



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica  
y del Medio Natural

Identificación y estudio de variantes patogénicas en  
humanos que afectan a la gravedad por COVID-19

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Ángel Moreno Serrano, Ana

Tutor/a: Cardona Serrate, Fernando

Cotutor/a externo: Pérez Tur, Jordi

CURSO ACADÉMICO: 2023/2024



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escuela Técnica Superior  
de Ingeniería Agronómica  
y del Medio Natural

Universidad Politécnica de Valencia

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

**Grado en Biotecnología**

Trabajo Final de Grado

# Identificación y estudio de variantes patogénicas en humanos que afectan a la gravedad por COVID-19



Autor: Ana Ángel - Moreno Serrano

Tutor experimental: Jordi Pérez Tur

Tutor académico: Fernando Cardona Serrate

Curso 2023/2024

Valencia, septiembre de 2024



# IDENTIFICACIÓN Y ESTUDIO DE VARIANTES PATOGENICAS EN HUMANOS QUE AFECTAN A LA GRAVEDAD POR COVID-19

## Resumen

La enfermedad de COVID-19 causada por el SARS-CoV-2, desencadenante de la pandemia global en 2020, destaca por su rango de sintomatología variable desde la ausencia de síntomas hasta la insuficiencia respiratoria grave y muerte incluso en individuos previamente sanos. Partiendo de ello, se sugiere que los factores genéticos del hospedador desempeñan un rol importante en la susceptibilidad y gravedad de la infección. En este contexto, se han utilizado estudios de asociación de genoma completo (GWAS) para identificar polimorfismos de nucleótido único (SNPs) que puedan estar asociados con la predisposición y/o severidad de infección por SARS-CoV-2.

Con el propósito de identificar genes cuyas variantes influyan en la gravedad de COVID-19, se realizó una revisión de estudios GWAS y se recopilaron variantes *missense* con predicción negativa y otras variantes potencialmente modificadoras de la pauta de lectura utilizando registros de secuenciación de variantes de pacientes. Posteriormente, se evaluó el impacto de dichas variantes utilizando *SNP-Set Kernel Association Test (SKAT)* y una prueba de Chi Cuadrado.

En base a los análisis estadísticos, se seleccionaron dos variantes en dos genes distintos: rs12720356 en el gen *TYK2* y rs17279437 en el gen *SLC6A20*. Así pues, se comprobó mediante mutagénesis *in silico* que las mutaciones producidas por estas variantes podían causar cambios estructurales significativos y alterar la disposición tridimensional original de sus proteínas. Por consiguiente, utilizando estrategias de transformación en *E. coli* y transfección de líneas celulares con los cDNAs de ambos genes, se diseñó un procedimiento para la obtención de las proteínas silvestres por *Western Blot*, que arrojó resultados incompletos y requiere el ajuste de las concentraciones de proteína. Paralelamente, se trazó un proceso de mutagénesis por PCR para introducir las mutaciones de las variantes, no obteniendo colonias que las incorporasen, lo que sugiere una revisión de los materiales y métodos empleados. Así pues, los estudios futuros permitirán comprobar diferencias en la expresión del genotipo silvestre y mutado y determinar si estas mutaciones influyen en la severidad de la infección por SARS-CoV-2.

**Palabras clave:** severidad; SARS-CoV-2; GWAS; variantes; mutagénesis; cDNA.

# IDENTIFICATION AND STUDY OF PATHOGENIC VARIANTS IN HUMANS THAT AFFECT SEVERITY OF COVID-19

## Abstract

COVID-19 disease caused by SARS-CoV-2, which triggered the global pandemic in 2020, is notable for its variable range of symptoms from the absence of symptoms to severe respiratory failure and death even in previously healthy individuals. Based on this, it has been proposed that genetic factors could play a crucial role in the susceptibility and severity of infection. In this context, genome-wide association studies (GWAS) have been used to identify single nucleotide polymorphisms (SNPs) that may be associated with the predisposition and/or severity of SARS-CoV-2 infection.

In order to identify genes whose variants influence the severity of COVID-19, a review of GWAS studies was performed and missense variants with negative prediction and other potentially modifying variants of the reading pattern were collected using patient variant sequencing records. The impact of these variants was subsequently assessed using the SNP-Set Kernel Association Test (SKAT) and a Chi Square test.

Based on statistical analyses, two variants were selected in two different genes: rs12720356 in the *TYK2* gene and rs17279437 in the *SLC6A20* gene. Thus, it was verified by *in silico* mutagenesis that the mutations produced by these variants could cause significant structural changes and alter the original three-dimensional arrangement of their proteins. Therefore, using strategies of transformation in *E. coli* and transfection of cell lines with the cDNAs of both genes, a procedure was designed to obtain the wild proteins by Western Blot, which yielded incomplete results and requires adjustment of protein concentrations. In parallel, a PCR mutagenesis process was designed to introduce the mutations of the variants, without obtaining colonies that incorporated them, which suggests a review of the materials and methods used. Future studies will therefore allow us to verify differences in the expression of the wild and mutated genotypes and determine whether these mutations influence the severity of SARS-CoV-2 infection.

**Key words:** severity; SARS-CoV-2; GWAS; variants; mutagenesis; cDNA.

*A mi familia, especialmente a Francisco y Cristina.  
Os guardo en mi corazón.*

## Agradecimientos

A mis padres, mi hermano y mi gran familia, por brindarme esta oportunidad y acompañarme hasta en la distancia. Os llevo conmigo allá donde vaya.

A Alex, por motivarme a seguir día a día y no dejar que me rinda.

A todos mis amigos y compañeros. Gracias por vivir esta experiencia conmigo.



# ÍNDICE DE CONTENIDOS

<b>Anexo I.</b>	Relación del trabajo con los Objetivos de Desarrollo Sostenible de la Agenda 2030.....	X
<b>1.</b>	<b>INTRODUCCIÓN</b> .....	1
1.1	El SARS-CoV-2 y la COVID-19.....	1
1.2	Genética del hospedador en COVID-19.....	2
1.3	El gen <i>SLC6A20</i> y su relación con el SARS-CoV-2.....	2
1.4	La quinasa <i>TYK2</i> y su variante rs12720356.....	3
<b>2.</b>	<b>OBJETIVOS</b> .....	4
2.1	Hipótesis.....	4
2.2	Objetivos.....	4
<b>3.</b>	<b>MATERIALES Y MÉTODOS</b> .....	4
3.1	Selección de variantes de genes candidatos.....	4
3.2	Análisis estadísticos de las variantes con predicción negativa.....	5
3.3	Análisis estructurales <i>in silico</i> de las proteínas.....	6
3.4	Transformación del cDNA, transfección en HEK293 y mutagénesis.....	7
3.4.1	Clones.....	7
3.4.2	Transformación en cepas de <i>E. coli DH5a</i> competentes. Purificación y cuantificación del DNA bacteriano.....	7
3.4.3	Cultivo de las líneas de HEK293.....	8
3.4.4	Transfección en líneas de HEK293.....	8
3.4.5	Extracción y cuantificación de proteínas.....	8
3.4.6	SDS-PAGE y Western Blot.....	9
3.4.7	Mutagénesis.....	10
<b>4.</b>	<b>RESULTADOS Y DISCUSIÓN</b> .....	11
4.1	Genes y regiones candidatas.....	11
4.2	Recopilación de variantes con potencial patogénico.....	12
4.3	Test estadístico SKAT-O.....	14
4.4	Prueba de chi-cuadrado.....	15
4.5	Análisis estructurales <i>in silico</i> del efecto de las variantes.....	16
4.5.1	Variante rs17279437 de <i>SLC6A20</i> .....	17
4.5.2	Variante rs12720356 de <i>TYK2</i> .....	20
4.6	Estudios <i>in vitro</i> en células HEK293.....	22
4.6.1	Preparación y mutagénesis de los cDNAs.....	22
4.6.2	Estudio de expresión de los cDNAs.....	23
<b>5.</b>	<b>CONCLUSIONES</b> .....	24

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Detalle de los anticuerpos primarios utilizados.....	9
<b>Tabla 2.</b> Reactivos y cantidades para la PCR de mutagénesis.....	10
<b>Tabla 3.</b> Cebadores diseñados para la introducción de las mutaciones p.I684S ( <i>TYK2</i> ) y p.T199M ( <i>SLC6A20</i> ) .....	10
<b>Tabla 4.</b> Genes individuales potenciales de estudio, breve descripción de su función y bibliografía asociada.....	12
<b>Tabla 5.</b> Compilación de genes de estudios <i>GWAS</i> como potenciales candidatos para su estudio.....	12
<b>Tabla 6.</b> Clasificación de variantes <i>missense</i> ordenadas según predicción de REVEL.....	13
<b>Tabla 7.</b> Clasificación de variantes potencialmente patogénicas según predicción de impacto genómico.....	14
<b>Tabla 8.</b> Conjunto de variantes para las que se realizó la Prueba de chi-cuadrado, o en su defecto, el Test exacto de Fisher.....	15
<b>Tabla 9.</b> Prueba de chi-cuadrado para las variantes de la Tabla 8.....	16
<b>Tabla 10.</b> Selección final de variantes para ensayos en cultivos celulares.....	16

## ÍNDICE DE FIGURAS

<b>Fig. 1.</b> Diagrama de la partícula del virión del SARS-CoV-2.....	1
<b>Fig. 2.</b> Estructura general del complejo <i>SIT1-ACE2-RBD</i> .....	3
<b>Fig. 3.</b> Estructura del dominio de una proteína quinasa típica mostrando el sitio de unión de ATP y los elementos conservados a su alrededor.....	3
<b>Fig. 4.</b> Hoja de ruta para la recolección de variantes candidatas a estudio.....	5
<b>Fig. 5.</b> Código empleado para la realización del SKAT en el entorno de R mediante RStudio.....	6
<b>Fig. 6.</b> Mapa físico de los plásmidos pCMV3 con el cDNA de <i>TYK2</i> y <i>SLC6A20</i> en la <i>ORF</i> .....	7
<b>Fig. 7.</b> Cultivo de células HEK293 en crecimiento para transfección.....	8
<b>Fig. 8.</b> Relación de la zona del genoma humano que engloba las variantes génicas encontradas y el número de estudios recopilados sobre su efecto en COVID-19.....	11
<b>Fig. 9.</b> Resultados obtenidos tras realizar el test estadístico de SKAT-O.....	15
<b>Fig. 10.</b> Resultados de VADAR para <i>SLC6A20</i> .....	17
<b>Fig. 11.</b> Resultados de predicción de fosforilación por Netphos para la Treonina en posición 199....	18
<b>Fig. 12.</b> Mapa de Crio-microscopía electrónica general y con vista en forma de <i>sticks</i> del complejo <i>ACE2-SIT1</i> .....	18
<b>Fig. 13.</b> Estructura del dominio transmembrana de <i>ACE2</i> y <i>SIT1</i> .....	19
<b>Fig. 14.</b> Detalle del dominio transmembrana de <i>ACE2</i> y <i>SIT1</i> en PyMOL.....	19
<b>Fig. 15.</b> Diferencias en PyMOL en la disposición de los puentes de hidrógeno en <i>SIT1</i> .....	20
<b>Fig. 16.</b> Resultados de las predicciones de fosforilación en Netphos para la serina en posición 684 en la variante p.I684S de <i>TYK2</i> .....	21
<b>Fig. 17.</b> Visualización en PyMOL de los dominios quinasa (JH1) y pseudoquinasa (JH2) de <i>TYK2</i> .....	21
<b>Fig. 18.</b> Disposición alterada de la hoja beta al comparar la estructuras silvestre con la mutada y fosforilada en la posición 684 en PyMOL.....	22
<b>Fig. 19.</b> Membrana de Western Blot tras tinción con Rojo Ponceau.....	23
<b>Fig. 20.</b> Imágenes capturadas a los 5 minutos del revelado y a los 15 minutos con Image Reader LAS-3000.....	23

## Anexo I. Relación del trabajo con los Objetivos de Desarrollo Sostenible de la Agenda 2030

A. Indicar el grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS)

	Alto	Medio	Bajo	No procede
ODS 1. Fin de la pobreza				X
ODS 2. Hambre cero				X
ODS 3. Salud y bienestar	X			
ODS 4. Educación de calidad				X
ODS 5. Igualdad de género				X
ODS 6. Agua limpia y saneamiento				X
ODS 7. Energía asequible y no contaminante				X
ODS 8. Trabajo decente y crecimiento económico				X
ODS 9. Industria, innovación e infraestructuras				X
ODS 10. Reducción de las desigualdades				X
ODS 11. Ciudades y comunidades sostenibles				X
ODS 12. Producción y consumo responsables				X
ODS 13. Acción por el clima				X
ODS 14. Vida submarina				X
ODS 15. Vida de ecosistemas terrestres				X
ODS 16. Paz, justicia e instituciones sólidas				X
ODS 17. Alianzas para lograr objetivos.				X

B. Describir brevemente la alineación del TFG con los ODS, marcados en la tabla anterior, con un grado alto

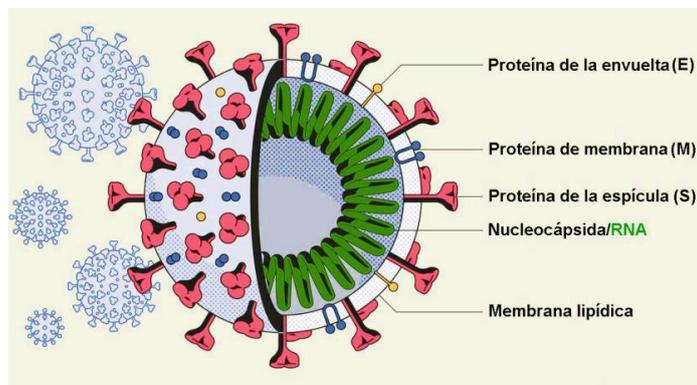
El presente trabajo se ajusta al Objetivo de Desarrollo Sostenible número 3: “Garantizar una vida sana y promover el bienestar para todos en todas las edades”. Las investigaciones sobre variantes que pueden afectar a la gravedad de la infección por SARS-CoV-2 posibilitan conocer las variantes genéticas más peligrosas, especialmente en poblaciones de alto riesgo. Este conocimiento facilita el desarrollo de pruebas de detección, así como la implementación de un plan de actuación que, desde estadios tempranos, prevenga la aparición de insuficiencia respiratoria grave y la saturación de las UCI de los hospitales. Asimismo, estos estudios permiten avanzar hacia la síntesis de fármacos que palíen o prevengan el desarrollo de COVID-19 grave en personas con variantes patogénicas.

