Deconvolución de datos de melanoma con quantiseqr: Datos bulk RNA-seq: GSE54467

Elena Eyre Sánchez, PhD

2024-10-27

Contents

1 Introducción y Objetivo

2 Paquetes y datos

Repositorio GitHub de : https://github.com/Danko-Lab/quantiseqr/blob/main/tutorial_deconvolution.pdf

```
knitr::opts_chunk$set(warning=FALSE)
#if (!requireNamespace("BiocManager", quietly = TRUE)) {
#  install.packages("BiocManager")
#}
#BiocManager::install("quantiseqr")
package_to_load <- c("readr", "dplyr", "ggplot2", "tidyr", "dplyr","RColorBrewer",
                     "Biobase","quantiseqr", "gplots")
for (package in package_to_load) {
  require(package, character.only = T); packageVersion(package)
}
extra_to_load <- c("knitr","stringr", "stringi", "ggrepel","ggpubr", "ggbreak",
                   "reshape2","ggfortify","cowplot","GEOquery","Seurat","data.table",
                   "limma","illuminaHumanv4.db","SummarizedExperiment","tibble")
for (package in extra_to_load) {
  require(package, character.only = T); packageVersion(package)
}
rm(package_to_load, extra_to_load)
```

#Datos

#Deconvolución

En este análisis utilizo los datos del estudio GSE54467 descargados mediante la función `getGEO` des de la base de datos GEO, del NCBI: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54467.

Las muestras consisten en 65 muestras analizadas con la plataformas GPL570, y con varios tratamientos: dabrafenib + trametinib.

Este estudio es de especial interés para el TFM debido a que los autores también proporcionan metadata la respuesta de los pacientes, cosa que permitirá estudiar posibles correlaciones con las poblaciones obtenidas de

