

RESEARCH

Red de los targets de SARS-CoV2

Sandra Castro Labrador*, Fiorella Piriz Sapio*, Adrián Segura Ortiz* and María del Rocío Valderrama Palacios*

*Correspondence:

sandra091999@gmail.com;
fiorellapiriz@uma.es;
adrianseor.99@gmail.com;
Rvalpalacios@gamil.com
ETSI Informática, Universidad de
Málaga, Málaga, España
La lista completa de información
de los autores está disponible al
final del artículo.

Abstract

En el presente proyecto de investigación procederemos a explorar las diversas interacciones que existen entre las proteínas humanas y las 29 proteínas conocidas que codifican el virus SARS-CoV-2. Con el fin de alcanzar este objetivo, realizaremos un intenso trabajo de recopilación de material bibliográfico acerca cómo interacciona el organismo comentado sobre el interactoma humano. Estos datos los podemos encontrar en artículos y libros en los que los experimentos realizados tienen como objetivo determinar e mapa de interacción de las proteínas humanas con las proteínas del SARS-Cov-2. En cuanto a la herramienta para obtener la red del sistema, emplearemos principalmente STRING y R, puesto que ya hemos visto a lo largo del curso la cantidad de información relevante que nos pueden proporcionar.

Keywords: SARS-CoV-2; interactoma humano; STRING; R

1 Introducción

A finales de 2019 se unió a la familia de los coronavirus el denominado síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV2), cuya infección da lugar al conocido coronavirus 2019 o COVID-19. Este virus ha provocado una pandemia y crisis mundial, dejando por delante un total de más de 2 millones de muertes y casi 96 millones de infectados en todo el mundo. Por ello, desde los equipos de investigación, se está trabajando inmensurablemente por identificar los componentes del mismo, así como sus genes, sus proteínas y las formas en las que estas interaccionan con el ser humano.

A día de hoy se sabe que el SARS-CoV-2 está formado por 29 proteínas que se asocian con las células humanas dando lugar a múltiples de síntomas que en su máxima expresión pueden llevar a la muerte del individuo con relativa facilidad.

Por otra parte, debemos tener en cuenta que el interactoma define el conjunto de interacciones moleculares que tiene lugar en el interior de una célula. El estudio de estas redes biológicas es realmente relevante en diversas áreas de investigación, ya que conocer cómo interaccionan las proteínas entre sí (PPI) permite el descubrimiento de patrones, la elaboración de fármacos o la explicación de procesos biológicos, entre otros.

Es por ello, por lo que en este proyecto pretendemos modelar y estudiar la red de interacción de las proteínas de este virus con las proteínas humanas, así como examinar brevemente si existe algún conjunto de fármacos con targets en las proteínas de la red, observar si la combinación de medicamentos afecta a las enzimas virales, determinar las funciones celulares a las que afecta, etc.

Finalmente, tanto los resultados como la presente memoria, serán proporcionados en el repositorio de github de la asignatura, accesible a todos los miembros del grupo.

2 Materiales y métodos

2.1 Carga de datos

El primer paso para poder modelar y analizar nuestra red es descargar los datos desde StringDB. Hemos obtenido la red de 332 mencionada por Gordon et al en la versión web de StringDB, ampliando el número de proteínas en la segunda capa hasta obtener una red completamente conectada, que hemos descargado en formato tsv. Este archivo lo hemos usado posteriormente para cargar la red en nuestro entorno de trabajo usando el paquete stringDB que proporciona R.

```

1 string_db <- STRINGdb$new(version = "11",
2                             species = 9606,
3                             score_threshold = 400,
4                             input_directory = "")
5
6 proteins.table <- read.delim(file = 'files/string-protein-annotations.
7   tsv',
8   sep = '\t', header = TRUE)
9 proteins.names <- data.frame(proteins.table[1])
10 colnames(proteins.names)[1] <- "genes"
11 proteins.mapped.table <- string_db$map(proteins.names, "genes",
12   removeUnmappedRows = TRUE)
13 proteins.mapped.string_ids <- proteins.mapped.table$STRING_id
14
15 string.network <- string_db$get_graph()
16 proteins.mapped.network <-
17   string_db$get_subnetwork(proteins.mapped.string_ids)
18
19 proteins.mapped.network.comp <- components(proteins.mapped.network)
20 nodes_to_remove <- names(proteins.mapped.network.comp$
21   membership[proteins.mapped.network.comp$membership!=1])
22
23 proteins.mapped.network <- delete_vertices(proteins.mapped.network,
24   nodes_to_remove)

```

2.2 Análisis inicial y robustez

Para el análisis inicial de la red nos hemos centrado en 3 parámetros básicos: distribución del grado, coeficiente de clústering y distancia de los nodos.

En cuanto a la robustez, se ha calculado tanto frente a ataques aleatorios como frente a ataques dirigidos, haciendo uso del código proporcionado en el campus virtual.

2.3 Linked Communities

Para la búsqueda de comunidades se ha empleado el paquete **linkcomm**, una herramienta que permita la creación, visualización y tratamiento de comunidades dentro de un grafo (igraph). Con la ayuda del documento *'The generation, visualization, and analysis of link communities in arbitrary networks with the R package linkcomm'*, hemos conseguido encontrar las comunidades más importantes (centralidad) y los módulos existentes en nuestra red. Además, se han realizado diversas

representaciones gráficas para visualizar las comunidades de una forma más atractiva.

2.4 Enriquecimiento funcional

Una vez obtenido el análisis de las comunidades queremos llevar a cabo un análisis de enriquecimiento funcional para poder obtener principalmente en qué procesos biológicos están involucrados los clústers más conectados obtenidos.

Comenzamos con un enriquecimiento funcional a partir de **STRINGdb**. En este caso se combinan la ontología **GO** y las base de datos **KEGG** y **Pfam**. Obtenemos los resultados del enriquecimiento de los diferentes clústers seleccionados mediante ficheros csv.

Elaboramos un segundo enriquecimiento funcional mediante el paquete **clusterProfiler**. Con este análisis corroboramos las funcionalidades obtenidas con el anterior enriquecimiento y mostramos gráficas para una visualización de los datos más intuitiva. Hemos elaborado una función con la cual mapeamos los identificadores de las proteínas para obtener los respectivos genes a partir de los cuales aplicamos las funciones de enriquecimiento mediante **GO** y **KEGG**. Hemos empleado el **método de simplificación** que ofrece clusterProfiler para evitar la redundancia de los datos resultantes del enriquecimiento. Esta función la hemos aplicado sobre el enriquecimiento mediante **enrichGO** siempre que indiquemos alguno de los dominios de la ontología GO (en los resultados hemos utilizado principalmente Biological Process (BP)).

Con este análisis obtenemos variedad de gráficas a partir de funciones como **cnetplot** y **heatmap** entre otras, con las cuales analizaremos la repercusión biológica de estos módulos.

2.5 Búsqueda en Drugbank

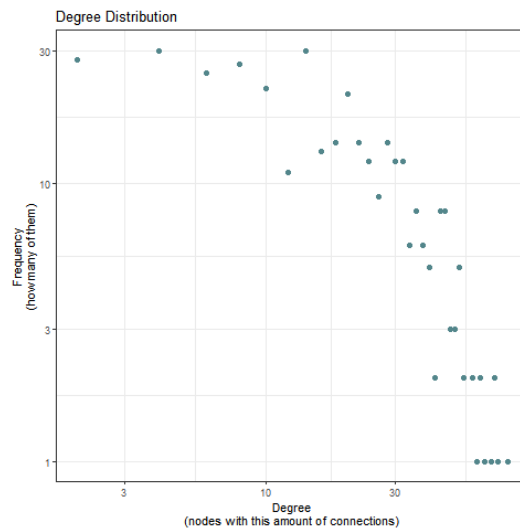
Por último, hemos buscado en la base de datos drugbank los identificadores de los medicamentos cuyos targets se encuentran entre las proteínas más conectadas de nuestra red. Finalmente, hemos elegido limitarnos a aquellos medicamentos que ya están aprobados, ya que la carga de la base de datos completa en el entorno de trabajo disparaba el tiempo de ejecución.

