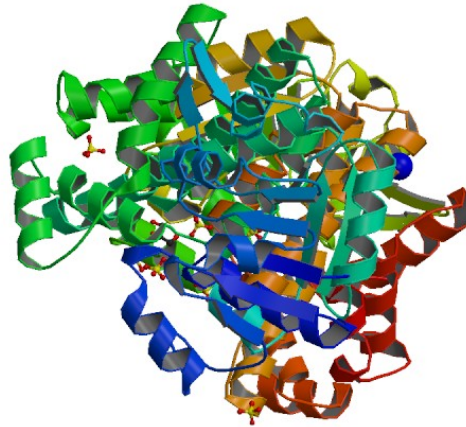


Proteínas moonlighting en la invasión del huésped y evasión de su Sistema Inmunitario por parte de bacterias, ¿son las mismas que las utilizadas por células metastásicas?



TRABAJO DE FIN DE MÁSTER

Autor: David Sánchez Redondo

Plan de Estudios del Estudiante: Máster de Bioestadística y Bioinformática

Área del trabajo final: Área 2. Bioinformática y proteómica en virulencia y enfermedades humanas

Director: Luis Franco Serrano

PRA: Marc Maceira Duch

Fecha: 24/06/2020

Universitat Oberta de Catalunya

Índice

0.- Ficha del trabajo final	7
1.- Introducción	8
1.1.- Contexto y justificación del Trabajo	8
1.1.1.- Función, regulación y otras características de las proteínas moonlighting.	9
1.1.2.- Proteínas moonlighting como factores de virulencia.	10
1.1.3.- Proteínas moonlighting en cáncer.	11
1.1.4.- Finalidad del Trabajo de Fin de Máster	11
1.2.- Recursos informáticos utilizados	11
1.3.- Objetivos del Trabajo	12
1.3.1.- Objetivos generales	12
1.3.2.- Objetivos específicos	12
1.4.- Enfoque y método seguido	12
1.5.- Planificación del Trabajo	12
1.5.1.- Tareas	12
1.5.2.- Hitos	13
1.6.- Sumario de productos obtenidos	13
1.7.- Descripción del resto de capítulos de la memoria	15
2.- Obtención de las listas de proteínas principales	15
2.1.- Proteínas involucradas en metástasis	15
2.2.- Proteínas moonlighting	15
3- Búsqueda de proteínas moonlighting que actúan como factores de virulencia	16
4.- Aplicación de filtros a la lista de proteínas involucradas en la metástasis	16
5.- Comparación de los códigos GO de las proteínas moonlighting y las listas filtradas	17
5.2.- Caracterización de los factores de virulencia en las distintas listas de resultados de la comparación entre proteínas moonlighting y listas filtradas de proteínas involucradas en metástasis	18
5.2.1.- Análisis de las proporciones de los factores de virulencia en las listas de resultados	19
5.3.- Presentación de los resultados de la comparación	20
6.- Análisis de los códigos GO compartidos entre proteínas involucradas en metástasis y factores de virulencia	23
6.1.- Unión a colágeno	28
6.2.- Unión a fibronectina	29
6.3.- Unión a laminina	32

7.- Conclusiones	33
7.1.- Descripción de las conclusiones del trabajo	33
7.2.- Valoración de los objetivos planteados inicialmente	34
7.3.- Valoración del seguimiento de la planificación	34
7.4.- Cuestiones a resolver	35
8.- Glosario	36
9.- Bibliografía	37



Esta obra está sujeta a una licencia de Reconocimiento-
NoComercial [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc/3.0/es/)

0.- Ficha del trabajo final

Título del trabajo: Proteínas moonlighting en la invasión del huésped y evasión de su Sistema Inmunitario por parte de bacterias, ¿son las mismas que las utilizadas por células metastásicas?

Nombre del autor: David Sánchez Redondo

Nombre del director/a: Luis Franco Serrano

Nombre del PRA: Marc Maceira Duch

Fecha de entrega (mm/aaaa): 06/2020

Titulación: Máster Universitario de Bioinformática y Bioestadística

Área del Trabajo Final: Área 2. Bioinformática y proteómica en virulencia y enfermedades humanas

Idioma del trabajo: Castellano

Palabras clave: Proteínas moonlighting, metástasis, matriz extracelular (o factor de virulencia)

Resumen del trabajo:

Tanto las bacterias patógenas como las células tumorales metastásicas luchan por diseminarse por el cuerpo humano y sobrevivir a la respuesta del Sistema Inmunitario. Al tener que superar los mismos obstáculos, el objetivo del estudio es encontrar los puntos comunes entre ambos procesos, partiendo de las proteínas moonlighting.

Las proteínas moonlighting son proteínas multifuncionales que en bacterias suelen actuar como factores de virulencia, por lo que las funciones asociadas a ellas están relacionadas con la infección y la supervivencia del patógeno. Se quiere comprobar qué funciones están compartidas en mayor medida entre proteínas moonlighting de virulencia y proteínas relacionadas con la metástasis.

Para realizar la comparación, las proteínas metastásicas se dividieron por procesos en las que estaban involucradas: metabolismo de glucosa, remodelación/regeneración de tejidos, interacción con plasminógeno e interacción con la matriz extracelular. Posteriormente, se compararon sus códigos GO con los de las proteínas moonlighting, diferenciando las que son factores de virulencia de las que no.

Los resultados obtenidos indican que los códigos GO relacionados con la unión a proteínas de la matriz extracelular, como el colágeno, fibronectina o laminina, están ampliamente representados entre las proteínas moonlighting de virulencia y proteínas metastásicas.

Son necesarios más estudios para comprobar si los sitios de unión que utilizan ambos tipos de proteínas son los mismos, si comparten secuencias de unión, descubrir nuevos fármacos bloqueadores de sitios de unión etc.

Abstract:

Pathogenic bacteria and metastatic tumor cells try to spread to new parts of the human body and survive to the immune response. Since both need to overcome the same obstacles, the aim of the study is to find common characteristics between both processes, starting from the so called moonlighting proteins.

Moonlighting proteins are multifunctional proteins and in bacteria are usually virulence factors. Therefore, functions associated with them are related to the infection and survival of the pathogen. It is wanted to identify the functions that are mostly shared between virulence moonlighting proteins and metastatic proteins.

In order to perform this comparison, metastatic proteins were divided according to the processes in which they were involved: glucose metabolism, remodeling/regeneration of tissues, interaction with plasminogen and interaction with extracellular matrix. Subsequently, their GO terms were compared to those of the moonlighting proteins, differentiating which of them were virulence factors.

The results obtained showed that GO terms related to the binding to extracellular matrix, such as collagen, fibronectin or laminin, are widely represented among virulence moonlighting proteins and metastatic proteins.

Further investigation is needed to prove if the binding sites used by both types of proteins are the same, if they share binding sequences or to discover new blocker drugs specific of these binding sites.

1.- Introducción

1.1.- Contexto y justificación del Trabajo

El genoma humano está compuesto por unos 20.000-25.000 genes [1], número muy inferior al número de funciones observadas en las células de nuestro organismo. Esta discrepancia es explicada en parte por el proceso de splicing alternativo, mediante el cuál un gen puede dar lugar a varias proteínas, que realizarían distintas funciones. Además, hay proteínas capaces de realizar múltiples funciones, relacionadas entre sí, las denominadas proteínas moonlighting o proteínas multitasking. Se tratan de un grupo de proteínas multifuncionales en las que una única cadena polipeptídica tiene la capacidad de realizar dos o más funciones distintas entre sí [2]. Mucho se desconoce sobre ellas, pero en los últimos años se están caracterizando nuevas proteínas moonlighting y se están relacionando con cada vez más procesos celulares y patologías. La función descubierta en primer lugar se denomina canónica y las demás moonlighting [3].

Piatigorsky y sus colaboradores descubrieron este tipo de proteínas en el año 1987, al demostrar que una determinada proteína del cristalino del pato, denominada cristalina delta2, era la misma que otra involucrada en el metabolismo, la arginosuccinato liasa [4,5]. La arginosuccinato liasa se trata de una enzima que está presente en casi todas las especies donde funciona en el ciclo de la urea y es necesaria para la biosíntesis del aminoácido arginina. Otros ejemplos de proteínas moonlighting son la cristalina epsilon, que se trata de la lactato deshidrogenasa, presente en algunas especies de reptiles y pájaros [5-7]. La cristalina zeta, presente en camellos, llamas, cerdos de guinea y ranas, que es la quinona oxidoreductasa [8-10], o la cristalina eta de las musarañas elefante, que es la aldehído deshidrogenasa [11].

El término moonlighting fue introducido por Jeffery, C.J., refiriéndose a aquellas proteínas multifuncionales cuya segunda función no derivase de una fusión génica [12]. En resumen, los casos que no se consideraría una proteína moonlighting serían los siguientes, aunque pueden variar entre distintos científicos dedicados a este campo:

- Proteínas obtenidas por el proceso de splicing alternativo o por proteólisis.
- Enzimas que tienen una gran cantidad de sustratos.
- Factores de transcripción que regulan la expresión de genes que participan en distintos procesos celulares.
- Enzimas que participan en distintas rutas metabólicas mediante la misma reacción.

Y los casos en los que sí se consideraría una proteína moonlighting:

- Cuando la proteína realiza distintas funciones dependiendo de las modificaciones post-traduccionales que sufra.
- Proteínas que se anclan a la membrana y realizan una función diferente a las que realizan en otros compartimentos celulares.

Una gran parte de las proteínas moonlighting se corresponden con aquellas que surgieron al principio de la evolución, como las involucradas en el metabolismo primario. Parece que el agrupamiento de distintas funciones en una misma proteína es una tendencia en los organismos, en lugar de generar nuevas proteínas para nuevas funciones [13]. El ejemplo más común de una proteína moonlighting en bacterias es una enzima, que cataliza una reacción bioquímica y además realiza otra función al ser secretada o unirse a otra proteína. Este es el caso de la enolasa de algunos patógenos, que actúa en la vía glucolítica en el citoplasma y como factor de virulencia al unirse al plasminógeno del huésped al ser secretada [14][15]. El plasminógeno también tiene un rol central en el inicio de la metástasis, ya que las células tumorales necesitan degradar la matriz extracelular para diseminarse [16].

Uno de los mayores subgrupos dentro de las proteínas moonlighting son aquellas enzimas o chaperonas intracelulares que realizan una función diferente en el exterior celular (ICSPs). A menudo son genes housekeeping, que están ampliamente extendidos a lo largo de la evolución y que están involucradas en glucólisis, ciclo del ácido cítrico, vía de las pentosas fosfato y metabolismo de proteínas y ácidos nucleicos. Este grupo de proteínas es muy común en bacterias, que utilizan proteínas moonlighting citosólicas en su superficie para formar y mantener interacciones con el huésped [17]. Muchas de ellas juegan un rol esencial en la infección, invasión, virulencia y formación de biofilms.

Cabe destacar las familias de proteínas “heat shock proteins” (HSPs), que han sido ampliamente estudiadas debido a su función como chaperonas ATP-dependientes, aunque muchas de ellas tienen importantes funciones en condiciones de baja concentración de ATP cuando son secretadas fuera de la célula. Se han descrito proteínas moonlighting dentro de las familias HSP60/HSP10, HSP70, HSP90, HSP100 y HSP110 [17]. Por ejemplo, la HSP60 HSP es una chaperona involucrada en el transporte de proteínas a la mitocondria, y además es un receptor en la superficie celular de lipoproteínas de alta afinidad debido a su alta afinidad por la apoA-II [18].

En los últimos años, las proteínas moonlighting están ganando importancia, y se han relacionado con procesos como mantenimiento del ADN, metabolismo, invasión de tejidos por patógenos, desarrollo de cáncer y otras enfermedades humanas, ser dianas de numerosos fármacos etc [19]. Por lo tanto, dilucidar cómo funcionan y cuáles son estas proteínas permitirá, entre otras cosas, tener un mejor entendimiento del origen de las proteínas, sobre el funcionamiento de los factores de virulencia de patógenos para poder utilizarlos como dianas de medicamentos, así como descubrir nuevas dianas para el tratamiento del cáncer.

1.1.1.- Función, regulación y otras características de las proteínas moonlighting.

Una cuestión que puede plantearse es si el fenómeno de las proteínas moonlighting es algo común o no, es decir, ¿es fácil que una proteína pueda realizar varias funciones?