la deconvolución.

```
setwd("~/Desktop/ELENA_UOC/TFM")

gset <- getGEO("GSE22155", GSEMatrix =TRUE, getGPL=FALSE)
if (length(gset) > 1) idx <- grep("GPL6102", attr(gset, "names")) else idx <- 1
gset_GPL6102 <- gset[[idx]]
#table(gset_GPL6102$characteristics_ch1.1) # OS (days)
#table(gset_GPL6102$characteristics_ch1.2) # event (0=alive, 1=dead)
#table(gset_GPL6102$characteristics_ch1.3) # sex
#table(gset_GPL6102$characteristics_ch1.4) # age at metastases
#table(gset_GPL6102$characteristics_ch1.5) # type of metastases: Lymohnode
#table(gset_GPL6102$characteristics_ch1.6) # age at primary diagnosis
#table(gset_GPL6102$characteristics_ch1.7) # localization of primary melanoma
#table(gset_GPL6102$characteristics_ch1.8) # type
#table(gset_GPL6102$characteristics_ch1.9) # breslow
#table(gset_GPL6102$characteristics_ch1.10) # clark
#table(gset_GPL6102$characteristics_ch1.11) # Stage (III and IV)
#table(gset_GPL6102$characteristics_ch1.12) # braf/nras
#table(gset_GPL6102$characteristics_ch1.13) # cdkn2a (hd=homozygous deletion, *=germline)
#table(gset_GPL6102$characteristics_ch1.14) # molecular subtype
#table(gset_GPL6102$characteristics_ch1.15)  # cd3 immunohistochemistry
#table(gset_GPL6102$characteristics_ch1.16) # cd20 immunohistochemistry
#table(gset_GPL6102$characteristics_ch1.17) # ki67 (0=<30%, 1=>30%)
#table(gset_GPL6102$`age at metastases:ch1`) # Age at metastases
#table(gset_GPL6102$`age at primary diagnosis:ch1`) # age at primary diagnosis
#table(gset_GPL6102$`localization of primary melanoma:ch1`) # localization of primary melanoma
#table(gset_GPL6102$`molecular subtype:ch1`) # molecular subtype

if (length(gset) > 1) idx <- grep("GPL6947", attr(gset, "names")) else idx <- 1
gset_GPL6947 <- gset[[idx]]
#table(gset_GPL6947$characteristics_ch1.1)# os (days)
#table(gset_GPL6947$characteristics_ch1.2) # event (0=alive, 1=dead):
#table(gset_GPL6947$characteristics_ch1.3) # sex
#table(gset_GPL6947$characteristics_ch1.4) # age at metastases
#table(gset_GPL6947$characteristics_ch1.5) # type of metastases
#table(gset_GPL6947$characteristics_ch1.6) # age at primary diagnosis
#table(gset_GPL6947$characteristics_ch1.7) # localization of primary melanoma
#table(gset_GPL6947$characteristics_ch1.8) # type
#table(gset_GPL6947$characteristics_ch1.9) # breslow
#table(gset_GPL6947$characteristics_ch1.10) # clark
#table(gset_GPL6947$characteristics_ch1.11) # stage
#table(gset_GPL6947$characteristics_ch1.12) # braf/nras
#table(gset_GPL6947$characteristics_ch1.13) # cdkn2a (hd=homozygous deletion, *=germline)
#table(gset_GPL6947$characteristics_ch1.14) # cdkn2a (hd=homozygous deletion, *=germline)
#table(gset_GPL6947$characteristics_ch1.15) # cd3 immunohistochemistry = NAs
#table(gset_GPL6947$characteristics_ch1.16) # cd20 immunohistochemistry = NAs
#table(gset_GPL6947$characteristics_ch1.17) # ki67 (0=<30%, 1=>30%) = NAs
#table(gset_GPL6947$`localization of primary melanoma:ch1`) # localization of primary melanoma
#table(gset_GPL6947$`molecular subtype:ch1`) # molecular subtype
#table(gset_GPL6947$`tissue:ch1`) # tissue
#table(gset_GPL6947$`stage:ch1`) # All IV
#table(gset_GPL6947$`type of metastases:ch1`)# Type if metastases
```

```r
# Debido a que los autores proporcionan los genes con la nomeclatura de Illumina, lo convierto a símbol
x <- illuminaHumanv4SYMBOL # cargado con el paquete illuminaHumanv4.db
mapped_probes <- mappedkeys(x) # Para sacar los símbolos
xx <- as.list(x[mapped_probes]) # Lo paso a listado
my_genes_GPL6102 <- as.data.frame(unlist(xx[(rownames(gset_GPL6102@assayData$exprs))])) # Lo convierto
my_genes_GPL6947 <- as.data.frame(unlist(xx[(rownames(gset_GPL6947@assayData$exprs))])) # Lo convierto
my_genes_GPL6102$gene <- rownames(my_genes_GPL6102)
my_genes_GPL6947$gene <- rownames(my_genes_GPL6947)

bulk_metadata_GPL6102 <- as.data.frame(gset_GPL6102@phenoData@data) # Paso la metadata disponible a una
bulk_metadata_GPL6947 <- as.data.frame(gset_GPL6947@phenoData@data) # Paso la metadata disponible a una

# Para usar los símbolos en lugar de nombres de ilumina, extraigo los datos de expresión:
bulk.mtx_GPL6102 <- as.data.frame(gset_GPL6102@assayData$exprs) # Los datos de expresión
bulk.mtx_GPL6947 <- as.data.frame(gset_GPL6947@assayData$exprs) # Los datos de expresión
bulk.mtx_GPL6102$gene <- rownames(bulk.mtx_GPL6102) # La columna que usaré para integrar
bulk.mtx_GPL6947$gene <- rownames(bulk.mtx_GPL6947) # La columna que usaré para integrar
bulk.mtx_GPL6102 <- inner_join(my_genes_GPL6102, bulk.mtx_GPL6102, by = "gene") # Integración de ambas
bulk.mtx_GPL6947 <- inner_join(my_genes_GPL6947, bulk.mtx_GPL6947, by = "gene") # Integración de ambas
bulk.mtx_GPL6102$gene <- NULL # Elimino la columna con nombres de Illumina
bulk.mtx_GPL6947$gene <- NULL # Elimino la columna con nombres de Illumina
colnames(bulk.mtx_GPL6102)[1] <- "symbols" # Nombro la columna de símbolos de los genes
colnames(bulk.mtx_GPL6947)[1] <- "symbols" # Nombro la columna de símbolos de los genes

# Agrego los posibles duplicados calculando la media:
bulk.mtx_GPL6102 <- aggregate(bulk.mtx_GPL6102, by = list(c(bulk.mtx_GPL6102$symbols)), mean) # Agregar
bulk.mtx_GPL6947 <- aggregate(bulk.mtx_GPL6947, by = list(c(bulk.mtx_GPL6947$symbols)), mean) # Agregar
rownames(bulk.mtx_GPL6102) <- bulk.mtx_GPL6102$Group.1 # Los nombres de genes únicos sin duplicados sir
rownames(bulk.mtx_GPL6947) <- bulk.mtx_GPL6947$Group.1 # Los nombres de genes únicos sin duplicados sir
bulk.mtx_GPL6102 <- bulk.mtx_GPL6102[,-c(1:2)] # Elimino las columnas usadas para conseguir los nombres
bulk.mtx_GPL6947 <- bulk.mtx_GPL6947[,-c(1:2)] # Elimino las columnas usadas para conseguir los nombres

# Convertir los datos de expresión del bulk RNA-seq a objeto ExpressionSet:
bulk.eset_GPL6102 <- Biobase::ExpressionSet(assayData = as.matrix(as.data.frame(bulk.mtx_GPL6102)))
bulk.eset_GPL6947 <- Biobase::ExpressionSet(assayData = as.matrix(as.data.frame(bulk.mtx_GPL6947)))
print("Object associated to platform GPL6102:")
```

```
## [1] "Object associated to platform GPL6102:"
```

```r
bulk.eset_GPL6102
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 16269 features, 57 samples
##   element names: exprs
## protocolData: none
## phenoData: none
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```