3 Resultados y discusión

3.1 Análisis inicial y robustez

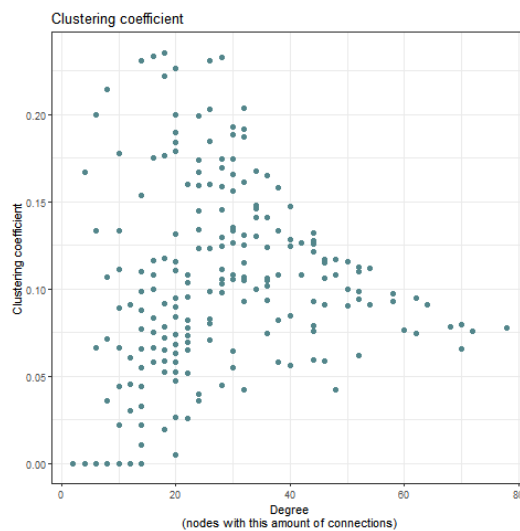
Como podemos observar, la distribución del grado de nuestros nodos sigue la ley de potencias, por lo que podemos decir que el modelo libre de escala es el que más se ajusta a nuestra red, algo que ya sabíamos de antemano, ya que estamos modelando una red real.

En cuanto a los hubs, podemos decir que hay dos grupos bastante diferenciados



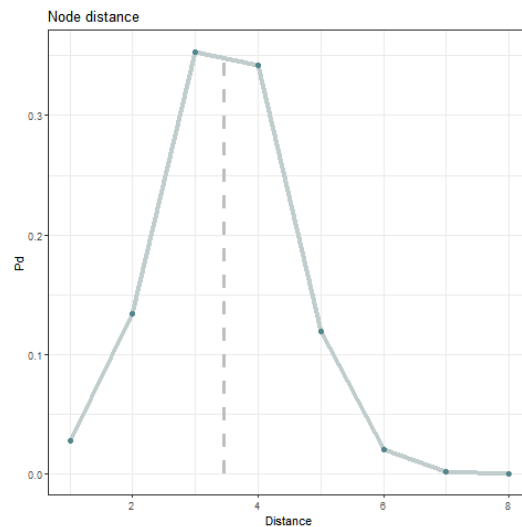
Distribución del grado en escala logarítmica

En este caso, el coeficiente medio de clustering de nuestra red es de 0.43. En la figura podemos observar que los nodos con un grado menor varían mucho en cuanto a C , teniendo nodos que están muy agregados y otros que están casi desconectados. Esto no pasa en los hubs, que se mantienen muy cercanos en cuanto a transitividad



Coeficiente de clustering

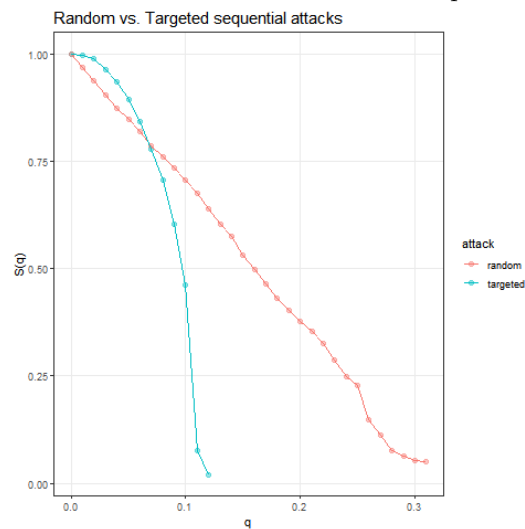
Si nos fijamos en la distancia entre los nodos de nuestra red, podemos comprobar que tiene propiedad de mundo pequeño, ya que la mayoría de nodos tienen distancia 3 o 4, siendo la media 3.46. Es decir, nuestra red está bien comunicada.



Distancia entre los nodos

En la siguiente figura observamos que la red es bastante robusta ante ataques aleatorios y bastante menos ante ataques dirigidos, lo que nos indica una vez más que el modelo libre de escala es el que se ajusta mejor.

Que disminuya tan rápido la conectividad eliminando una fracción tan pequeña de nos indica que debemos centrarnos en el estudio de esas proteínas principalmente.



Robustez frente a ataques dirigidos y aleatorios

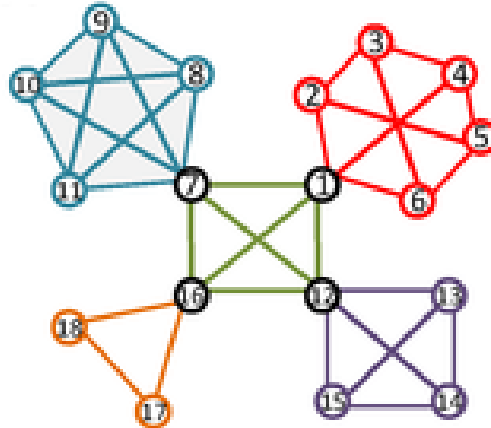
3.2 Linked Communities

En esta sección procedemos a obtener las comunidades existentes dentro de nuestra red de proteínas, obtenida en la sección anterior.

Antes debemos tener claro que las comunidades son conjuntos de nodos relacionados entre sí que poseen funciones semejantes y que buscan conseguir un objetivo común. Es por ello por lo que la identificación de las comunidades es un problema relevante para muchas áreas de investigación como la sociología, la biología o la informática.

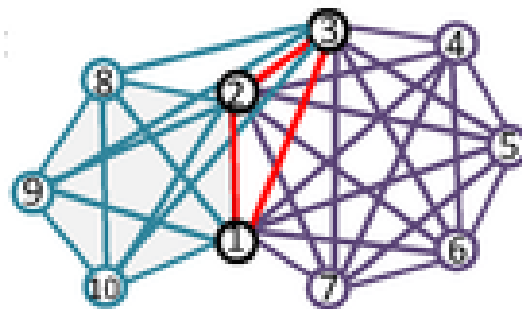
En general, las comunidades se pueden clasificar en dos tipos, las que están formadas por nodos y las que están formadas por enlaces.

- Comunidad de nodos: Se tratan de subgrafos formados por nodos densamente conectados entre ellos, pero muy poco conectados con los nodos de alrededor, como se muestra en la imagen.



Comunidad de nodos

- Link community: Consiste en un subgrafo en el que existe una gran cantidad de enlaces pero muy pocos con las comunidades externas. La forma de detectar estos conjuntos es mediante la división de los enlaces de la red. En estas particiones, las conexiones determinan una comunidad, pero los nodos pueden pertenecer a varias comunidades.



Link community

Con esto queremos remarcar que la importancia de la detección de estos subgrafos se debe a que en el campo de la biología, permiten encontrar módulos proteicos con una misma función celular o predecir las funciones de las proteínas.

Por tanto, diseñaremos un código en R que nos permite la detección de las comunidades con el fin de filtrar aquellas que son más importantes (centralidad). De esta forma, podemos llevar a cabo un enriquecimiento funcional de las mismas y así obtener algunas de las funciones celulares que determinan red del SARS-CoV-2.

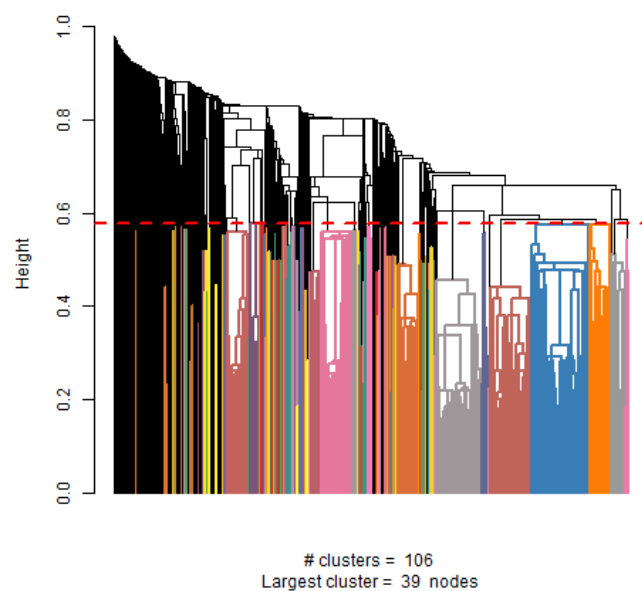
Además, en el código siguiente se incluyen diferentes gráficas para la visualización de las comunidades.