El sitio catalítico de una enzima está formado por unos pocos aminoácidos con una determinada orientación espacial entre ellos, suponiendo un bajo porcentaje respecto al número total de aminoácidos de una proteína. Por lo tanto, es fácil que a lo largo de la evolución la secuencia que no forma el sitio catalítico cambie hasta adquirir nuevas funciones sin que se vea afectada la actividad catalítica primaria de la enzima. Ehinger y colaboradores encontraron que en el caso de la enolasa de *Streptococcus*, un pequeño motivo consistente en 9 aminoácidos, rico en lisina y residuos cargados negativamente era suficiente para unirse al plasminógeno del huésped. Los cambios en este sitio de unión no afectaban a la capacidad de la enolasa de catalizar su reacción, y además, la conformación de esta proteína no era diferente a otras enolasas que no unen plasminógeno [20]. Otro ejemplo es el de la fosfoglucoasa isomerasa (PGI), los aminoácidos que están altamente conservados entre las variantes de esta proteína en los distintos organismos son los que se agrupan alrededor del sitio activo de la enzima, es decir, los necesarios para su función en la glucólisis. Por otra parte, se observó que diferentes dominios de la superficie de esta proteína son sustancialmente diferentes entre distintas especies, no alterando su capacidad para catalizar su reacción enzimática [21].

En otros casos la proteína moonlighting experimenta un cambio conformacional para realizar su función moonlighting, como es el caso de la aconitasa, que cataliza una reacción en el Ciclo de Krebs y además, tras un cambio conformacional, adquiere la capacidad de unirse a moléculas de RNA [22].

En resumen, son muchos los estudios que apoyan la hipótesis de que es fácil que una proteína pueda albergar dos o más funciones en su secuencia, ya que suelen ser pocos los aminoácidos necesarios para realizar una determinada función. Además, existen otros mecanismos como pueden ser los cambios conformacionales que permiten a la proteína realizar otra función.

Una gran incógnita es cómo se regulan las distintas funciones de una proteína moonlighting, pese a no haber una respuesta clara, se conocen varios mecanismos:

- A veces realizan las distintas funciones a la vez.

- Diferentes compartimentos celulares para distintas funciones. El mecanismo mediante el cual una proteína citoplasmática es seleccionada y secretada al exterior o se ancla a la membrana plasmática sin péptido señal no es conocido [17].
- Las regulaciones post-traduccionales (PTMs) son importantes para la regulación de las funciones y la translocación de las proteínas a diferentes compartimentos celulares.
- Cambios en el ambiente o necesidades de la célula pueden provocar cambios en la proteína moonlighting, como situaciones de estrés celular.

Encontrar características comunes entre las proteínas moonlighting es otro de los objetivos a los que se dedican más esfuerzos en este campo de investigación, aunque quedan muchas cuestiones por resolver aún. No obstante, determinados factores suelen asociarse con la multifuncionalidad:

- Tener más de una localización celular.
- Estar presente tanto en el citoplasma como de manera extracelular al ser secretada.
- Tener una expresión diferencial en distintas situaciones.
- Capacidad para formar oligómeros.
- Poseer diferentes sitios de unión, lo que le permitirá interactuar con diferentes proteínas.

1.1.2.- Proteínas moonlighting como factores de virulencia.

Se han descrito muchas proteínas moonlighting citosólicas en bacterias que además actúan como adhesinas en la superficie de la membrana celular. No hay una respuesta clara para esto, una hipótesis es que como muchas de sus secuencias están altamente conservadas entre el patógeno y el huésped, el sistema inmunitario no genera una respuesta frente a ellas para evitar una reacción cruzada que provocaría una enfermedad autoinmune [23]. Otra incógnita sobre este proceso es que se desconoce como son seleccionadas y secretadas estas proteínas, puesto que no se ha detectado ninguna secuencia señal en ellas.

La primera enzima de este tipo que fue descubierta fue la GAPDH, que cataliza la conversión de D-gliceraldehído 3-fosfato a 3-fosfo-D-glicerol fosfato en la glucólisis dentro del citoplasma. Una de sus más de 20 funciones moonlighting descritas es que en la superficie celular actúa como un receptor para la transferrina [24].

Una misma proteína moonlighting puede tener diferentes funciones adicionales dependiendo de la especie que se trate. Este es el caso de la enolasa, que además de catalizar la reacción reversible de 2-fosfoglicerato a fosfoenolpiruvato en el citoplasma tiene diversas funciones adicionales aparte de funcionar como adhesina en la superficie de la bacteria [25]. Muchas de estas proteínas también han sido descritas como moduladores del sistema inmunitario y juegan un papel clave en la infección y virulencia de microorganismos patógenos [14,15].

Algunos ejemplos de proteínas moonlighting que permiten a la bacteria unirse a las células del huésped son la fructosa-1,6-bifosfato aldolasa de *Neisseria meningitidis* [26] y *Streptococcus pneumoniae* [27], o la chaperona Hsp60 de *Clostridium difficile* [28], *Helicobacter pylori* [29], *Chlamydia pneumoniae* [30], *Legionella pneumophila* [31] y otras especies.

Otras proteínas moonlighting se unen a elementos de la matriz extracelular del huésped o mucinas secretadas en la capa mucosa del intestino o vías aéreas. Las proteínas EF-Tu y piruvato deshidrogenasa de *Mycoplasma pneumoniae*, la malato sintasa de *Mycobacterium tuberculosis*, la autolisina AltA de *Streptococcus mutans*, la autolisina AltC de *Staphylococcus caprae* y la autolisina Aaa de *Staphylococcus aureus* se unen a uno o más componentes de la matriz extracelular como la fibronectina, laminina o el colágeno [32-36].

También existen proteínas moonlighting que actúan como inmunomoduladores, como la GAPDH de *Haemonchus contortus*, un nemátodo que infecta el tracto intestinal de ovejas y cabras, se une a la proteína C3

de la vía alternativa del complemento, inhibiendo la cascada del complemento y ayudando así al patógeno a evadir la respuesta Inmunitaria [37].

Una proteína que es utilizada por numerosas bacterias patógenas como diana de sus adhesinas es el plasminógeno, que es utilizado para invadir los tejidos del huésped. Cuando el plasminógeno se encuentra en la superficie del patógeno puede ser convertido a la forma activa de la proteasa, la plasmina. Esta activación puede llevarse a cabo por una proteasa propia del patógeno o tomando una del huésped, como el activador tisular del plasminógeno o el activador del plasminógeno tipo uroquinasa [16]. La plasmina activa degrada la matriz extracelular y la membrana basal, siendo utilizada por el patógeno para invadir nuevos tejidos. El organismo *Mycoplasma hyopneumoniae* utiliza la plasmina para degradar proteínas del huésped y obtener así aminoácidos esenciales que es incapaz de sintetizar [38,39].

1.1.3.- Proteínas moonlighting en cáncer.

Como ya se ha comentado, las proteínas moonlighting suelen realizar sus funciones en distintos compartimentos celulares o tras sufrir una modificación post-traducciona, si la traslocación de estas proteínas o su regulación se ve alterada podría favorecer la tumorigénesis o contribuir a un peor pronóstico en determinados tumores. El entendimiento de estos procesos puede resultar esencial para comprender la fisiopatología de los tumores, así como para la determinación de nuevas estrategias de tratamiento de distintos cánceres. Algunos ejemplos de proteínas moonlighting que están involucradas en tumores se recogen en una review de Kyung-Won M. y colaboradores [40]:

- Hsp90. Es una chaperona molecular, que ayuda al correcto plegamiento de proteínas en el citoplasma. Además, se ha observado que esta proteína es secretada y tiene un papel pro-tumorigénico. Se ha visto que altos niveles en suero de esta proteína está correlacionada con peor pronóstico en pacientes con cáncer [41]. Hsp90 estabiliza a la metaloproteasa de matriz-2 (MMP-2), lo cuál favorece la angiogénesis y la invasión tumoral. También interacciona con la E-cadherina en el cáncer de próstata, favoreciendo la transición epitelio-mesenquimal [42]. Hsp90 también se ha localizado en el núcleo [43]. Inhibidores de Hsp90 inducen la degradación de múltiples proteínas oncogénicas, como HER2/neu [44,45].
- Transglutaminasa 2. Se ha visto que TG2 activa la vía de señalización NF-KB favoreciendo señales anti-apoptóticas en cáncer de ovario [46].
- Otros ejemplos son las proteínas nucleares HMGB1, p53, ESE-1 o b-catenina y proteínas secretadas como varias metaloproteínas de la matriz extracelular o NAG-1.

1.1.4.- Finalidad del Trabajo de Fin de Máster

En la bibliografía hasta la fecha se pueden observar numerosos estudios sobre las proteínas moonlighting en patógenos y en cáncer, poniendo de manifiesto su rol en ellos. Lo innovador de este trabajo es que se va a intentar ver si hay alguna relación entre las proteínas moonlighting que actúan como factor de virulencia en patógenos y aquellas proteínas involucradas en la metástasis de células tumorales en humanos. El porqué de esta hipótesis es que tanto las bacterias patógenas como las células tumorales intentan diseminarse por el cuerpo humano, así como luchar contra el Sistema Inmunitario para no ser eliminadas, por lo que podrían utilizar mecanismos y/o proteínas parecidas para lograrlo. Como se ha descrito que un gran número de proteínas moonlighting de bacterias poseen una segunda función moonlighting como factor de virulencia, son un buen punto de partida para esta tarea.

1.2.- Recursos informáticos utilizados

Para visualizar y analizar los datos elegidos para el presente estudio se ha utilizado RStudio [47], así como para programar todo el código que se muestra en el resto del informe. El PDF del informe se ha generado utilizando Rmarkdown [48]. Además, se han utilizado los paquetes de R dplyr [49], knitr [50], kableExtra [51], data.table [52], stringr [53], plyr [54], readxl [55] y googledrive [56].

1.3.- Objetivos del Trabajo

1.3.1.- Objetivos generales

1. Comprobar si las proteínas moonlighting que actúan como factores de virulencia en bacterias son las mismas, o actúan mediante los mismos mecanismos, que aquellas utilizadas por las células metastásicas para invadir nuevos tejidos y evadir al Sistema Inmunitario.

1.3.2.- Objetivos específicos

1. Conocer si las proteínas obtenidas de la comparación están relacionadas con el metabolismo de la glucosa, el plasminógeno, remodelación de tejidos o la matriz extracelular.
2. Comprobar si las funciones moonlighting de las proteínas obtenidas de la comparación son las mismas en el ser humano que en la bacteria donde actúa como factor de virulencia.
3. En el caso de obtener proteínas moonlighting involucradas en la metástasis, estudiar su papel en los distintos tumores y comprobar si son diana de fármacos.

1.4.- Enfoque y método seguido

En la comparación de las proteínas relacionadas con la virulencia y las involucradas con la metástasis hay dos puntos clave que pueden ser afrontados de diversas maneras: la obtención de las distintas listas de proteínas y el modo de compararlas. Respecto a la primera se plantean las siguientes opciones:

- Llevar a cabo un análisis sin ningún tipo de filtro, lo cuál resultaría un trabajo demasiado amplio para el tiempo que se dispone en el Trabajo de Fin de Máster.
- Otro posible enfoque se centra en determinadas funciones, como la reparación/remodelación tisular, procesos clave para la diseminación tanto de bacterias como de células tumorales en el cuerpo humano. Se trataría de escoger aquellas proteínas dentro del listado de metástasis cuyas funciones estuvieran relacionadas con estos procesos y compararlas con el listado de proteínas moonlighting.

Por otra parte, la comparación se puede realizar de diversas formas:

- Comparar los interactomas, para ver si interactúan con proteínas parecidas.
- Comparar los códigos GO de las proteínas para ver si comparten segundas funciones.

La estrategia que se seguirá será la selección de determinadas proteínas relacionadas con la metástasis atendiendo a distintos filtros, como puede ser su papel en el metabolismo de la glucosa, reparación tisular... y compararlas con las proteínas moonlighting involucradas en la virulencia. Se elige la aplicación de filtros para hacer más tangible la comparación, ya que el tiempo disponible es limitado.