```r
print("Object associated to platform GPL6947:")
```

```
## [1] "Object associated to platform GPL6947:"
```

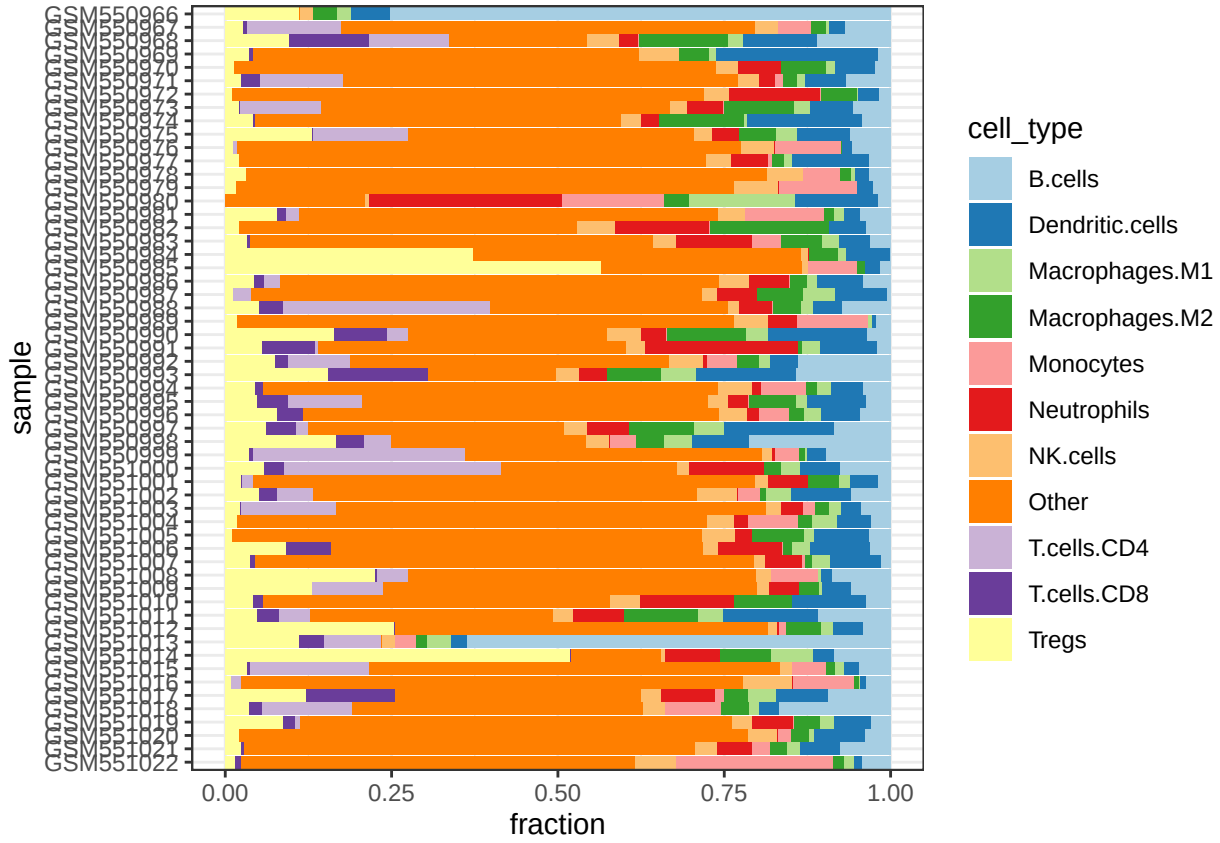
```r
bulk.eset_GPL6947
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 19130 features, 22 samples
##   element names: exprs
## protocolData: none
## phenoData: none
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```

```
ti_racleGPL6102 <- quantiseqr::run_quantiseq(
  expression_data = bulk.eset_GPL6102@assayData$exprs,
  signature_matrix = "TIL10",
  is_arraydata = FALSE,
  is_tumordata = TRUE,
  scale_mRNA = TRUE
)
ti_racleGPL6947 <- quantiseqr::run_quantiseq(
  expression_data = bulk.eset_GPL6947@assayData$exprs,
  signature_matrix = "TIL10",
  is_arraydata = FALSE,
  is_tumordata = TRUE,
  scale_mRNA = TRUE
)
print("Plataforma GPL6102:")
```

```
## [1] "Plataforma GPL6102:"
```