# 1. INTRODUCCIÓN

## 1.1 EL SARS-CoV-2 y la COVID-19

En diciembre de 2019, se reportaron los primeros casos de la enfermedad producida por el SARS-CoV-2 (COVID-19), un coronavirus de tipo 2 causante de síndrome respiratorio agudo severo (Guan et al., 2020). Dicha enfermedad desembocó en una pandemia global con más de 670 millones de casos confirmados y 6,8 millones de muertes relacionadas (*Coronavirus Resource Center of Johns Hopkins University*, octubre de 2023). Los coronavirus (*Fig. 1*) son virus de RNA monocatenario de polaridad positiva envueltos por cápsida, pertenecientes a la familia *Coronaviridae* y el orden *Nidovirales*, estando ampliamente distribuidos en humanos y mamíferos (Huang et al., 2020). Estos virus cuentan aproximadamente con una docena de marcos de lectura abierta (*ORF*), encontrándose seis de ellos muy conservados, concretamente los que codifican los enzimas de la maquinaria de replicación y las proteínas estructurales (N: nucleocápsida, E: envuelta, M: membrana, S: *spike*).

El proceso de penetración celular del SARS-CoV-2 tiene lugar mediante endocitosis y fusión con la pared celular, mediando la unión de la proteína S a la enzima convertidora de angiotensina 2 (*ACE2*), y posteriormente, la escisión en subunidades S1 y S2 por proteasas como *TMPRSS2/4* y furinas. El virus es reconocido por receptores *Toll-like* y sensores citoplasmáticos, que inducen la síntesis de Interferón  $\alpha$  y  $\beta$  que se unen a su receptor (*IFNAR*) y activan vías de tirosina quinasa a través de la quinasa *TYK2* y el traductor de señal y activador de la transcripción *JAK-STAT*. Estas vías promueven la expresión de genes como *OAS1*, *OAS2* y *OAS3*, codificadores de 2'-5'-oligoadenilato sintetetas que activan la respuesta antiviral celular innata (Gamero de Luna y Sánchez Jaén, 2024).



**Fig. 1.** Diagrama de la partícula del virión del SARS-CoV-2. Se muestran las distintas proteínas estructurales (envuelta, membrana, espícula y nucleocápsida) y la membrana lipídica. Imagen de Manuel Bortoletti, adaptada por la Sociedad Española de Virología

En lo que respecta a las manifestaciones clínicas de la COVID-19, estas varían desde la ausencia de síntomas hasta la insuficiencia respiratoria grave y la muerte. La COVID-19 grave requiere cuidados intensivos con soporte respiratorio y puede causar daños a largo plazo. La variedad de manifestaciones clínicas de la infección por SARS-CoV-2 no está determinada únicamente por la variabilidad de carga viral, dado que se ha relacionado con factores de riesgo como la edad avanzada, el sexo masculino y comorbilidades como diabetes, hipertensión o enfermedad coronaria (Degenhardt et al., 2022). No obstante, estos factores de riesgo por separado no explican toda la variabilidad en la gravedad de la enfermedad observada entre distintos individuos (*COVID-19 Host Genetics Initiative*, 2021). Por este motivo, las manifestaciones graves de la infección, incluidas las muertes en individuos

sanos, sugieren que los factores genéticos pueden tener un papel clave. Dada la relevancia de la genética en el pronóstico y prevención de la enfermedad, el campo de la genómica en COVID-19 ha crecido de manera rápida, publicándose gran cantidad de estudios en los últimos cuatro años. La información sobre variantes de riesgo generales o específicas de cada población es crucial para el desarrollo de futuras pruebas de detección, que permitirían identificar a individuos de alto riesgo para la COVID-19 (Kovanda et al., 2024).

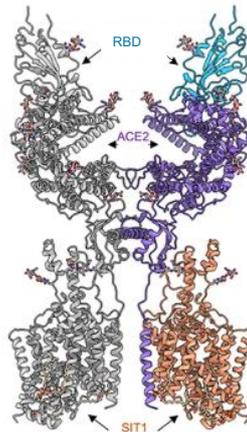
## 1.2 Genética del hospedador en COVID-19

Con el objetivo de comparar y recopilar datos mundiales de COVID-19, en el año 2020 se creó la COVID-19 *Host Genetics Initiative*, el mayor consorcio genético de esta enfermedad, el cual tiene el propósito de mapear de manera robusta los determinantes genéticos de susceptibilidad, severidad y consecuencias de la infección por SARS-CoV-2 (COVID-19 *Host Genetics Initiative*, 2021). Actualmente, se llevan a cabo diversos tipos de estudios como los Estudios de asociación de Genoma Completo (GWAS), que se emplean para predecir qué SNPs (Polimorfismos de Nucleótido Único) pueden estar asociados en susceptibilidad y gravedad de distintas enfermedades como la COVID-19 (Eshetie et al., 2023).

El primer estudio GWAS realizado en la investigación de esta enfermedad (*The Severe Covid-29 GWAS Group*, 2020) identificó dos loci genéticos (3p21.31 y 9q34.2) asociados con la susceptibilidad a la COVID-19. El estudio publicado por la COVID-19 *Host Genetics Initiative* en 2021 confirmó estos resultados y encontraron asociaciones adicionales, sugiriendo que ciertas mutaciones genéticas aumentan el riesgo de COVID-19 severo o grave. Concretamente, el *locus* multigénico 3p21.31 del cromosoma 3 en humanos destaca por mostrar significancia estadística elevada en diferentes estudios de GWAS que buscan asociación de distintos SNPs con el riesgo de gravedad de infección por SARS-CoV-2 (*The Severe Covid-19 GWAS Group*, 2020; Degenhardt et al., 2022). En estos estudios, la señal de asociación cubre un *cluster* de seis genes diferentes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* y *XCR1*) que codifican para distintos receptores de quimiocinas y los cuales se estudian de forma individual o conjunta para encontrar variantes que puedan afectar a la gravedad de la enfermedad.

## 1.3 El gen *SLC6A20* y su relación con el SARS-CoV-2

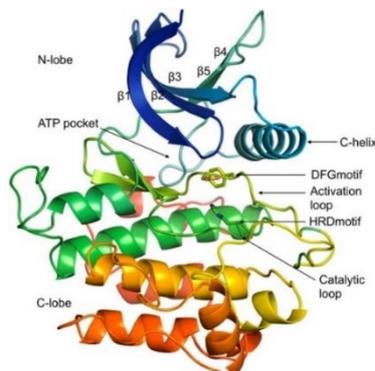
La superfamilia génica SLC o *Solute Carrier* del genoma humano está compuesta por transportadores de un gran número de sustratos a través de membranas en diversos procesos fisiológicos. Estos genes codifican proteínas *carriers* que transportan aminoácidos para regular el mantenimiento de la homeostasis intracelular. Dentro de la familia SLC6 se encuentra el gen *SLC6A20*, que codifica el transportador de sodio-imino ácido 1 (*SIT1*), conocido también como *IMINO* o *XTRP3*. Este transportador se encarga de la absorción de aminoácidos en riñón e intestino (Xia et al., 2024). *SIT1* heterodimeriza e interactúa funcionalmente con la enzima convertidora de la angiotensina 2 (*ACE2*) en la membrana plasmática, y esta enzima actúa como receptor celular crítico del SARS-CoV-2 uniéndose directamente al dominio de unión al receptor (RBD) de la glicoproteína de la espícula (S) (Fig. 2) (Shen et al., 2022). Estudios recientes sugieren que la infección por SARS-CoV-2 podría regular la actividad de transporte en el complejo *ACE2-SIT1* (Li et al., 2023). Además, se ha propuesto que *SIT1* es un regulador de los niveles de glicina y que este aminoácido tiene efectos beneficiosos contra la secreción de citocinas proinflamatorias inducida por la infección de SARS-CoV-2 (Semiz, 2021).



**Fig. 2.** Modificada de Shen et al., 2022, *Cell discovery*, 8(1), 123. <https://doi.org/10.1038/s41421-022-00488-x>. Estructura general del complejo SIT1-ACE2-RBD. Uno de los dos protómeros se marca en color: azul para RBD, violeta para ACE2 y salmón en el caso de SIT1.

#### 1.4 La quinasa *TYK2* y su variante rs12720356

En lo que respecta al cromosoma 19, en el *locus* 19p13.2 del mismo se encuentra el gen que codifica para la proteína *TYK2*, una tirosina quinasa citoplasmática de la familia Janus (JAK), que se expresa constitutivamente en células autoinmunes. Esta tirosina quinasa participa en la cascada de señalización por citocinas, involucradas en respuesta inmune e inflamación, como el Interferón (IFN) de tipo I y III, las citocinas proinflamatorias IL-6, -12, -22, -23, -26 y anti-inflamatoria IL-10 (Enerbäck et al., 2018). Las JAK se caracterizan por poseer un dominio quinasa dual, compuesto por un dominio quinasa JH1 precedido por uno pseudoquinasa (JH2). Gran parte de las mutaciones que se asocian a enfermedades en las JAK quinastas se encuentran en el dominio JH2 (Min et al., 2015). La variante no sinónima rs12720356, que consiste en un cambio aminoacídico de isoleucina por serina en la posición 684 (p.I684S) en el exón 15, podría influir en el dominio catalítico basándose en su localización en la hoja  $\beta$  dentro de lóbulo N (Fig. 3) del dominio pseudoquinasa de *TYK2* (Li et al., 2013). Esta variante ha sido asociada en algunos estudios con susceptibilidad a enfermedad autoinmune, ya que el cambio a serina, un aminoácido pequeño y polar, podría afectar al lóbulo N y comprometer al posicionamiento de la hélice  $\alpha$ C. Este deterioro catalítico es funcionalmente significativo en desordenes inmunes complejos (Enerbäck et al., 2018), haciendo de esta variante concreta una candidata prometedora para su estudio.



**Fig. 3.** De Modi et al., 2019, *Proceedings of the National Academy of Sciences of the United States of America* vol. 116(14): 6818-6827. <https://doi.org/10.1073/pnas.1814279116>. Estructura del dominio de una proteína quinasa típica mostrando el sitio de unión de ATP y los elementos conservados a su alrededor, destacando el lóbulo N, donde se encuentra una estructura en lámina  $\beta$  así como la hélice  $\alpha$ C; y el lóbulo C, incluyendo el loop catalítico, bolsillo de ATP, loop de activación, motivos DFG y HDR. La mutación p.I684S se localiza dentro de la lámina  $\beta$  del lóbulo N. (*INSR* quinasa, PDB 1GAG).

En resumen, los factores genéticos del hospedador en COVID-19 parecen desempeñar un papel relevante en la variabilidad en sintomatología y gravedad de la enfermedad. Por lo tanto, el estudio de variantes génicas seleccionadas y relacionadas con la entrada del SARS-CoV-2 en las células (en el caso de las variantes de SLC6A20) o el desencadenamiento de la respuesta inmune (en las vinculadas a TYK2), conforma una estrategia viable para detectar variantes patogénicas asociadas a la gravedad de infección por COVID-19.

## 2. OBJETIVOS

### 2.1 Hipótesis

Este trabajo parte de la hipótesis de que la gravedad con la que un individuo infectado por el SARS-CoV-2 manifiesta la COVID-19 viene mediada, en parte, por variantes genéticas específicas del hospedador.

### 2.2 Objetivos

- Identificar variantes genéticas relacionadas con la gravedad de la COVID-19 en función de su significatividad estadística en estudios de GWAS en relación de la gravedad de la enfermedad.
- Priorizar las variantes utilizando predicciones bioinformáticas.
- Estudiar *in silico* el efecto estructural de las variantes identificadas.
- Iniciar el análisis *in vitro* del efecto de las variantes de interés.

## 3. MATERIALES Y MÉTODOS

### 3.1 Selección de variantes de genes candidatos

Primeramente, se realizó una búsqueda exhaustiva de bibliografía de GWAS sobre COVID-19 con la finalidad de recopilar artículos científicos que identificaran distintos *locus* dentro del genoma humano con relación con la gravedad de la infección por SARS-CoV-2. En estos estudios, se seleccionaron genes con variantes asociadas que destacasen por su significancia estadística elevada en GWAS.

Tras establecer un compendio de genes potenciales para este estudio, se amplió la búsqueda a las variantes de estos genes y sus colindantes en el genoma dentro del archivo AllGenes\_AllVariants\_220713.xlsx, el cual contiene datos de todas las variantes observadas en exomas de pacientes y recopilados en experimentos de COVID-19 anteriores de este laboratorio. Este archivo incluye la alteración génica que causa cada variante detectada en cada gen, así como su posición, consecuencia, cambio aminoacídico y posición, predicción de patogenicidad (por REVEL, SIFT y PolyPhen), así como el intervalo de frecuencia del alelo menor y las muestras de pacientes control y afectadas (clasificadas como heterocigotas u homocigotas) para la variante.

Dentro de todos los tipos de variantes génicas registradas en el archivo, este estudio se centró en las variantes *missense* (que provocan cambio de un aminoácido) y variantes que desembocasen en modificaciones en la pauta de lectura (*frameshift*), puesto que son las que más fácilmente pueden indentificar un cambio que afecte negativamente a la actividad de la proteína original.

En lo que respecta a la recopilación de variantes *missense*, la lista se ordenó de mayor a menor predicción patogénica según la herramienta REVEL, siguiendo la guía de predicciones de patogenicidad en variantes de la web de Ensembl ([https://www.ensembl.org/info/genome/variation/prediction/protein\\_function.html](https://www.ensembl.org/info/genome/variation/prediction/protein_function.html)). Como este predictor integra scores de otros como SIFT y PolyPhen, sus pronósticos también se tuvieron en cuenta de forma individual en este estudio.

Con relación a la agrupación de variantes que cambiasen la pauta de lectura, se tuvieron en cuenta las que producían cambios en la región de *splicing*, pérdida o generación de codón de stop, desplazamiento de la pauta de lectura, variantes de miRNA y degradación del RNA mensajero mediada por mutación terminadora (NMD, *Nonsense-mediated decay*). Este grupo se ordenó según la predicción de consecuencias de variantes en la web de Ensembl según su “impacto genómico”, considerando la severidad de las consecuencias de los términos definidos por *Sequence Ontology* (SO) junto con otras herramientas de anotación de variantes ([https://www.ensembl.org/info/genome/variation/prediction/predicted\\_data.html](https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html)).

Después de reunir ambas clasificaciones (reflejadas en el apartado de resultados y discusión), se hizo una búsqueda de herramientas que permitiesen hacer un *Gene Enrichment* y análisis estadísticos para destacar un número reducido de variantes potenciales de estudio (Fig. 4).

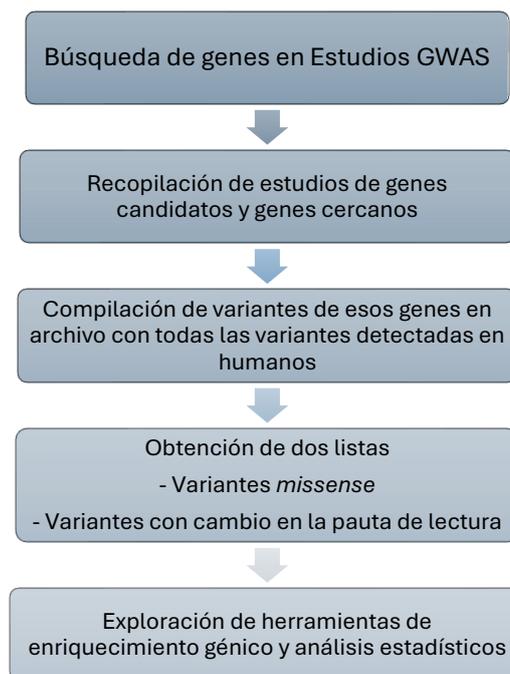


Fig. 4. Hoja de ruta para la recolección de variantes candidatas a estudio.