A lo largo de la ejecución del código hemos añadido distintas gráficas para la representación de las comunidades. A continuación se muestra un dendrograma de las *link communities*, en el que se puede apreciar una gran cantidad de comunidades. En esta imagen es difícil ver la centralidad de las comunidades, por lo que haremos más representaciones.

```

1 #link communities dendogram
2 png(file="linkcomm_dend.png")
3 plot(proteins.mapped.network.lc, type = "dend")
4 dev.off()

```



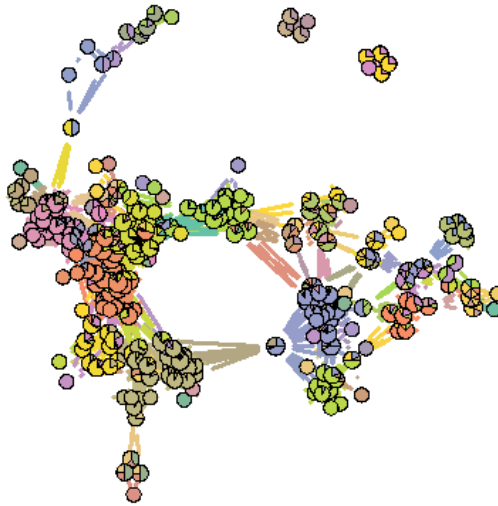
Dendograma de Linked Communities

Empleando el diseño **Fruchterman Reingold** se puede obtener una visión más amplia de las comunidades.

```

1 #link communities Fruchterman Reingold layout
2 png(file="linkcomm_layout_fruchterman_reingold.png")
3 plot(proteins.mapped.network.lc, type = "graph", layout =
4 layout.fruchterman.reingold, vlabel=FALSE)
5 dev.off()

```



Linked Communities con Fruchterman Reingold layout

Sin embargo, lo que nos interesa en este caso es obtener la '**centralidad comunitaria**' pues se define como la suma de las áreas con mayor influencia sobre los nodos vecinos de la red. Por tanto, nos permite obtener aquellos clusters que ejercen una mayor influencia sobre la red, pudiendo ser determinantes en la funcionalidad de la misma.

Por otra parte, la **modularidad** consiste en el grado de separación y recombinación existente entre los componentes de una red. Es decir, se considera como una medida de la presencia de estructura comunitaria. Esto permite la búsqueda de comunidades, quedándonos con aquellas que tengan un valor de modularidad positivo y lo más grande posible (modularidad optimizada).

Por ello, hemos calculado la modularidad de las comunidades de nuestra red y hemos representado el valor de la modularidad para cada una de ellas.

```

1 # Community centrality
2 community.centralidad <- getCommunityCentrality(proteins.mapped.network
   .lc)
3
4 #modularity of the communities
5 community.connectedness <- getCommunityConnectedness(
6 proteins.mapped.network.lc, conn = "modularity")
7
8 png(file="communities_modularity.png")
9 plot(proteins.mapped.network.lc, type = "commsumm", summary = "
   modularity")
10 dev.off()

```

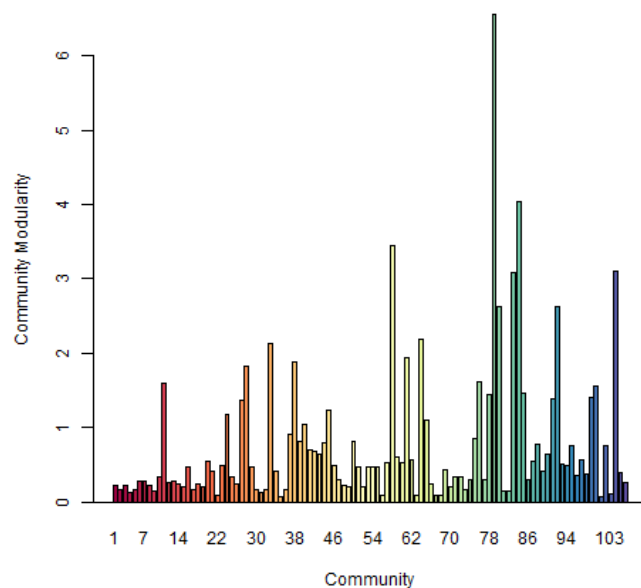



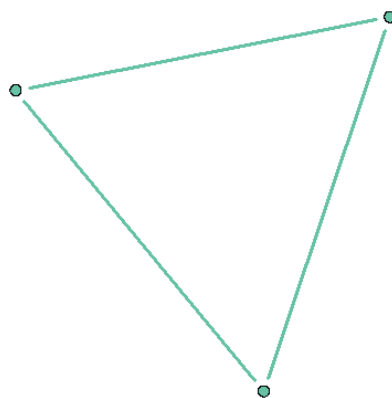
Diagrama de barras de la modularidad de las comunidades

En la imagen se observa claramente como hay una comunidad con una modularidad muy alta, lo que indica que esta es más influyente en la red que el resto, concretamente la comunidad 79.

```

1 # Focus on one linkcomm
2 #plot one cluster with maximun community modularity
3 png(file="cluster12_graph.png")
4 plot(proteins.mapped.network.lc, type = "graph", clusterids =
5 community.connectedness.maximum, vlabel=FALSE)
6 dev.off()

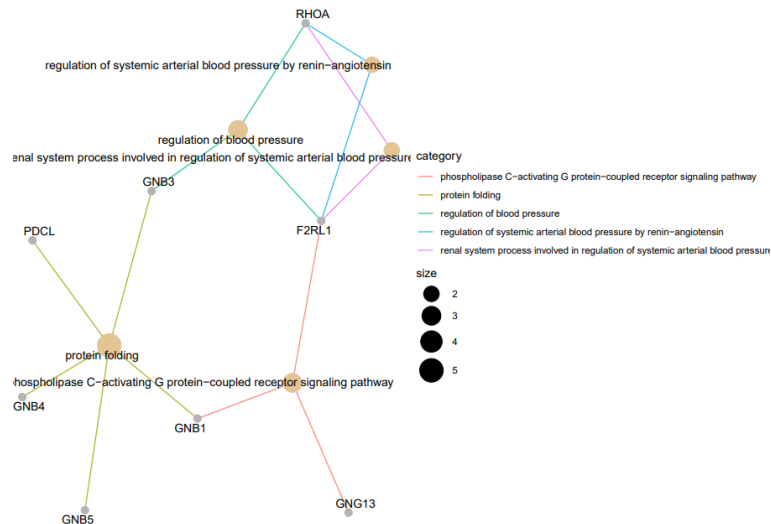
```



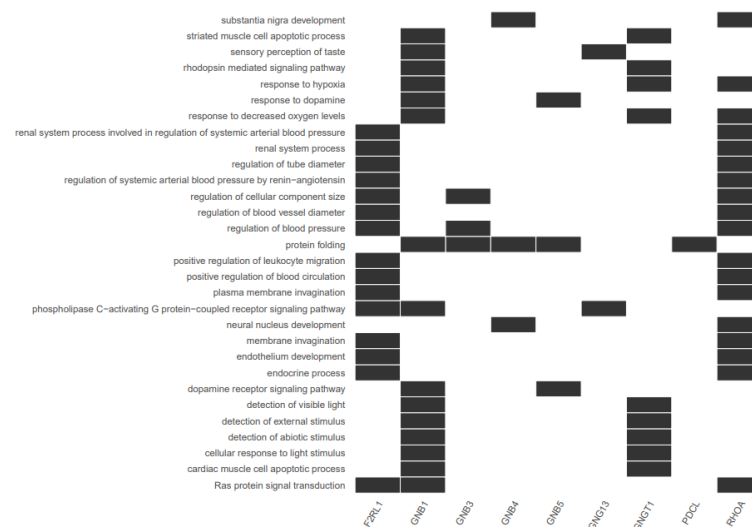
Por tanto, realizaremos un análisis funcional centrándonos en aquellas comunidades con una modularidad mayor, para determinar si sus funciones moleculares son determinantes o no.

3.3 Enriquecimiento funcional

3.3.1 Cluster 104



En la figura superior podemos observar mediante un mapa de calor o heatmap, los 5 términos de la GO predominantes de la subred, además de otras funciones asociadas a cada proteína. Para el clúster 104 vemos que el término con un mayor tamaño en el mapa es ‘protein folding’. Esto indica que su GeneRatio (porcentaje de DEGs asociado al término) es grande y el q valor o p valor ajustado, es pequeño.