1.5.- Planificación del Trabajo

1.5.1.- Tareas

El listado de tareas a realizar y sus fechas adjudicadas se encuentra en el documento adjunto como material suplementario S4, realizado con Gantt Project [57]. Las tareas asociadas al TFM terminan el 25/05/2020, 15 días antes del plazo para entregar la memoria. Este lapso se ha reservado para posibles contratiempos durante el trabajo o por si surgen nuevos objetivos que merezcan ser estudiados. Además, dentro de los plazos de cada tarea, se ha añadido algún tiempo extra, tiempo contemplado para la realización de las memorias de las distintas PECs.

1.5.2.- Hitos

- Obtención de las dos listas de proteínas con la que se empezará el trabajo: proteínas cuya función moonlighting sea factor de virulencia y proteínas que estén involucradas en la metástasis.
- Obtención de códigos GO que permitan comparar estas proteínas más adelante.
- Aunque no sea un hito como tal, ya que se podría continuar el trabajo sin él, la información obtenida del análisis de los códigos GO de las dos listas principales de proteínas podría aportar información valiosa que repercutiría en distintos aspectos del trabajo, como podría ser la aplicación de nuevos filtros.
- Obtención de resultados tras la aplicación de los diversos filtros a la lista de proteínas relacionadas con metástasis, de lo contrario habría que cambiar la estrategia de estudio.
- Obtención de resultados tras las comparaciones de las distintas listas. Esto permitirá realizar posteriores análisis estadísticos y funcionales de las proteínas de interés.

1.6.- Sumario de productos obtenidos

Los resultados obtenidos a lo largo del trabajo están disponibles en Google Drive, pudiéndose acceder en el siguiente enlace: [Carpeta de resultados](#).

A continuación se detallan las carpetas que contiene, así como los archivos adjuntados:

1. **Listas primarias:** contiene las listas con las que se comenzó el trabajo.
 - **metastasis.xlsx:** lista de proteínas involucradas en metástasis. Material suplementario [S1](#).
 - **moonlighting proteins.xlsx:** lista de proteínas moonlighting. Material suplementario [S2](#).
 - **moons de virulencia listado.xlsx:** lista de proteínas moonlighting que son factores de virulencia. Material suplementario [S3](#).
2. **Filtros lista metastasis:** contiene las listas de proteínas relacionadas con la metástasis agrupadas por las funciones en las que están involucradas.
 - **filtro glucosa.xlsx:** proteínas relacionadas con el metabolismo de la glucosa. Material suplementario [S4](#).
 - **filtro matriz extracelular.xlsx:** proteínas relacionadas con la matriz extracelular. Material suplementario [S5](#).
 - **filtro plasminogeno.xlsx:** proteínas relacionadas con el plasminógeno. Material suplementario [S6](#).
 - **filtro tejido.xlsx:** proteínas relacionadas con la remodelación de tejidos. Material suplementario [S7](#).
 - **filtro totales.xlsx:** proteínas totales sin filtro. Material suplementario [S8](#).
3. **Codigos GO compartidos:** contiene los códigos GO compartidos entre las proteínas moonlighting y las distintas listas filtradas de proteínas relacionadas con la metástasis.
 - **glucosa GO compartido.xlsx:** códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en la metástasis relacionadas con el metabolismo de la glucosa. Material suplementario [S9](#).
 - **matriz extracelular GO compartido.xlsx:** códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en la metástasis relacionadas con la matriz extracelular. Material suplementario [S10](#).

- **plasminogeno GO compartido.xlsx:** códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en la metástasis relacionadas con el plasminógeno. Material suplementario [S11](#).
 - **tejido GO compartido.xlsx:** códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en la metástasis relacionadas con la remodelación de tejidos. Material suplementario [S12](#).
 - **totales GO compartido.xlsx:** códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en la metástasis. Material suplementario [S13](#).
4. **Factores virulencia:** contiene las listas de códigos GO compartidos y describe cuáles de las proteínas moonlighting son factores de virulencia, así como algunas de sus características.
- **glucosa virulencia.xlsx:** Material suplementario [S14](#).
 - **matriz extracelular virulencia.xlsx:** Material suplementario [S15](#).
 - **plasminogeno virulencia.xlsx:** Material suplementario [S16](#).
 - **tejido virulencia.xlsx:** Material suplementario [S17](#).
 - **totales virulencia.xlsx:** Material suplementario [S18](#).
5. **Listas comparacion definitivo:** contiene las listas definitivas de las comparaciones entre proteínas moonlighting y proteínas involucradas en metástasis. Hay una única entrada por cada proteína metastásica.
- **glucosa definitivo.xlsx:** Material suplementario [S19](#).
 - **matriz extracelular definitivo.xlsx:** Material suplementario [S20](#).
 - **plasminogeno definitivo.xlsx:** Material suplementario [S21](#).
 - **tejido definitivo.xlsx:** Material suplementario [S22](#).
 - **totales definitivo.xlsx:** Material suplementario [S23](#).
6. **Tablas comparativas GO:** Contiene las frecuencias con las que comparten cada código GO proteínas moonlighting y proteínas metastásicas, haciendo distinción entre factores de virulencia y proteínas moonlighting totales.
- **glucosa GO comparativa.xlsx:** Material suplementario [S24](#).
 - **matriz extracelular GO comparativa.xlsx:** Material suplementario [S25](#).
 - **plasminogen GO comparativa.xlsx:** Material suplementario [S26](#).
 - **tejido GO comparativa.xlsx:** Material suplementario [S27](#).
 - **totales GO comparativa.xlsx:** Material suplementario [S28](#).
 - **resumen GO comparativa.xlsx:** Esta tabla contiene los códigos GO más representados de cada lista anterior, así como sus definiciones. Material suplementario [S29](#).
7. **planificacion proyecto.pdf:** Contiene la planificación temporal y tareas del proyecto. Material suplementario [S30](#).
8. **Memoria final.Rmd:** archivo rmarkdown con el que se generó el PDF de la memoria del Trabajo de Fin de Máster.

1.7.- Descripción del resto de capítulos de la memoria

La memoria está organizada como el flujo de trabajo que se ha seguido, primero se detalla como se consiguieron las listas primarias de proteínas moonlighting y de metástasis (Apartado 2), seguidamente se caracterizan cuáles de las proteínas moonlighting actúan como factores de virulencia (Apartado 3) y se aplican filtros para agrupar por procesos en los que están involucradas las proteínas relacionadas con la metástasis (Apartado 4).

Una vez que se tienen las listas anteriores, se realiza la comparación de los códigos GO entre proteínas moonlighting, ya sean factores de virulencia o no, con las distintas listas de proteínas metastásicas filtradas (Apartado 5). Por último, se analizan los datos de la comparación (Apartado 6) y se detallan las conclusiones del trabajo (Apartado 7).

2.- Obtención de las listas de proteínas principales

2.1.- Proteínas involucradas en metástasis

Los genes involucrados en metástasis se obtuvo de la base de datos “Human Cancer Metastasis Database” (HCMBD) [58]. Se trata de una base de datos en la que se almacena información obtenida a partir de análisis de transcriptomas de células tumorales, con el fin de caracterizar genes diferencialmente expresados en células metastásicas.

Tras obtener el listado con los genes diferencialmente expresados en metástasis, se procesó para eliminar entradas repetidas y los micro-RNAs, ya que no son de interés en este estudio. Seguidamente, se utilizó la herramienta “Retrieve/ID mapping tool” [59] de UniProtKB para adquirir las proteínas codificadas por estos genes y sus códigos GO. En una primera búsqueda se obtuvieron 13958 resultados, correspondiendo casi 12000 de ellas a entradas sin revisar. Una vez descartadas quedaron un total de 1975 entradas, que se adjuntan en el material suplementario S1.

A continuación se describen las principales variables del listado de proteínas involucradas en metástasis:

- **Entry:** código Uniprot de cada proteína.
- **Protein names:** recoge los distintos nombres de la proteína.
- **Gene names:** recoge los distintos nombres del gen que codifica la proteína.
- **Organismo:** organismo al que pertenece la proteína. En este caso, como es obvio, todas pertenecen al ser humano.
- **Gene ontology (GO):** descripción de las funciones de la proteína.
- **Gene Ontology IDs:** recoge todos los códigos GO asociados a cada proteína. Esta variable será utilizada para la comparación de los códigos GO de estas proteínas con los de las proteínas moonlighting.

2.2.- Proteínas moonlighting

El listado de proteínas moonlighting se obtienen de la base de datos “Multitasking Proteins DataBase” [60]. En ella se almacenan proteínas moonlighting, entre las que se encuentran numerosos factores de virulencia, con diversas de sus características. La tabla que las describe se adjunta en el material suplementario S2. Las variables más representativas de la tabla son las siguientes:

- **Uni Prot:** código Uniprot de la proteína moonlighting.
- **Protein Name:** nombre de la proteína moonlighting.
- **Canonical Function:** función canónica de la proteína moonlighting, es decir, la función que fue descrita en primer lugar.
- **GO:** código GO de la función canónica.

- **Moonlighting Function:** función moonlighting de la proteína moonlighting.
- **GO Moon:** código GO de la función moonlighting. Estos serán utilizados para la comparación con los códigos GO de las proteínas involucradas en metástasis.
- **Organism:** organismo al que pertenece la proteína moonlighting.
- **Human_Disease :** enfermedad en la que se ha descrito una alteración de la proteína.
- **Drugs :** drogas que tienen como diana la proteína en cuestión.
- **PDB :** identificador de la proteína en Protein DataBank.

3- Búsqueda de proteínas moonlighting que actúan como factores de virulencia

Se procesa la lista de proteínas moonlighting, eliminando aquellos casos en los que existen NAs en la variable “GO Moon”, que contiene los códigos GO de la función moonlighting de las distintas proteínas. Estos serán utilizados más adelante para comprobar qué proteínas moonlighting comparten un código GO de algunas de sus funciones moonlighting con aquellos códigos GO de las proteínas involucradas en metástasis. También se eliminan los espacios de los nombres de las columnas para evitar complicaciones innecesarias.

Se intentó caracterizar qué proteínas eran factores de virulencia en la lista de “moonlighting_proteins” y apenas se obtuvieron resultados. Para realizarlo se buscaron las palabras “virulence factor” en la variable “GO_Moon”, pero esta función no está descrita explícitamente en muchos factores de virulencia.

Finalmente, esta tarea se hizo utilizando una lista que fue proporcionada por Franco-Serrano, L., co-autor de la base de datos de las proteínas moonlighting y tutor de este trabajo. En ella están anotadas manualmente las proteínas moonlighting de la lista que son factores de virulencia, que se adjunta en el material suplementario S3. Las variables de la tabla son las siguientes:

- **UNIPROT:** código Uniprot de la proteína moonlighting que actúa como factor de virulencia.
- **PROTEIN NAME:** nombre del factor de virulencia.
- **CANONICAL:** función canónica del factor de virulencia.
- **MOONLIGHTING:** función moonlighting del factor de virulencia.
- **ORGANISM:** organismo al que pertenece el factor de virulencia.

4.- Aplicación de filtros a la lista de proteínas involucradas en la metástasis

Para agrupar las diferentes proteínas de metástasis según los procesos en los que están involucradas se utilizan filtros. Se buscan, con la función grep, determinadas palabras en la variable “Gene_ontoly_(GO)” de la lista S1 relacionadas con alguna función. Un ejemplo de filtrado de aquellas proteínas relacionadas con el metabolismo de la glucosa sería el siguiente:

```
glucose <- metastasis[grep("glucose", metastasis$`Gene_ontology_(GO)`),]
glycolytic <- metastasis[grep("glycolytic", metastasis$`Gene_ontology_(GO)`),]
glycolysis <- metastasis[grep("glycolysis", metastasis$`Gene_ontology_(GO)`),]
todo_glucosa <- unique(rbind(glucose, glycolysis, glycolytic))
```

La variable “Entry” contiene los códigos Uniprot, “Gene_ontology_(GO)” contiene las funciones de las proteínas y “Gene_ontology_IDs” un resumen de todos los códigos GO de cada proteína.