### 3.2 Análisis estadísticos de las variantes con predicción negativa

Con el objetivo de evaluar el impacto de las variantes recogidas, se realizó un estudio de enriquecimiento génico *in silico* utilizando el SNP-Set Kernel Association Test (SKAT) (<http://cran.nexr.com/web/packages/SKAT>) en el entorno estadístico R, con el fin de buscar asociación entre el conjunto de SNPs y los fenotipos de gravedad de COVID-19. Este análisis comparó distintos fenotipos, tales como pacientes hospitalizados en unidades de cuidados intensivos (“UCI”), hospitalizados fuera de UCI (“NoUCI”) o no hospitalizados (“NoHosp”) para la identificación de

diferencias significativas entre los mismos. Esta clasificación fenotípica se obtuvo del archivo Pacientes.xlsx, que establecía la relación de cada fenotipo con cada muestra de pacientes del archivo AllGenes\_AllVariants\_220713.xlsx. Se generaron los distintos ficheros input que incluían las variantes con la clasificación de la muestra según si eran homocigotos o heterocigotos para ese *SNP*, junto a su fenotipo correspondiente. Los tres grupos se compararon dos a dos, añadiendo un análisis que agrupara todos los pacientes hospitalizados (incluyendo “UCI” y “NoUCI”) y los distinguiera de los no hospitalizados (“NoHosp”). Por último, se generó un código en R para efectuar el análisis (Fig. 5).

```

1 install.packages("SKAT")
2 library(SKAT)
3
4 Y <- "C:\\Users\\Propietario\\Desktop\\Y.txt"
5 Z <- "C:\\Users\\Propietario\\Desktop\\Z.txt"
6
7 Y <- read.table(Y, header = FALSE)
8 Z <- read.table(Z, header = FALSE)
9
10 Z <- as.matrix(Z)
11 Y <- Y[,-1]
12
13 lista <- list(
14   Z = Z,
15   Y = Y
16 )
17
18 obj <- SKAT_Null_Model(Y ~ 1, out_type = "D", data = lista)
19
20 resultado_SKAT <- SKAT(Z, obj)
21
22 SKAT(Z, obj, method="SKATO")$p.value
23
22:37 (Top Level)
R Script

```

Console Terminal Background Jobs

R 4.3.3 · C:/Users/Propietario/Desktop/ ·  
Platform: x86\_64-w64-mingw32/x64 (64-bit)

**Fig. 5.** Código empleado para la realización del SKAT en el entorno de R mediante RStudio. El objeto Y se refiere a un vector numérico compuesto por la lista de individuos/muestras asociadas a su fenotipo. Z está vinculado a una matriz numérica en la que cada fila representa un individuo y cada columna representa un *SNP* diferente.

Para finalizar esta parte y poder proceder al análisis estructural, se decidió seleccionar dos genes con un número importante de variantes asociadas y elevada patogenicidad predicha para las mismas, que se recopilaron del archivo AllGenes\_AllVariants\_220713.xlsx, observándose una disparidad en el número de individuos para las distintas categorías fenotípicas asociadas en Pacientes.xlsx (UCI, NoUCI y NoHosp). Esta desigualdad consistía en un número muy elevado de muestras de pacientes NoUCI respecto al número de muestras dentro de los otros dos grupos. Por consiguiente, se procedió a hacer la prueba de chi-cuadrado para las variantes de interés. En el caso de las variantes con menos de 5 pacientes para alguno de los tres grupos, se estableció el p-valor mediante la realización del Test exacto de Fisher.

### 3.3 Análisis estructurales *in silico* de las proteínas

Tras la selección de las variantes a estudiar, se inició un procedimiento de análisis estructural *in silico* del posible efecto de las mutaciones aminoacídicas en las dos proteínas de interés.

En primer lugar, se buscaron las estructuras de *TYK2* y *SLC6A20* (este último unido a *ACE2* y el RBD de la S) en la base de datos de Protein Data Bank (<https://www.rcsb.org/>) tomando los archivos formato pdb 3ZON y 7y75 respectivamente. Después se introdujeron las mutaciones p.I684S (rs17279437) en 3ZON y p.T199M (rs12720356) en 7y75, utilizando la herramienta para mutagénesis del programa PyMOL. A continuación, se minimizó la energía de las estructuras mutadas con el servidor YASARA <https://www.yasara.org/minimizationserver.htm> (Krieger et al., 2009).

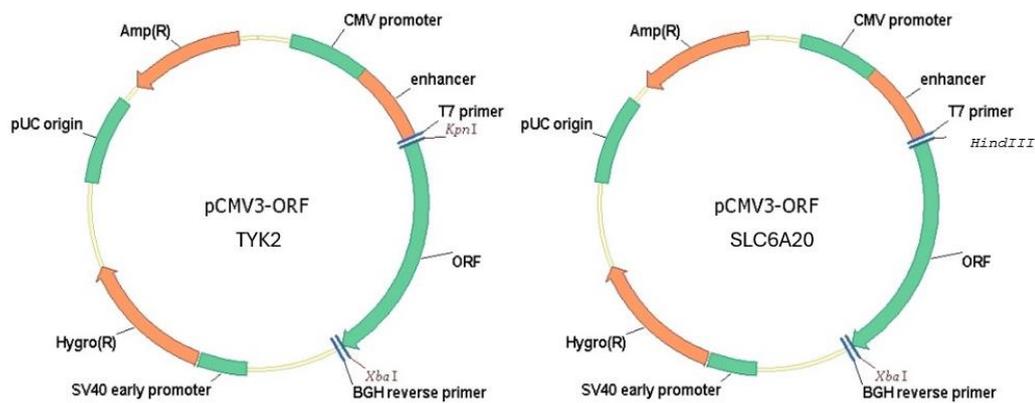
Con la herramienta VADAR 1.8 (<http://vadar.wishartlab.com/>) de Willard et al. (2003) se compararon los gráficos de Ramachandran del genotipo silvestre y mutado en el caso de *SLC6A20*, y con NetPhos 3.1 (<https://services.healthtech.dtu.dk/services/NetPhos-3.1/>) se predijo la fosforilación del residuo mutado (serina) de *TYK2* y el silvestre (treonina) de *SLC6A20*.

Por último, se fosforiló el genotipo silvestre de *SLC6A20* y mutado para *TYK2* mediante el plugin PyTMS de PyMOL para comparar diferencias en disposición de los aminoácidos en sus cadenas y la posible repercusión que pudieran tener en el bolsillo catalítico y otras zonas claves dentro de la proteína mediante PyMOL.

### 3.4 Transformación del cDNA, transfección en HEK293 y mutagénesis

#### 3.4.1 Clones

Se adquirió cDNA de *TYK2* y *SLC6A20* de la empresa Sino Biological clonado en plásmidos con la estructura que se indica en la Fig. 6. En el caso de *TYK2* la ORF contenía 3564 pb y en *SLC6A20*, 1779 pb incluyendo la secuencia del gen y la resistencia a los antibióticos correspondientes, así como otras secuencias implicadas en la replicación del plásmido y la expresión del cDNA. Se transformaron bacterias con estos plásmidos para obtener los plásmidos en cantidad para hacer mutagénesis y transfectar células en cultivo.



**Fig. 6.** Mapa físico de los plásmidos pCMV3 con el cDNA de *TYK2* (izquierda) y *SLC6A20* (derecha) en la ORF. El plásmido de *TYK2* está flanqueado por *KpnI* y *XbaI* y *SLC6A20* por *HindIII* y *XbaI*. Ambos contienen promotor de citomegalovirus (CMV) y los genes de resistencia a ampicilina y a higromicina.

#### 3.4.2 Transformación en cepas de *E. coli* DH5α competentes. Purificación y cuantificación del DNA bacteriano.

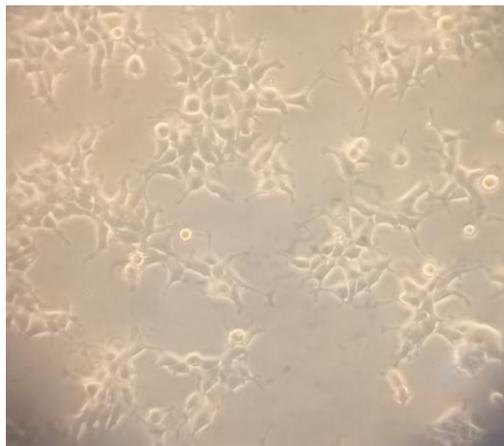
En primer lugar, se diluyeron 1/10 con agua mili-Q los plásmidos para transformar en bacterias *E. coli* cepa DH5α. Para ello, se añadieron 0,5 µl de cada uno en un tubo con 40 µL de DH5α electrocompetentes previamente preparados en este laboratorio y conservados a -20°C, y se transfirió la mezcla a una cubeta de electroporación. Ambas muestras se electroporaron a 1700 voltios. A continuación, se añadieron 120 µl de LB y se mantuvieron en agitación 30 minutos a 37°C para que las bacterias expresaran la resistencia al antibiótico. Transcurrida la media hora, se sembraron en placas de LB agar con Ampicilina a 50 µg/ml.

Tras 24 horas, se observó que habían crecido colonias para ambos genes y se tomó una de cada gen para inocular en tubos con 2 mL de LB-Ampicilina, que tras 6 horas de incubación se añadieron a dos matraces con 50 mL de LB Ampicilina, conservándose en agitación a 37 °C *overnight*. El cultivo bacteriano crecido al día siguiente se centrifugó a 4000 rpm 30 minutos a 4°C y se conservó el pellet de bacterias para la posterior purificación.

El DNA bacteriano se purificó siguiendo el protocolo de ZymoPURE™ del Kit *Plasmid Midiprep* de Zymo Research. Una vez finalizado el proceso de purificación de ambas muestras, se utilizó el espectrofotómetro NanoDrop 2000 para medir la calidad y cantidad de DNA obtenido.

#### 3.4.3 Cultivo de las líneas de HEK293

Con el objetivo de expresar las proteínas de interés mediante transfección temporal con los plásmidos obtenidos, se cultivó la línea celular HEK293 (*Fig. 7*), una línea epitelial inmortalizada a partir de células de riñón embrionario humano que se ha demostrado que expresan ACE2 (Hamming et al., 2004). Para ello se utilizó medio DMEM 1X + GutaMAX™ con glucosa y piruvato sódico, al que se añadieron 100 U/ml de antibiótico Penicilina-Estreptomicina, 10% de FBS (suero bovino fetal) y 1% de L- Glutamina. Se utilizaron frascos T25 y T75 para mantenimiento de los cultivos, los cuales se tripsinizaron con Tripsina-EDTA 0,05 % (Gibco) para pasar a nuevos frascos cuando la confluencia era demasiado elevada o se requería pasar las células al recipiente apropiado para efectuar la transfección. Finalmente, se cultivaron 2 placas P60 para transfectar con los cDNAs de *TYK2* y *SLC6A20*.



*Fig. 7. Cultivo de células HEK293 en crecimiento para transfección.*

#### 3.4.4 Transfección en líneas de HEK293

Para el proceso de transfección, se preparó una mezcla de 1,5 µg de DNA con el doble de cantidad de reactivo X-treme (X-treme GENE HP DNA Transfection Reagent de Roche) y el resto de Medio de suero reducido Opti-MEM (Gibco) hasta los 500 µL, siguiendo las recomendaciones del fabricante. La mezcla se dispensó gota a gota en 1 placa P60 por cada uno de los dos genes, con una confluencia celular del 80%. A continuación, las placas se incubaron a 37°C y 5% de CO<sub>2</sub> durante 48 horas.

#### 3.4.5 Extracción y cuantificación de proteínas

Transcurridos dos días, se llevó a cabo la extracción de las proteínas. Para ello, se retiró el medio de cultivo y se lavaron las placas con solución salina PBS 1X. Se añadieron 0,5 mL de tampón RIPA con 20 µl de inhibidores de proteasas y se mantuvieron en frío durante 2 minutos. Posteriormente, se

rasparon las células utilizando un *cell-scraper*. El lisado celular recogido se incubó en frío en baño de ultrasonidos 30 minutos y a continuación, se centrifugó 10 minutos a 4°C a 13000xg. Por último, se recuperó el sobrenadante y se midió la cantidad de proteínas con el NanoDrop 2000.

#### 3.4.6 SDS-PAGE y Western Blot

Dado que en futuros ensayos de expresión podría ser necesaria la proteína ACE2 para comprobar su interacción con SIT1 (*SLC6A20*), se adquirieron anticuerpos primarios de las casas ThermoFisher y Proteintech (Tabla 1) para poder detectar las tres proteínas (*TYK2*, *SLC6A20* y *ACE2*) mediante Western Blot.

**Tabla 1.** Detalle de los anticuerpos primarios utilizados

Anticuerpo	Marca comercial
<i>SLC6A20</i> Policlonal de conejo PA5-68332	ThermoFisher
<i>TYK2</i> Monoclonal de ratón 67411-1-Ig	Proteintech
<i>ACE2</i> Monoclonal de ratón 66699-1-Ig	Proteintech

Para la polimerización del gel de electroforesis SDS- PAGE (*Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis*) previo al Western Blot, se preparó acrilamida al 8% partiendo de una solución de acrilamida al 40% mezclada 29:1 con bisacrilamida de ITW Reagents. Se utilizó esta concentración debido a que el rango de proteínas a detectar era de tamaño considerable, desde los 66 kDa en el caso de *SLC6A20*, pasando por los 92 kDa de *ACE2* y hasta los 133 kDa de *TYK2*. Seguidamente, se montó el gel y se prepararon las muestras incorporando tampón de carga Laemmli 5X (250 mM Tris-HCl pH 6.8, 40% glicerol, 4% SDS, 0.005% azul de bromofenol disueltos en ddH<sub>2</sub>O) con β-mercaptoetanol en una concentración de 40 μL por cada mililitro de Laemmli. Se preparó un total de 25 μL para *TYK2* y *SLC6A20* y 35 μL para *ACE2* con el tampón de carga y agua mili-Q, de manera que la concentración de proteína en el pocillo fuera de 1 μg/μL. Se incorporó la misma cantidad de tampón que de marcador de peso molecular *Protein Marker V (pre-stained) peqGOLD* de VWR Chemicals, diluido con agua mili-Q hasta llegar a 20 μL por pocillo de marcador. Posteriormente, se calentaron las muestras a 100°C durante 5 minutos, y se cargaron en el gel junto con el marcador de pesos moleculares. Después, se incorporó el *Running Buffer* en una concentración 1X (250 mM Tris-HCl pH 8, 1920 mM Glicina y 1% SDS disueltos en ddH<sub>2</sub>O), y se aplicó un voltaje constante de 100V para el *Stacking gel* y 120V para el *Resolving gel*.

