

## RESEARCH

# Respuesta a la infección por SARS-CoV2

Carmen Lucía Arrabalí Cañete<sup>\*</sup>  
, Oleg Brezitsky  
, Juan Antonio Herrera Conde  
and Sergio Martín Vera

<sup>\*</sup>Correspondence:  
[carmenarrabali@uma.es](mailto:carmenarrabali@uma.es)  
ETSI Informática, Universidad de  
Málaga, Málaga, España  
Full list of author information is  
available at the end of the article

## Abstract

En el presente proyecto de investigación procederemos a explorar un análisis en profundidad la respuesta transcripcional al SARS-CoV-2, comparándolo con otros virus atacantes del sistema respiratorio. Los modelos celulares y animales de la infección por SARS-CoV-2 revelaron una respuesta en forma de inflamación única y nunca antes analizada, debido a niveles bajos de interferones tipo I y III, yuxtapuestas a quimiocinas elevadas y alta expresión de IL -6. En cuanto a la herramienta para obtener la red del sistema, emplearemos principalmente R y WGCNA, puesto que ya hemos visto a lo largo del curso la cantidad de información relevante que nos pueden proporcionar.

**Keywords:** SARS-CoV-2; COVID-19; Coronavirus

## 1 Introducción

Los coronavirus son un grupo diverso de virus de ARN monocatenario de sentido positivo con una amplia gama de huéspedes vertebrados. Cuatro géneros comunes de coronavirus (alfa, beta, gamma y delta) circulan entre los vertebrados y causan enfermedades leves del tracto respiratorio superior en humanos y gastroenteritis en animales. Sin embargo, en las últimas dos décadas han surgido tres betacoronavirus humanos altamente patógenos a partir de eventos zoonóticos. En 2002-2003, el coronavirus 1 relacionado con el síndrome respiratorio agudo severo (SARS-CoV-1) infectó a  $\approx 8000$  personas en todo el mundo con una tasa de letalidad de  $\approx 10\%$ , seguido por el coronavirus relacionado con el síndrome respiratorio de Oriente Medio (MERS-CoV), que ha infectado a  $\approx 2500$  personas con una tasa de letalidad de  $\approx 36\%$  desde 2012. En la actualidad, el mundo sufre una pandemia de SARS-CoV-2, causante de la enfermedad por coronavirus 2019 (COVID-19) y tiene una tasa de mortalidad global que aún está por determinar.

La infección por SARS-CoV-2 se caracteriza por una variedad de síntomas que incluyen fiebre, tos y malestar general en la mayoría de los casos, pero en los casos más graves, pueden llegar a desarrollar un síndrome de dificultad respiratoria aguda y lesión pulmonar aguda, lo que provoca morbilidad y mortalidad causadas por daños en la luz alveolar que conducen a inflamación y neumonía.

Entendiendo qué respuesta puede tener el cuerpo cuando se infecta por SARS-CoV-2, ahora se estudiará esta respuesta, pero desarrollada en las células del epitelio del pulmón mediante el análisis de perfiles de expresión génica publicados en el dataset GEO GSE147507 [1].

## 2 Materiales y métodos

### 2.1 Carga de datos

Para poder hacer uso de los datos que se anexan al artículo científico en el que está basado el proyecto, se han de descargar desde la página web del NCBI, con un archivo llamado *setup.sh*. Por otro lado, hay un archivo llamado *launch.sh* que se encarga de ejecutar todos los archivos de R. Se obtienen un total de 21797 genes con 79 muestras cada uno con los que se empieza el análisis.

### 2.2 Análisis inicial

Se crea el archivo llamado *Análisis.EG.dataInput.R*, el cual carga los datos y se crea la limpieza y preprocesamiento de los datos. En primer lugar se crea una configuración del entorno de R y se procesan los datos. Se hace uso de la librería de *WCGNA* obtenida desde *Bioconductor*.

Por otro lado se hace un reajuste de datos, se transforman para que sean las filas las correspondientes a los genes y, las columnas, a las muestras de los mismos.