Estas son las palabras utilizadas para cada filtro:

- **Proteínas totales:** no se utiliza ningún filtro. Esta lista se utilizará a modo de control en los análisis posteriores.
- **Metabolismo de la glucosa:** “glucose”, “glycolysis” y “glycolytic”. La razón de este filtro es que muchos factores de virulencia de bacterias son proteínas moonlighting cuya función canónica está relacionada con el metabolismo de la glucosa. La hipótesis que se plantea es que las proteínas relacionadas con el metabolismo de la glucosa también actúen como proteínas moonlighting en el ser humano y favorezcan la progresión de células tumorales.
- **Proteínas relacionadas con el plasminógeno:** “plasminogen”. En la bibliografía está ampliamente registrado que tanto bacterias patógenas como células tumorales utilizan la plasmina y otras proteasas para degradar la matriz extracelular y otras estructuras y diseminarse por el cuerpo humano.
- **Proteínas relacionadas con la matriz extracelular:** “extracellular matrix”. Como se ha explicado, la destrucción y/o remodelación del entorno celular es de vital importancia para la propagación de los patógenos o las células tumorales.
- **Proteínas relacionadas con la remodelación/reparación de tejidos:** “tissue remodeling” y “tissue regeneration”.

Las listas obtenidas al aplicar cada filtro descrito se adjuntan en el material suplementario S4-S8, que están contenidos en la carpeta [Filtros lista metastasis](#).

5.- Comparación de los códigos GO de las proteínas moonlighting y las listas filtradas

Para esta comparación se tomaron los códigos GO descritos en la variable “Gene_Ontology_IDs” de cada lista de proteínas involucradas en metástasis filtrada (S4-S8) y la variable “GO Moon” de la lista de proteínas de virulencia. Como se ha comentado en la introducción, las proteínas moonlighting que actúan como factores de virulencia suelen ser proteínas del metabolismo básico que a lo largo de la evolución han adquirido nuevas funciones además de su función canónica. Por lo tanto, si una proteína moonlighting está relacionada con la virulencia de una bacteria, su principal función no va a ser aquella por la cual es un factor de virulencia, sino que se tratará de una función moonlighting. Por ello se utilizan los códigos GO asociados a las funciones moonlighting.

La comparación se realiza con el siguiente código, que será explicado a continuación:

```
resultados <-
  setNames(data.frame(matrix(ncol = 4, nrow = 100000)),
           c("Entry", "Uni_prot", "GO_compartido", "Organismo"))
t <- 1
for (i in 1:nrow(matriz)){
  a <- matriz$Gene_ontology_IDs[i]
  c <- str_remove_all(a, "; ")
  d <- substring(text = c, seq(1, nchar(c)-1, 10), seq(10, nchar(c), 10))
  for(j in 1:length(d)){
    m <- grep(d[j], moonlighting_proteins$GO_Moon)
    if (length(m) > 0){
      for(k in 1:length(m)){
        resultados$Entry[t] <- matriz$Entry[i]
        resultados$Uni_prot[t] <- moonlighting_proteins$Uni_Prot[m[k]]
        resultados$GO_compartido[t] <- d[j]
        resultados$Organismo[t] <- moonlighting_proteins$Organism[m[k]]
        t<- t+1
      }
    }
  }
}
```

```

    else{
      next
    }
  }
}
resultados <- resultados[complete.cases(resultados),]

```

Al comienzo se crea una dataframe de resultados vacía, denominada “resultados”, cuyas variables son:

- **Entry:** recoge el código Uniprot de las proteínas involucradas en metástasis.
- **Uni_prot:** recoge los enlaces a Uniprot de las proteínas moonlighting.
- **GO_compartido:** código GO compartido entre la proteína de metástasis y la proteína moonlighting.
- **Organismo:** organismo al que pertenece la proteína moonlighting.

En el primer bucle “for” se incluye el nombre de alguna de las listas filtradas obtenidas anteriormente (en este caso la de la matriz extracelular), y recorrerá fila a fila la lista tomando la variable “Gene_ontology_IDs”. Como cada proteína posee numerosos códigos, es necesario separarlos e irlos utilizando uno a uno. Para ello se utilizan las funciones “str_remove_all” y “substring”, con las que se consigue un vector “d” que contiene todos los códigos GO de la variable “Gene_ontology_IDs” de la proteína involucrada en metástasis. El hecho de que los códigos GO en las listas filtradas estén separados siempre por “;” y que todos ellos tengan la misma longitud de caracteres ha permitido sistematizar la comparación.

En el segundo bucle “for” se recorren todos los códigos GO del vector d, y para cada uno de ellos se busca una coincidencia en la variable “GO Moon” de la lista de proteínas moonlighting. En la variable m se guardan las posiciones de esta última lista dónde ha habido alguna coincidencia.

Por último, si hay alguna coincidencia, es decir, si la longitud de m es mayor que 0, se guarda en la variable resultados los códigos Uniprot de la proteína involucrada en metástasis (en la variable “Entry”) y de la proteína moonlighting (en la variable “Uni_prot”), el código GO que comparten y el organismo al que pertenece la proteína moonlighting.

Los resultados obtenidos con esta función se adjuntan como material suplementario S9-S13, contenidos en la carpeta denominada [Codigos GO compartidos](#).

5.2.- Caracterización de los factores de virulencia en las distintas listas de resultados de la comparación entre proteínas moonlighting y listas filtradas de proteínas involucradas en metástasis

Una vez hechas las comparaciones, se añade la variable “virulence_factor” a las listas de resultados (S9-S13), siendo un 1 cuando la proteína moonlighting (variable “Uni_prot” de las listas) es un factor de virulencia y 0 cuando no lo es. La función es la siguiente:

```

for(i in 1:nrow(lista_resultados)){
  a <- as.character(lista_resultados[[3]][i])
  b <- strsplit(a, ";")[[1]]
  c <- paste(b[41:46], collapse = "")
  d <- grep(c, lista_virulencia[[2]])
  if(length(d)!=0){
    lista_virulencia[[5]][i] <- 1
    lista_virulencia[[6]][i] <- lista_virulencia[[4]][d]
    lista_virulencia[[7]][i] <- lista_virulencia[[5]][d]
  }
}

```

```

else{
  lista_virulencia[[5]][i] <- 0
}
}

```

En el bucle “for” se observan dos variables:

- **Lista_resultados** -> se trata de cualquiera de las listas resultados descritas en el apartado 5.1 (material suplementario S9-S13).
- **Lista_virulencia** -> es la lista adjuntada como material suplementario S3, descrita en el apartado 3, que describe las proteínas moonlighting que son factores de virulencia.

Primero se extrae el código Uniprot del enlace de la variable “Uni_prot” de las listas S9-S13. Por ejemplo, del enlace http://www.uniprot.org/uniprot/P05455%22target=%22_blank el código Uniprot sería P05455. Se pudo hacer de manera automática programando una función gracias a que los enlaces tienen una longitud fija, y el código Uniprot siempre está en la misma posición dentro de ellos. La función toma este código y busca alguna coincidencia en la variable “UNIPROT” de la lista S3, y si el código Uniprot está en ambas listas se añade un 1 a la variable “virulence_factor” de las listas de resultados, y si no se añade un 0. Además, si se trata de un factor de virulencia, se añade su función canónica y sus funciones moonlighting, tomándolas también de la lista S3.

Los resultados obtenidos se adjuntan como material suplementario S14-S18, contenidos en la carpeta [Factores virulencia](#).

5.2.1.- Análisis de las proporciones de los factores de virulencia en las listas de resultados

Dentro de las funciones compartidas, se cuantifica cuántas de ellas son compartidas con factores de virulencia. Se analizan las proporciones de factores de virulencia en las distintas listas (S14-S18) con una función, en la que se eliminan las repeticiones de la variable “Uni_prot” que contiene las proteínas moonlighting, ya que puede haber varias proteínas involucradas en metástasis que compartan un código GO con la misma proteína moonlighting. Seguidamente se divide el número de filas con un 1 en la variable “virulence_factor” por el número total de filas. La función utilizada es la siguiente:

```

proporcion_virulencia <- function(conjunto){
  a <- conjunto[unique(conjunto[[4]]),]
  nrow(a[a[[6]] == 1,])/nrow(a)*100
}

```

Cuadro 1: Análisis de los códigos GO compartidos

	Proporción de códigos GO compartidos con factores de virulencia (%)
Proteínas Totales	8.695652
Glucosa	16.829746
Tejido	21.487603
Matriz extracelular	28.157350
Plasminógeno	13.304721

Como se puede observar en el Cuadro 1, los factores de virulencia de las listas filtradas (S14-S17) comparten un porcentaje mayor de funciones con proteínas involucradas en la metástasis que aquellas sin filtrar (S18). Caben destacar los filtros de proteínas relacionadas con la matriz extracelular y el de proteínas relacionadas con la reparación/remodelación tisular, cuyos porcentajes superan el 20%, muy superiores al de proteínas

totales, que supone un 8,6%. Estos resultados parecen indicar que tanto bacterias patógenas como células tumorales metastásicas necesitan interactuar y modificar la matriz extracelular para su diseminación por el huésped. Si lo hacen utilizando las mismas proteínas, si se unen a las mismas dianas o si estas proteínas comparten motivos comunes son preguntas sin respuesta aún, se necesitarían más análisis para clarificarlo.

Otro aspecto a destacar es el también alto porcentaje de las proteínas relacionadas con el metabolismo de la glucosa (16,8%). Está descrito en la bibliografía que muchas proteínas del metabolismo básico actúan también como factores de virulencia en bacterias, y por otra parte, en células tumorales su expresión y función puede verse alterada, produciéndose el denominado efecto Warburg. El efecto Warburg consiste en un aumento en el consumo y metabolismo de la glucosa de manera anaerobia en el citosol de las células tumorales [61]. Por lo tanto, el alto porcentaje entre ambas listas podría ser debido a lo explicado anteriormente, pero también podría ser indicativo de que las células cancerosas utilizan proteínas del metabolismo de la glucosa de la misma manera que las bacterias patógenas para su diseminación o modulación del Sistema Inmunitario.

No obstante, lo más destacado de estos resultados es que proteínas relacionadas con la matriz extracelular y/o la remodelación tisular son claves en la diseminación tanto de células tumorales como de bacterias patógenas. Las potenciales implicaciones de este hecho son muchas:

- Que la fisiopatología de la colonización de tejidos por patógenos y células metastásicas sea parecido en algunas ocasiones.
- Que fármacos bloqueadores de sitios de unión a proteínas de la matriz extracelular puedan servir tanto para tratar infecciones como para prevenir metástasis en tumores.
- Caracterización de secuencias de unión a proteínas de la matriz extracelular utilizando estas proteínas. Posteriormente, estas secuencias podrían utilizarse para predecir potenciales factores de virulencia o proteínas que pueden favorecer la progresión de tumores.
- Entender como interaccionan estas proteínas con la matriz extracelular o como remodelan el entorno que las rodea puede ayudar a descubrir nuevas dianas terapéuticas, tanto en infecciones como en cáncer.
- Hay algún artículo que describe como infecciones bacterianas graves favorecen la metástasis. Mirar si tendría algo que ver con la remodelación de la matriz extracelular y la hipótesis “seed and soil”.
- Algunas de las proteínas involucradas en metástasis pueden ser también proteínas moonlighting, y que actúen de manera parecida a factores de virulencia de bacterias.

Como puede verse, son muchas las hipótesis que pueden plantearse, pero que por motivos de tiempo están fuera del alcance de este trabajo.