La transferencia a la membrana de nitrocelulosa se realizó utilizando el *iBlot 3 Western Blot Transfer System* de ThermoFisher. La carga de proteínas y la correcta transferencia a la membrana se comprobó utilizando la tinción con Rojo Ponceau durante 2 minutos en agitación y lavada con TBST 1X en agitación. Luego, la membrana se mantuvo 1 hora en agitación con TBS Tween 1X 5% leche. A continuación, se recortó la membrana en tres partes para incubar cada una de ellas en una dilución 1:1000 del anticuerpo primario correspondiente en TBS-Tween 20 (TBST; 150 mM NaCl, 20 mM Tris-HCl pH 8, 0.1% Tween 20 disueltos en ddH<sub>2</sub>O) con 5% leche en agitación a 4°C o/n. Al día siguiente, se lavaron los 3 fragmentos de membrana 3 veces con TBST en agitación durante 5 minutos. Después, las membranas se incubaron en una dilución 1/5000 de anticuerpo secundario anti-ratón (*ACE2* y *TYK2*) o anti-conejo (*SLC6A20*) durante 1 hora. Posteriormente, se volvió a lavar por triplicado con TBST, se reveló la membrana en un *Hypercassette* de Amersham con una solución de Peróxido y Luminol a partes iguales utilizando el kit *ECL Select™ Western Blotting Detection Reagent* de Cytiva. La membrana impregnada en solución de detección se mantuvo en oscuridad 5 minutos y seguidamente, se capturaron imágenes con el software de Image Reader LAS-3000 de Fujifilm en intervalos de 30 segundos.

### 3.4.7 Mutagénesis

Con el fin de introducir las variantes de interés en los cDNA de *TYK2* y *SLC6A20*, se realizó un proceso de mutagénesis basado en la reacción en cadena de la polimerasa (PCR) y el método del círculo rodante, utilizando los reactivos detallados en la *Tabla 2*.

**Tabla 2.** Reactivos y cantidades para la PCR de mutagénesis.

Componente	Cantidad
Cebador Forward 10 $\mu$ M	1,5 $\mu$ L
Cebador Reverse 10 $\mu$ M	1,5 $\mu$ L
DNA molde (50 ng/ $\mu$ L)	1 $\mu$ L
dNTPs 2,5 mM	1,5 $\mu$ L
Enzima Pfu Turbo (2.5 U/ $\mu$ L)	1 $\mu$ L
Tampón Pfu 10X	5 $\mu$ L
H <sub>2</sub> O Mili-Q	38,5 $\mu$ L
<b>Total</b>	<b>50 <math>\mu</math>L</b>

La secuencia de los cebadores para la mutagénesis (*Tabla 3*) se diseñaron para generar los cambios de nucleótido específicos en el cDNA utilizando el software GeneRunner y fueron sintetizados por Integrated DNA Technologies.

**Tabla 3.** Cebadores diseñados para la introducción de las mutaciones p.I684S (*TYK2*) y p.T199M (*SLC6A20*).

Cebador	Secuencia
<i>TYK2</i> rs12720356 Forward	5' GCGCGGCCCTGAAAATAGCATGGTGACAGAGTACGTGG 3'
<i>TYK2</i> rs12720356 Reverse	5' CCA CGT ACTCTGTCACCATGCTATTTTCAGGGCCGCGC 3'
<i>SLC6A20</i> rs17279437 Forward	5' GCAAGGTGGTGTATTTTCATGGCGTCACTGCCCTATTGC 3'
<i>SLC6A20</i> rs17279437 Reverse	5' GCAATAGGGCAGTGACGCCATGAAATACACCACCTTGC 3'

En inicio, se programó una PCR con las siguientes condiciones:

1 ciclo inicial de 3 minutos a 95 °C;

35 ciclos de:

30 segundos a 95 °C, 30 segundos a 55 °C y 8 minutos a 72 °C para desnaturalización, alineamiento y extensión respectivamente;

1 ciclo de elongación de 10 minutos a 72 °C.

Debido a que no se obtuvieron colonias tras dos intentos con las mismas condiciones, se realizó una PCR con las siguientes características:

1 ciclo de 2 minutos a 95°C;

50 ciclos de 30 segundos a 95°C, 8 minutos a 68°C para *SLC6A20* y 10 minutos a 68°C para *TYK2*;

1 ciclo de 10 minutos a 68°C.

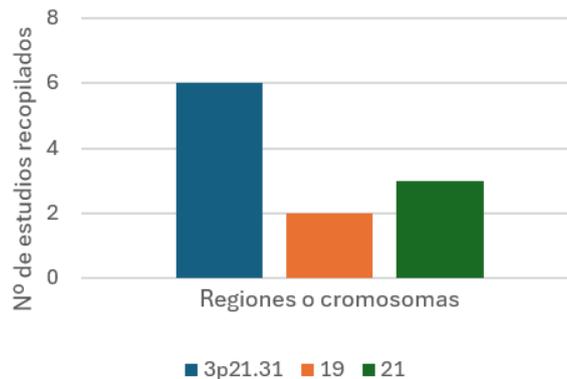
Cuando la PCR hubo finalizado, se realizó la digestión del DNA incorporando 1  $\mu$ L de enzima DpnI (ThermoFisher Scientific) que digiere únicamente DNA metilado, para eliminar el DNA molde y obtener únicamente el DNA mutado, manteniéndose la mezcla durante 1 hora a 37°C. Tras el proceso de digestión, se purificó el DNA con GENE CLEAN™ Turbo Kit de MP Biomedicals.

Nuevamente, se llevó a cabo el proceso de transfección del cDNA por electroporación (detallado en el apartado 3.4.2, “Transformación en cepas de *E. coli* DH5 $\alpha$  competentes”) para sembrar las bacterias en placas de LB agar con Ampicilina a 50  $\mu$ g/ml, incubándose a 37°C *o/n*. Al día siguiente, las colonias crecidas en las placas se inocularon en tubos con LB y ampicilina a 100mg/mL y se mantuvieron en agitación 37 °C *o/n*.

## 4 RESULTADOS Y DISCUSIÓN

### 4.1 Genes y regiones candidatas

En lo que respecta a los trabajos científicos seleccionados para el desarrollo de este trabajo, se recopilaron estudios y meta-análisis de GWAS que encontraron genes en distintas regiones del genoma humano cuyas variantes estuvieran asociadas con la gravedad de COVID-19. En definitiva, fueron acotadas 3 regiones dentro del genoma de *Homo sapiens* para proceder a un estudio con mayor profundidad, incluyendo el cromosoma 3 *locus* 3p21.31, y los cromosomas 19 y 21 (Fig. 8).



**Fig. 8.** Relación de la zona del genoma humano que engloba las variantes génicas encontradas y el número de estudios recopilados sobre su efecto en COVID-19

En el *locus* 3p21.31 resaltan los genes *SLC6A20*, *LZTFL1* y *CXCR6*, con funciones relacionadas con la COVID-19 (Tabla 4). Estos tres genes se mencionan en el trabajo de *The Severe Covid-19 GWAS Group* (2020), el cual acota este *locus* como de susceptibilidad genética en pacientes con fallo respiratorio por COVID-19. De igual forma, Gamero de Luna y Sánchez Jaen (2024) encontraron que *TYK2*, en el cromosoma 19, que codifica para una tirosina quinasa, contribuye a un mayor riesgo a la infección grave por COVID-19. Por otra parte, en el cromosoma 21, el gen *TMPRSS2* codifica una proteasa clave para la entrada viral (Pairo-Castineira et al., 2023); e *IFNAR2*, un receptor de Interferón  $\alpha$  y  $\beta$ , está significativamente asociado a COVID-19 severo en individuos jóvenes (Degenhardt et al., 2022).

**Tabla 4.** Genes individuales potenciales de estudio, breve descripción de su función y bibliografía asociada.

Cromosoma	Gen/locus	Función principal de la proteína que codifica	Bibliografía
3	3p21.31		Severe Covid-19 GWAS Group (2020) Breno et al. (2023)
	<b>SLC6A20</b>	Transportador IMINO de prolina dependiente de sodio	Eshetie et al. (2023)
	<b>LZTFL1</b>	Factor de transcripción de cremallera de leucina	Downes et al. (2021) Degenhardt et al. (2022)
	<b>CXCR6</b>	Receptor de quimiocina con motivo C-X-C	Smieszek et al. (2022)
19	<b>TYK2</b>	Tirosina quinasa	Degenhardt et al. (2022) Pairo-Castineira et al. (2023)
21	<b>IFNAR2</b>	Receptor de interferón $\alpha$ y $\beta$	Degenhardt et al. (2022) Jalkanen et al. (2023)
	<b>TMPRSS2</b>	Serina proteasa transmembrana	Ravikanth et al. (2021)

Por consiguiente y tras la recopilación de genes colindantes a los genes reflejados en la *Tabla 4*, se obtuvo la siguiente lista de genes potenciales para el estudio (*Tabla 5*). Estos genes se registraron por cercanía en los diferentes *plots* de *GWAS*, con la finalidad de no pasar por alto ningún gen que, a pesar de no aparecer reflejado en la bibliografía estudiada, pudiese contener alguna variante patogénica.

**Tabla 5.** Compilación de genes de estudios *GWAS* como potenciales candidatos para su estudio

Cromosoma	Genes	Bibliografía
3	SLC6A20, CCR5, LARS2, FYCO1, CCR3, CCR1, LZTFL1, CXCR6, XCR1, SACM1L, LIMD1	Breno et al. (2023) Degenhardt et al. (2022) Severe Covid-19 GWAS Group (2020)
19	CCR2, ABCA7, TYK2, P2PRY11, PPAN, PPAN-P2RY11, TYK2, AP1M2, SLC44A2, S1PR2, MRPL4, ICAM1, ICAM3, DNMT1, KRI1, ANGPTL6, PDE4A, S1PR5, ILF3, CDC37, RAVER1, FDX1L, MIR199A1, C19orf38, CTD-2369P2.12	Degenhardt et al. (2022) Pairo-Castineira et al. (2023)
21	KCNE2, MX1, GART, OLIG1, IFNAR2, AP000295.9, MX2, IFNGR2, TMPRSS2, SON, C21orf62, TMEM50B, DNAJC28, AP000304.12, IL10RB, FAM3B, DONSON, BACE2, SMIM11, IFNAR1	Pairo-Castineira et al. (2023)

#### 4.2 Recopilación de variantes con potencial patogénico

Una vez determinados los genes a estudiar se procedió a la compilación de variantes de interés, que se localizaron introduciendo los genes mencionados dentro del archivo de datos de secuenciación de exomas de pacientes del laboratorio, buscando que la consecuencia producida por dichas variantes fuera un cambio en la pauta de lectura o un cambio aminoacídico. Cabe destacar que, tras recoger todas esas variantes, se observó que todas se encontraban en un rango de frecuencia del alelo menor entre 0.001 y 0.05, por lo que pueden considerarse variantes raras (de baja frecuencia).

En primer lugar, se estableció la lista de variantes *missense* (Tabla 6), ordenando las variantes de mayor a menor predicción negativa de REVEL. Asimismo, se conservaron algunas variantes que, pese a no estar predichas como patogénicas, tenían un número de heterocigotos para la variante mayor a 10 individuos, o un número de homocigotos mayor a 2 (en rojo).

Se registraron en total 45 variantes y en la Tabla 6 se indica el gen, cromosoma, posición, cambio de aminoácido, nombres establecidos para la variante, valores de predictores de patogenicidad y número de heterocigotos y homocigotos que poseen la variante. Se consideraron “probables causantes de la enfermedad” aquellas variantes mayores a un valor de 0.5 en REVEL (en rojo) y “probablemente benignas” las inferiores a este valor (en negro). También, junto a este orden se tuvo en cuenta la clasificación de SIFT en cuanto a variantes “deletéreas” (*deleterious*, con valor de SIFT menor de 0.05, en rojo) sobre “toleradas” (*tolerated*, mayor o igual a 0.05, en negro), y la de PolyPhen primando las “probablemente dañinas” (marcadas como *probab\_damag*, valor de PolyPhen mayor a 0.908, en rojo) frente a las “posiblemente dañinas” (*possib\_damag*, entre 0.446 y 0.908, en marrón) o “benignas” (*benign*, menor o igual a 0.446, en negro).

Dentro de esta lista destacan variantes como rs17279437 de *SLC6A20*, por ser detectada en 40 individuos heterocigotos y ser “deletérea” y “probablemente dañina” según los predictores.

**Tabla 6.** Clasificación de variantes *missense* ordenadas según predicción de REVEL. Se indica Cromosoma (“Crom”), Gen, Posición, Cambio de aminoácido, Nombre(s) de la variante, datos de predictores SIFT, PolyPhen y REVEL junto a número de Heterocigotos y Homocigotos con la alteración.

