

PAC 1 Dades òmiques Raquel Torres

Raquel Torres

2024-11-06

1. Seleccionar un dataset de metabolòmica

Ens baixem un dels datasets del repositori de github i el carreguem a R studio

```
Dades_PAC1<-read.csv("C:/Users/Raque/Desktop/Modul B -Boring/UOC/Anàlisi Dades Òmiques/DataValues_S013.csv")
info_general<-read.csv("C:/Users/Raque/Desktop/Modul B -Boring/UOC/Anàlisi Dades Òmiques/DataInfo_S013.csv")
```

2. Un cop descarregades les dades crea un contenidor del tipus `SummarizedExperiment` que contingui les dades i les metadades (informació sobre el dataset, les files i les columnes). La classe `SummarizedExperiment` és una extensió d'`ExpressionSet` i moltes aplicacions o bases de dades (com `metabolomicsWorkbench`) l'utilitzen en comptes d'usar `expressionSet`.

Dividim el dataset en metadades i dades. Si analitzem les dades, veiem que la primera columna identifica la mostra, fins a la columna 10 s'especifiquen les característiques de cada mostra i a partir de la columna 10 es mostren els diferents metabòlits. Fem la divisió i preparem les dades de la següent forma:

```
metadades_PAC1<-Dades_PAC1[,1:10]
dades_metabolits<-Dades_PAC1[,11:ncol(Dades_PAC1)]
#Marquem com a identificador de la mostra la primera columna
rownames(metadades_PAC1)<-as.character(metadades_PAC1[,1])
#Per evitar duplicitats, treiem la primera columna
metadades<-metadades_PAC1[,1]
#Convertim les metadades en un data frame per colData
colData<-DataFrame(metadades)
#Transposem la matriu de metabòlits per a que cada columna sigui una mostra
dades_metabolits_matrix <- t(as.matrix(dades_metabolits))
```

Creem el `SummarizedExperiment`:

```
SumExp_PAC1<-SummarizedExperiment(assays = list(counts=dades_metabolits_matrix),
                                   colData = colData)
#Afegim la informació adicional que es trobava al repositori de github
metadata(SumExp_PAC1)<-info_general
SumExp_PAC1
```

```
## class: SummarizedExperiment
## dim: 686 39
## metadata(4): X VarName varTpe Description
## assays(1): counts
## rownames(686): GLU_T0 INS_T0 ... SM.C24.0_T5 SM.C24.1_T5
## rowData names(0):
## colnames(39): 1 2 ... 38 39
## colData names(9): SUBJECTS SURGERY ... MEDINF_TO MEDHTA_TO
```

3. Porteu a terme una exploració del dataset que us proporcioni una visió general en la línia del que hem vist en les activitats

Les dades formen part d'un estudi que analitza la resposta de diferents metabolits abans, al mes, als 3 mesos i als 6 mesos després de la cirurgia bariàtrica independentment de la pèrdua de pes. S'analitzen 39 pacients que han estat sotmesos a bypass en Y de Roux (26) o sleeve gàstric (13). La mitja d'edat dels pacients és de 40.8 i la SD 9.8 anys. Si els dividim pels grups assignats (1 i 2), la mitjana del grup 1 és 40.87 (SD 9.5) anys i la del grup 2 40.7 (SD 10.8), sense diferències estadísticament significatives (p 0.06).

```
#Podem fer un anàlisi del contenidor
```

```
length(SumExp_PAC1)
```

```
## [1] 686
```

```
head(metadata(SumExp_PAC1)) #podem comprovar que la informació general s'hagi guardat correctament i qu
```

```
##           X VarName      varType Description
## 1 SUBJECTS SUBJECTS    integer  dataDesc
## 2 SURGERY  SURGERY  character  dataDesc
## 3   AGE    AGE      integer  dataDesc
## 4  GENDER  GENDER  character  dataDesc
## 5   Group   Group    integer  dataDesc
## 6 MEDDM_TO MEDDM_TO    integer  dataDesc
```

```
head(rownames(SumExp_PAC1)) #podem saber els noms i codificació dels metabòlits analitzats
```

```
## [1] "GLU_TO"          "INS_TO"          "HOMA_TO"
## [4] "HBA1C_TO"        "HBA1C.mmol.mol_TO" "PESO_TO"
```

```
names(colData(SumExp_PAC1)) #podem demanar els noms de les diferents metadades recollides
```

```
## [1] "SUBJECTS" "SURGERY" "AGE"      "GENDER" "Group" "MEDDM_TO"
## [7] "MEDCOL_TO" "MEDINF_TO" "MEDHTA_TO"
```

```
#Podem analitzar alguna dada, com les diferències entre edats entre els diferents grups.
```

```
ages<-colData(SumExp_PAC1)$AGE
groups<-colData(SumExp_PAC1)$Group
tapply(ages, groups, mean)
```

```
##           1           2
## 40.87500 40.66667
```

```
tapply(ages, groups, sd)
```

```
##           1           2
##  9.511149 10.794619
```

```
t.test(AGE ~ Group, data = colData)
```

```
##
## Welch Two Sample t-test
##
## data: AGE by Group
## t = 0.061334, df = 27.011, p-value = 0.9515
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
## -6.760964 7.177631
## sample estimates:
## mean in group 1 mean in group 2
## 40.87500 40.66667
```

4. Elabora un informe que descrigui el procés que heu realitzat, incloent-hi la descàrrega de les dades, la creació del contenidor, l'exploració de les dades i la reposició de les dades en github. El nom del repositori ha de ser el següent: APELLIDO1-Cognom2-Nom-PEC1. Per exemple en el meu cas el repositori s'anomenaria: "Sanchez-Pla-Alex-PEC1"

Generem el present informe en Rmarckdown amb els passos seguits per a la creació del SummarizedExperiment i l'exploració de les dades.

5. Creau un repositori de github2 que contingui l'informe, l'objecte contenidor amb les dades i les metadades en format binari de R (arxiu amb extensió .Rda), el codi R per a l'exploració de les dades o dades en format text i o les metadades sobre el dataset en un arxiu markdown.

Primer he generat el repositori a Github i, en crear el projecte a RStudio l'he enllaçat amb el repositori via la URL. He anat envaint els diferents documents al repositori a mesura que anava treballant la PAC. He generat el document de R, i he guardat el contenidor en un document amb extensió .Rda de la següent forma:

```
save(SumExp_PAC1, file = "SumExp_PAC1.Rda")
```

URL per accedir al repositori: <https://github.com/rtorresi/Torres-Iglesias-Raquel-PAC1.git>