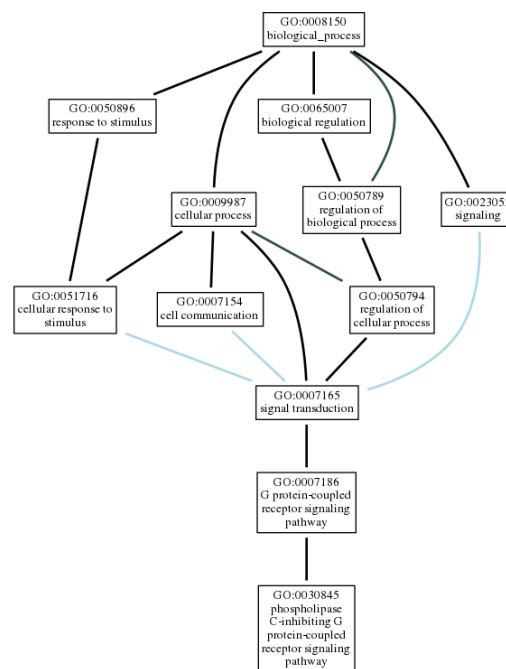


Investigando en la ontología, obtenemos que el **plegamiento de proteínas** es un proceso biológico que facilita el ensamblaje de proteínas para dar lugar a una estructura terciaria correcta.

De ahí podemos deducir el papel que juega esta función pues un fallo en plegamiento puede provocar un desorden celular con amplias consecuencias. Es más,

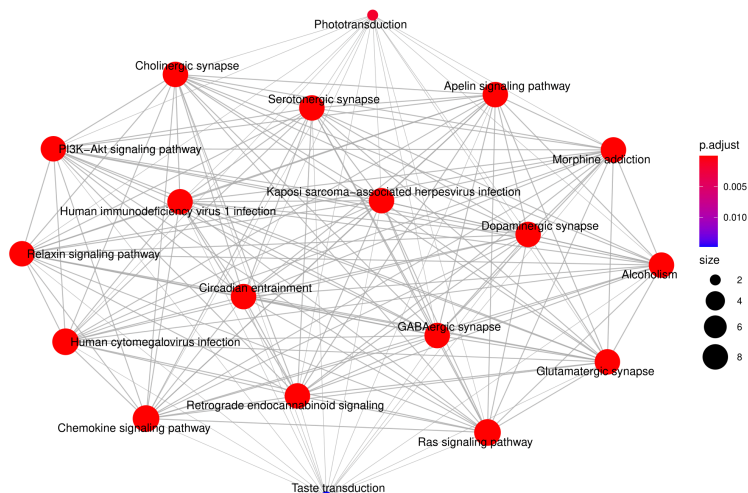
existen una gran cantidad de enfermedades generadas por este tipo de fallos como el Alzheimer, el Parkinson, la fibrosis quística y muchos otros trastornos degenerativos, tal y como se mencionan en el artículo *‘Protein-misfolding diseases and chaperone-based therapeutic approaches’*. Según este estudio, “más de la mitad de las enfermedades humanas podrían estar relacionadas con un plegamiento incorrecto de las proteínas”. Este proceso parece afectar a las chaperonas, pues son las encargadas de reparar las mutaciones que causan las patologías.

Por otra parte, la **‘ruta de señalización del receptor acoplado a la proteína G que inhibe la fosfolipasa C’**, es otro de los procedimientos identificados en mayor medida con el clúster en cuestión. Concretamente, esta vía además de inhibir la actividad de la fosfolipasa C, conlleva una disminución de los niveles de DAG y IP3. Este último es un mensajero de señalización celular cuya disminución parece estar relacionada con la **autofagia** o vía de degradación de proteínas, orgánulos y material citoplasmático. Por su parte, los segundos mensajeros DAG dan lugar a la activación de la proteína quinasa C, la cual permite la activación de una serie de rutas metabólicas que inducen la expresión de proteínas que activan los linfocitos T. Estos desempeñan un papel fundamental en la regulación del sistema inmune, por lo que una alteración de los mismo puede generar una inmunodeficiencia severa.



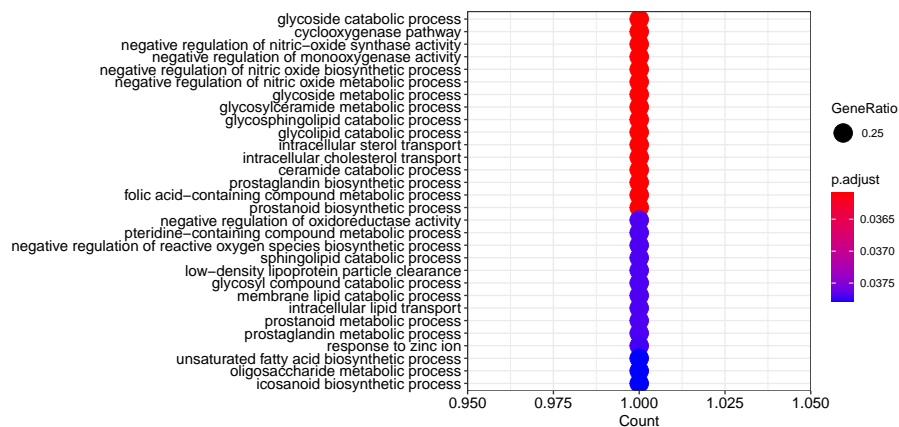
Otra de las funciones biológicas que aparecen en el enriquecimiento de este módulo, está relacionada con la **regulación de la presión sanguínea**, y, por tanto, de los niveles de oxígeno. Este es un proceso muy complejo que viene determinado por el Sistema Nervioso Autónomo, el SNC y el riñón. El objetivo del SN es mantener la presión arterial (PA) mediante la regulación de los niveles de oxígeno. Sin embargo, un nivel alto de PA o **hipertensión**, puede llegar a ser nocivo para el corazón, puesto que lo obliga a bombear más sangre, contribuyendo así al endurecimiento de las arterias, a la producción de accidentes cerebrovasculares, enfermedades renales o insuficiencia cardíaca.

Según un estudio reciente, '*COVID-19 and Hypertension: What We Know and Don't Know*', el SARS-CV-2 interactúa con el Sistema renina-angiotensina aldosterona (RAAS), actuando sobre el receptor ACE2 induciendo una desregulación de la misma, lo que genera la acumulación local de angiotensina II. Debido a que el sistema RAAS se encarga de regular la presión arterial, la alteración del mismo da lugar a la hipertensión de los pacientes.



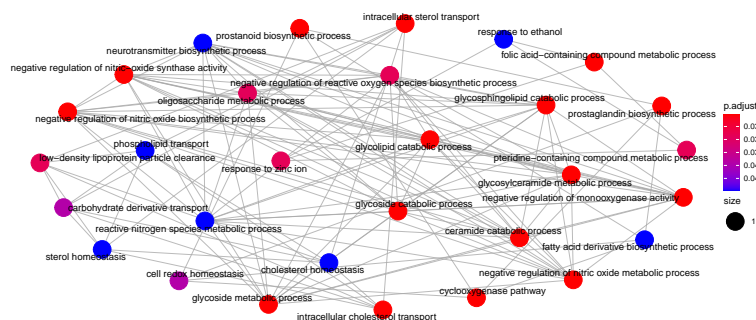
Para concluir, en la figura superior podemos observar como alguno de los procesos mencionados se encuentran en el mapa de vías metabólicas. Además, en el mapa de calor obtenido, también se observan muchos de los términos más presentados en el clúster 104.

3.3.2 Clúster 84



Este clúster está formado por 4 proteínas altamente interconectadas cuyas funciones biológicas se relacionan con 4 aspectos principalmente. El primero de ellos y el más abundante es el **metabolismo y catabolismo de glúcidos y lípidos**, los cuales son nuestra principal fuente de energía en el organismo. El hecho de que las proteínas del virus interactúen con las de nuestro clúster puede provocar alteraciones en el desarrollo de esta función tan esencial. El siguiente tema a tratar es la **regulación negativa del óxido nítrico**. Este compuesto es mencionado en