5.3.- Presentación de los resultados de la comparación

Con la siguiente función se transforman las listas de resultados (S14-S18) descritas en el apartado 5.2, de tal forma que hay una única entrada por cada proteína involucrada en metástasis de la variable “Entry”, y por cada proteína moonlighting que comparte función con ella hay 5 columnas, que describen el código GO compartido, código de Uniprot, función canónica, función moonlighting y organismo al que pertenece.

```
conversion_organismo <- function(datos1, nombre_fichero){  
  
  datos <- datos1[datos1$virulence_factor==1,]  
  
  datos$Entry <- as.character(datos$Entry)  
  datos$Uni_prot <- as.character(datos$Uni_prot)  
  datos$GO_compartido <- as.character(datos$GO_compartido)  
  datos$canonical <- as.character(datos$canonical)
```

```

datos$moonlighting <- as.character(datos$moonlighting)
datos$Organismo <- as.character(datos$Organismo)
for (i in 1:nrow(datos)){

  a <- as.character(datos[[2]][i])
  b <- strsplit(a,"")[[1]]
  c <- paste(b[41:46], collapse = "")
  datos[[2]][i] <- c

}
x <-c("Entry_1", "GO_compartido_1", "Proteina_moon_1",
      "canonical_1", "moonlighting_1", "Organismo_1")
resultado <- data.frame(character(10000), character(10000), character(10000),
                        character(10000), character(10000), character(10000),
                        stringsAsFactors = FALSE)
colnames(resultado) <- x
t <- 2
j <- 1

resultado$Entry_1[1] <- datos$Entry[1]
resultado$GO_compartido_1[1] <- datos$GO_compartido[1]
resultado$Proteina_moon_1[1] <- datos$Uni_prot[1]
resultado$canonical_1[1] <- datos$canonical[1]
resultado$moonlighting_1[1] <- datos$moonlighting[1]
resultado$Organismo_1[1] <- datos$Organismo[1]
resultado[resultado==""] <- NA

for (i in 2:nrow(datos)){

  if (datos$Entry[i] == datos$Entry[i-1]){

    if(is.na(resultado[,ncol(resultado)][j]) == TRUE){

      posicion_NAs <- which(is.na(resultado[j,]))
      primer_NA <- min(posicion_NAs)
      posicion_GO <- primer_NA
      posicion_uniprot <- primer_NA + 1
      posicion_canonical <- primer_NA + 2
      posicion_moonlighting <- primer_NA + 3
      posicion_organismo <- primer_NA + 4

      resultado[j, posicion_GO] <- datos$GO_compartido[i]
      resultado[j, posicion_uniprot] <- datos$Uni_prot[i]
      resultado[j, posicion_canonical] <- datos$canonical[i]
      resultado[j, posicion_moonlighting] <- datos$moonlighting[i]
      resultado[j, posicion_organismo] <- datos$Organismo[i]

    }

    else{

      q <- paste("GO_compartido",t,sep = "_" )
      w <- paste("Proteina_moon",t,sep = "_" )

```

```

e <- paste("canonical",t,sep = "_" )
r <- paste("moonlighting",t,sep = "_" )
o <- paste("Organismo", t, sep = "_")
t <- t+1
resultado <- resultado %>% mutate(!q := rep(NA,10000), !w := rep(NA,10000),
                                !e := rep(NA,10000), !r := rep(NA,10000),
                                !o := rep(NA,10000))

resultado[j, ncol(resultado)-4] <- datos$GO_compartido[i]
resultado[j, ncol(resultado)-3] <- datos$Uni_prot[i]
resultado[j, ncol(resultado)-2] <- datos$canonical[i]
resultado[j, ncol(resultado)-1] <- datos$moonlighting[i]
resultado[j, ncol(resultado)] <- datos$Organismo[i]

}
}

else{
  j <- j+1
  resultado$Entry_1[j] <- datos$Entry[i]
  resultado$GO_compartido_1[j] <- datos$GO_compartido[i]
  resultado$Proteina_moon_1[j] <- datos$Uni_prot[i]
  resultado$canonical_1[j] <- datos$canonical[i]
  resultado$moonlighting_1[j] <- datos$moonlighting[i]
  resultado$Organismo_1[j] <- datos$Organismo[i]
}
}
resultado <- resultado[is.na(resultado$Entry_1) == FALSE, ]
View(resultado)
write.csv(resultado,
file = paste("/home/david/Escritorio/Máster/TFM/listas con organismo/listas a enviar",
nombre_fichero, sep = "/"))
}

```

En primer lugar, se seleccionan las filas de la lista de resultados en las que hay un factor de virulencia, es decir, donde el valor de la variable “virulence_factor” sea 1. Para que sea más fácil de interpretar, se sustituye el enlace de la proteína moonlighting a la página Uniprot por su código Uniprot, extrayéndolo del propio enlace con el primer bucle “for”.

Seguidamente se crea una data frame vacía para almacenar los resultados reordenados, denominada “resultado”. Se comienza con 5 columnas:

- **Entry_1**: código Uniprot de la proteína involucrada en metástasis.
- **GO_compartido_1**: código GO compartido con la proteína moonlighting de virulencia.
- **Proteina_moon_1**: proteína moonlighting.
- **canonical_1**: función canónica de la proteína moonlighting.
- **moonlighting_1**: función moonlighting de la proteína moonlighting.
- **Organismo_1**: organismo al que pertenece el factor de virulencia.

Tras lo anterior se añade manualmente la primera fila a la data frame “resultado”, y el resto de filas se rellenan con NAs. Conforme se vaya avanzando en la lista que se introduce en la función vamos a tener dos posibles escenarios:

1. Que el código Uniprot de la proteína relacionada con la metástasis (variable "Entry") sea la primera vez que aparece, por lo que hay que ponerla en una nueva fila en la data frame "resultado". Para ello,

se inicializa un contador "j", al que se le suma 1 cada vez que aparece un código Uniprot nuevo en la variable "Entry" de la lista de resultados.

2. Que el código Uniprot de la variable "Entry" ya haya aparecido. Como la variable "Entry" de las listas obtenidas en el apartado 5.2 están ordenadas alfabéticamente, si se repite el nombre significa que es el mismo que el último descrito en la dataframe "resultado". En este caso hay, a su vez, dos posibles escenarios:
 - Que no haya columnas creadas previamente para poner los datos de la proteína moonlighting de virulencia con la que comparte un código GO nuestra proteína relacionada con la metástasis. Por lo tanto, habría que crear 5 nuevas variables: "GO compartido X", "Proteína moon X", "canonical X", "moonlighting X", "Organismo X", donde X representa un número. Para numerar las distintas columnas se utiliza el contador "t", que comienza por 2. A este contador se le suma 1 cada vez que se crean columnas nuevas.
 - Que sí haya columnas creadas. Esto se detecta si en la fila j hay algún NA, si lo hay es que hay variables disponibles para rellenar. Para poner los resultados en las columnas adyacentes que sí tienen datos, se toma la posición del primer NA de la fila. Como el orden de las variables creadas siempre es el mismo (GO compartido X - Proteína moon X - canonical X - moonlighting X - Organismo X), a partir de la posición del primer NA, que correspondería a GO compartido X se saca la posición de las demás variables.

Los resultados obtenidos con esta función se adjuntan como material suplementario S19-S23, contenidos en la carpeta [Listas comparacion definitivo](#).

6.- Análisis de los códigos GO compartidos entre proteínas involucradas en metástasis y factores de virulencia

Para realizar esta tarea se han utilizado las listas S14-S18, que describen los códigos GO compartidos entre proteínas moonlighting y proteínas involucradas en metástasis. Con la siguiente función se calcula, en las distintas listas por separado, el número de veces que aparece repetido cada código GO, dividiendo por el número total de filas, representándose la frecuencia de cada código en porcentaje. La frecuencia de cada GO se calcula dos veces, una teniendo en cuenta solamente a los factores de virulencia y otra contando con todas las proteínas moonlighting. De esta manera se obtiene una tabla en la que para cada código GO hay dos frecuencias: una entre las proteínas moonlighting que son factores de virulencia y otra entre las proteínas moonlighting totales.

```
creador_tablas <- function(lista_virulencia, nombre_fichero){

  x <-count(lista_virulencia[lista_virulencia$virulence_factor==1,], 'GO_compartido')
  x_sorted<- x[order(-x$freq),]
  y <- count(lista_virulencia, "GO_compartido")
  y_sorted <- y[order(-y$freq),]
  frecuencias <- setNames(data.frame(matrix(ncol = 3, nrow = 10000)), c("GO_compartido",
    "Frecuencia en lista total", "Frecuencia en lista con factores de virulencia"))
  o <- 1
  x_sorted$GO_compartido <- as.character(x_sorted$GO_compartido)

  for (i in 1:nrow(x_sorted)){

    frecuencias$GO_compartido[o] <- x_sorted$GO_compartido[i]
    frecuencias$'Frecuencia en lista con factores de virulencia'[o] <-
      x_sorted$freq[i]/sum(x_sorted$freq)*100
```

```

m <- grep(x_sorted$GO_compartido[i], y_sorted$GO_compartido)
frecuencias$'Frecuencia en lista total'[o] <-
  y_sorted$freq[m]/sum(y_sorted$freq)*100
o <- o+1
}

frecuencias <- frecuencias[complete.cases(frecuencias),]
frecuencias$'Frecuencia en lista total' <-
  round(frecuencias$'Frecuencia en lista total',2)
frecuencias$'Frecuencia en lista con factores de virulencia' <-
  round(frecuencias$'Frecuencia en lista con factores de virulencia',2)
write.csv(frecuencias,
file = paste("/home/david/Escritorio/Máster/TFM/Memoria_final/Tablas_comparativas_GO/",
nombre_fichero, sep = "/"))
}

```

Los resultados se adjuntan como material suplementario S24-S28, contenidos en la carpeta [Tablas comparativas GO](#).

Para ver si las diferencias entre los porcentajes descritos anteriorente son significativamente diferentes se utiliza la función de R “prop.test”, que utiliza la prueba de Chi-cuadrado. La función utilizada para obtener el p.value del test estadístico es la siguiente:

```

test_freq <- function(lista, GO){
  virulencia <- count(lista[lista$virulence_factor == 1,], "GO_compartido")
  totales <- count(lista, "GO_compartido")
  a <- totales[totales$GO_compartido == GO,]
  b <- virulencia[virulencia$GO_compartido == GO,]
  filas_totales <- nrow(lista)
  filas_virulencia <- nrow(lista[lista$virulence_factor==1,])
  x <- prop.test(x = c(a$freq,b$freq), n = c(filas_totales, filas_virulencia),
                alternative = "less")
  return(x$p.value)
}

```

La función toma dos variables, “lista”, que corresponde a alguna lista que describe los códigos GO compartidos entre proteínas moonlighting y proteínas metastásicas (S14-S18) y “GO”, que será un determinado código GO. Al comienzo, se obtienen las frecuencias de cada código GO teniendo en cuenta solo los factores de virulencia (“virulencia”) y también de la lista total(“totales”). Seguidamente se extrae la frecuencia del código GO introducido en las listas “totales” y “virulencia”, y serán utilizados en el test estadístico. El número de casos en las dos listas se obtiene con la función “nrow”.

A continuación, se representan los códigos GO que están más representados entre los factores de virulencia de las diferentes listas filtradas:

Cuadro 2: Representación de los Códigos GO compartidos

GO_compartido	Frecuencia en lista total (%)	Frecuencia en lista con factores de virulencia (%)	Lista
GO:0005102	6.93	35.56	Totales
GO:0007155	6.02	31.66	Totales

Cuadro 2: Representación de los Códigos GO compartidos (*cont-
nued*)

GO_compartido	Frecuencia en lista total (%)	Frecuencia en lista con factores de virulencia (%)	Lista
GO:0019899	3.45	10.22	Totales
GO:0002020	0.72	5.12	Totales
GO:0001968	0.39	3.06	Totales
GO:0005518	0.22	1.96	Totales
GO:0050840	0.23	1.89	Totales
GO:0043236	0.15	1.32	Totales
GO:0030246	0.29	0.96	Totales
GO:0005102	27.78	54.17	Plasminogeno
GO:0002020	4.12	11.11	Plasminogeno
GO:0007155	5.56	11.11	Plasminogeno
GO:0001968	3.29	9.72	Plasminogeno
GO:0043236	1.44	4.86	Plasminogeno
GO:0019899	3.7	4.17	Plasminogeno
GO:0005539	1.03	1.39	Plasminogeno
GO:0008201	0.82	1.39	Plasminogeno
GO:0043394	0.41	0.69	Plasminogeno
GO:0005102	6.03	31.71	Tejido
GO:0007155	5.43	29.27	Tejido
GO:0019899	4.82	14.63	Tejido
GO:0002020	1.34	9.76	Tejido
GO:0005518	0.33	3.05	Tejido
GO:0030246	0.54	1.83	Tejido
GO:0003779	0.67	1.22	Tejido
GO:0005539	0.33	1.22	Tejido
GO:0008201	0.27	1.22	Tejido
GO:0043237	0.13	1.22	Tejido
GO:0001819	0.27	0.61	Tejido
GO:0004888	0.2	0.61	Tejido
GO:0008289	0.2	0.61	Tejido
GO:0043067	0.47	0.61	Tejido
GO:0043208	0.13	0.61	Tejido
GO:0050663	0.07	0.61	Tejido
GO:0051015	0.07	0.61	Tejido
GO:0007155	16.69	37.16	Matriz
GO:0005102	11.55	25.07	Matriz
GO:0002020	2.57	7.71	Matriz
GO:0001968	2.24	7.36	Matriz
GO:0050840	1.63	5.7	Matriz
GO:0005518	1.28	4.82	Matriz
GO:0043236	0.82	3.07	Matriz
GO:0019899	1.89	2.37	Matriz
GO:0008201	0.63	1.18	Matriz
GO:0030246	0.56	0.79	Matriz

