

gse1397

Guillermo Ayala

2025-03-11

Table of contents

gse1397	1
Packages	1
Adding Entrez and Ensembl identifiers	2
gse1397ed	3
From ExpressionSet to ExpressionData	3
gse1397.gsc	3
Making the gene set collection	3

gse1397

Packages

```
pacman::p_load("GEOquery","Biobase","AnnotationDbi","BiocGenerics","tami")
```

Download dataset from GEO.

```
gse1397raw = GEOquery::getGEO("GSE1397")[[1]]
```

Setting anotation.

```
Biobase::annotation(gse1397raw) = "hgu133a.db"
```

Select *samples* corresponding to cerebrum and cerebellum.

```
(ts21=c(grep("T.*21.*cerebrum", as.character(pData(gse1397raw)[,"title"]))),
  grep("TS21.*cerebellum", as.character(pData(gse1397raw)[,"title"]))))
(eu=c(grep("Euploid.*cerebrum", as.character(pData(gse1397raw)[,"title"]))),
  grep("Euploid.*[Cc]erebellum", as.character(pData(gse1397raw)[,"title"]))))
gse1397raw = gse1397raw[,c(eu,ts21)]
colnames0 = colnames(gse1397raw)
```

- Modify phenotype.

```
tissue = factor(rep(c(1,2,1,2),c(4,3,4,3)),levels=1:2,
  labels=c("Cerebrum","Cerebellum"))
type = factor(c(rep(1,7),rep(2,7)),levels=1:2,
  labels=c("Euploid","TS21"))
pData(gse1397raw) = data.frame(tissue,type)
colnames(gse1397raw) = colnames0
```

- Saving data

Normalize the samples (quantile normalization).

```
gse1397 = gse1397raw
exprs(gse1397) = limma::normalizeBetweenArrays(exprs(gse1397))
```

Adding Entrez and Ensembl identifiers

```
pacman::p_load("hgu133a.db")
probeid2ee = AnnotationDbi::select(hgu133a.db,keys=featureNames(gse1397),
  columns=c("ENTREZID","ENSEMBL"),keytype="PROBEID")
```

Note that the correspondence between PROBEID and ENTREZID is not 1-1. We have to select one correspondence.

```
indices = BiocGenerics::match(featureNames(gse1397),probeid2ee$PROBEID)

fData(gse1397) = probeid2ee[indices,]
```

Saving normalized data

```
save(gse1397,file='gse1397.rda')
```

This is the data set is tamidata::gse1397.

gse1397ed

From ExpressionSet to ExpressionData

Usually, we will start our analysis with a data set organized using a Biobase::ExpressionSet. For instance, tamidata::gse1397. It is easy to construct an ExpressionData.

```
x = ExpressionData(exprm = exprs(gse1397),  
                   groups = pData(gse1397)[,"type"],  
                   type="microarray")
```

gse1397.gsc

Making the gene set collection

```
pacman::p_load(GSEABase)  
gse1397.gsc = GeneSetCollection(gse1397, setType=GOCollection())  
names(gse1397.gsc) = unlist(lapply(gse1397.gsc, setName))
```