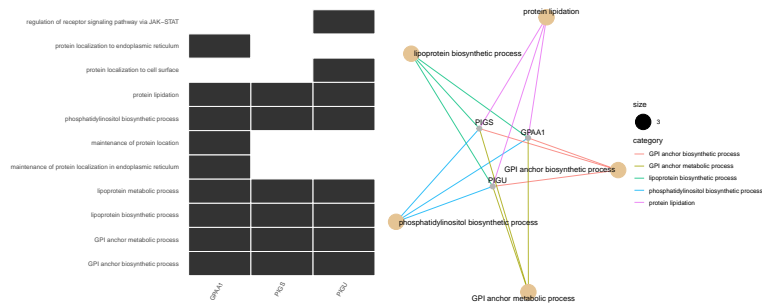
varias de las funciones biológicas de nuestras proteínas y se trata de un gas que se genera en el endotelio. Se caracteriza por tener propiedades vasodilatadoras y contribuir al mantenimiento de la presión arterial baja. Por tanto, una regulación negativa del mismo significaría un déficit de este gas en el cuerpo, lo que puede producir entre otras cosas, hipertensión arterial. En tercer lugar en nuestras funciones se hace mención a la **ruta de la ciclooxigenasa (COX)** y por tanto a sus principales productos que son las **prostaglandinas**. Éstas son un conjunto de sustancias de carácter lipídico derivadas de los ácidos grasos que conllevan diversos efectos en nuestro organismo, a menudo contrapuestos. Las prostaglandinas afectan y actúan sobre diferentes sistemas, incluyendo el sistema nervioso, el músculo liso y el sistema reproductor. Además, juegan un papel importante en regular diversas funciones como la presión sanguínea, la coagulación de la sangre, la respuesta inflamatoria alérgica y la actividad del aparato digestivo. Por tanto, la interacción del virus con estos compuestos pueden suponer un ataque en nuestro cuerpo. Finalmente aparecen varias funciones relacionadas con el **transporte intracelular**, lo cual es de vital importancia en las células para expulsar de su interior los desechos del metabolismo o trasladar sustancias que sintetiza como hormonas entre otras actividades.



Además, se puede apreciar como la clasificación de nuestras funciones en estos 4 aspectos se encuentra claramente reflejada en la figura anterior, donde las proteínas pertenecientes a un mismo tema presentan un mayor número de conexiones entre ellas.

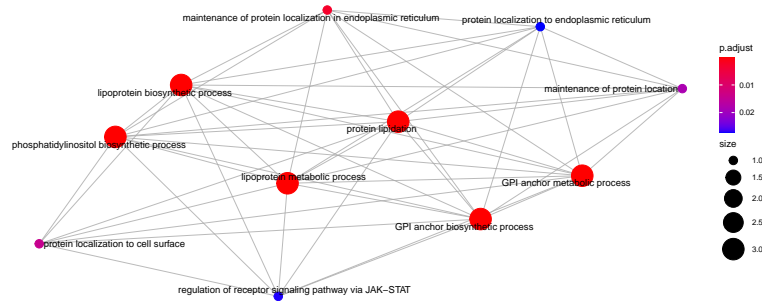
3.3.3 Clúster 79

La principal funcionalidad que podemos ver es la biosíntesis de las proteínas de membrana GPI: Proteínas de la superficie celular que se pueden unir a la membrana mediante una estructura de glicolípidos denominadas **anclaje de glicosil-fosfatidilinositol (GPI)**. Las proteínas que son detectadas en el cluster con una mayor fiabilidad estadística son las necesarias para la construcción de estas estructuras. Procesos celulares asociados a la biosíntesis de lipoproteínas son atribuidos a dichas proteínas junto con procesos metabólicos de la formación de estos complejos GPI.



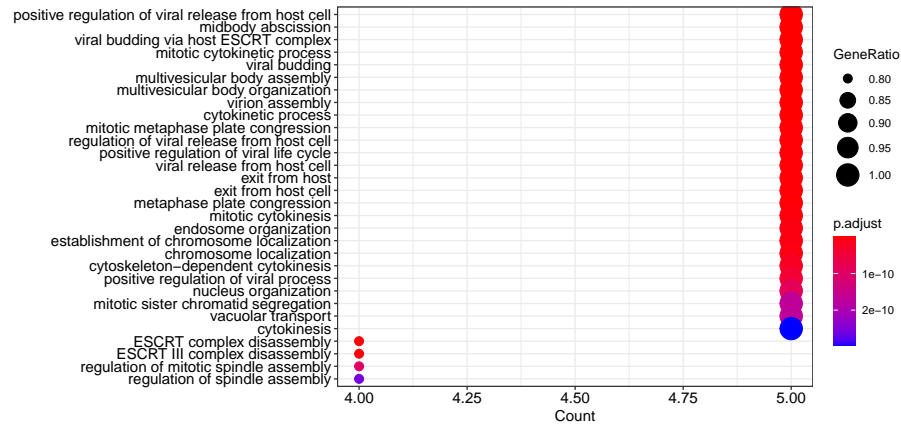
Estas estructuras conforman las balsas lipídicas en las membranas celulares las cuales son zonas accedidas por ciertas proteínas del SARS-CoV2, en concreto la proteína **ORF9c**, la cual podría tratarse de una de las proteínas del coronavirus humano que adquiere un dominio de transmembrana mediante la interacción con estos anclajes y que podría estar asociada con el ataque hacia procesos de señalización inmunológicos. En recientes estudios se muestra cómo ORF9c interactúa con las proteínas de membrana y altera los procesos antivirales en líneas celulares epiteliales pulmonares.

Además se observa que estos componentes son incluso mecanismo de entrada directa en otros virus y se aprecian diferencias en cuanto a estos procedimientos en distintos tipos de coronavirus.



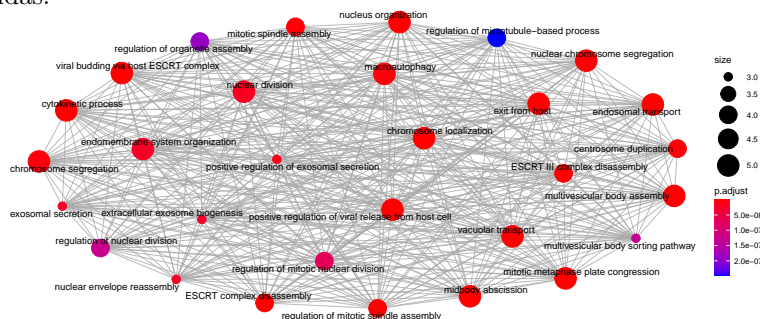
En este caso nuestro análisis de las comunidades proteicas nos ha llevado a concretar la importancia funcional que podría ayudar a estudiar más en profundidad la acción del SARS-CoV2 a partir de un nivel topológico.

3.3.4 Clúster 58



Este grupo de 5 proteínas es sin duda el más interesante para nuestro estudio. Sus funciones están altamente vinculadas a la actividad viral en nuestro organismo, lo que ya explica de entrada el motivo de su interacción con las proteínas del covid. Tras la detenida lectura de todas sus funciones se pueden extraer una serie de elementos fundamentales que aparecen en repetidas ocasiones. A continuación, muestro una breve explicación de las más relevantes:

- 1 **Liberación viral de la célula huésped:** Esta función biológica se asocia con la replicación de los virus en nuestro organismo, de tal forma que permite la salida de viriones de una célula infectada para promover la propagación del virus de célula a célula. Una vez ensambladas en el sitio de replicación, las partículas virales pueden liberarse por gemación, exocitosis, extrusión o lisis de la célula huésped. Algunos virus también median el transporte directo de su genoma viral a las células adyacentes gracias a las proteínas de movimiento.
- 2 **Incipiente viral:** Proceso viral mediante el cual los virus envueltos adquieren una membrana derivada del hospedador enriquecida en proteínas virales para formar su envoltura externa. El proceso comienza cuando las nucleocápsidas, ensambladas o en proceso de construcción, inducen la formación de una curvatura de la membrana en el plasma del huésped o la membrana del orgánulo y se envuelven en la yema en formación. El proceso termina cuando la yema es finalmente pellizcada por la escisión de la membrana para liberar la partícula envuelta en el espacio luminal o extracelular.
- 3 **Ensamblaje del cuerpo medio y citocinesis:** La citocinesis consiste en la separación física del citoplasma en dos células hijas durante la división celular. El cuerpo medio es una estructura transitoria que conecta a ambas células al final de este proceso con la función principal de localizar el sitio de abscisión que separa físicamente a las dos.
- 4 **Cuerpos multivesiculares:** Tipo especial de lisosoma recubierto por una membrana que contiene en su interior un variable número de pequeñas vesículas. Estos cuerpos también se denominan endosomas tardíos y son la antesala de la degradación de las moléculas endocitadas, la cual se realiza finalmente en los lisosomas gracias a unas enzimas denominadas hidrolasas ácidas.



Las fuertes conexiones entre estos conceptos y la interacción del covid en nuestro organismo son evidentes. Además, cabe destacar el hecho de que las funciones de este clúster presentan grandes vínculos entre sí. Son funciones extremadamente relacionadas que comparten un mismo fin o que participan en localizaciones y lapsos de tiempo muy cercanos. Es por esto que la red presentada en la figura anterior mues-

tra una gran cantidad de enlaces donde todos los nodos poseen un grado bastante elevado.