Cuadro 2: Representación de los Códigos GO compartidos (*cont-
nued*)

GO_compartido	Frecuencia en lista total (%)	Frecuencia en lista con factores de virulencia (%)	Lista
GO:0005539	0.35	0.53	Matriz
GO:0044406	0.22	0.48	Matriz
GO:0004888	0.28	0.35	Matriz
GO:0043394	0.14	0.26	Matriz
GO:0032722	0.09	0.18	Matriz
GO:0043237	0.05	0.18	Matriz
GO:0005102	6.89	38.81	Glucosa
GO:0007155	3.38	19.54	Glucosa
GO:0019899	5.01	16.28	Glucosa
GO:0001968	0.89	7.6	Glucosa
GO:0002020	0.7	5.43	Glucosa
GO:0043236	0.19	1.9	Glucosa
GO:0044406	0.26	1.49	Glucosa
GO:0008201	0.17	0.81	Glucosa
GO:0008289	0.25	0.81	Glucosa
GO:0004888	0.17	0.54	Glucosa
GO:0006897	0.11	0.54	Glucosa
GO:0003779	0.21	0.41	Glucosa
GO:0030246	0.11	0.41	Glucosa
GO:0050715	0.08	0.41	Glucosa
GO:0005539	0.07	0.27	Glucosa
GO:0030247	0.1	0.27	Glucosa
GO:0032722	0.06	0.27	Glucosa
GO:0050714	0.06	0.27	Glucosa
GO:0001819	0.06	0.14	Glucosa

Atendiendo a los códigos GO más compartidos en la lista total sin filtros representados en el Cuadro 2, se tratan de [GO:0005102](#) y [GO:0007155](#), cuyas definiciones se encuentran en el Cuadro 3. Son códigos bastante inespecíficos, que describen la interacción no covalente con moléculas receptoras y la unión célula-célula o célula-sustrato, respectivamente. Los que les siguen, [GO:0019899](#) y [GO:0002020](#) están relacionados entre sí, el primero describe la interacción con alguna enzima y el segundo, que es algo más específico, la interacción con alguna proteasa. Como era de esperar, este último código está más representado entre los factores de virulencia en la lista filtrada del plasminógeno (11.1%) que en la lista total (5.12%), ya que el plasminógeno es una proteasa.

Los siguientes códigos de la lista, que son [GO:001968](#), [GO:0005518](#), [GO:0050840](#) y [GO:0043236](#), describen la interacción con diferentes componentes de la matriz extracelular, como el colágeno, fibronectina y laminina. Estos resultados indican el rol esencial de la matriz extracelular en la diseminación de patógenos y metástasis, posteriormente se hablará de ellos.

Hay algunos términos que no están tan representados entre los factores de virulencia a comparación de las proteínas moonlighting totales que están relacionados con la secreción de citoquinas, como [GO:0050715](#) y [GO:0032722](#). Podría ser que estas funciones estén relacionadas con la remodelación del entorno celular y/o con la modulación del Sistema Inmunitario, pero con los datos disponibles no pueden sacarse más conclusiones.

Cuadro 3: Definición de los Códigos GO compartidos

GO:0007155	The attachment of a cell, either to another cell or to an underlying substrate such as the extracellular matrix, via cell adhesion molecules.
GO:0019899	Interacting selectively and non-covalently with any enzyme.
GO:0002020	Interacting selectively and non-covalently with any protease or peptidase.
GO:0001968	Interacting selectively and non-covalently with a fibronectin, a group of related adhesive glycoproteins of high molecular weight found on the surface of animal cells, connective tissue matrices, and in extracellular fluids.
GO:0005518	Interacting selectively and non-covalently with collagen, a group of fibrous proteins of very high tensile strength that form the main component of connective tissue in animals. Collagen is highly enriched in glycine (some regions are 33 % glycine) and proline, occurring predominantly as 3-hydroxyproline (about 20 %)
GO:0050840	Interacting selectively and non-covalently with a component of the extracellular matrix
GO:0043236	Interacting selectively and non-covalently with laminins, glycoproteins that are major constituents of the basement membrane of cells
GO:0030246	Interacting selectively and non-covalently with any carbohydrate, which includes monosaccharides, oligosaccharides and polysaccharides as well as substances derived from monosaccharides by reduction of the carbonyl group (alditols), by oxidation of one or more hydroxy groups to afford the corresponding aldehydes, ketones, or carboxylic acids, or by replacement of one or more hydroxy group(s) by a hydrogen atom. Cyclitols are generally not regarded as carbohydrates.
GO:0005539	Interacting selectively and non-covalently with any glycan (polysaccharide) containing a substantial proportion of aminomonosaccharide residues.
GO:0008201	Interacting selectively and non-covalently with heparin, any member of a group of glycosaminoglycans found mainly as an intracellular component of mast cells and which consist predominantly of alternating alpha-(1->4)-linked D-galactose and N-acetyl-D-glucosamine-6-sulfate residues.
GO:0043394	Interacting selectively and non-covalently with a proteoglycan, any glycoprotein in which the carbohydrate units are glycosaminoglycans.
GO:0003779	Interacting selectively and non-covalently with monomeric or multimeric forms of actin, including actin filaments.
GO:0043237	Interacting selectively and non-covalently with laminin-1, a glycoprotein trimer with the subunit composition alpha1, beta1, gamma1.
GO:0001819	Any process that activates or increases the frequency, rate or extent of production of a cytokine.
GO:0004888	Combining with an extracellular or intracellular signal and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity or state as part of signal transduction.
GO:0008289	Interacting selectively and non-covalently with a lipid.
GO:0043067	Any process that modulates the frequency, rate or extent of programmed cell death, cell death resulting from activation of endogenous cellular processes.
GO:0043208	Interacting selectively and non-covalently with glycosphingolipid, a compound with residues of sphingoid and at least one monosaccharide.
GO:0050663	The regulated release of cytokines from a cell. Cytokines are any of a group of proteins that function to control the survival, growth and differentiation of tissues and cells, and which have autocrine and paracrine activity.
GO:0051015	Interacting selectively and non-covalently with an actin filament, also known as F-actin, a helical filamentous polymer of globular G-actin subunits.
GO:0044406	The attachment of a symbiont to its host via adhesion molecules, general stickiness etc., either directly or indirectly. The host is defined as the larger of the organisms involved in a symbiotic interaction.

GO:0032722	Any process that activates or increases the frequency, rate, or extent of chemokine production.
GO:0006897	A vesicle-mediated transport process in which cells take up external materials or membrane constituents by the invagination of a small region of the plasma membrane to form a new membrane-bounded vesicle.
GO:0050715	Any process that activates or increases the frequency, rate or extent of the regulated release of cytokines from a cell.
GO:0030247	Interacting selectively and non-covalently with any polysaccharide, a polymer of many (typically more than 10) monosaccharide residues linked glycosidically.
GO:0050714	Any process that activates or increases the frequency, rate or extent of the controlled release of a protein from a cell.

6.1.- Unión a colágeno

El colágeno es una proteína fibrilar, se trata del principal componente de la matriz extracelular y la proteína más abundante en mamíferos. Es esencial para el desarrollo de tejidos, su mantenimiento y su regeneración [62]. En el ser humano se han descrito más de 20 tipos de colágeno, distribuidos diferencialmente dependiendo del tipo de estructura y tejido del que forman parte.

Cuando las proteínas se desnaturalizan, debido a la acción de proteasas por ejemplo, se muestran nuevos epítopos que en su forma íntegra son inaccesibles. Este es el caso del epítipo HUI77 del colágeno, que contiene una secuencia de aminoácidos del tipo PGxPG, donde P es prolina, G es glicina y x puede ser cualquier otro aminoácido [63,64]. Se ha demostrado que utilizando anticuerpos específicos de este epítipo se inhibe el crecimiento y la metástasis de células de melanoma [65]. Este anticuerpo también ha demostrado que puede inhibir la angiogénesis y el crecimiento del cáncer de mama y ovario [64,66].

Se han descrito proteínas en bacterias que comparten motivos característicos del colágeno humano [67]. Algunas de estas proteínas son factores de virulencia, como Scl1 y Scl2 de *S.pyogenes*, descritas también en las tablas de resultados. Dependiendo del serotipo, el dominio V no relacionado con el colágeno de Scl1 puede unir lipoproteínas de alta densidad (LDL), factor H, CFHR1 o proteínas de la matriz extracelular como laminina o fibronectina. El dominio colágeno-like de Scl1 mimetiza al colágeno de los mamíferos, permitiéndole interactuar con integrinas (ligandos del colágeno) como a2b1 y a11b1 a través de la secuencia de unión GLPGER [68]. Otros ejemplos de estas proteínas son BclA y BclB de *Bacillus anthracis*, que forman filamentos en la superficie de la bacteria [69].

El hecho de que bacterias patógenas utilicen estas proteínas que comparten secuencias con otras del huésped puede tener distintas explicaciones:

- Evitar una respuesta del Sistema Inmunitario frente al factor de virulencia.
- Mimetizar proteínas del huésped y utilizar sus sitios de unión, como sería el caso de la unión de Scl1 de *S.pyogenes* a la integrina a2b1, que en condiciones normales une colágeno.

El código GO que describe la interacción con el colágeno es el [GO:0005518](#). En la lista filtrada de matriz extracelular aparece representado en un 4.82% entre los factores de virulencia y en un 1.28% en las proteínas moonlighting totales (p.value = 1.99e-26). En la lista filtrada de tejido aparece representado en un 3.05% entre los factores de virulencia y en un 0.33% en las proteínas moonlighting totales (p.value = 9.63e-05). Algunas de las proteínas que interactúan con el colágeno que aparecen en las listas de resultados se muestran en el Cuadro 4.

Cuadro 4: Ejemplos de proteínas que unen colágeno

Proteína	Organismo	Nombre de la proteína	Función
Q9Y653	Homo sapiens	Adhesion G-protein coupled receptor G1	Receptor involved in cell adhesion and probably in cell-cell interactions. Mediates cell matrix adhesion in developing neurons and hematopoietic stem cells. Receptor for collagen III/COL3A1 in the developing brain and involved in regulation of cortical development,
P16070	Homo sapiens	CD44	The protein encoded by this gene is a cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration. It is a receptor for hyaluronic acid (HA) and can also interact with other ligands, such as osteopontin, collagens, and matrix metalloproteinases (MMPs).
P07858	Homo sapiens	Cathepsin B	Thiol protease which is believed to participate in intracellular degradation and turnover of proteins
P84022	Homo sapiens	Mothers against decapentaplegic homolog 3	The protein encoded by this gene belongs to the SMAD. SMAD proteins are signal transducers and transcriptional modulators that mediate multiple signaling pathways. This protein functions as a transcriptional modulator activated by transforming growth factor-beta and is thought to play a role in the regulation of carcinogenesis
Q24803	Entamoeba histolytica	Aldehyde-alcohol dehydrogenase 2	Alcohol dehydrogenase. Binds fibronectin, laminin and type II collagen
A2DHT2	Trichomonas vaginalis	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis. Binds fibronectin, plasminogen and collagen.
Q70YI1	Legionella pneumophila	Outer membrane protein MIP	Essential virulence factor associated with macrophage infectivity. Exhibits PPIase activity.

Un posible ejemplo de proteína moonlighting involucrada en metástasis es P84022, ya que está involucrada en la transducción de señales y transcripción, y además, une colágeno. Una revisión de estas proteínas sería útil para ver cómo se comportan dependiendo de las distintas situaciones celulares, y en este caso en concreto, si la unión a colágeno es determinante en el desarrollo de la metástasis.

