

Deconvolución de datos de melanoma con quantiseqr: Datos bulk RNA-seq: GSE91061

Elena Eyre Sánchez, PhD

2024-11-16

Contents

1	Introducción y Objetivo	1
2	Paquetes y datos	1

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 GSE91061 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=GSE91061>.

Las muestras consisten en 109 muestras analizadas con la plataformas GPL9052, y con varios tratamientos: anti-CTLA4 y anti-PD1.

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")

GSE91061_BMS038109Sample_hg19KnownGene_fpkms_csv <- read_csv("Deconvolution_analysis/GSE91061_BMS038109S
gex <- as.data.frame(GSE91061_BMS038109Sample_hg19KnownGene_fpkms_csv[,-1])
rownames(gex) <- GSE91061_BMS038109Sample_hg19KnownGene_fpkms_csv$...1
gex$ENTREZID <- rownames(gex)

anno.result <- select(org.Hs.eg.db, keys=as.character(rownames(gex)), columns="SYMBOL", keytype="ENTREZID")
anno.result <- anno.result[!is.na(anno.result$ENTREZID),]

gex <- inner_join(gex, anno.result, by = "ENTREZID")

gset <- getGEO("GSE91061", GSEMatrix = TRUE, getGPL = FALSE)
if (length(gset) > 1) idx <- grep("GPL9052", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]
#table(gset$characteristics_ch1.1) # response
#table(gset$visit (pre or on treatment):ch1) # Visit: Pre / On
bulk_metadata <- as.data.frame(gset@phenoData@data) # Paso la metadata disponible a una tabla

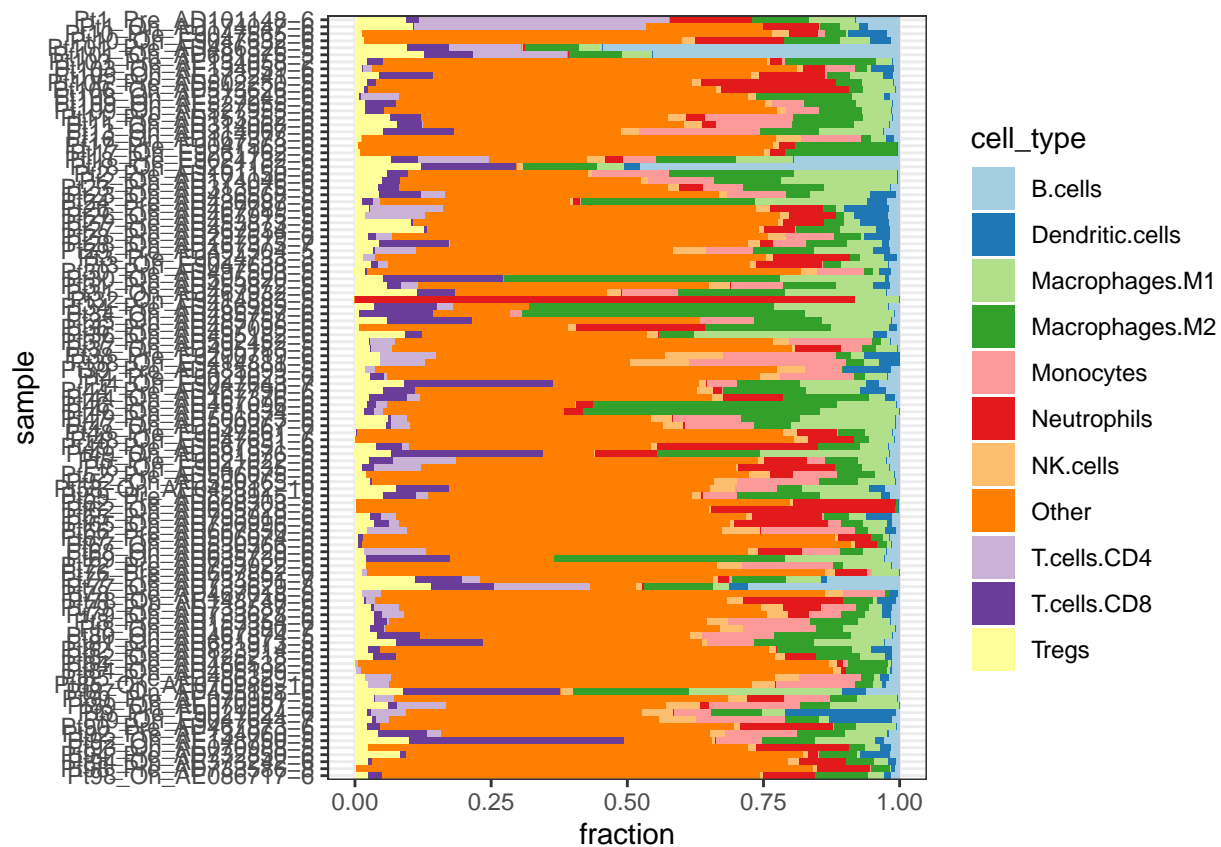
# Para usar los símbolos en lugar de nombres de ilumina, extraigo los datos de expresión:
bulk.mtx <- as.data.frame(gset[,1:109]) # Los datos de expresión
bulk.mtx$symbols <- gset$SYMBOL # La columna que usaré para integrar

# Agrego los posibles duplicados calculando la media:
bulk.mtx <- aggregate(bulk.mtx, by = list(c(bulk.mtx$symbols)), mean) # Agregar
rownames(bulk.mtx) <- bulk.mtx$Group.1 # Los nombres de genes únicos sin duplicados sirven para dar nom
Probes <- bulk.mtx$Group.1
bulk.mtx$Group.1 <- NULL # Elimino las columnas usadas para conseguir los nombres
bulk.mtx$symbols <- NULL # Elimino las columnas usadas para conseguir los nombres

# Convertir los datos de expresión del bulk RNA-seq a objeto ExpressionSet:
bulk.eset <- Biobase::ExpressionSet(assayData = as.matrix(as.data.frame(bulk.mtx)))
bulk.eset

## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 22070 features, 109 samples
## element names: exprs
## protocolData: none
## phenoData: none
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:

ti_racle <- quantiseqr::run_quantiseqr(
  expression_data = bulk.eset@assayData$exprs,
  signature_matrix = "TIL10",
  is_arraydata = FALSE,
  is_tumordata = TRUE,
  scale_mRNA = TRUE
)
quantipLOT(ti_racle)
```



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 <- as.data.frame(ti_racle)
ref.based.estimates$title <- rownames(ref.based.estimates)
ref.based.estimates <- inner_join(ref.based.estimates, bulk_metadata, by = "title")
knitr::kable(head(ref.based.estimates[,1:7]), digits=2, caption = "Sección de las primeras muestras como ejemplo del resultado")
```

Table 1: Sección de las primeras muestras como ejemplo del resultado

Sample	B.cells	Macrophages.M1	Macrophages.M2	Monocytes	Neutrophils	NK.cells
Pt1_Pre_AD101148-6	0.08	0.09	0.11	0.00	0.15	0.00
Pt1_On_AD174047-6	0.06	0.04	0.04	0.00	0.10	0.00
Pt10_Pre_E9047565-6	0.02	0.00	0.04	0.01	0.05	0.03
Pt10_On_E9047632-6	0.02	0.06	0.10	0.00	0.16	0.02
Pt101_Pre_AD486328-5	0.55	0.04	0.10	0.00	0.00	0.00
Pt101_On_AD681975-5	0.45	0.06	0.10	0.00	0.00	0.00

```
write.csv(ref.based.estimates, "./quantiseqr_GSE91061.csv", row.names = FALSE)
```