3.4 Búsqueda en Drugbank

La búsqueda en drugbank arroja los siguientes resultados:

```

1
2 IDs de los medicamentos cuyos targets se encuentran entre las
3 prote nas m s conectadas de la red DB00753; DB01189 DB01133
4 DB03147 DB05015; DB06603 DB00175; DB00227; DB00277; DB00313;
5 DB00641; DB01076; DB01095; DB01223; DB01303; DB02546; DB06176;
6 DB09091 DB00157 DB12010 DB00570

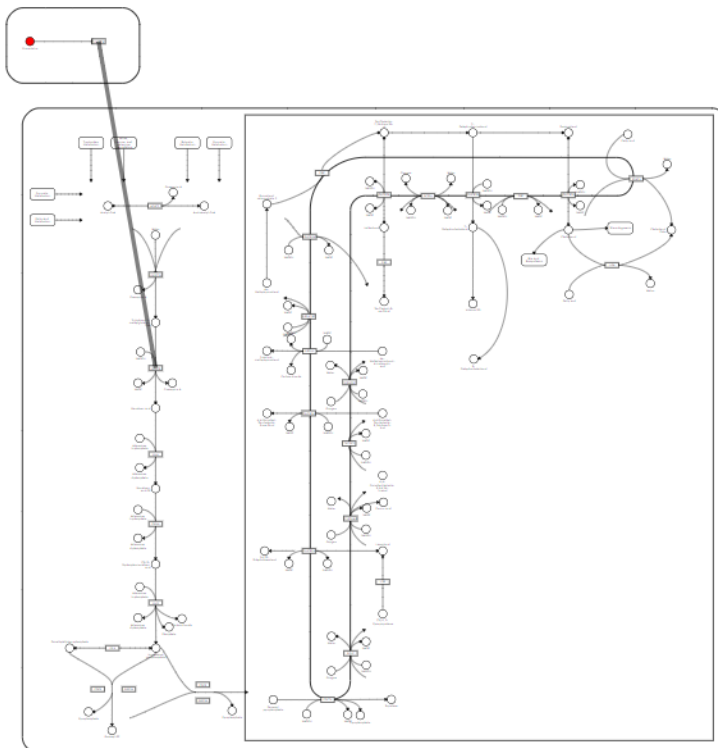
```

A continuación, analizaremos algunos de ellos.

- Enfermedades cardiovasculares

3.4.1 Pravastatin, Lovastatin

La pravastatina es un medicamento que se suele usar para tratar y prevenir los eventos coronarios, disminuyendo la mortalidad en caso de que estos se den.



Ruta metabólica de la pravastatina

Existen evidencias de que el SARS-Cov-2 puede afectar al corazón[1], agravando la condición de los pacientes y dejando secuelas que podrían ser de por vida, por lo que el tratamientos de estos eventos coronarios y su prevención podría ser clave para reducir la mortalidad y mejorar la calidad de vida del enfermo.

3.4.2 Atorvastatin, Rosuvastatin, Fluvastatin, Simvastatin

Ligand	Affinity [kcal/mol]	No. of interactions	Binding residues	Interaction with 6LU7
Rosuvastatin	-7.7	7	GLU 166 MET 165 GLU 166 CYS 145 THR 26 GLN 192 CYS 145	
Pravastatin	-6.6	4	THR 190 GLU 166 LEU 141 GLU 166	
Pitavastatin	-8.2	2	ARG 188 GLU 166	
Lovastatin	-7.4	2	HIS 163 HIS 164	
Fluvastatin	-7.7	2	GLU 166 HIS 163	
Atorvastatin	-6.8	3	GLN 189 GLU 166 GLN 189	

Tabla de afinidad de las estatinas tomada de [2]

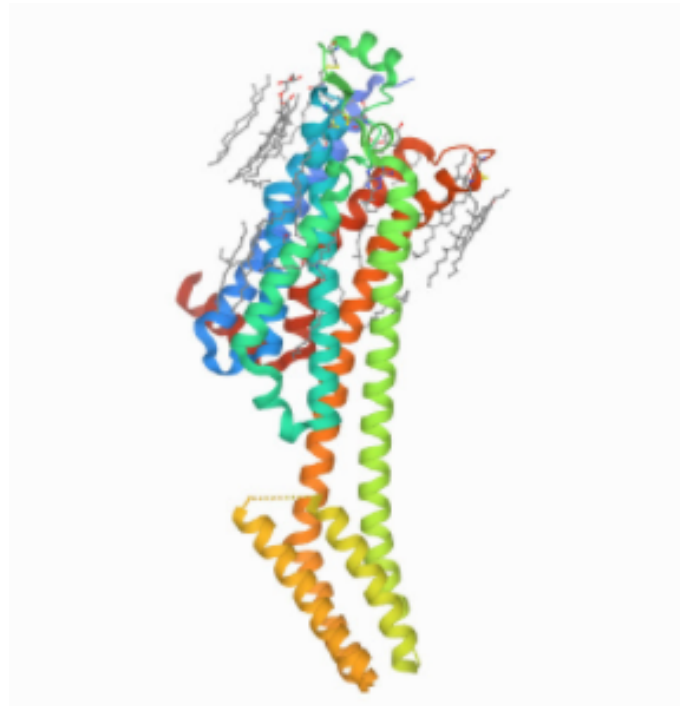
Las **estatinas** o **statins** son una clase de medicamento que se emplean para el tratamiento de la dislipemia, concentración elevada lípidos o baja en lipoproteínas. En concreto, es son medicamentos muy usados para tratar las enfermedades cardiovasculares (ECV) como la aterosclerosis o la angina, puesto que los niveles elevados de colesterol (LDL) suponen un gran riesgo en el desarrollo de estas enfermedades, y este tipo de medicamentos dan lugar a una reducción significativa de los niveles de LDL.

Además, en el artículo *'Statins and the COVID-19 main protease: in silico evidence on direct interaction'* se llegó a la conclusión de que las estatinas podrían ser inhibidores eficaces del SARS-CoV-2. Por tanto al igual que la pravastatina, cualquiera de estos medicamentos podrían ser útiles para la lucha contra el Covid-19, y más específicamente, contra los efectos cardíacos que provoca.

- Enfermedades respiratorias

3.4.3 Theophylline, Oxtriphylline

Medicamentos como la teofilina o oxtrifilina están indicados para tratar los síntomas de asma y bronquitis, entre otras patologías. Dado que comparten algunos síntomas en común con nuestro caso de estudio, este tipo de medicamentos podrían ser útiles para su tratamiento.



Vista 3D del receptor A2A adenosina

En el caso de la oxtrifilina, se trata de un broncodilatador cuyo mecanismo de acción implica al recepto A2A adenosina, que ya se había propuesto como sujeto de estudio. [3]

3.4.4 Aminophylline

La aminofilina es un fármaco formado a partir de teofilina y etilendiamina que es inyectado por vía intravenosa y tiene una gran aplicación para el tratamiento de **edemas pulmonares** (exceso de líquido en los pulmones). Su mecanismo de acción consiste en libera la teofilina que actúa relajando el músculo liso (broncodilatación) y suprimiendo las respuestas de las vías respiratorias a los estímulos. Estos resultados sobre el organismo permiten el control de los síntomas y un mejor funcionamiento del sistema respiratorio.

Sin embargo, según estudios recientes, este medicamento provoca diversos efectos adversos provocados por la inhibición de PDE III y el receptor antagonista de la adenosina, que dan lugar a dolores de cabeza, mareos, vómitos, arritmias, etc. Por tanto, se sugiere el empleo de otros broncodilatadores como los anteriormente mencionados.

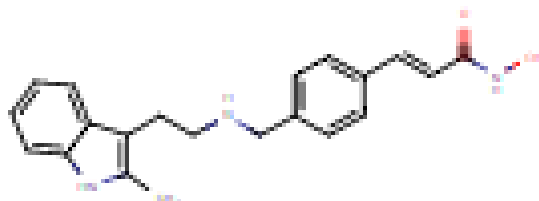
- Tratamiento de tumores

3.5 Panobinostat

Además del panobiostat, la búsqueda en drugbank muestra varios medicamentos más cuya función principal es el tratamiento de diversos tipos de cánceres.

En este caso, inhibe a las histonas DACs, que son responsables de diversos procesos celulares, entre ellos expresión génica y metabolismo de las proteínas.

