

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

Elena Eyre Sánchez, PhD

2024-10-27

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", "biomaRt", "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 TCGA_SKCM 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=TCGA_SKCM.

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")

gex <- read_delim("Datasets/TCGA-SKCM.htseq_fpkm.tsv", delim = "\t",
                  escape_double = FALSE, trim_ws = TRUE)
bulk_metadata <- read_delim("Datasets/TCGA-SKCM.GDC_phenotype.tsv", delim = "\t",
                            escape_double = FALSE, trim_ws = TRUE)
#os_data <- read_delim("Datasets/TCGA-SKCM.survival.tsv", delim = "\t",
#                      escape_double = FALSE, trim_ws = TRUE)

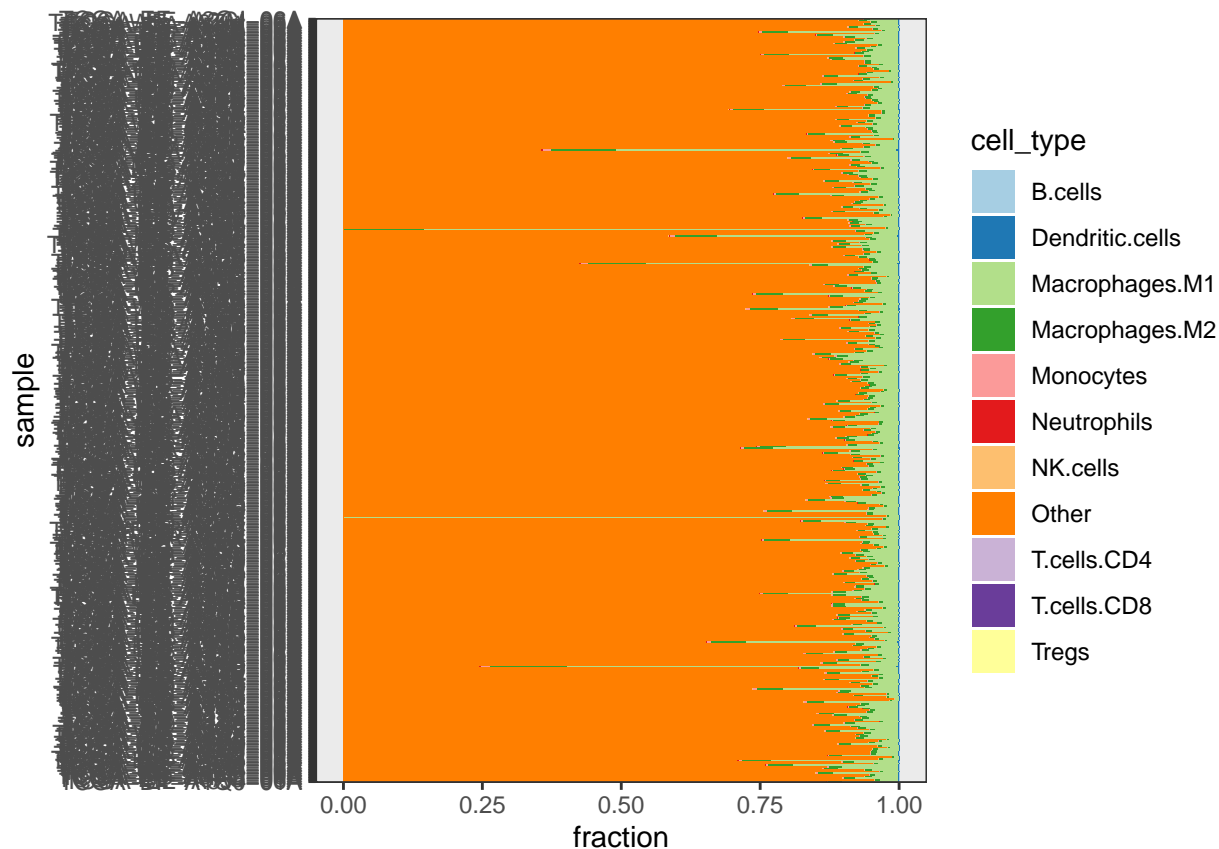
# Debido a que los autores proporcionan los genes con la nomenclatura de ENSEMBL, lo convierto a símbolos
mart <- useDataset("hsapiens_gene_ensembl", useMart("ensembl"))
genes <- gex$Ensembl_ID
G_list <- getBM(filters= "ensembl_gene_id_version", attributes= c("ensembl_gene_id_version","hgnc_symbol"))

# Para usar los símbolos en lugar de nombres de ENSEMBL:
gex2 <- merge(gex,G_list,by.x="Ensembl_ID",by.y="ensembl_gene_id_version")
gex2<-gex2[,-1]
gex2 <- aggregate(gex2, by = list(c(gex2$hgnc_symbol)), mean) # Agrego los posibles duplicados calculando la media
gex2 <- gex2[,-c(ncol(gex2))]
rownames(gex2) <- gex2$Group.1 # Los nombres de genes únicos sin duplicados sirven para dar nombre a la matriz
Probes <- gex2$Group.1
gex2$Group.1 <- NULL # Elimino las columnas usadas para conseguir los nombres
bulk.mtx <- as.data.frame(gex2) # Los datos de expresión

# 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: 15584 features, 472 samples
## element names: exprs
## protocolData: none
## phenoData: none
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:

ti_racle <- quantiseqr::run_quantiseq(
  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 cual 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$submitter_id.samples <- rownames(ref.based.estimates)
ref.based.estimates <- inner_join(ref.based.estimates, bulk_metadata, by = "submitter_id.samples")
knitr::kable(head(ref.based.estimates[,1:7]), digits=2, caption = "Sección de las primeras muestras y c
```

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

Sample	B.cells	Macrophages.M1	Macrophages.M2	Monocytes	Neutrophils	NK.cells
TCGA-EE-A2GJ-06A	0	0.05	0.01	0	0	0
TCGA-EE-A2GI-06A	0	0.06	0.01	0	0	0
TCGA-WE-A8ZM-06A	0	0.04	0.01	0	0	0
TCGA-DA-A1IA-06A	0	0.05	0.01	0	0	0
TCGA-D3-A51H-06A	0	0.03	0.01	0	0	0
TCGA-XV-A9VZ-01A	0	0.07	0.02	0	0	0

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