Crom	Gen	Posición	Cambio de aminoácido	Nombre(s) de la variante	SIFT	PolyPhen	REVEL	Heterocigotos con la alteración	Homocigotos con la alteración
21	KCNE2	57/123	I/T	rs74315448&CM993509&COSV51710065	deleterious(0.04)	probab_damag(0.999)	0.863	1	0
19	ABCA7	1629/2146	S/L	rs184590335&COSV104548149	deleterious(0)	probab_damag(0.992)	0.823	5	0
19	TYK2	928/1187	A/V	rs35018800&CM153199&COSV53386305	deleterious(0.03)	probab_damag(1)	0.817	4	0
3	SLC6A20	199/592	T/M	rs17279437&CM086964&COSV62072026	deleterious(0)	probab_damag(1)	0.725	40	0
3	SLC6A20	164/592	G/R	rs745541068&COSV62072514	deleterious(0)	probab_damag(1)	0.691	1	0
21	MX1	151/662	A/T	rs140022520&COSV99063909	deleterious(0.04)	probab_damag(0.999)	0.645	1	0
21	GART	93/1010	C/F	rs145820555&COSV63113823	deleterious(0)	probab_damag(1)	0.610	1	0
19	P2RY11	261/374	Y/C	rs75945356&CM176688&COSV99461162	deleterious(0)	probab_damag(1)	0.603	2	0
19	PPAN-P2RY11	681/794	Y/C	rs75945356&CM176688&COSV99461162	delet_low_conf(0)	probab_damag(1)	0.603	2	0
19	PPAN	681/794	Y/C	rs75945356&CM176688&COSV99461162	delet_low_conf(0)	probab_damag(1)	0.603	2	0
19	TYK2	1104/1187	P/A	rs34536443&CM070300&COSV53387399	deleterious(0)	probab_damag(1)	0.586	15	1
19	ABCA7	1599/2146	W/M	rs117187003&CM140122	deleterious(0.02)	probab_damag(0.987)	0.576	3	0
19	ANGPTL6	96/470	L/P	rs559282550&COSV104582528	deleterious(0)	probab_damag(1)	0.573	7	0
21	KCNE2	8/123	T/A	rs22349168&CM003449&COSV51710582	deleterious(0)	probab_damag(0.999)	0.567	5	0
21	GART	93/1010	C/S	rs138272749&COSV100738149	deleterious(0.01)	probab_damag(0.987)	0.553	1	0
19	KEAP1	369/624	V/A	rs149721624&COSV50290796	tolerated(0.46)	benign(0.127)	0.544	1	0
3	CCR5	73/352	A/V	rs56198941&CM995341	deleterious(0)	probab_damag(0.971)	0.537	1	0
19	ABCA7	758/2146	Q/H	rs76282929	deleterious(0)	probab_damag(0.989)	0.510	2	0
19	TYK2	53/1187	A/T	rs55762744&CM1210260&COSV53385459	deleterious(0)	probab_damag(0.997)	0.458	10	0
3	CCR5	55/352	L/Q	rs17998638&CM995340&COSV99068718	deleterious(0)	probab_damag(1)	0.432	18	2
19	TYK2	684/1187	I/S	rs12720356&CM050352&COSV53384675	deleterious(0.04)	probab_damag(0.998)	0.343	36	2
19	ABCA7	718/2146	N/T	rs37522398&COSV54029608	deleterious(0.01)	possib_damag(0.529)	0.291	15	2
21	IFNGR2	58/337	T/R	rs49869588&COSV51639478	deleterious(0.04)	possib_damag(0.875)	0.245	21	0
19	ABCA7	463/2146	R/H	rs37522338&COSV54030179	deleterious(0.03)	probab_damag(0.997)	0.243	16	2
19	ANGPTL6	358/470	R/C	rs147149731	tolerated(0.07)	benign(0.173)	0.238	11	0
19	PDE4A	331/886	P/S	rs150660796	tolerated(0.1)	benign(0.227)	0.208	25	0
19	ABCA7	877/2146	A/T	rs74176364	deleterious(0.04)	benign(0.389)	0.198	42	4
19	S1PR5	318/398	L/Q	rs35483143	deleterious(0)	probab_damag(0.995)	0.165	20	2
19	ABCA7	215/2146	G/S	rs72973581&CM1610867	tolerated(0.65)	benign(0.029)	0.164	25	1
19	ABCA7	1686/2146	Q/R	rs4147918&COSV54029338	tolerated(0.44)	benign(0)	0.152	15	2
19	SLC44A2	153/706	L/F	rs147820753&CM111321&COSV100213425	deleterious(0.04)	benign(0.092)	0.134	12	0
19	ABCA7	319/2146	T/A	rs37522328&COSV54030168	tolerated(1)	benign(0)	0.133	21	2
3	LZTFL1	246/299	D/N	rs1129183&COSV56111224	tolerated(0.1)	possib_damag(0.939)	0.126	43	0
3	CXCR6	3/342	E/K	rs2234355	tolerated(0.15)	possib_damag(0.682)	0.112	9	2
21	AP000304.12	47/191	K/R	rs4842&COSV51708983	tolerated(0.25)	benign(0)	0.107	58	3
21	FAM3B	182/235	M/I	rs111988437&COSV63612943	tolerated(0.51)	benign(0)	0.073	17	1
19	KRI1	327/709	W/M	rs45518133&COSV57264865	tolerated(0.09)	probab_damag(0.994)	0.064	14	0
21	FAM3B	14/235	W/M	rs2838012&COSV63614213	deleterious(0.01)	benign(0.023)	0.062	46	2
19	KRI1	185/709	E/A	rs11545166	tolerated(0.06)	benign(0.094)	0.060	21	0
19	ICAM1	352/532	P/L	rs1801714&CM088314	tolerated(0.08)	probab_damag(0.986)	0.057	29	0
3	LARS2	324/903	H/Q	rs71645922&CM051944	tolerated(0.59)	benign(0.072)	0.041	15	0
19	ICAM3	525/547	S/T	rs2230399&COSV50328770	tolerated(0.13)	possib_damag(0.675)	0.041	51	2
19	FDX1L	26/183	R/G	rs62640397&COSV5346401	tolerated(0.06)	benign(0)	0.028	13	0
21	SM111	51/58	K/R	rs34016792&COSV67892274	tolerated(0.34)	benign(0)	0.028	20	1
19	TYK2	963/1187	G/S	rs2304255&CM083204&COSV53384815	tolerated(0.88)	benign(0.001)	0.027	55	0

\*Se marcan en rojo las predicciones negativas (*deleterious*, *probab\_damag*, *REVEL*>0.5) y en marrón las posiblemente negativas (*possib\_damag*). Los datos de la columna “Heterocigotos con la alteración” se marcan en rojo cuando la alteración se registra en más de 10 individuos, y los de “Homocigotos con la alteración” se marcan en rojo cuando se anotan 2 o más.

Posteriormente se buscaron variantes que pudiesen provocar cambios en la pauta de lectura, obteniendo 12 variantes que se ordenaron según la predicción de grado de impacto genómico (Tabla 7). Este impacto es “Alto” si se asume que la variante puede tener un alto impacto disruptivo en la proteína como truncamiento de esta, pérdida de función o desencadenamiento de NMD; “Bajo”, si se trata de una variante inocua o de impacto improbable para la proteína y “Modificador”, si se trata de una variante o gen no codificante donde las predicciones son complicadas o no hay evidencias de impacto. A diferencia de la clasificación anterior, se incluyó la consecuencia producida por el cambio. Cabe destacar la variante de desplazamiento de marco de lectura rs333 del gen CCR5 registrada en más de 30 individuos, con una predicción de impacto genómico alto.

**Tabla 7.** Clasificación de variantes potencialmente patogénicas según predicción de impacto genómico. Se indica Cromosoma (“Crom”), Gen, Consecuencia del cambio, Posición, Cambio, Nombre(s) de la variante, número de Heterocigotos y Homocigotos con la alteración (“Het alter” y “Hom alter”, respectivamente), así como el impacto genómico.

Crom	Gen	Consecuencia	Posición	Cambio	Nombre(s) de la variante	Het alter	Hom alter	Impacto genoma
21	FAM3B	Ganancia de codón de stop	125/235	R/*	rs139331735	1	0	Alto
19	ABCA7	Variante de desplazamiento del marco de lectura	577-591/2146	LRDTMRAMGLS RAVL/X	rs142076058	1	0	Alto
3	CCR5	Variante de desplazamiento del marco de lectura	184-194/352	YSQYQFWKNF Q/X	rs333	31	2	Alto
21	MX2	Variante <i>missense</i> y de la región de splicing	83/715	M/T	rs56680307& COSV58099509	6	1	Bajo
19	KRI1	Variante <i>missense</i> y de la región de splicing	37/709	R/Q	rs200924770	5	0	Bajo
3	CCR1	Variante de retención de stop	356/355	*	rs41413045	3	0	Bajo
19	CTD-2369P2.12	Variante intrónica y del transcrito NMD	.	.	rs79442975& COSV53348119	13	0	Modificador
19	CTD-2369P2.12	Variante intrónica y del transcrito NMD	.	.	rs62640397& COSV53346401	13	0	Modificador
19	CTD-2369P2.12	Variante intrónica y del transcrito NMD	.	.	rs150936648	3	0	Modificador
19	CTD-2369P2.12	Variante intrónica y del transcrito NMD	.	.	rs538550251& COSV53344918	1	0	Modificador
19	CCR2	Variante sinónima y del transcrito NMD	83/157	I	rs370172423& COSV99532541	3	0	Modificador
19	MIR199A1	Variante de miRNA maduro	.	.	rs183664554	1	0	Modificador

\* Los datos de la columna “Heterocigotos con la alteración” se marcan en rojo cuando la alteración se registra en más de 10 individuos, y los de “Homocigotos con la alteración” se marcan en rojo cuando se anotan 2 o más.

#### 4.3 Test estadístico SKAT-O

Tras la realización de la prueba estadística de SKAT-O sobre las distintas variantes y fenotipos de gravedad de COVID-19, en definitiva, no hubo resultados concluyentes, ya que el p-value obtenido era mayor al  $\alpha = 0,05$  (establecido un intervalo de confianza del 95%) en todos los casos tras realizar el ajuste óptimo (Fig. 9). Luego de analizar los resultados, se observó que el análisis omitió algunos SNPs por altas tasas de pérdidas o no variación de las muestras, por lo que se decidió buscar otra aproximación distinta para determinar las variantes potenciales de estudio.

UCI + NoUCI vs NoHosp  
 p-value = 0.03778464  
 optimal adjust/SKATO = 0.07168325

UCI vs NoUCI  
 (8 SNPs with either high missing rates or no-variation are excluded!)  
 p-value = 0.6042817  
 optimal adjust/SKATO = 0.781515

UCI vs NoHosp  
 (11 SNPs with either high missing rates or no-variation are excluded!)  
 p-value = 0.5065007  
 optimal adjust/SKATO = 0.4691807

NoUCI vs NoHosp  
 p-value = 0.03167871  
 optimal adjust/SKATO = 0.0652541

**Fig. 9.** Resultados obtenidos tras realizar el test estadístico de SKAT-O. Se indica el p valor recogido junto con el ajuste óptimo posterior.

#### 4.4 Prueba de chi-cuadrado

Siguiendo con la metodología establecida, el estudio se enfocó en los dos genes con más variantes y peor predicción de las mismas, resultando en *TYK2* y *SLC6A20*. Al observar sus variantes con detenimiento, se vio que algunas de ellas se relacionaban a numerosos individuos distribuidos aparentemente de forma irregular en las distintas categorías fenotípicas (UCI, NoUCI y NoHosp), por lo que, para determinar la presencia de algún sesgo en la distribución de los grupos, se efectuó la prueba de chi-cuadrado y Test exacto de Fisher en esas variantes en concreto. Las variantes intrónicas o sinónimas, a pesar de que no producían cambio aminoacídico, también se registraban en numerosos individuos y se distribuían de manera dispar entre grupos, por lo que se mantendrían como control en el estudio estadístico (Tabla 8).

**Tabla 8.** Conjunto de variantes para las que se realizó la Prueba de chi-cuadrado, o en su defecto, el Test exacto de Fisher. Se indican las consecuencias que ocasiona cada variante (marcadas en rojo las variantes que causan pérdida de sentido), así como la Posición y el Aminoácido que produce. Se reflejan también los valores de los predictores SIFT, PolyPhen y REVEL y se marcan en rojo en las variantes predichas como deletéreas (deleterious), probablemente dañinas (probab\_damag) o con valor de REVEL sobre 0.5.

Gen	SLC6A20		TYK2					
Consecuencia	Variante <i>missense</i>	Variante sinónima	Variante <i>missense</i>	Variante <i>missense</i>	Variante sinónima	Variante <i>missense</i>	Variante sinónima	Variante intrónica
Posición	199/592	94/592	1104/1187	684/1187	651/1187	363/1187	172/1187	.
Aminoácido	T/M	V	P/A	I/S	I	G/S	T	.
Variante	rs17279437	rs2742399	rs34536443	rs12720356	rs12720355	rs2304255	rs280523	rs12720248
SIFT	<i>deleterious</i> (0)	.	<i>deleterious</i> (0)	<i>deleterious</i> (0.04)	.	tolerated(0.88)	.	.
PolyPhen	<i>probab_</i> <i>damag</i> (1)	.	<i>probab_</i> <i>damag</i> (1)	<i>probab_</i> <i>damag</i> (0.998)	.	benign(0.001)	.	.
REVEL	0.725	.	0.586	0.343	.	0.027	.	.

Tras la realización de la Prueba de chi-cuadrado sobre las variantes de la *Tabla 8*, se observó que para 3 de las 4 variantes missense se rechaza la hipótesis nula que afirma que no hay diferencias entre las proporciones de los grupos para más de tres de las comparaciones fenotípicas establecidas: UCI/No UCI/No Hospitalizados, Hospitalizados/No Hospitalizados, UCI/No Hospitalizados y No UCI/ No Hospitalizados (*Tabla 9*). Considerando los resultados obtenidos, se decidió proseguir los estudios con dos de esas variantes, debido a que reunían los requisitos de predicción patogénica elevada y diferencias de distribución entre fenotipos de gravedad de COVID-19: rs17279437, dentro del gen *SLC6A20*, en el cromosoma 3 de *Homo sapiens* y rs12720356 de *TYK2*, localizada en el cromosoma humano número 19 (*Tabla 10*).

**Tabla 9.** Prueba de chi-cuadrado para las variantes de la *Tabla 8*. Las filas de Distribución Fenotípica clasifican el número de individuos con/sin la variante, indicando el Total de Portadores de la variante (Heterocigotos y Homocigotos) y de Homocigotos de referencia. Se indican las distintas comparaciones fenotípicas de los grupos Hospitalizados, No Hospitalizados, UCI y No UCI con los grados de libertad correspondientes y el valor límite a partir del cual se rechaza la hipótesis nula. En negrita se marcan los resultados de la Prueba de chi-cuadrado (aquellos con el resultado de chi cuadrado y su p valor correspondiente) y del Test exacto de Fisher (aquellos solo con p valor) cuyos valores son significativos.

Gen	SLC6A20		TYK2					
	rs17279437	rs2742399	rs34536443	rs12720356	rs12720355	rs2304255	rs280523	rs12720248
Distribución fenotip. con variante	6 UCI, 9 NoUCI, 12 NoHosp	0 UCI, 10 NoUCI, 1 NoHosp	3 UCI, 8 NoUCI, 3 NoHosp	3UCI, 16 NoUCI, 15 NoHosp	1 UCI, 5 NoUCI, 4 NoHosp	4 UCI, 28 NoUCI, 18 NoHosp	5 UCI, 16 NoUCI, 8 NoHosp	4 UCI, 12 NoUCI, 7NoHosp
Total portadores variante (Heterocigotos/Homocigotos)	27 Het.	11 Het.	13 Het. 1 Hom.	32 Het. 2 Hom.	10Het.	50 Het.	28 Het. 1 Hom.	23 Het.
Distribución fenotípica sin variante	54 UCI, 180 NoUCI, 68 NoHosp	60 UCI, 179 NoUCI, 79 NoHosp	55 UCI, 177 NoUCI, 76 NoHosp	57 UCI, 173 NoUCI, 65 NoHosp	55 UCI, 178 NoUCI, 67 NoHosp	56 UCI, 161 NoUCI, 62NoHosp	55 UCI, 173 NoUCI, 71 NoHosp	41UCI, 117 NoUCI, 61 NoHosp
Total homocigotos de referencia	302	318	308	295	300	279	299	219
UCI/NoUCI/NoHosp 2g.l. (5.99)	<b>X<sup>2</sup>=8.13</b> (p= <b>0.01713</b> )	p=0.1057	p=0.867	<b>X<sup>2</sup>=9.95</b> (p= <b>0.02027</b> )	p=0.3492	<b>X<sup>2</sup>=6.71</b> (p= <b>0.03571</b> )	p=0.8962	p=0.9566
UCI/NoUCI 1g.l. (3.84)	X <sup>2</sup> =2.20 (p=0.1374)	<b>p=0</b>	p=0.7268	p=0.577	p=1	p=0.1227	p=1	p=1
Hosp/NoHosp 1g.l. (3.84)	<b>X<sup>2</sup>=6.47</b> (p= <b>0.01094</b> )	p=0.3071	p=1	<b>X<sup>2</sup>=8.07</b> (p= <b>0.004478</b> )	p=0.245	<b>X<sup>2</sup>=4.37</b> (p= <b>0.0365</b> )	X <sup>2</sup> =0.213 (p=0.6442)	X <sup>2</sup> =0.0686 (p=0.7934)
UCI/NoHosp 1g.l (3.84)	X <sup>2</sup> =0.765 (p= 0.3818)	p=1	p=0.6978	<b>p=0.02</b>	p=0.3828	<b>p=0.017</b>	p=0.7771	p=1
NoUCI/NoHosp 1g.l. (3.84)	<b>X<sup>2</sup>=8.18</b> (p= <b>0.004222</b> )	p=0.1825	p=1	<b>X<sup>2</sup>=5.83</b> (p= <b>0.01575</b> )	p=0.2711	X <sup>2</sup> =2.34 (p=0.1259)	X <sup>2</sup> =0.1885 (p=0.6642)	X <sup>2</sup> =0.05 (p=0.8226)

**Tabla 10.** Selección final de variantes para ensayos en cultivos celulares. Se indica Variante, Gen, Posición en la secuencia de DNA, Base nitrogenada de referencia, Alteración de la base, Posición en la proteína y Cambio de aminoácido.