Tal y como se describe en [4] *'Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A'*, sería interesante ver como podría aplicarse la acción de este tipo de fármacos al tratamiento del SARS-CoV-2.



Estructura molecular del panbiostat

3.6 Belinostat

El belinostat es un fármaco huérfano relativamente "joven" que inhibe la enzima histona desacetilasa (HDAC), atacando a **linfoma de células T periféricas** que ha recidido o que no ha podido ser tratado con otro tratamiento. Actúa bloqueando la multiplicación celular y destruyendo células cancerosas. Además, previene el crecimiento de los vasos sanguíneos, impidiendo así el crecimiento del tumor.

- Sedación e intubación

3.7 Isoflurane, Desflurane

Normalmente, los pacientes gravemente infectados son sedados e intubados, pero todavía no se están empleando anestésicos volátiles para ello. Sin embargo, se han realizado varios estudios que sugieren que la administración de anestésicos volátiles como el isoflurano o desflurano a pacientes con síndrome de dificultad respiratoria aguda (SDRA), mejora significativamente la oxigenación y el control de la sedación profunda y rápida. Es más, en el estudio *'Sevoflurane, a sigh of relief in COVID-19?'* ha obtenido como resultados que la mortalidad se reduce en un 35 por ciento (confianza del 95 por ciento: 0,18-0,68) en pacientes sedados con isoflurano, en comparación con propofolmidazolam.

Estos beneficios potenciales se deben principalmente a que los anestésicos volátiles presentan propiedades inmunomoduladoras que reducen las señales de muerte y, por tanto, la respuesta inflamatoria. Por tanto, el uso de este tipo de fármacos cuyos targets son algunas de las proteínas de nuestra red, podrían contribuir notablemente a la reducción de la lesión pulmonar.

4 Discusión

En este estudio hemos partido de la red de interacciones de proteínas humanas con el SARS-cov2 obtenida en *'A SARS-CoV-2 protein interaction map reveals targets for drug repurposing'* y la hemos **ampliado hasta obtener 425 interacciones**. Tras eliminar los nodos no conectados, hemos obtenido una componente conexa de 395 nodos. A partir de este grafo hemos aplicado diversas herramientas

con el objetivo de obtener más información acerca de su naturaleza, como se ha comentado en el apartado de resultados.

Mediante un **enriquecimiento funcional** de las 4 comunidades más centrales de nuestra red hemos podido conocer los principales procesos biológicos, funciones moleculares y componentes celulares a los que se asocian las proteínas estudiadas. Tal y como se comenta en [5], diversos procesos se relacionan con la actividad cardiovascular del organismo, llegando a provocar desordenes en la presión arterial o en el sistema renina-angiotensina-aldosterona. Esto también lo hemos podido comprobar al estudiar el clúster 84, pues provoca una regulación negativa del óxido nítrico, un compuesto que contribuye a la regulación de la presión arterial.

Otra función que hemos podido descubrir es la participación en la síntesis de lipoproteínas formando los complejos GPI, que permiten el ataque a los procesos de señalización inmunológica. De ahí podemos deducir la efectividad de este virus que es capaz de provocar una inmunodeficiencia severa. Además, la forma en que el virus se propaga en el organismo es particular del SARS-CoV-2.

Finalmente, gracias a la **búsqueda en Drugbank** no solo hemos podido analizar los medicamentos que se dirigen a las proteínas de nuestra red, sino también nos permite interpretar el mecanismo de acción de los mismos. Por ejemplo, la pravastina podría disminuir las probabilidades de mortalidad del paciente pues es un fármaco que previene el ataque del virus sobre el corazón, mientras que la teofilina podría ayudar a combatir la grave neumonía que provoca el virus en cuestión.

Además, hemos comprobado cómo los anestésicos volátiles pueden ser una nueva alternativa para la sedación e intubación de pacientes gravemente afectados por el Covid-19 y que efectos puede tener sobre la probabilidad de muerte de los mismos.

En general, los resultados expuestos en esta memoria representan un enfoque hacia el descubrimiento de propiedades funcionales de las proteínas humanas atacadas por el SARS-CoV-2 y de fármacos que podrían facilitar la lucha contra el virus y otras enfermedades similares.

5 Conclusiones

Tras contemplar y discutir todos los resultados obtenidos durante el desarrollo de nuestro estudio cabe mencionar la alta conectividad y **mantenimiento de coherencia** entre las distintas fases del proceso. Se puede apreciar como los módulos más importantes de nuestra red han sido los que finalmente han permitido descubrir las funciones moleculares y procesos biológicos más trascendentales respecto a la interacción del virus en nuestro organismo. Además, estas funciones fueron claramente confirmadas en apartados posteriores cuando el estudio llevado a cabo sobre DrugBank nos proporcionaba una serie de medicamentos cuyos objetivos se centraban en tratar muchas de las consecuencias que estas funciones podían acarrear sobre nuestra salud. Es decir, los descubrimientos llevados a cabo durante este trabajo han mantenido en todo momento una dirección muy definida que nos ha permitido realizar afirmaciones sólidas y cómodamente contrastadas.

Además, la conclusión más importante que cabe mencionar de este trabajo es sin duda la gran cantidad de información útil que nos puede ofrecer el estudio de un interactoma concreto cuando ha sido correctamente acotado y seleccionado bajo condiciones coherentes. En nuestro caso, estudiar la porción del **interactoma**

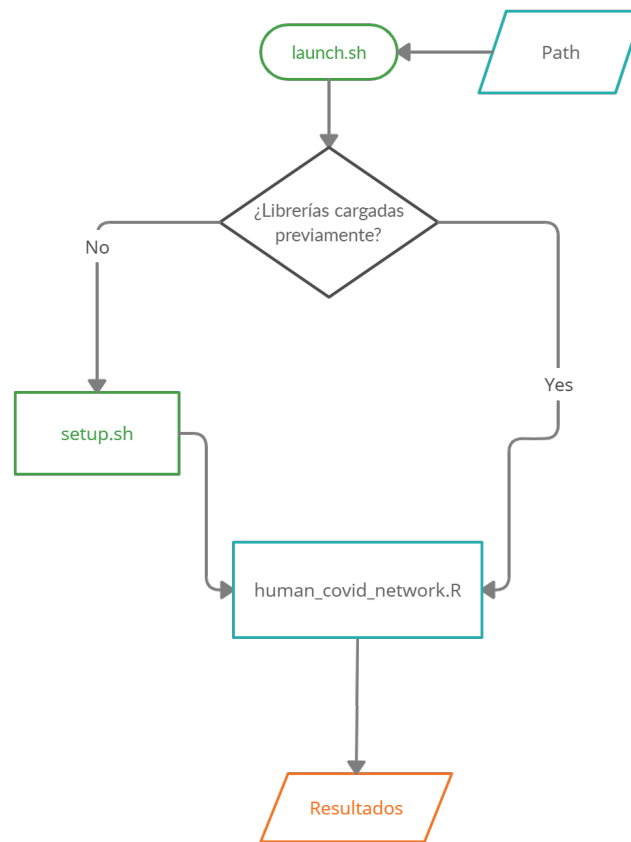
humano asociado al virus SARS-CoV-2, nos ha permitido conocer no sólo las funciones de las proteínas humanas, sino que a través de dicha información ser capaces de descubrir las posibles interacciones del virus con nuestro organismo, las posibles consecuencias clínicas de dicha interacción, características del propio virus y por último qué medicamentos pueden llegar a ser más efectivos para luchar contra el mismo. Es decir, la red estudiada en este trabajo ha conseguido actuar como una especie de espejo sobre el que hemos podido analizar el reflejo de las partes más esenciales de nuestro virus de interés.

Es evidente que cuantos más trozos de este espejo metafórico tengamos, mayor será la claridad y la precisión con la que podamos analizar el virus y por tanto mayor será el conocimiento que tengamos sobre el mismo. Con esto, queremos señalar la gran importancia que tiene seguir descubriendo nuevos elementos y conexiones del interactoma humano, ya que actualmente se afirma que nuestro conocimiento sobre el mismo es aproximadamente de un 10%. Ampliar esta gran red o incluso llegar a completarla nos permitiría sin lugar a dudas obtener una mayor cantidad de información con la que poder realizar análisis más profundos y efectivos. Esto también se extendería a estudios más concretos como el nuestro, ya que poseer el interactoma completo asociado al SARS-CoV-2 nos abriría las puertas a nuevos descubrimientos incluidos aquellos que nos permitan eliminar su actual amenaza.