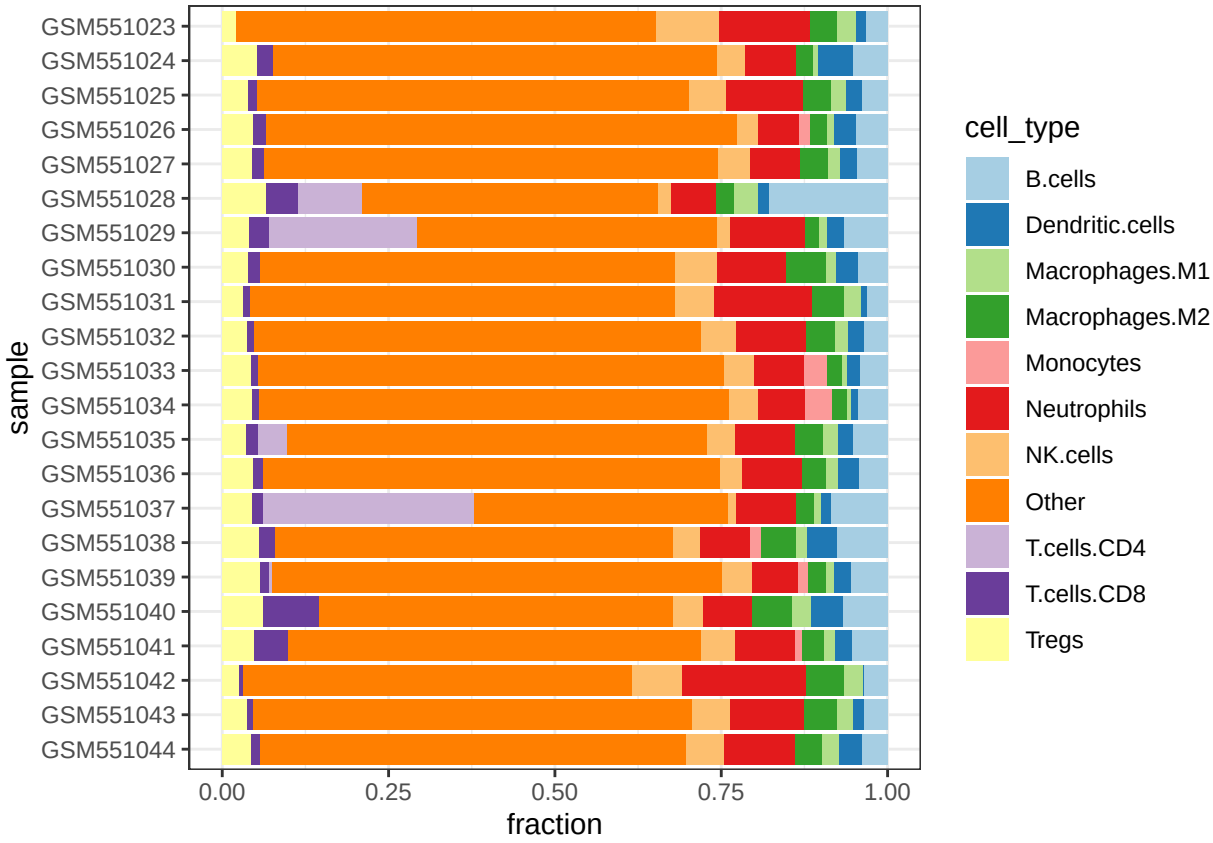
```
quantiplot(ti_racleGPL6102)
```



```
print("Plataforma GPL6947:")
```

```
## [1] "Plataforma GPL6947:"
```

```
quantiplot(ti_racleGPL6947)
```



