

# **Proyecto final – Genómica computacional: Análisis de las diferencias de ORFs en proteínas del SARS-CoV-2 para determinar el impacto en la capacidad de transmisión del virus**

## **1. Introducción**

El COVID-19, identificado por primera vez en Wuhan, China, en diciembre de 2019, es causado por el virus SARS-CoV-2 y ha llevado a una crisis sanitaria global sin precedentes. Desde su aparición, el virus ha provocado millones de muertes y ha tenido un impacto devastador en las economías y sociedades alrededor del mundo. Investigaciones iniciales y continuas son esenciales para entender la patogénesis, transmisión y variabilidad del virus, así como para desarrollar estrategias efectivas de intervención, incluyendo vacunas y tratamientos antivirales. El estudio de su estructura genética y la evolución de sus cepas es fundamental para anticipar y mitigar sus efectos a largo plazo (World Health Organization, 2020).

## **2. Estructura y Genómica del SARS-CoV-2**

El SARS-CoV-2 pertenece a la familia Coronaviridae y es un virus de ARN de cadena simple positiva. Su genoma, que es uno de los más largos entre los virus de ARN, incluye una compleja organización de ORFs que codifican tanto proteínas estructurales como no estructurales. Las proteínas estructurales incluyen la proteína spike (S), envelope (E), membrana (M), y nucleocapsid (N), mientras que las proteínas no estructurales juegan roles en la replicación y modulación de la respuesta inmune del huésped. El conocimiento detallado de la secuencia genómica del virus ha sido crucial para el desarrollo de diagnósticos y vacunas en respuesta a la pandemia (Zhou et al., 2020).

## **3. Función de los ORFs**

Los ORFs del SARS-CoV-2 codifican al menos 29 proteínas, que son esenciales para diversas funciones virales. Los ORFs más grandes, 1a y 1b, son traducidos en poliproteínas que se escinden en proteínas no estructurales implicadas en la replicación del ARN viral y la inhibición de las respuestas inmunitarias. Estas proteínas no estructurales incluyen, entre otras, la ARN polimerasa dependiente de ARN y proteínas que manipulan la respuesta inmune del huésped. El estudio de estos ORFs y sus productos es vital para entender cómo el virus copia su genoma y evade el sistema inmunitario (Kim et al., 2020).

### **3.1. Diferencias entre Cepas**

Las variaciones en los ORFs entre distintas cepas del SARS-CoV-2 pueden afectar características clave del virus, como su transmisibilidad y virulencia. La emergencia de variantes como la Alfa y Delta ha mostrado diferencias significativas en términos de escape inmune y eficacia de las vacunas. Estas variantes resultan de mutaciones puntuales, deletions, y otras reorganizaciones genéticas que pueden alterar la

estructura y función de las proteínas codificadas. La vigilancia genómica global es crucial para detectar y caracterizar estas variantes a medida que surgen (Volz et al., 2021).

Sin embargo, para detectar diferencias entre proteínas, no podemos analizar los open reading frames pues en el caso de las proteínas el término ORF no se aplica de la misma manera porque las proteínas son ya el producto de la traducción de ORFs del ARN mensajero. Sin embargo, cuando se analizan las secuencias de proteínas, se puede buscar dominios funcionales o motivos que son regiones de la secuencia de aminoácidos que tienen una estructura y función específicas. Esto será lo que intentaremos encontrar, motivos de glicosilación.

### **3.2 Sitios de glicosilación**

La proteína Spike está extensamente glicosilada, lo que significa que hay muchos sitios donde se añaden cadenas de azúcar a la proteína. Estos sitios de glicosilación pueden afectar la estabilidad de la proteína, su reconocimiento por parte del sistema inmunitario del huésped, y su capacidad para unirse al receptor ACE2.

### **4. Técnicas de Secuenciación y Análisis Genómico**

La secuenciación de próxima generación (NGS) ha revolucionado la capacidad de monitorizar y estudiar variantes del SARS-CoV-2. Las técnicas de NGS permiten una rápida caracterización de las cepas virales en circulación y han sido fundamentales en la identificación de mutaciones que podrían conferir ventajas adaptativas al virus. Además, la bioinformática juega un rol esencial en el análisis de los datos de secuenciación, permitiendo comparaciones detalladas entre cepas y la identificación de patrones evolutivos (Lu et al., 2020).

