Actividad 1 revisada contra la solución

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* **que corresponde a cada posible dataset del estudio**. 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\RtmpaAUyHp/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.

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.

Con Columns() podemos ver la descripción de cada una de las 6 muestras del estudio. Así sabemos que las primeras tres muestras corresponden a vesículas germinales, y las demás a ovocitos en metafase II.

```
Columns(gds)
```

```
    sample development.stage
1 GSM132592     immature GV
2 GSM132657     immature GV
3 GSM132658     immature GV
4 GSM132659      mature MII
```

```
5 GSM132660        mature MII
6 GSM132661        mature MII
                                                      description
1          Value for GSM132592: Fully grown GV-stage oocytes; src: oocyte
2 Value for GSM132657: Fully grown GV-stage oocytes sample #2; src: oocyte
3  Value for GSM132658: Fully grown GV-stage oocyte sample #3; src: oocyte
4                    Value for GSM132659: MII-stage oocytes; src: oocyte
5           Value for GSM132660: Mature MII-oocyte sample #2; src: oocyte
6          Value for GSM132661: Mature MII-oocytes Sample #3; src: oocyte
```

Con pData() sobre el GSE también podemos obtener algo de información sobre el experimento.

```
pData(eSetFromGEO)
```

```
                                           title geo_accession
GSM132592         Fully grown GV-stage oocytes       GSM132592
GSM132657 Fully grown GV-stage oocytes sample #2     GSM132657
GSM132658  Fully grown GV-stage oocyte sample #3     GSM132658
GSM132659                    MII-stage oocytes       GSM132659
GSM132660          Mature MII-oocyte sample #2       GSM132660
GSM132661         Mature MII-oocytes Sample #3       GSM132661
                       status submission_date last_update_date type
GSM132592 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
GSM132657 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
GSM132658 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
GSM132659 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
GSM132660 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
GSM132661 Public on Sep 05 2006     Aug 29 2006      Aug 28 2018  RNA
          channel_count source_name_ch1 organism_ch1
GSM132592             1          oocyte Mus musculus
GSM132657             1          oocyte Mus musculus
GSM132658             1          oocyte Mus musculus
GSM132659             1          oocyte Mus musculus
GSM132660             1          oocyte Mus musculus
GSM132661             1          oocyte Mus musculus

GSM132592                    GV-stage oocytes isolated from eCG primed (44-46h) 22-24d-old :
GSM132657              GV-stage oocytes isolated from eCG-primed (44-46h) 22-24 days old :
GSM132658              GV-stage oocytes isolated from eCG-primed (44-46h) 22-24 days old :
GSM132659    Mature MII-oocytes isolated from eCG(44-46h)/hCG (14h) treated 22-24 days old :
GSM132660 Mature MII-oocyte isolated eCG944-46h)/hCG(14h)-treated female 22-24 days old (C57)
```

```
GSM132661      Mature MII-stage oocytes isolated eCG(44-46h)/hCG(14h)-treated female 22-24 day
            biomaterial_provider_ch1
GSM132592   The Jackson Laboratory
GSM132657   The Jackson Laboratory
GSM132658   The Jackson Laboratory
GSM132659   The Jackson Laboratory
GSM132660   The Jackson Laboratory
GSM132661   The Jackson Laboratory


GSM132592
GSM132657
GSM132658
GSM132659 Ovarian follicular development was stimulated by intraperitoneal injection of 5 IU
GSM132660 Ovarian follicular development was stimulated by intraperitoneal injection of 5 IU
GSM132661 Ovarian follicular development was stimulated by intraperitoneal injection of 5 IU


GSM132592 Released oocyte-cumulus cell complexes (OCCs) were collected and cumulus cells sur
GSM132657 Released oocyte-cumulus cell complexes (OCCs) were collected and cumulus cells sur
GSM132658 Released oocyte-cumulus cell complexes (OCCs) were collected and cumulus cells sur
GSM132659
GSM132660
GSM132661
            molecule_ch1
GSM132592    total RNA
GSM132657    total RNA
GSM132658    total RNA
GSM132659    total RNA
GSM132660    total RNA
GSM132661    total RNA


GSM132592 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
GSM132657 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
GSM132658 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
GSM132659 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
GSM132660 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
