison
  URL:
  PMIDs: 17022963

  Abstract: A 36 word abstract is available. Use 'abstract' method.
  notes:
   channel_count:
      1
   dataset_id:
      GDS2300
      GDS2300
   description:
      Comparison of oocytes at the germinal vesicle stage (GV) to those at the m
etaphase II (MII) stage. GV-stage oocytes are transcriptionally silent. Re
sults show that the destruction of transcripts that occurs during oocyte m
aturation is selective.
      immature GV
      mature MII
   email:
      geo@ncbi.nlm.nih.gov
   feature_count:
      45101
   institute:
      NCBI NLM NIH
   name:
      Gene Expression Omnibus (GEO)
   order:
      ordered
   platform:
      GPL1261
   platform_organism:
      Mus musculus
   platform_technology_type:
      in situ oligonucleotide
   pubmed_id:
      17022963
```

```
ref:
    Nucleic Acids Res. 2005 Jan 1;33 Database Issue:D562-6
reference_series:
    GSE5668
sample_count:
    6
sample_id:
    GSM132592,GSM132657,GSM132658
    GSM132659,GSM132660,GSM132661
sample_organism:
    Mus musculus
sample_type:
    RNA
title:
    Germinal vesicle stage and metaphase II stage oocyte comparison
type:
    Expression profiling by array
    development stage
    development stage
update_date:
    Mar 27 2007
value_type:
    transformed count
web_link:
    http://www.ncbi.nlm.nih.gov/geo
```

Con esta información, podemos saber que este estudio buscaba comparar la expresión génica entre dos estadíos de ovocitos de ratón: ovocitos en vesícula germinal y ovocitos en metafase II. Para ello, usaron la plataforma GPL1261, que es un microarray de síntesis de oligonucleótidos *in situ* diseñado especialmente para *Mus musculus,* y evaluaron 6 muestras.