Encontramos las proporciones del bulk RNA-seq en el apartado `ti_racle`, el qual puedo integrar en la metadata que ya tenía y almacenar en un archivo para posteriores análisis.

```
ref.based.estimates_GPL6102 <- as.data.frame(ti_racleGPL6102)
ref.based.estimates_GPL6947 <- as.data.frame(ti_racleGPL6947)
ref.based.estimates_GPL6102$geo_accession <- rownames(ref.based.estimates_GPL6102)
ref.based.estimates_GPL6947$geo_accession <- rownames(ref.based.estimates_GPL6947)
ref.based.estimates_GPL6102 <- inner_join(ref.based.estimates_GPL6102, bulk_metadata_GPL6102, by = "geo
ref.based.estimates_GPL6947 <- inner_join(ref.based.estimates_GPL6947, bulk_metadata_GPL6947, by = "geo
knitr::kable(head(ref.based.estimates_GPL6102[,1:7]), digits=2, caption = "Sección de las primeras mues
```

Table 1: Sección de las primeras muestras como ejemplo del resultado con la plataforma GPL6102

Sample	B.cells	Macrophages.M1	Macrophages.M2	Monocytes	Neutrophils	NK.cells
GSM550966	0.75	0.02	0.04	0.00	0.00	0.02
GSM550967	0.07	0.01	0.02	0.05	0.00	0.04
GSM550968	0.11	0.02	0.14	0.00	0.03	0.05
GSM550969	0.02	0.01	0.05	0.00	0.00	0.06
GSM550970	0.02	0.01	0.07	0.00	0.06	0.03
GSM550971	0.07	0.01	0.02	0.01	0.02	0.03

```
knitr::kable(head(ref.based.estimates_GPL6947[,1:7]), digits=2, caption = "Sección de las primeras muest
```

Table 2: Sección de las primeras muestras como ejemplo del resultado con la plataforma GPL6947

Sample	B.cells	Macrophages.M1	Macrophages.M2	Monocytes	Neutrophils	NK.cells
GSM551023	0.03	0.03	0.04	0.00	0.14	0.09
GSM551024	0.05	0.01	0.03	0.00	0.08	0.04
GSM551025	0.04	0.02	0.04	0.00	0.12	0.05
GSM551026	0.05	0.01	0.03	0.02	0.06	0.03
GSM551027	0.05	0.02	0.04	0.00	0.08	0.05
GSM551028	0.18	0.04	0.03	0.00	0.07	0.02

```
write.csv(ref.based.estimates_GPL6102,"./quantiseqr_GSE22155_GPL6102.csv", row.names = FALSE)
write.csv(ref.based.estimates_GPL6947,"./quantiseqr_GSE22155_GPL6947.csv", row.names = FALSE)
```