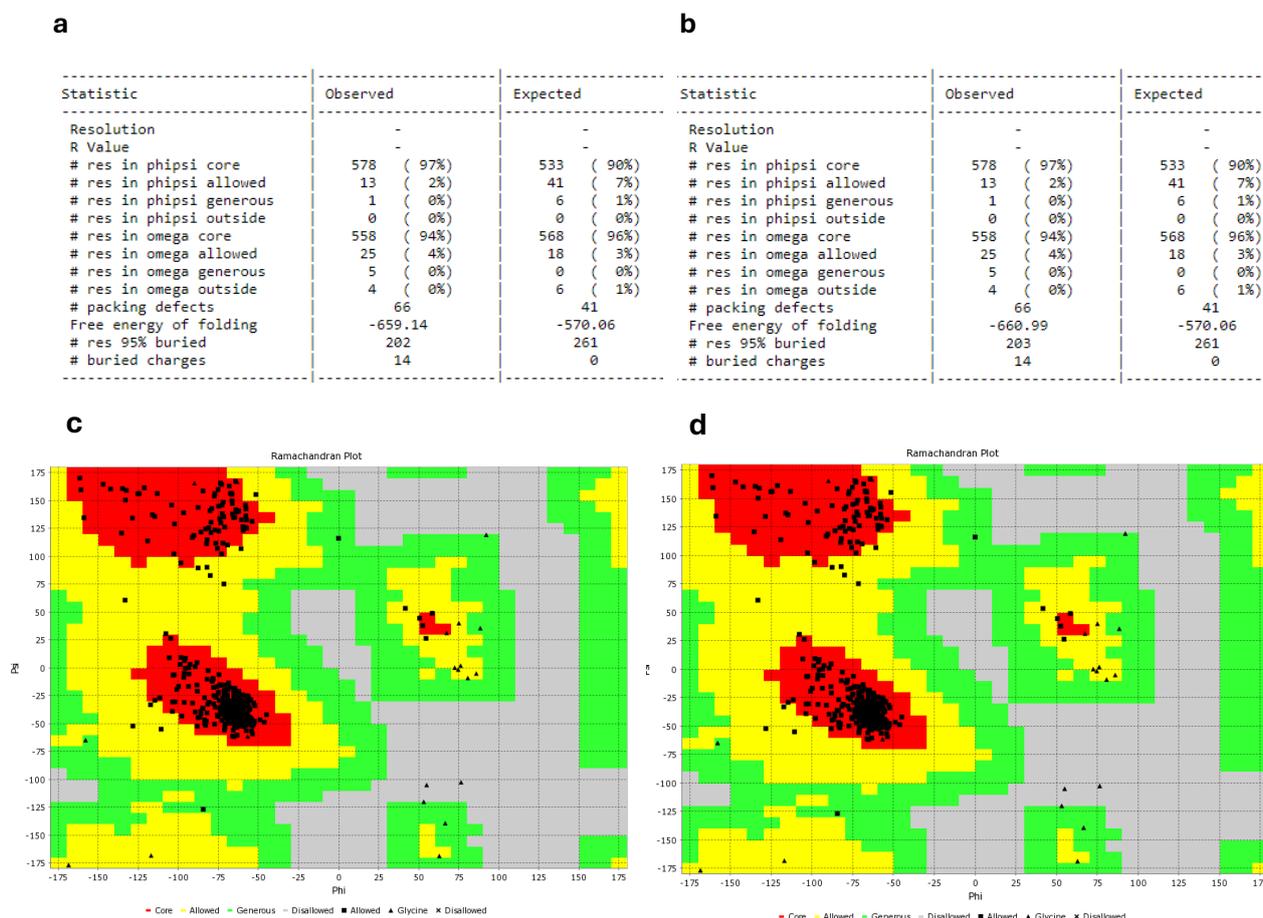
Variante	Gen	Posición en DNA	Base de referencia	Alteración de la base	Posición en la proteína	Cambio de aminoácido
rs17279437	SLC6A20	45814094	G	A	199/592	T/M
rs12720356	TYK2	10469975	A	C	684/1187	I/S

#### 4.5 Análisis estructurales *in silico* del efecto de las variantes

A continuación, se obtuvieron los siguientes resultados en cuanto a los análisis estructurales *in silico* para las dos variantes citadas.

#### 4.5.1 Variante rs17279437 de SLC6A20

En lo que se refiere a SLC6A20, se recogen los siguientes resultados de las gráficas de Ramachandran (Fig. 10). No se observan diferencias importantes entre los valores de genotipo silvestre (T199) y mutado (M199) en lo que respecta a defectos de ensamblaje y a la energía libre de plegamiento (Fig. 10 a y b). Tampoco se observan diferencias a simple vista en las gráficas (Fig. 10 c y d). Esto indicaría que *a priori* no se predicen cambios importantes en la estructura de la proteína mutada.



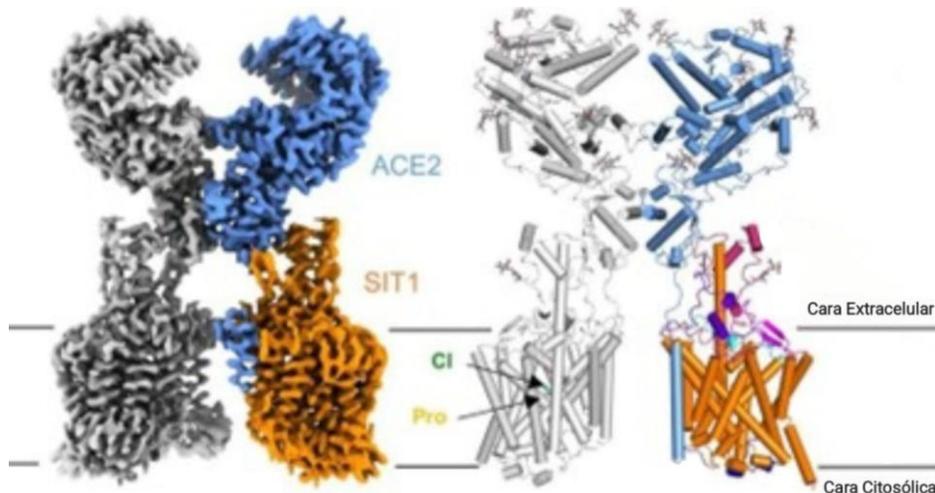
**Fig. 10.** Resultados de VADAR para SLC6A20. Estadísticas observadas y esperadas para las estructuras de SLC6A20 silvestre (a) y mutada (b) en la posición 199. Los valores observados de cantidad de defectos de ensamblaje así como la energía libre resultados son muy similares para ambos casos. Gráficas de Ramachandran para el fenotipo silvestre (c) y mutado en 199 (d) de SLC6A20. No se observan diferencias estructurales para la variante estudiada ya que no se marca presencia de aminoácidos en nuevas zonas prohibidas.

Al buscar predicción para fosforilación de la treonina silvestre en Netphos, se vio que el score de probabilidad de fosforilación se acerca a 0.5 para la proteína quinasa dependiente de cGMP (PKG), lo que no es concluyente pero si apunta a que ese residuo podría estar fosforilado (Fig. 11). Esto podría ser un cambio estructural importante dado que el residuo mutado carecería de grupo fosfato.

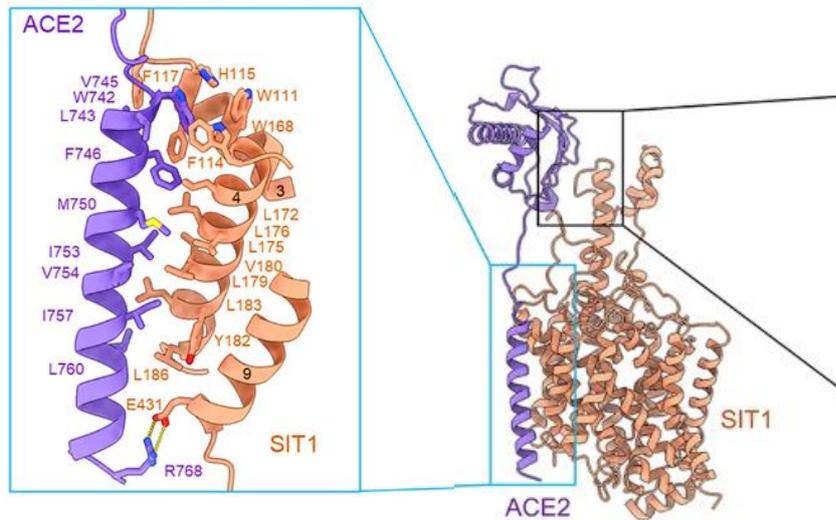
```
# netphos-3.1b prediction results
#
# Sequence          # x  Context    Score  Kinase  Answer
# -----
# Sequence          199 T  VVYFTASLP  0.475  PKG     .
# Sequence          199 T  VVYFTASLP  0.448  GSK3    .
# Sequence          199 T  VVYFTASLP  0.447  cdc2    .
# Sequence          199 T  VVYFTASLP  0.438  CaM-II  .
# Sequence          199 T  VVYFTASLP  0.361  CKI     .
# Sequence          199 T  VVYFTASLP  0.344  DNAPK   .
# Sequence          199 T  VVYFTASLP  0.320  p38MAPK .
# Sequence          199 T  VVYFTASLP  0.303  PKA     .
# Sequence          199 T  VVYFTASLP  0.291  CKII    .
# Sequence          199 T  VVYFTASLP  0.242  ATM     .
# Sequence          199 T  VVYFTASLP  0.205  cdk5    .
# Sequence          199 T  VVYFTASLP  0.201  RSK     .
# Sequence          199 T  VVYFTASLP  0.195  PKC     .
# Sequence          199 T  VVYFTASLP  0.084  PKB     .
# Sequence          199 T  VVYFTASLP  0.020  unsp    .
#
```

**Fig. 11.** Resultados de predicción de fosforilación por Netphos para la Treonina en posición 199. Se indica en la columna “Context” la secuencia diana, en “Score” la puntuación obtenida (probabilidad) y en “Kinase” la quinasa que potencialmente fosforila el residuo. El valor obtenido más alto es 0.475 para la quinasa PKG.

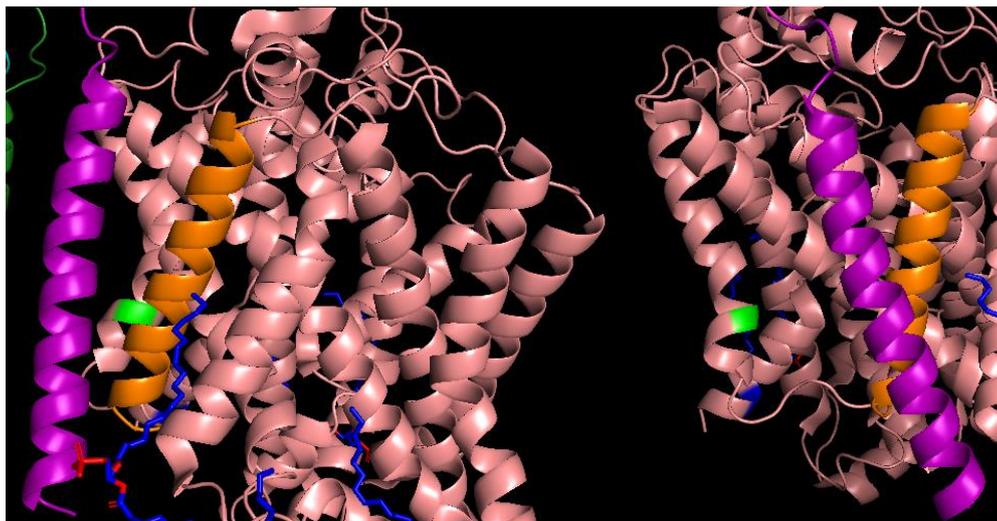
Según estudios previos, la proteína codificada por *SLC6A20*, también conocida como SIT1, interacciona con la parte extracelular del receptor transmembrana ACE2, que es además el receptor de la proteína S que media la entrada del virus SARS-CoV-2 (Yan et al., 2020) (Fig. 12). En este caso, la mutación estudiada, pT199M es un cambio de un aminoácido pequeño, polar y sin carga por uno apolar, grande y con un átomo de azufre. Este residuo se encuentra en una hélice transmembrana de la parte externa de SIT1 relativamente cercano a la única hélice transmembrana de ACE2. Esta hélice de ACE2 está en contacto directo con una hélice de SIT1, si bien el residuo estudiado no está en la zona de interacción (Figs. 13 y 14).



**Fig. 12.** Modificada de Chen et al., 2023, Cell discovery, 9(1), 93. (<https://doi.org/10.1038/s41421-023-00596-2>). Mapa de Criomicroscopía electrónica general (panel izquierdo) y con vista en forma de sticks (derecho) del complejo ACE2-SIT1. Se muestra la parte transmembrana. Un protómero de ACE2 y SIT1 está coloreado de azul y naranja, respectivamente. El otro protómero está coloreado de gris.