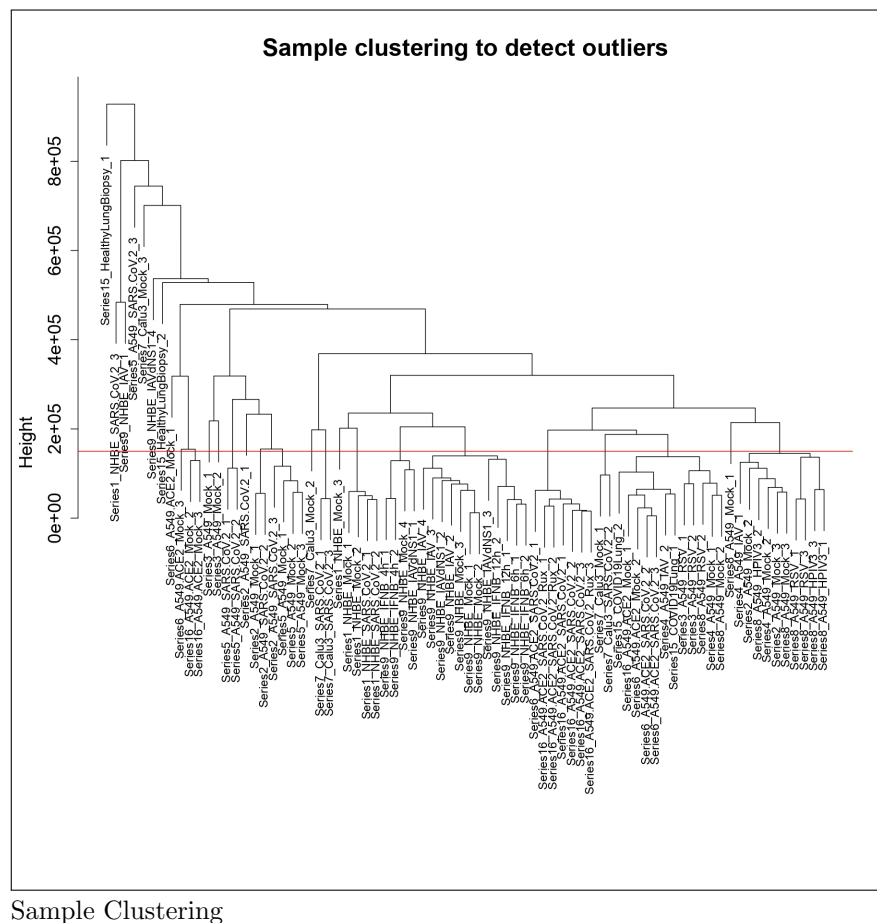
Como se desconoce si hay alguna falta de datos, se comprueba si hay algún dato que falta o que es nulo. En el caso de que si haya datos no válidos, se eliminan los genes y las muestras de estos.

Luego, se agrupan para poder ver si hay o no valores atípicos y luego se eliminan aquellos que si que lo sean con un corte de altura.

Todos los datos mencionados anteriormente se almacenan para poder hacer así el análisis de red.

## 3 Resultados

Una vez instalados todos los paquetes y librerías necesarias, debemos empezar con la limpieza de datos y su preprocesamiento, por lo que cargamos el dataset de perfiles de expresión génica. Transformamos los datos para poder trabajar con filas y columnas, descubriendo así si hay genes con valores perdidos. Si la última comprobación devuelve *TRUE*, no necesitaremos ejecutar el siguiente código. De lo contrario, se eliminan los genes, mostrándose prescindibles. Agrupamos las muestras para ver si hay valores atípicos y una vez detectados, los eliminamos eligiendo un corte de altura.



## 4 Discusión

## 5 Conclusiones

### Abreviaciones

- NCBI: National Center for Biotechnology Information
- WGCNA: Weighted Correlation Network Analysis

### Disponibilidad de datos y materiales

[Proyecto en GitHub.](#)

### Contribución de los autores

C.L.A.C y J.A.H.C.: Encargados de la escritura de la memoria y resultados.

O.B. y S.M.V : Encargados de la escritura del código en R.

### Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

### References

1. Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., Wang, T.T., Schwartz, R.E., Lim, J.K., Albrecht, R.A., tenOever, B.R.: Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**(5), 1036–10459 (2020). doi:[10.1016/j.cell.2020.04.026](https://doi.org/10.1016/j.cell.2020.04.026)