

RNA-seq COVID

Antonio Osamu Katagiri Tanaka - A01212611@itesm.mx

April 17, 2020

D. Blanco-Melo, B. Nilsson-Payant, W.-C. Liu, R. Moeller, M. Panis, D. Sachs, R. Albrecht, B.R. TenOever, SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems, *BioRxiv*. (2020) 2020.03.24.004655. <https://doi.org/10.1101/2020.03.24.004655>.

```
# Clear all objects (from the workspace)
```

```
rm(list = ls())
```

```
# Suppress Warning messages
```

```
options(warn = -1)
```

```
# Turn off scientific notation like 1e+06
```

```
options(stringsAsFactors = F)
```

```
# INSTALL with:
```

```
# if (!requireNamespace("BiocManager", quietly = TRUE))
```

```
#   install.packages("BiocManager")
```

```
# BiocManager::install("GEOquery")
```

```
# BiocManager::install("DESeq2")
```

```
# BiocManager::install("limma")
```

```
# library(devtools)
```

```
# install_github("tpq/exprso")
```

```
# LOAD Libs
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse :
```

```
## v ggplot2 3.3.0      v purrr   0.3.3
```

```
## v tibble  2.1.3      v stringr 1.4.0
```

```
## v tidyr   1.0.2      v forcats 0.5.0
```

```
## v readr   1.3.1
```

```
## -- Conflicts ----- tidyverse_confli
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
library(DESeq2)
```

```
## Loading required package: S4Vectors
```

```

## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:dplyr':
##
##   combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
##   IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##   anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##   dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##   grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##   order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##   rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##   union, unique, unsplit, which, which.max, which.min
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:tidyr':
##
##   expand
## The following objects are masked from 'package:dplyr':
##
##   first, rename
## The following object is masked from 'package:base':
##
##   expand.grid
## Loading required package: IRanges
##
## Attaching package: 'IRanges'
## The following object is masked from 'package:purrr':
##
##   reduce
## The following objects are masked from 'package:dplyr':
##
##   collapse, desc, slice
## The following object is masked from 'package:grDevices':
##
##   windows
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: Biobase

```

```
## Welcome to Bioconductor
##
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: DelayedArray
## Loading required package: matrixStats

##
## Attaching package: 'matrixStats'

## The following objects are masked from 'package:Biobase':
##
## anyMissing, rowMedians

## The following object is masked from 'package:dplyr':
##
## count

## Loading required package: BiocParallel

##
## Attaching package: 'DelayedArray'

## The following objects are masked from 'package:matrixStats':
##
## colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

## The following object is masked from 'package:purrr':
##
## simplify

## The following objects are masked from 'package:base':
##
## aperm, apply, rowsum

library(limma)

##
## Attaching package: 'limma'

## The following object is masked from 'package:DESeq2':
##
## plotMA

## The following object is masked from 'package:BiocGenerics':
##
## plotMA

# LOAD provided functions
source("../script_ejercicios.R")
```

Download GDS file, put it in the current directory, and load it

```
datos <- load("../GSE147507_datos_covid.Rdata")
# head(datos_covid)
names(datos_covid)

## [1] "SARS004_mock_3" "SARS004_mock_2"
## [3] "SARS004_mock_1" "SARS004_CoV2_3"
## [5] "SARS004_CoV2_2" "SARS004_CoV2_1"
## [7] "CoV002.mock3.indexG3" "CoV002.mock2.indexG2"
## [9] "CoV002.mock1.indexG1" "CoV002.CoV2.3.indexG6"
## [11] "CoV002.CoV2.2.indexG5" "CoV002.CoV2.1.indexG4"
## [13] "svRNA184.mock.3.indexF3" "svRNA184.mock.1.indexF1"
## [15] "svRNA184.RSV.3.indexH9" "svRNA184.RSV.1.indexF4"
## [17] "X3_9_mock1_13" "X3_9_mock2_14"
```

```
## [19] "X3_9_wt1_15"                "X3_9_wt2_16"

# https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147507

download.file(url = 'https://ftp.ncbi.nlm.nih.gov/geo/series/GSE147nnn/GSE147507/suppl/GSE147507_RawReadCounts_Human.tsv.gz',
              destfile = './NCBI_GEO/GSE147507_RawReadCounts_Human.tsv.gz')

gse147507 <-
  read.table(
    './NCBI_GEO/GSE147507_RawReadCounts_Human.tsv.gz',
    header = T,
    stringsAsFactors = F
  )
```

Cell Culture: * Normal human bronchial epithelial (NHBE) * Human adenocarcinomic alveolar basal epithelial (A549) cells

Viruses: * SARS-related coronavirus 2 (SARS.CoV.2) * influenza A/Puerto Rico/8/1934 (H1N1) virus (IAV) * human respiratory syncytial virus (RSV)

```
# DROP extra data
gse147507 <-
  gse147507 %>% select(-contains(c(
    "ACE2", "Calu3", "HPIV3", "NS1", "IFNB", "Lung"
  )))

# head(gse147507)
names(gse147507)
```

```
## [1] "Series1_NHBE_Mock_1"      "Series1_NHBE_Mock_2"
## [3] "Series1_NHBE_Mock_3"      "Series1_NHBE_SARS.CoV.2_1"
## [5] "Series1_NHBE_SARS.CoV.2_2" "Series1_NHBE_SARS.CoV.2_3"
## [7] "Series2_A549_Mock_1"      "Series2_A549_Mock_2"
## [9] "Series2_A549_Mock_3"      "Series2_A549_SARS.CoV.2_1"
## [11] "Series2_A549_SARS.CoV.2_2" "Series2_A549_SARS.CoV.2_3"
## [13] "Series3_A549_Mock_1"      "Series3_A549_Mock_2"
## [15] "Series3_A549_RSV_1"       "Series3_A549_RSV_2"
## [17] "Series4_A549_Mock_1"      "Series4_A549_Mock_2"
## [19] "Series4_A549_IAV_1"       "Series4_A549_IAV_2"
## [21] "Series5_A549_Mock_1"      "Series5_A549_Mock_2"
## [23] "Series5_A549_Mock_3"      "Series5_A549_SARS.CoV.2_1"
## [25] "Series5_A549_SARS.CoV.2_2" "Series5_A549_SARS.CoV.2_3"
## [27] "Series8_A549_Mock_1"      "Series8_A549_Mock_2"
## [29] "Series8_A549_Mock_3"      "Series8_A549_RSV_1"
## [31] "Series8_A549_RSV_2"       "Series8_A549_RSV_3"
## [33] "Series9_NHBE_Mock_1"      "Series9_NHBE_Mock_2"
## [35] "Series9_NHBE_Mock_3"      "Series9_NHBE_Mock_4"
## [37] "Series9_NHBE_IAV_1"       "Series9_NHBE_IAV_2"
## [39] "Series9_NHBE_IAV_3"       "Series9_NHBE_IAV_4"
```