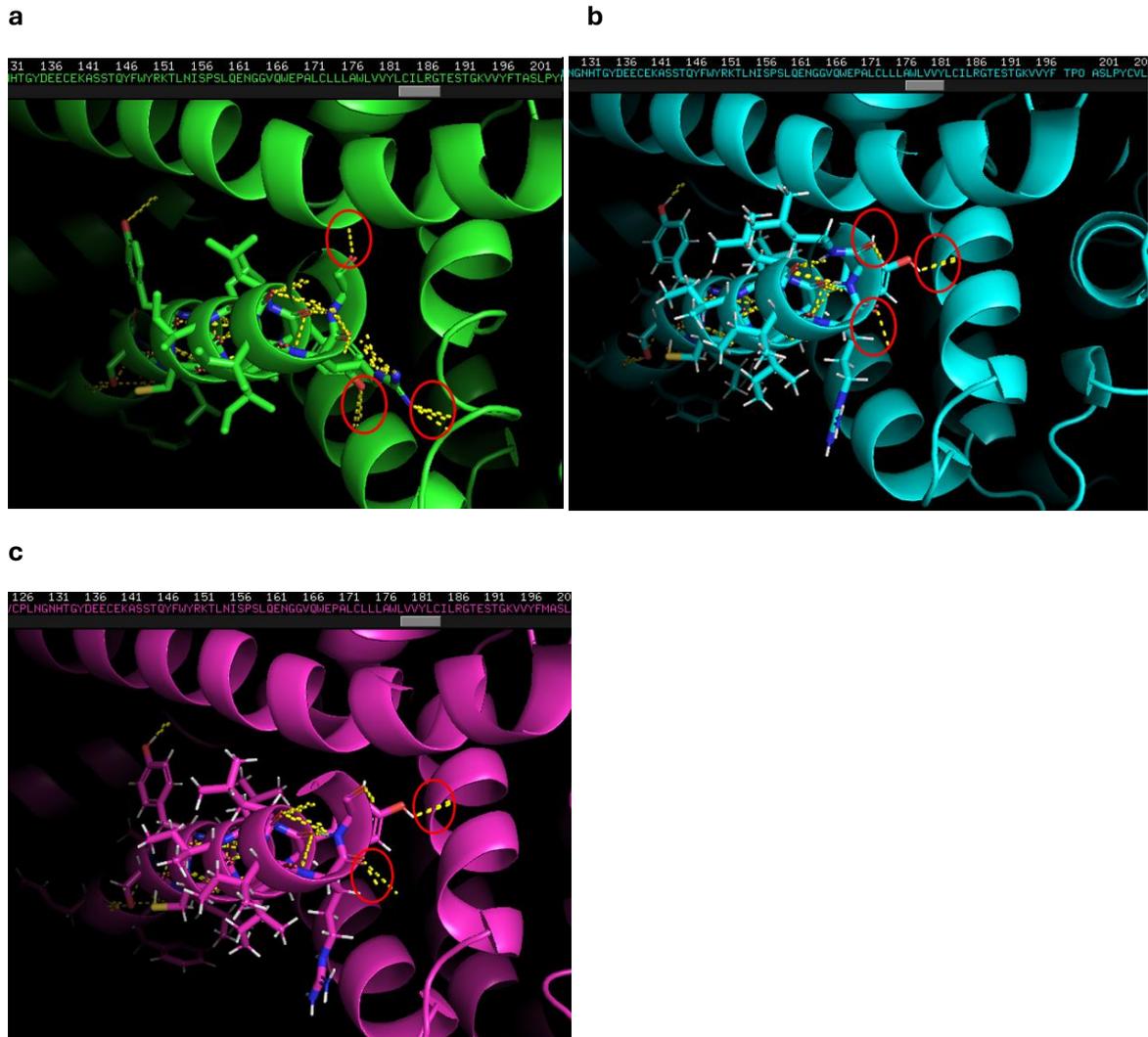


**Fig. 13.** Modificado de Shen et al., 2022, *Cell discovery*, 8(1), 123. (<https://doi.org/10.1038/s41421-022-00488-x>). Estructura del dominio transmembrana de ACE2 y SIT1 (derecha). ACE2 y SIT1 están coloreadas de violeta y salmón, respectivamente. Vistas ampliadas de la interacción entre la hélice transmembrana de ACE2 y SIT1 (izquierda) indicando los aminoácidos implicados.



**Fig. 14.** Detalle del dominio transmembrana de ACE2 y SIT1 en PyMOL. Perspectiva frontal (izquierda) y lateral (derecha). ACE2 está coloreada de violeta, y la hélice de SIT1 que interactúa con ella en la zona inferior, de naranja. Se observa como el residuo 199 (resaltado en verde) se encuentra alejado de la zona de interacción con la única hélice transmembrana de ACE2 (PDB: 7y75).

Mediante la superposición de las estructuras silvestre (PDB: 7y75), mutada (75y75 p.T199M) y fosforilada (7y75 fosforilada en T199) en PyMOL, se observó que hay un cambio en los puentes de hidrógeno para los tres casos (Fig. 15). En la hélice en la que se sitúa el aminoácido 199, se observan puentes intercatenarios de forma natural (Fig. 15 a), pero para la estructura mutada y fosforilada cambia la dirección hacia otras cadenas de SIT1 (Fig. 15 b, c) y se reduce del número de los mismos (Fig. 15 b, c). Debido a que todos estos cambios podrían afectar en la rigidez de la hélice y desembocar en una estructura cuaternaria diferente, este cambio aminoacídico es un buen candidato para realizar ensayos experimentales que determinen las diferencias entre la forma silvestre y mutada.



**Fig. 15.** Diferencias en PyMOL en la disposición de los puentes de hidrógeno (líneas discontinuas amarillas) en SIT1. Se muestra la hélice que contiene del aminoácido 193 al 214 en los modelos estructurales de a) la proteína silvestre (p.T199), b) la proteína fosforilada en TPO199 y c) la proteína mutada (p.M199).

#### 4.5.2 Variante rs12720356 de TYK2

Con respecto a *TYK2*, al mutar con la herramienta de PyMOL y buscar predicción de fosforilación en Netphos en la variante p.I684S, se observa un score máximo de 0.455 para la fosforilación de la serina por la quinasa GSK3. Aunque se puede considerar susceptible de fosforilación con un score partir de 0.5, se tuvo en cuenta, al igual que en el caso de la treonina de *SLC6A20*, ya que con un valor tan cercano al límite no se puede descartar completamente la fosforilación (Fig. 16).

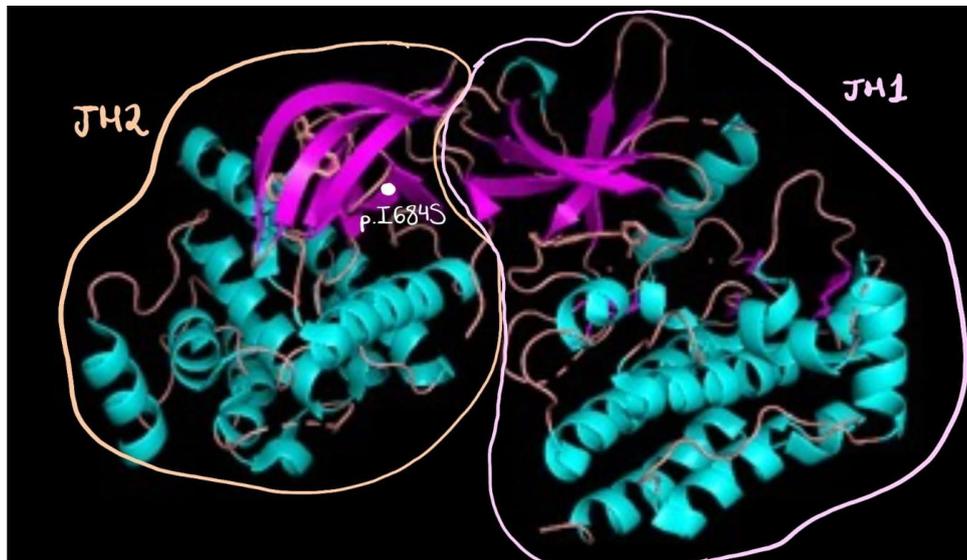
```

>Sequence      1187 amino acids
#
# netphos-3.1b prediction results
#
# Sequence      # x  Context    Score  Kinase    Answer
# -----
# Sequence      684 S  GPENSMVTE  0.455  GSK3     .
# Sequence      684 S  GPENSMVTE  0.447  CaM-II   .
# Sequence      684 S  GPENSMVTE  0.439  CKII     .
# Sequence      684 S  GPENSMVTE  0.431  DNAPK    .
# Sequence      684 S  GPENSMVTE  0.386  cdc2     .
# Sequence      684 S  GPENSMVTE  0.383  CKI      .
# Sequence      684 S  GPENSMVTE  0.346  PKG      .
# Sequence      684 S  GPENSMVTE  0.257  ATM      .
# Sequence      684 S  GPENSMVTE  0.254  RSK      .
# Sequence      684 S  GPENSMVTE  0.247  p38MAPK  .
# Sequence      684 S  GPENSMVTE  0.231  PKA      .
# Sequence      684 S  GPENSMVTE  0.190  cdk5     .
# Sequence      684 S  GPENSMVTE  0.171  PKB      .
# Sequence      684 S  GPENSMVTE  0.117  unsp     .
# Sequence      684 S  GPENSMVTE  0.046  PKC      .

```

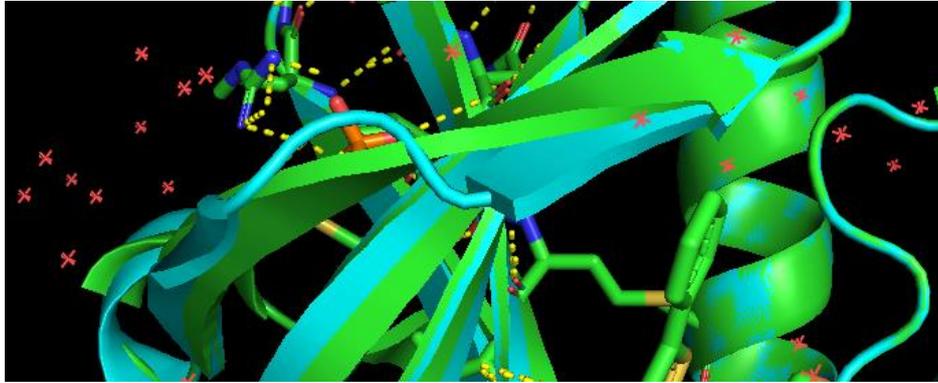
**Fig. 16.** Resultados de las predicciones de fosforilación en Netphos para la serina en posición 684 en la variante p.I684S de TYK2. El valor más alto corresponde a 0.455 a la quinasa GSK3. Se indica en la columna "Context" la secuencia diana, en "Score" la puntuación obtenida (probabilidad) y en "Kinase" la quinasa que potencialmente fosforila el residuo.

Según el estudio de Li et al. (2013), la mutación p.I684S de la proteína TYK2 (Fig. 17) se sitúa en la hélice beta del dominio pseudoquinasa (JH2), similar y adyacente al dominio quinasa (JH1). Este dominio JH2 contiene una zona de unión de ATP catalíticamente inactiva, de manera que el ATP lo estabiliza y modula la actividad de la quinasa (Min et al., 2015). Es por ello que esta variación estructural no parece alterar directamente el sitio catalítico del dominio quinasa, pero podría tener alguna repercusión indirecta.



**Fig. 17.** Visualización en PyMOL de los dominios quinasa (JH1) y pseudoquinasa (JH2) de TYK2. Ambos están unidos y formados ambos por alfa hélices (en azul) y lámina beta (en violeta) donde se localizaría la mutación p.I684S. (PDB: 4OLI).

Tras incorporar la mutación y fosforilación en el aminoácido 684 del dominio JH2 de *TYK2* (PDB:3ZON) y observar las diferencias en la estructura detenidamente (*Fig. 18*), encontramos un cambio en la misma hoja  $\beta$  en torno a los aminoácidos 577 a 610, que queda más corta para la mutada y fosforilada SEP684 (en azul) que para la secuencia original (en verde), lo que podría afectar al ordenamiento de la estructura terciaria de la proteína.



**Fig. 18.** Disposición alterada de la hoja beta al comparar la estructuras silvestre con la mutada y fosforilada en la posición 684 en PyMOL. I684 corresponde a la secuencia en verde Y SEP684 a la secuencia en azul. (PDB: 3ZON)

Por consiguiente, se confirma que la correcta disposición de la proteína podría ser alterada por esta mutación. Hay una distorsión en la hélice beta que podría estar causada por la introducción del grupo fosfato, lo que puede provocar una desorganización de la estructura llegando a afectar al dominio quinasa. Debemos tener en cuenta que se está produciendo un cambio de aminoácidos de isoleucina, que es apolar alifático a serina, aminoácido polar, sin carga y muy pequeño. Esto explicaría, junto con su posible fosforilación, las diferencias estructurales en esta hoja beta.

## 4.6 Estudios *in vitro* en células HEK293

### 4.6.1 Preparación y mutagénesis de los cDNAs

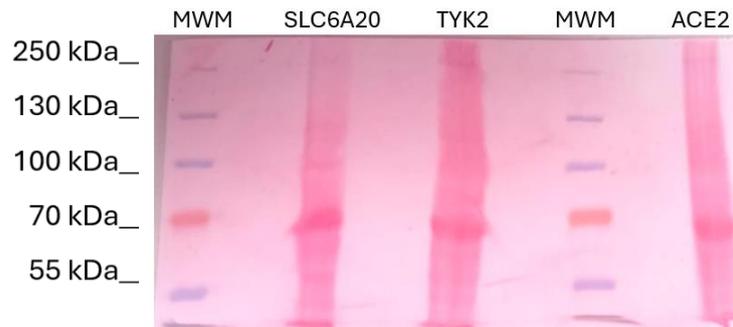
Teniendo el DNA bacteriano que incorporaba los plásmidos purificados, se efectuaron medidas de la concentración en un espectrofotómetro NanoDrop, obteniendo las siguientes: para *TYK2* se registró una concentración de 1711,5 ng/ $\mu$ l y para *SLC6A20*, de 1803 ng/ $\mu$ l. Las ratios 260/280 y 260/230 fueron respectivamente, de 1,92 y 2,25 para *TYK2* y 1,91 y 2,27 para *SLC6A20*, por lo que se puede afirmar que se trata de muestras puras y sin contaminantes.

Al transformar el producto purificado de la mutagénesis, y tras efectuar varias repeticiones con condiciones de PCR diferentes, no se obtuvieron colonias de bacterias que hubiesen incorporado las mutaciones en cuestión, pues no hubo crecimiento bacteriano en los tubos inoculados. Esto se pudo deber a un mal diseño de los *primers* de mutagénesis en primer momento o unas condiciones inadecuadas de PCR. También cabe la posibilidad de que las placas de medio LB en las que se sembró después de transformar el producto de PCR o las colonias de DH5competentes estuvieran en mal estado.

#### 4.6.2 Estudio de expresión de los cDNAs

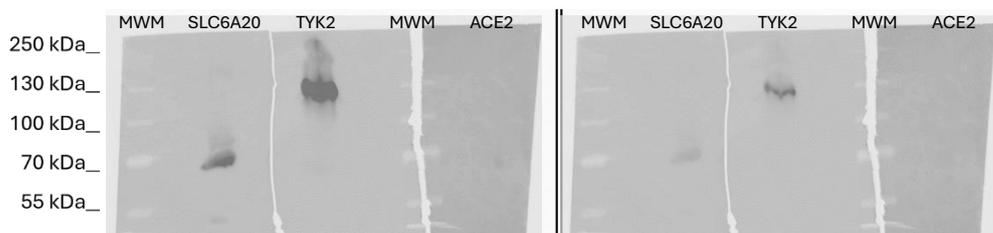
Tras la extracción de proteínas realizada a partir de la transfección de las células HEK293 con el DNA de interés, se obtuvieron los siguientes datos de cuantificación: 1,185 mg/mL para *TYK2* y 1,877 mg/mL para *SLC6A20*. A pesar de que las concentraciones detectadas no eran muy elevadas, se trataba de suficiente cantidad para realizar los ensayos posteriores.