### **5. Proteína Spike (S)**

La proteína Spike del SARS-CoV-2 es el principal antígeno en la mayoría de las vacunas COVID-19 desarrolladas y tiene un rol crucial en la patogenicidad viral, ya que media la entrada del virus en las células huésped. La unión de la proteína Spike al receptor ACE2 en las células humanas es el primer paso en el proceso de infección. Mutaciones en los motivos de Spike pueden aumentar la afinidad por ACE2, potenciando la transmisión o alterar la antigenicidad del virus, afectando la eficacia de las vacunas. Estudiar estas mutaciones es esencial para prever y mitigar su impacto en la salud pública (Hoffmann et al., 2020).

### **6. Proteasa Principal (Mpro)**

La proteasa principal, también conocida como Mpro o 3CLpro, es responsable de la escisión de las poliproteínas virales en proteínas funcionales más pequeñas que son esenciales para la replicación viral. Esta proteasa es un objetivo terapéutico prometedor porque inhibirla podría detener la replicación del virus. Los inhibidores de Mpro están siendo investigados como potenciales tratamientos antivirales para el COVID-19, con varios candidatos en ensayos clínicos (Jin et al., 2020).

## 7. Proteína Nucleocápside (N)

La proteína Nucleocápside del SARS-CoV-2 es crítica para múltiples etapas del ciclo de vida viral, incluyendo el ensamblaje del virión, la replicación del genoma y la modulación de la respuesta inmune celular. Mutaciones en los motivos de esta proteína pueden tener impactos significativos en la eficiencia de la replicación viral y en la capacidad del virus para evadir la respuesta inmunitaria del huésped. Por lo tanto, es un área de interés tanto para el desarrollo de diagnósticos como de terapias (Kang et al., 2020).

## 8. Implicaciones para Vacunas y Terapias

El entendimiento de cómo las variaciones en los motivos afectan las características del SARS-CoV-2 es crucial para el diseño de vacunas y estrategias terapéuticas. Las vacunas deben ser capaces de inducir una respuesta inmunitaria amplia que pueda efectivamente neutralizar múltiples cepas del virus, mientras que los tratamientos antivirales deben ser eficaces contra las enzimas virales esenciales. La evolución continua del virus requiere una adaptación constante de estas intervenciones para mantener su eficacia a lo largo del tiempo (Corey et al., 2020).

## 9. Propuesta de proyecto

Basándonos en la información recopilada anteriormente, la propuesta del proyecto será analizar las diferencias en los motivos de glicosilación de la proteína spike, la proteína Mpro y la proteína nucleocápside en diferentes cepas del SARS-Cov-2. Sin embargo, el proyecto se dividirá en tres fases, cada una corresponde al análisis de una de las proteínas mencionadas anteriormente.

Para la fase que se presentará al final del semestre, lo que se espera es terminar un análisis completo y detallado de la proteína spike en diferentes cepas pues, consideramos, es la que más puede darnos información para concluir o descartar la hipótesis planteada.

## 10. Referencias

- World Health Organization. (2020). Coronavirus disease (COVID-19) pandemic. Recuperado de <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273. <https://doi.org/10.1038/s41586-020-2012-7>
- Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, 181(4), 914-921.e10. <https://doi.org/10.1016/j.cell.2020.04.011>
- Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., ... & Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593(7858), 266-269. <https://doi.org/10.1038/s41586-021-03470-x>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., ... & Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565-574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)

- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., ... & Zhang, B. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811), 289-293. <https://doi.org/10.1038/s41586-020-2223-y>
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., ... & Deng, A. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*, 10(7), 1228-1238. <https://doi.org/10.1016/j.apsb.2020.04.009>
- Corey, L., Mascola, J. R., Fauci, A. S., & Collins, F. S. (2020). A strategic approach to COVID-19 vaccine R&D. *Science*, 368(6494), 948-950. <https://doi.org/10.1126/science.abc5312>
- Morens, D. M., Breman, J. G., & Fauci, A. S. (2020). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996), 242-249. <https://doi.org/10.1038/nature02759>