GSM132661 Total RNA was extracted from 300 oocytes using the Picopure RNA isolation Kit (Arc
            label_ch1
GSM132592    Biotin
GSM132657    Biotin
GSM132658    Biotin
GSM132659    Biotin
GSM132660    Biotin
GSM132661    Biotin
```

```
GSM132592 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
GSM132657 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
GSM132658 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
GSM132659 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
GSM132660 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
GSM132661 Total RNA was extracted from oocytes using the PicoPure RNA Isolation Kit according
          taxid_ch1                 hyb_protocol        scan_protocol
GSM132592     10090 Standard AffyMetrix procedures Standard AffyMetrix
GSM132657     10090 Standard AffyMetrix procedures Standard AffyMetrix
GSM132658     10090 Standard AffyMetrix procedures Standard AffyMetrix
GSM132659     10090 Standard AffyMetrix procedures Standard AffyMetrix
GSM132660     10090 Standard AffyMetrix procedures Standard AffyMetrix
GSM132661     10090 Standard AffyMetrix procedures Standard AffyMetrix
                                                                    descripti
GSM132592              Fully grown GV-stage oocytes isolated from 22-24d old eCG-primed mi
GSM132657              Fully grown GV-stage oocytes isolated from 22-24d old eCG-primed mi
GSM132658              Fully grown GV-stage oocytes isolated from 22-24d old eCG-primed mi
GSM132659 Mature MII-stage oocytes isolated from 22-24d old eCG(44-46h)/hCG(14h)- treated mi
GSM132660 Mature MII-stage oocytes isolated from 22-24d old eCG(44-46h)/hCG(14h)- treated mi
GSM132661 Mature MII-stage oocytes isolated from 22-24d old eCG(44-46h)/hCG(14h)- treated mi

GSM132592 Probe level data were imported into the R software environment and expression value
GSM132657 Probe level data were imported into the R software environment and expression value
GSM132658 Probe level data were imported into the R software environment and expression value
GSM132659 Probe level data were imported into the R software environment and expression value
GSM132660 Probe level data were imported into the R software environment and expression value
GSM132661 Probe level data were imported into the R software environment and expression value
          platform_id contact_name      contact_email     contact_institute
GSM132592     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
GSM132657     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
GSM132658     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
GSM132659     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
GSM132660     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
GSM132661     GPL1261 You-Qiang,,Su youqiang.su@jax.org The Jackson Laboratory
          contact_address contact_city contact_state contact_zip/postal_code
GSM132592 600 Main Street   Bar Harbor            ME                   04609
GSM132657 600 Main Street   Bar Harbor            ME                   04609
GSM132658 600 Main Street   Bar Harbor            ME                   04609
GSM132659 600 Main Street   Bar Harbor            ME                   04609
GSM132660 600 Main Street   Bar Harbor            ME                   04609
GSM132661 600 Main Street   Bar Harbor            ME                   04609
          contact_country
```

```
GSM132592              USA
GSM132657              USA
GSM132658              USA
GSM132659              USA
GSM132660              USA
GSM132661              USA
                                                        supplementary_file
GSM132592 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132592/suppl/GSM132592.CEL.gz
GSM132657 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132657/suppl/GSM132657.CEL.gz
GSM132658 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132658/suppl/GSM132658.CEL.gz
GSM132659 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132659/suppl/GSM132659.CEL.gz
GSM132660 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132660/suppl/GSM132660.CEL.gz
GSM132661 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM132nnn/GSM132661/suppl/GSM132661.CEL.gz
          data_row_count                     relation
GSM132592           21702 Reanalyzed by: GSE119085
GSM132657           21702 Reanalyzed by: GSE119085
GSM132658           21702 Reanalyzed by: GSE119085
GSM132659           21702 Reanalyzed by: GSE119085
GSM132660           21702 Reanalyzed by: GSE119085
GSM132661           21702 Reanalyzed by: GSE119085
```