6.2.- Unión a fibronectina

La fibronectina es una glicoproteína con múltiples dominios, que se encuentra ampliamente distribuida por fluidos y matrices extracelulares de tejidos animales y humanos. Esta tiene un rol importante en varios procesos biológicos, como la adhesión a la matriz extracelular, diferenciación, crecimiento y migración de células. En el ser humano existen dos formas de fibronectina, la soluble y la insoluble. La primera está presente en el plasma sanguíneo y es secretada por los hepatocitos. La segunda es uno de los mayores componentes de la matriz extracelular y membranas basales, es producida por fibroblastos y células endoteliales.

Alteraciones en la expresión, degradación u organización de la fibronectina se han relacionado con patologías como transformaciones oncológicas y fibrosis [70, 71]. También está ampliamente descrito en la bibliografía la adhesión de bacterias patógenas a la fibronectina, como las proteínas FnBPA y FnBPB de *Staphylococcus aureus* [72,73], SepB de *Streptococcus agalactiae* [74], Fnm de *Enterococcus faecium* [75] etc.

Se ha probado en modelos animales que utilizar proteínas que unen fibronectina como componentes de vacunas puede proporcionar inmunidad frente a la infección de patógenos. Un ejemplo es la vacunación de ratas con la proteína recombinante purificada EfbA de *Enterococcus faecalis*, con la que se consiguió protección frente a endocarditis y prevención de la infección por este patógeno [76].

Las proteínas de bacterias que unen fibronectina pueden ser o no ser moonlighting, determinararlo puede ser de gran ayuda a la hora de generar vacunas, ya que las proteínas moonlighting altamente conservadas entre especies no logran buenas respuestas vacunales. La prueba de ello es que no hay ninguna vacuna en el mercado basada en una proteína moonlighting. [23].

Una cuestión a resolver es si las células metastásicas se unen de igual manera a la fibronectina que las bacterias patógenas, ya que si fuera así, podrían utilizarse los mismos fármacos para ambas. Un ejemplo podría ser la heparina. Diversos estudios describen que la heparina inhibe la unión de adhesinas de *Staphylococcus* a la fibronectina de manera específica [77]. Este hallazgo es de gran interés en el ámbito de los implantes de prótesis clínicas, que estarán recubiertas de fibronectina al ser introducidas en el organismo y una de sus principales complicaciones es la infección por *Staphylococcus*. Esto explica las observaciones previas en las que el tratamiento con heparina de los biomateriales de las prótesis clínicas disminuían el riesgo de infección por *Staphylococcus aureus* [78]. Por otra parte, utilizando ratones como modelo, se observa una disminución del ratio de metástasis de células de tumor gástrico humano al administrar heparina N-desulfatada como tratamiento. Este tratamiento no afectó al crecimiento del tumor primario y no provocó efectos secundarios relacionados con el sangrado en los ratones [79]. El efecto anti-metastásico de la heparina podría deberse, por lo menos parcialmente, a la inhibición de la unión de las células tumorales a la fibronectina.

El código GO que describe la interacción con la fibronectina es el [GO:0001968](#). En la lista filtrada de matriz extracelular aparece representado en un 7.36 % entre los factores de virulencia y en un 2.24 % en las proteínas moonlighting totales (p.value = 6.87e-34). Estos resultados son consistentes con la bibliografía, se observa de manera significativa una mayor unión a la fibronectina por parte de los factores de virulencia que de las proteínas moonlighting totales. Algunas de las proteínas que interaccionan con la fibronectina que aparecen en las listas de resultados se muestran en el Cuadro 5.

Cuadro 5: Ejemplos de proteínas que unen fibronectina

Proteína	Organismo	Nombre de la proteína	Función
P29279	Homo sapiens	CCN family member 2	Major connective tissue mitogen secreted by vascular endothelial cells. Promotes proliferation and differentiation of chondrocytes. Mediates heparin- and divalent cation-dependent cell adhesion in many cell types including fibroblasts, myofibroblasts, endothelial and epithelial cells. Enhances fibroblast growth factor-induced DNA synthesis
P43235	Homo sapiens	Cathepsin K	Thiol protease involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling. Displays potent endoprotease activity against fibrinogen at acid pH. May play an important role in extracellular matrix degradation

P26006	Homo sapiens	Integrin alpha-3	Integrin alpha-3/beta-1 is a receptor for fibronectin, laminin, collagen, epiligrin, thrombospondin and CSPG4. Integrin alpha-3/beta-1 provides a docking site for FAP (seprase) at invadopodia plasma membranes in a collagen-dependent manner and hence may participate in the adhesion, formation of invadopodia and matrix degradation processes, promoting cell invasion.
Q96HF1	Homo sapiens	Secreted frizzled-related protein 2	Soluble frizzled-related proteins (sFRPS) function as modulators of Wnt signaling through direct interaction with Wnts. They have a role in regulating cell growth and differentiation in specific cell types
P9WQP1	Mycobacterium tuberculosis	Diacylglycerol acyltransferase/mycolyltransferase Ag85B	The antigen 85 proteins (FbpA, FbpB, FbpC) are responsible for the high affinity of mycobacteria for fibronectin, a large adhesive glycoprotein, which facilitates the attachment of M.tuberculosis to murine alveolar macrophages (AMs). They also help to maintain the integrity of the cell wall by catalyzing the transfer of mycolic acids to cell wall arabinogalactan and through the synthesis of alpha,alpha-trehalose dimycolate
J7M6M6	Streptococcus pyogenes	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis. Fibronectin and plasminogen binding protein
C0S1J2	Paracoccidioides brasiliensis	Enolase	Glycolysis. Fibronectin binding protein
Q2G0U9	Staphylococcus aureus	N-acetylmuramoyl-L-alanine amidase sle1	Peptidoglycan hydrolase involved in the splitting of the septum during cell division. Binds to both alpha and beta-chains of human fibrinogen as well as fibronectin, which suggests a role in the colonization of host factor-coated material or host tissue. Also exhibits lytic activity against S.carnosus and S.aureus cells but not against M.luteus cells

Entre las proteínas involucradas en metástasis se observan numerosas catepsinas (algunos ejemplos son P25774, P07711 o P43235), que son proteasas ampliamente distribuidas por los animales, dependiendo de su estructura, tienen como ligando diversas proteínas. Participan en procesos como reabsorción ósea, degradación y remodelación de la matriz extracelular, secreción de hormonas tiroideas etc. Esto pone de manifiesto la necesidad de modificar la matriz extracelular para colonizar nuevos tejidos, tal y como ocurre con las bacterias patógenas.

Un claro ejemplo de proteína moonlighting de virulencia es P9WQP1 de *Mycobacterium tuberculosis*, representada en el Cuadro 5. Su función en condiciones normales es la transferencia de ácidos micólicos, ayudando al mantenimiento de la pared celular. Además, tiene como función moonlighting la unión a fibronectina, facilitando a la bacteria infectar al huésped. Otra proteína que aparece como moonlighting en muchas especies bacterianas es la gliceraldehído 3-fosfato deshidrogenasa, como la de *Streptococcus pyogenes*, que une plasminógeno y fibronectina.

6.3.- Unión a laminina

La laminina constituye el mayor componente no colagenoso de la membrana basal, que es un tipo de matriz extracelular que se encuentra en la base de las células epiteliales. Son glicoproteínas triméricas que interactúan tanto con otros elementos de la membrana basal como con receptores de la superficie celular.

En la bibliografía se describen asociaciones entre distintos tipos de lamininas y cáncer, un ejemplo es la laminina-511. Se ha correlacionado el aumento de su expresión con una mayor progresión y metástasis en cáncer de mama avanzado y otros tipos de tumores, señalando los ligandos de esta proteína como posibles dianas terapéuticas [80]. También hay estudios que demuestran que determinados dominios de la laminina alfa2 favorecen la infección por bacterias, como *Streptococcus* patógenos del grupo A y B o *Staphylococcus aureus* [81]. Este tipo de descubrimientos son de vital importancia, ya que identifican nuevas estrategias de tratamiento de infecciones que no requieren de antibióticos, solamente de “bloqueadores” de las dianas a las que se unen las bacterias patógenas. Es un modo de luchar contra la amenaza de los microorganismos multirresistentes a antibióticos.

Como se ha visto en la comparación de los códigos GO, tanto proteínas metastásicas como proteínas moonlighting que actúan como factores de virulencia tienen como factor común la matriz extracelular (no el único, sino el más representativo). Es fácil encontrar investigaciones que lo demuestran por separado, pero no alguna que las intente relacionar. Podría haber casos en los que un patógeno utilice la misma secuencia de una proteína como diana para unirse que una célula tumoral al colonizar un nuevo tejido y producir metástasis. Si fuera así, un mismo fármaco podría servir para evitar ambos procesos.

El código GO que describe la interacción con la laminina es [GO:0043236](#). En la lista filtrada de matriz extracelular aparece representado en un 3.07% entre los factores de virulencia y en un 0.82% en las proteínas moonlighting totales ($3.01e-17$). En la lista filtrada de tejido el código GO que aparece es el [GO:0043237](#), cuya definición es la interacción selectiva con la laminina 1, aparece representado en un 1.22% entre los factores de virulencia y en un 0.13% en las proteínas moonlighting totales ($p.value = 0.057$), en este caso el $p.value$ no es significativo, pero sí que está próximo a serlo, quizá con una muestra mayor podría bajar del umbral 0.05. Algunas de las proteínas que interactúan con la laminina que aparecen en las listas de resultados se muestran en el Cuadro 6.

Cuadro 6: Ejemplos de proteínas que unen laminina

Proteína	Organismo	Nombre de la proteína	Función
P25774	Homo sapiens	Cathepsin S	Thiol protease. Key protease responsible for the removal of the invariant chain from MHC class II molecules. The bond-specificity of this proteinase is in part similar to the specificities of cathepsin L
P09382	Homo sapiens	Galectin-1	Lectin that binds beta-galactoside and a wide array of complex carbohydrates. Plays a role in regulating apoptosis, cell proliferation and cell differentiation. Inhibits CD45 protein phosphatase activity and therefore the dephosphorylation of Lyn kinase. Strong inducer of T-cell apoptosis

P07996	Homo sapiens	Thrombospondin-1	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. Binds heparin. May play a role in dentinogenesis and/or maintenance of dentin and dental pulp (By similarity). Ligand for CD36 mediating antiangiogenic properties. Plays a role in ER stress response, via its interaction with the activating transcription factor 6 alpha (ATF6) which produces adaptive ER stress response factors
O69174	Staphylococcus aureus	Enolase	Catalyzes the reversible conversion of 2-phosphoglycerate into phosphoenolpyruvate. It is essential for the degradation of carbohydrates via glycolysis (By similarity). Binds laminin when expressed on the bacterial cell surface; this probably induces destruction of the extracellular matrix, favoring invasion and dissemination
A8AWF1	Streptococcus gordonii	Putative pyruvate, phosphate dikinase regulatory protein	Bifunctional serine/threonine kinase and phosphorylase involved in the regulation of the pyruvate, phosphate dikinase (PPDK) by catalyzing its phosphorylation/dephosphorylation
Q5K117	Lactobacillus crispatus	Enolase	Catalyzes the reversible conversion of 2-phosphoglycerate into phosphoenolpyruvate. It is essential for the degradation of carbohydrates via glycolysis (By similarity). Binds laminin when expressed on the bacterial cell surface; this probably induces destruction of the extracellular matrix, favoring invasion and dissemination

Se puede observar que también hay catepsinas entre las proteínas que unen laminina. Cabe destacar también la enolasa, que de la misma manera que la gliceraldehído 3-fosfato deshidrogenasa, actúa como proteína moonlighting de virulencia en distintas especies de patógenos.

7.- Conclusiones

7.1.- Descripción de las conclusiones del trabajo

Las principales conclusiones que se han obtenido en el trabajo son:

- Determinados procesos parecen estar más representados entre proteínas moonlighting que actúan como factor de virulencia y proteínas relacionadas con metástasis. Los procesos que se han probado en el estudio han sido metabolismo de la glucosa, interacción con plasminógeno, remodelación/regeneración de tejidos e interacción con la matriz celular.
- La matriz extracelular parece ser clave tanto para la diseminación de patógenos como para la metástasis. Varios códigos GO relacionados con la interacción con proteínas de la matriz extracelular, como

el colágeno, fibronectina o laminina, aparecen altamente representados. Por como se ha realizado el análisis, se puede estar seguro de que estas funciones son comunes entre las funciones moonlighting de factores de virulencia, pero no si son comunes entre las proteínas metastásicas, y si este fuera el caso, si se tratan de las funciones mediante las cuales favorecen el proceso metastásico. Sería necesario continuar el estudio en este sentido, comprobar si los códigos GO que describen las funciones relacionadas con la unión a proteínas de la matriz extracelular están más representados, y posteriormente, comprobar caso a caso cuál es la función normal de las proteínas metastásicas y cuál es su expresión en tejidos normales. Si su expresión está aumentada en células tumorales, podría sugerir que está favoreciendo su diseminación. O por otra parte, si en condiciones normales no está involucrada en la unión a matriz extracelular, podría ser indicativo de que es una proteína moonlighting que está facilitando el inicio de la metástasis.