6 Anexos

6.1 Flujo de trabajo en BASH

Hemos planteado el flujo de BASH de la siguiente manera: Tendremos un archivo principal *launch.sh* que se ejecuta desde la consola pasándole como parámetro la ruta de la carpeta donde se guardarán los resultados. Este comprobará si están instalados los paquetes necesarios para la ejecución del código. Si estos están instalados en la carpeta *software/deps-r* (Porque ya se haya ejecutado previamente el archivo *setup.sh*), se ejecutará directamente el archivo *human-covid-network.R*. La ejecución del archivo mostrará resultados numéricos, como el diámetro de la red, por pantalla y generará gráficos en formato PNG y PDF.



6.2 Ejecución

Es necesaria la ejecución desde un terminal Linux, ya que algunas de las librería que se han utilizado para realizar este trabajo solo pueden instalarse desde ahí. Además, deberemos disponer de la versión 4.0 o superior de R para poder acceder a la versión más actualizada de STRINGdb (Si se usa una versión anterior se generaran resultados distintos a los expuestos en esta memoria).

Abreviaciones**Disponibilidad de datos y materiales**

<https://github.com/Fiorellaps/project.template>

Contribución de los autores

S.C.L. : Encargada del Enriquecimiento funcional de los clústers más centrales de la red en R y en el documento y creación de launch.sh; F.P.S: Escritura de abstract, introducción, resultados (Linked Communities), discusión y obtención de LC en R; A.S.O: Búsqueda de la red de interacción PPI, escritura de resultados (Enriquecimiento Funcional) y conclusión.; MR.V:P: Búsqueda de medicamentos en Drugbank, análisis inicial y robustez, creación de setup en bash y escritura de resultados (Búsqueda de Drugbank y análisis inicial y robustez).

Author details

ETSI Informática, Universidad de Málaga, Málaga, España.

References

- Oudit GY, Kassiri Z, Jiang C, Liu PP, Poutanen SM, Penninger JM, et al. SARS-coronavirus modulation of myocardial ACE2 expression and inflammation in patients with SARS. *European Journal of Clinical Investigation*. 2009 7;39(7):618–625. Available from: <https://pubmed.ncbi.nlm.nih.gov/19453650/>.
- Reiner Hatamipour M, Banach M, Pirro M, Al-Rasadi K, Jamialahmadi T, et al. Statins and the Covid-19 main protease: In silico evidence on direct interaction. *Archives of Medical Science*. 2020;16(2):490–496. Available from: <https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC7212226/>; [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7212226/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7212226/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7212226/).
- Abouelkhair MA. Targeting adenosineric pathway and adenosine A2A receptor signaling for the treatment of COVID-19: A hypothesis. *Medical Hypotheses*. 2020 11;144:110012.
- White KM, Rosales R, Yildiz S, Kehrer T, Miorin L, Moreno E, et al. Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A. *Science (New York, NY)*. 2021 1; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33495306>.
- Gysi DM, Valle , Zitnik M, Ameli A, Gan X, Varol O, et al. Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19. *arXiv*. 2020 4; Available from: <http://arxiv.org/abs/2004.07229>.
- Nair P, Milan SJ, Rowe BH. Addition of intravenous aminophylline to inhaled beta 2 -agonists in adults with acute asthma . *Cochrane Database of Systematic Reviews*. 2012 12;.
- (PDF) Wrappers for Performance Enhancement and Oblivious Decision Graphs;. Available from: https://www.researchgate.net/publication/36368197_Wrappers_for_Performance_Enhancement_and_Oblivious_Decision_Graphs.
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020 7;583(7816):459–468. Available from: <https://doi.org/10.1038/s41586-020-2286-9>.
- Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Elsevier B.V.*; 2020. Available from: <https://pubmed.ncbi.nlm.nih.gov/32272173/>.
- Definición de belinostat - Diccionario de cáncer - Instituto Nacional del Cáncer;. Available from: <https://www.cancer.gov/espanol/publicaciones/diccionario/def/belinostat>.
- Li Z, Zhang XS, Wang RS, Liu H, Zhang S. Discovering link communities in complex networks by an integer programming model and a genetic algorithm. *PLoS ONE*. 2013 12;8(12):83739. Available from: www.plosone.org.
- DrugBank Online — Detailed Drug and Drug Target Information;. Available from: <https://go.drugbank.com/>.
- Schnepf E. From Prey via Endosymbiont to Plastid: Comparative Studies in Dinoflagellates. In: *Origins of Plastids*. Springer US; 1992. p. 53–76. Available from: https://link.springer.com/chapter/10.1007/978-1-4615-2818-0_5.
- ISSN — ;. Available from: <https://www.issn.org/>.
- Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2006 6;103(23):8577–8582. Available from: www.pnas.org/cgi/doi/10.1073/pnas.0601602103.
- L Margulis. *Origin of Eukaryotic Cells: Evidence and Research Implications for a Theory ...* - Lynn Margulis - Google Libros. Yale University Press, editor. New Haven; 1970. Available from: https://books.google.es/books/about/Origin_of_Eukaryotic_Cells.html?id=mrBzQgAACAAJ&redir_esc=y.
- Dominguez Andres A, Feng Y, Rosa Campos A, Yin J, Yang CC, James B, et al. SARS-CoV-2 ORF9c Is a Membrane-Associated Protein that Suppresses Antiviral Responses in Cells. *bioRxiv*. 2020 8;p. 2020.08.18.256776. Available from: <https://doi.org/10.1101/2020.08.18.256776>.
- Lu F. SARS-CoV-2 ORF9c: a mysterious membrane-anchored protein that regulates immune evasion? *Nature Reviews Immunology*. 2020 11;20(11):648. Available from: <https://www.nature.com/articles/s41577-020-00449-z>.
- Nieuwenhuijs-Moeke GJ, Jainandunsing JS, Struys MMRF. Sevoflurane, a sigh of relief in COVID-19? *Elsevier Ltd*; 2020. Available from: <https://www.who.int/docs/default->.
- Reiner Hatamipour M, Banach M, Pirro M, Al-Rasadi K, Jamialahmadi T, et al. Statins and the Covid-19 main protease: In silico evidence on direct interaction. *Archives of Medical Science*. 2020 4;16(2):490–496. Available from: <https://doi.org/10.5114/aoms.2020.94655>.
- Kalinka AT. The generation, visualization, and analysis of link communities in arbitrary networks with the R package linkcomm; 2020.
- Cooper GM. *The Origin and Evolution of Cells*. 2000; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK9841/>.

23. Schmidt N, Lareau CA, Keshishian H, Ganskih S, Schneider C, Hennig T, et al. The SARS-CoV-2 RNA–protein interactome in infected human cells. *Nature Microbiology*. 2020 12;p. 1–15. Available from: <https://doi.org/10.1038/s41564-020-00846-z>.
24. Hoffmann HH, Schneider WM, Sánchez-Rivera FJ, Luna JM, Ashbrook AW, Soto-Feliciano YM, et al. Title: Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. *bioRxiv*. 2020 9;p. 2020.09.11.291716. Available from: <https://doi.org/10.1101/2020.09.11.291716>.
25. Flinspach AN, Zacharowski K, Ioanna D, Adam EH. Volatile Isoflurane in Critically Ill Coronavirus Disease 2019 Patients—A Case Series and Systematic Review. *Critical Care Explorations*. 2020 10;2(10):e0256. Available from: <https://pubmed.ncbi.nlm.nih.gov/33134946/>.
26. Feijen EJP, Martens JA, Jacobs PA. Zeolites and their Mechanism of Synthesis. *Studies in Surface Science and Catalysis*. 1994 1;84(C):3–21.
27. ¿Cuáles son Statins?;. Available from: [https://www.news-medical.net/health/What-are-Statins-\(Spanish\).aspx](https://www.news-medical.net/health/What-are-Statins-(Spanish).aspx).
28. ¿En una crisis asmática la inyección de aminofilina agregada a los broncodilatadores mejora la función pulmonar y otros resultados o causa efectos perjudiciales? — Cochrane;. Available from: https://www.cochrane.org/es/CD002742/AIRWAYS_en-una-crisis-asmatica-la-inyeccion-de-aminofilina-agregada-los-broncodilatadores-mejora-la-funcion.
29. Koonin EV, Altschul SF, Bork P. ...functional motifs.... *Nature Genetics*. 1996;13(3):266–268. Available from: <https://pubmed.ncbi.nlm.nih.gov/8673121/>.