Después de cargar las muestras en el gel SDS-PAGE y transferirlo a la membrana de *Western Blot*, se registró la siguiente imagen al teñir con Rojo Ponceau (*Fig. 19*). Los tres pocillos correspondientes a las tres proteínas cargadas se tiñeron como se observa en la imagen.



**Fig. 19.** Membrana de *Western Blot* tras tinción con Rojo Ponceau. Se indica en cada pocillo la proteína o marcador de peso molecular (MWM) y el tamaño estimado para cada banda.

Tras efectuar las distintas incubaciones con los anticuerpos primarios y secundarios, se obtuvieron las siguientes imágenes (*Fig. 20*). Como se puede comprobar, la proteína *TYK2* se marca cerca de los 130 kDa acorde a su peso molecular (133 kDa), en una banda inespecífica a los 5 primeros minutos y en una franja gruesa e intensa con elevado ruido de fondo a los 15 minutos. También se observa que la proteína *SLC6A20* de 66 kDa figura en torno a los 70 kDa, pero se detecta ruido de fondo y se marca como una banda gruesa e inespecífica. Por otro lado, en el caso de *ACE2*, no se detectó señal suficiente y no se visualizó una banda definida. Estos resultados podrían estar causados por una dilución inadecuada de las proteínas en cuestión o de sus respectivos anticuerpos.



**Fig. 20.** Imágenes capturadas a los 5 minutos del revelado (derecha, menos intensa) y a los 15 minutos (izquierda, más intensa) con *Image Reader LAS-3000*. La captura de los marcadores moleculares se encuentra superpuesta en cada una. Se indica el pocillo correspondiente a cada proteína o marcador y el tamaño estimado a cada banda.

En vista a los resultados obtenidos tras la realización del ensayo *Western Blot*, dichos problemas podrían solventarse mediante la dilución de las proteínas *TYK2* y *SLC6A20* o bien sus anticuerpos primarios correspondientes, para obtener bandas menos gruesas y con menor ruido de fondo. Además, se debería elevar la concentración de *ACE2* para conseguir detectar señal suficiente, y se debería supervisar manipulación de la membrana en los procesos entre incubaciones para evitar uniones inespecíficas y errores en el proceso de detección.

## 5 CONCLUSIONES

Las conclusiones extraídas del presente trabajo se pueden resumir en lo que sigue:

- Las variantes génicas rs17279437 del gen *SLC6A20* y rs12720356 del gen *TYK2* que corresponden a las mutaciones p.T199M y p.I684S respectivamente, son óptimas candidatas de estudio de acuerdo a su elevada predicción de patogenicidad, así como a la diferencia de las distribuciones fenotípicas de gravedad de COVID-19 observadas en pacientes.
- Los estudios estructurales de las proteínas codificadas por *TYK2* y *SLC6A20* apuntan a que las mutaciones p.T199M y p.I684S podrían desembocar en alteraciones estructurales que afecten al correcto ordenamiento en el espacio para ambas proteínas, perturbando sus funciones e interacciones normales.
- En el *Western Blot* realizado, se detectó una señal saturada e inespecífica para los pocillos que contenían *TYK2* y *SLC6A20*. Además, no se detectó señal para *ACE2*, necesaria para realizar estudios de interacción entre *SIT1-ACE2*. Este problema se podría solventar ajustando las diluciones de cada anticuerpo a concentraciones más adecuadas, es decir, aumentando la concentración de *ACE2* y disminuyendo la concentración de los anticuerpos de *TYK2* y *SLC6A20*, junto a un mayor control de la manipulación de la membrana en cada momento para evitar contaminaciones o uniones no específicas.
- Mediante las distintas repeticiones del proceso de mutagénesis por PCR, no se obtuvieron colonias de DH5 $\alpha$  que pudieran incorporar las mutaciones estudiadas para poder realizar una transfección en líneas celulares y comparar con las líneas no mutadas. Para futuros estudios, sería conveniente la revisión de los *primers* utilizados, así como las condiciones de PCR y los medios a emplear para el cultivo bacteriano.

## Bibliografía

- BRENO, M., NORIS, M., RUBIS, N., PARVANOVA, A. I., MARTINETTI, D., GAMBA, S., LIGUORI, L., MELE, C., PIRAS, R., ORISIO, S., VALOTI, E., ALBERTI, M., DIADEI, O., BRESIN, E., RIGOLDI, M., PRANDINI, S., GAMBA, T., STUCCHI, N., CARRARA, F., DAINA, E., ... ORIGIN STUDY GROUP (2023). A GWAS in the pandemic epicenter highlights the severe COVID-19 risk locus introgressed by Neanderthals. *iScience*, 26(10), 107629. <https://doi.org/10.1016/j.isci.2023.107629>
- COVID-19 HOST GENETICS INITIATIVE (2021). Mapping the human genetic architecture of COVID-19. *Nature*, 600(7889), 472–477. <https://doi.org/10.1038/s41586-021-03767-x>
- DEGENHARDT, F., ELLINGHAUS, D., JUZENAS, S., LERGA-JASO, J., WENDORFF, M., MAYA-MILES, D., UELLEND AHL-WERTH, F., ELABD, H., RÜHLEMANN, M. C., ARORA, J., ÖZER, O., LENNING, O. B., MYHRE, R., VADLA, M. S., WACKER, E. M., WIENBRANDT, L., BLANDINO ORTIZ, A., DE SALAZAR, A., GARRIDO CHERCOLES, A., PALOM, A., ... FRANKE, A. (2022). Detailed stratified GWAS analysis for severe COVID-19 in four European populations. *Human molecular genetics*, 31(23), 3945–3966. <https://doi.org/10.1093/hmg/ddac158>
- DOWNES, D. J., CROSS, A. R., HUA, P., ROBERTS, N., SCHWESSINGER, R., CUTLER, A. J., MUNIS, A. M., BROWN, J., MIELCZAREK, O., DE ANDREA, C. E., MELERO, I., COVID-19 MULTI-OMICS BLOOD ATLAS (COMBAT) CONSORTIUM, GILL, D. R., HYDE, S. C., KNIGHT, J. C., TODD, J. A., SANSOM, S. N., ISSA, F., DAVIES, J. O. J., & HUGHES, J. R. (2021). Identification of *LZTFL1* as a candidate effector gene at a COVID-19 risk locus. *Nat Genet* 53, 1606–1615 (2021). <https://doi.org/10.1038/s41588-021-00955-3>
- ENERBÄCK, C., SANDIN, C., LAMBERT, S., ZAWISTOWSKI, M., STUART, P. E., VERMA, D., TSOI, L. C., NAIR, R. P., JOHNSTON, A., & ELDER, J. T. (2018). The psoriasis-protective TYK2 I684S variant impairs IL-12 stimulated pSTAT4 response in skin-homing CD4+ and CD8+ memory T-cells. *Scientific reports*, 8(1), 7043. <https://doi.org/10.1038/s41598-018-25282-2>
- ESHETIE, S., JULLIAN, P., BENYAMIN, B., & LEE, S. H. (2023). Host genetic determinants of COVID-19 susceptibility and severity: A systematic review and meta-analysis. *Reviews in medical virology*, 33(5), e2466. <https://doi.org/10.1002/rmv.2466>
- GAMERO-DE-LUNA, E. J., & SÁNCHEZ-JAÉN, M. R. (2024). Factores genéticos asociados a long COVID. *Semergen*, 50(2), 102187. <https://doi.org/10.1016/j.semerg.2023.102187>
- GUAN, W. J., NI, Z. Y., HU, Y., LIANG, W. H., OU, C. Q., HE, J. X., LIU, L., SHAN, H., LEI, C. L., HUI, D. S. C., DU, B., LI, L. J., ZENG, G., YUEN, K. Y., CHEN, R. C., TANG, C. L., WANG, T., CHEN, P. Y., XIANG, J., LI, S. Y., ... CHINA MEDICAL TREATMENT EXPERT GROUP FOR COVID-19 (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*, 382(18), 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>
- HAMMING, I., TIMENS, W., BULTHUIS, M. L., LELY, A. T., NAVIS, G., & VAN GOOR, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *The Journal of pathology*, 203(2), 631–637. <https://doi.org/10.1002/path.1570>
- HUANG, C., WANG, Y., LI, X., REN, L., ZHAO, J., HU, Y., ZHANG, L., FAN, G., XU, J., GU, X., CHENG, Z., YU, T., XIA, J., WEI, Y., WU, W., XIE, X., YIN, W., LI, H., LIU, M., XIAO, Y., ... CAO, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)*, 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

- JALKANEN, J., KHAN, S., ELIMA, K., HUTTUNEN, T., WANG, N., HOLLMÉN, M., ELO, L. L., & JALKANEN, S. (2023). Polymorphism in interferon alpha/beta receptor contributes to glucocorticoid response and outcome of ARDS and COVID-19. *Crit Care* 27, 112. <https://doi.org/10.1186/s13054-023-04388-8>
- KOVANDA, A., LUKEŽIČ, T., MAVER, A., VOKAČ KRIŽAJ, H., ČIŽEK SAJKO, M., ŠELB, J., RIJAVEC, M., BIDOVEC-STOJKOVIĆ, U., BITEŽNIK, B., RITUPER, B., KOROŠEC, P., & PETERLIN, B. (2024). Genomic Landscape of Susceptibility to Severe COVID-19 in the Slovenian Population. *International Journal of Molecular Sciences*; 25(14):7674. <https://doi.org/10.3390/ijms25147674>
- KRIEGER, E., JOO, K., LEE, J., LEE, J., RAMAN, S., THOMPSON, J., TYKA, M., BAKER, D., & KARPLUS, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins*, 77 Suppl 9(Suppl 9), 114–122. <https://doi.org/10.1002/prot.22570>
- LI, Y., CHEN, Y., ZHANG, Y., SHEN, Y., XU, K., LIU, Y., WANG, Z., & YAN, R. (2023). Structural insight into the substrate recognition and transport mechanism of amino acid transporter complex ACE2-B<sup>0</sup>AT1 and ACE2-SIT1. *Cell Discov* 9, 93. <https://doi.org/10.1038/s41421-023-00596-2>
- LI, Z., GAKOVIC, M., RAGIMBEAU, J., ELORANTA, M. L., RÖNNBLUM, L., MICHEL, F., & PELLEGRINI, S. (2013). Two rare disease-associated Tyk2 variants are catalytically impaired but signaling competent. *Journal of immunology (Baltimore, Md. : 1950)*, 190(5), 2335–2344. <https://doi.org/10.4049/jimmunol.1203118>
- MIN, X., UNGUREANU, D., MAXWELL, S., HAMMARÉN, H., THIBAUT, S., HILLERT, E. K., AYRES, M., GREENFIELD, B., EKSTEROWICZ, J., GABEL, C., WALKER, N., SILVENNOINEN, O., & WANG, Z. (2015). Structural and Functional Characterization of the JH2 Pseudokinase Domain of JAK Family Tyrosine Kinase 2 (TYK2). *The Journal of biological chemistry*, 290(45), 27261–27270. <https://doi.org/10.1074/jbc.M115.672048>
- MODI, V., & DUNBRACK, R. L., JR (2019). Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14), 6818–6827. <https://doi.org/10.1073/pnas.1814279116>
- PAIRO-CASTINEIRA, E., RAWLIK, K., BREATHERICK, A. D., QI, T., WU, Y., NASSIRI, I., MCCONKEY, G. A., ZECHNER, M., KLARIC, L., GRIFFITHS, F., OOSTHUYZEN, W., KOUSATHANAS, A., RICHMOND, A., MILLAR, J., RUSSELL, C. D., MALINAUSKAS, T., THWAITES, R., MORRICE, K., KEATING, S., MASLOVE, D., ... BAILLIE, J. K. (2023). GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature*, 617(7962), 764–768. <https://doi.org/10.1038/s41586-023-06034-3>
- RAVIKANTH, V., SASIKALA, M., NAVEEN, V., LATHA, S. S., PARSA, K. V. L., VIJAYASARATHY, K., AMANCHY, R., AVANTHI, S., GOVARDHAN, B., RAKESH, K., KUMARI, D. S., SRIKARAN, B., RAO, G. V., & REDDY, D. N. (2021). A variant in *TMPRSS2* is associated with decreased disease severity in COVID-19. *Meta gene*, 29, 100930. <https://doi.org/10.1016/j.mgene.2021.100930>
- SEMIZ, S. (2021). SIT1 transporter as a potential novel target in treatment of COVID-19. *Biomolecular Concepts*, 12(1), 156-163. <https://doi.org/10.1515/bmc-2021-0017>
- SEVERE COVID-19 GWAS GROUP, ELLINGHAUS, D., DEGENHARDT, F., BUJANDA, L., BUTI, M., ALBILLOS, A., INVERNIZZI, P., FERNÁNDEZ, J., PRATI, D., BASELLI, G., ASSELTA, R., GRIMSRUD, M. M., MILANI, C., AZIZ, F., KÄSSENS, J., MAY, S., WENDORFF, M., WIENBRANDT, L., UELLEND AHL-WERTH, F., ZHENG, T., ... KARLSEN, T. H. (2020). Genomewide Association Study

- of Severe Covid-19 with Respiratory Failure. *The New England journal of medicine*, 383(16), 1522–1534. <https://doi.org/10.1056/NEJMoa2020283>
- SHEN, Y., WANG, J., LI, Y., ZHANG, Y., TIAN, R., & YAN, R. (2022). Structures of ACE2-SIT1 recognized by Omicron variants of SARS-CoV-2. *Cell discovery*, 8(1), 123. <https://doi.org/10.1038/s41421-022-00488-x>
- SMIESZEK, S. P., POLYMEROPOULOS, V. M., POLYMEROPOULOS, C. M., PRZYCHODZEN, B. P., BIRZNIKES, G., & POLYMEROPOULOS, M. H. (2022). Elevated plasma levels of CXCL16 in severe COVID-19 patients. *Cytokine*, 152, 155810. <https://doi.org/10.1016/j.cyto.2022.155810>
- WILLARD, L., RANJAN, A., ZHANG, H., MONZAVI, H., BOYKO, R. F., SYKES, B. D., & WISHART, D. S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic acids research*, 31(13), 3316–3319. <https://doi.org/10.1093/nar/gkg565>
- XIA, R., PENG, H. F., ZHANG, X., & ZHANG, H. S. (2024). Comprehensive review of amino acid transporters as therapeutic targets. *International journal of biological macromolecules*, 260(Pt 2), 129646. <https://doi.org/10.1016/j.ijbiomac.2024.129646>
- YAN, R., ZHANG, Y., LI, Y., XIA, L., GUO, Y., & ZHOU, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, 367(6485), 1444–1448. <https://doi.org/10.1126/science.abb2762>