- A pesar de las incógnitas mencionadas, en la bibliografía se puede comprobar el importante rol que tiene la matriz extracelular en estos procesos. El bloqueo de los sitios de unión de proteínas como el colágeno, fibronectina. . . puede ser una nueva estrategia de tratamiento para evitar infecciones y diseminación de células tumorales. Incluso se puede estudiar si algún fármaco bloqueador anti-tumoral puede ser reutilizado para evitar la adhesión de bacterias. Un ejemplo de ello es la heparina, aunque no se utilice en humanos. Un mejor entendimiento de cómo se produce la unión a la matriz extracelular ayudará a descubrir nuevas dianas que podrán ser utilizadas tanto para tratar el cáncer como infecciones bacterianas.
- El análisis a gran escala de las funciones de proteínas puede ayudar a dirigir investigaciones más específicas. A raíz de las conclusiones de este estudio podrían plantearse nuevos objetivos, como hallar epítomos de unión dentro de las proteínas de la matriz extracelular, estudiar qué proteínas metastásicas son moonlighting, probar si determinados fármacos bloqueantes de la unión a la matriz extracelular son efectivos en la prevención de infecciones etc.
- Por motivos de tiempo, el trabajo se ha enfocado a la matriz extracelular, ya que los resultados indicaban que era el elemento en común más representativo. No obstante, sería necesario estudiar si las proteínas relacionadas con el metabolismo de la glucosa metastásicas son moonlighting y funcionan de una manera similar a los factores de virulencia de bacterias, teniendo su función metabólica principal y además, actuar como elemento de adhesión a la matriz extracelular.

7.2.- Valoración de los objetivos planteados inicialmente

Los objetivos planteados para el trabajo fueron bastante acertados, ya que los filtros que se aplicaron aparecieron más representados entre las proteínas moonlighting de virulencia. Si bien es cierto que en un principio se pensaba que los resultados nos iban a dirigir más hacia las proteínas involucradas en el metabolismo de la glucosa y el plasminógeno, no fue así. También se planteó la comparación de los interactomas de las diferentes listas de proteínas, pero no dio ningún resultado, debido a que la mayoría de los factores de virulencia de bacterias no devolvían ningún resultado, haciendo imposible una comparación objetiva. Otro aspecto que se esperaba es que salieran representados más códigos GO relacionados con la modulación del Sistema Inmunitario, y si bien es cierto que algunos códigos relacionados con la modulación de expresión de citoquinas han aparecido, está lejos de lo planeado.

En general el trabajo ha sido productivo y se han obtenido los resultados deseados, un punto en común entre proteínas moonlighting de virulencia y proteínas metastásicas, aunque no hayan sido los que se pensaron en un principio.

7.3.- Valoración del seguimiento de la planificación

En líneas generales se ha ido seguido el plan de trabajo. Algunas tareas, como el análisis de los código GO ha resultado ser más extensa de lo que se esperaba, de hecho, han quedado cuestiones que por tiempo no se han podido abordar. El tiempo extra debido a la situación excepcional que se ha vivido por el COVID-19 se ha utilizado para completar determinados aspectos del análisis de los códigos GO.

7.4.- Cuestiones a resolver

Aunque ya se han ido comentando las cuestiones a resolver, se resumen a continuación además de otras preguntas que han ido surgiendo a lo largo del proyecto:

- Comparar secuencias de proteínas moonlighting de virulencia e intentar predecir secuencias generales de unión a determinadas proteínas de la matriz extracelular para poder predecir qué proteínas de patógenos son factores de virulencia. Este mismo proceso puede realizarse con proteínas metastásicas y ver si hay secuencias comunes con los factores de virulencia.
- Comprobar cuáles de las proteínas metastásicas son moonlighting. Este hecho es de vital importancia para evitar efectos secundarios en fármacos, ya que al inhibir una determinada proteína por una función en concreto, se está inhibiendo otras que no participan en la enfermedad.
- Estudiar si los códigos GO relacionados con la unión a la matriz extracelular están muy representados entre las proteínas metastásicas, y si fuera así, si es esa la función responsable de que favorezcan la metástasis.
- Las adhesinas de los patógenos suelen ser inmunogénicas, y tras la primera fase de la infección baja su expresión, ya que dejan de ser tan necesarias y así se evita la respuesta del Sistema Inmunitario. Como se ha comentado en el trabajo, muchas proteínas moonlighting de virulencia están altamente conservadas evolutivamente, haciendo más difícil una respuesta inmunitario frente a ellas, ya que produciría reacciones cruzadas con el huésped. Una pregunta interesante es si la expresión durante la infección de un patógeno de adhesinas no moonlighting y moonlighting son parecidas o difieren, o si las proteínas moonlighting se transportan en menor medida a la superficie bacteriana para evitar una respuesta del Sistema Inmunitario.
- Investigar si hay algún fármaco que pudiera ser utilizado tanto en infecciones como para prevenir metástasis.
- Hacer estudios sobre cuáles son los epítomos diana de las proteínas moonlighting de virulencia y proteínas metastásicas que se unen a la matriz extracelular.

8.- Glosario

- **Adhesina:** factores de unión producidos por bacterias para adherirse a sus hospedanted.
- **Catepsina:** son un tipo de proteasas compuesta por más de 12 familias que están presentes en animales y otros organismos. Son esenciales para el recambio proteico.
- **Código GO:** término que describe una determinada función, proceso celular o componente celular relacionado con un gen.
- **Colágeno:** proteína fibrosa que forma parte de lamatriz extracelular. Es la proteína más abundante en los mamíferos, consta de más de 25 tipos y está presente en la mayoría de los tejidos.
- **Factor de virulencia:** producto generado por bacterias patógenas que facilitan su infección al huésped, o bien favorecen su supervivencia ante diversas amenazas, como el Sistema Inmunitario o antibióticos.
- **Fibronectina:** glucoproteína presente en todos los vertebrados. En humanos existe en forma soluble en plasma e insoluble en la matriz extracelular.
- **Función canónica:** la función canónica de una proteína moonlighting es aquella que se descubrió en primer lugar.
- **Función moonlighting:** la/s función/es moonlighting de una proteína moonlighting son aquellas descubiertas después de la canónica.
- **Heat Shock Protein (HSP):** proteínas muy conservadas a lo largo de la evolución, presente tanto en organismos unicelulares como pluricelulares. Actúan como chaperonas ayudando el correcto plegamiento de otras proteínas.
- **ICSPs:** subgrupo de proteínas moonlighting que se caracterizan por ser enzimas o chaperonas que en el citoplasma celular realizan una determinada función, diferente a aquella que ejercen en el exterior celular.
- **Laminina:** glucoproteínas de elevado peso molecular que son un componente clave de las membranas basales.
- **Matriz extracelular:** es una entidad estructuralmente compleja que forma una red tridimensional que engloba todos los órganos, tejidos y células del organismo. Entre sus muchas funciones están la de proteger a las células contenidas en ella, nutrición, angiogénesis, fibrosis o regeneración tisular.
- **Plasminógeno:** zimógeno precursor de la proteasa plasmina, cuya prncipal función es la disolución de coágulos sanguíneos.
- **Proteína moonlighting:** proteína multifuncional que realiza dos o más funciones no relacionadas entre sí, a menudo en diferentes compartimentos celulares.

9.- Bibliografía

Para realizar la bibliografía se ha utilizado el gestor bibliográfico Zotero [82], y el estilo de citación elegido ha sido Vancouver.

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 21 de octubre de 2004;431(7011):931-45.
2. Jeffery JC. An introduction to protein moonlighting. *Biochem Soc Trans*. diciembre de 2014;42(6):1679-83.
3. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci USA*. 6 de abril de 2010;107(14):6544-9.
4. Wistow G, Piatigorsky J. Recruitment of enzymes as lens structural proteins. *Science*. 19 de junio de 1987;236(4808):1554-6.
5. Piatigorsky J. Multifunctional lens crystallins and corneal enzymes. More than meets the eye. *Ann N Y Acad Sci*. 15 de abril de 1998;842:7-15.
6. Hendriks W, Mulders JW, Bibby MA, Slingsby C, Bloemendal H, de Jong WW. Duck lens epsilon-crystallin and lactate dehydrogenase B4 are identical: a single-copy gene product with two distinct functions. *Proc Natl Acad Sci USA*. octubre de 1988;85(19):7114-8.
7. Wistow GJ, Mulders JW, de Jong WW. The enzyme lactate dehydrogenase as a structural protein in avian and crocodylian lenses. *Nature*. 9 de abril de 1987;326(6113):622-4.
8. Rao PV, Krishna CM, Zigler JS. Identification and characterization of the enzymatic activity of zeta-crystallin from guinea pig lens. A novel NADPH:quinone oxidoreductase. *J Biol Chem*. 5 de enero de 1992;267(1):96-102.
9. Rao PV, Zigler JS. Zeta-crystallin from guinea pig lens is capable of functioning catalytically as an oxidoreductase. *Arch Biochem Biophys*. enero de 1991;284(1):181-5.
10. Huang QL, Russell P, Stone SH, Zigler JS. Zeta-crystallin, a novel lens protein from the guinea pig. *Curr Eye Res*. mayo de 1987;6(5):725-32.
11. Bateman OA, Purkiss AG, van Montfort R, Slingsby C, Graham C, Wistow G. Crystal structure of eta-crystallin: adaptation of a class 1 aldehyde dehydrogenase for a new role in the eye lens. *Biochemistry*. 22 de abril de 2003;42(15):4349-56.
12. Jeffery CJ. Moonlighting proteins. *Trends in Biochemical Sciences*. 1 de enero de 1999;24(1):8-11.
13. Jeffery CJ. Moonlighting proteins: old proteins learning new tricks. *Trends Genet*. agosto de 2003;19(8):415-7.
14. Henderson B, Martin A. Bacterial Virulence in the Moonlight: Multitasking Bacterial Moonlighting Proteins Are Virulence Determinants in Infectious Disease. *Infection and Immunity*. 1 de septiembre de 2011;79(9):3476-91.
15. Henderson B, Martin A. Bacterial moonlighting proteins and bacterial virulence. *Curr Top Microbiol Immunol*. 2013;358:155-213.
16. Danø K, Andreasen PA, Grøndahl-Hansen J, Kristensen P, Nielsen LS, Skriver L. Plasminogen activators, tissue degradation, and cancer. *Adv Cancer Res*. 1985;44:139-266.
17. Jeffery CJ. Protein moonlighting: what is it, and why is it important? *Philos Trans R Soc Lond, B, Biol Sci*. 19 de enero de 2018;373(1738).

18. Bocharov AV, Vishnyakova TG, Baranova IN, Remaley AT, Patterson AP, Eggerman TL. Heat shock protein 60 is a high-affinity high-density lipoprotein binding protein. *Biochem Biophys Res Commun.* 14 de octubre de 2000;277(1):228-35.
19. Franco-Serrano L, Huerta M, Hernández S, Cedano J, Perez-Pons J, Piñol J, et al. Multifunctional Proteins: Involvement in Human Diseases and Targets of Current Drugs. *Protein J.* 2018;37(5):444-53.
20. Ehinger S, Schubert W-D, Bergmann S, Hammerschmidt S, Heinz DW. Plasmin(ogen)-binding alpha-enolase from *Streptococcus pneumoniae*: crystal structure and evaluation of plasmin(ogen)-binding sites. *J Mol Biol.* 29 de octubre de 2004;343(4):997-1005.
21. Jeffery CJ, Bahnson BJ, Chien W, Ringe D, Petsko GA. Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator. *Biochemistry.* 8 de febrero de 2000;39(5):955-64.
22. Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla-Camps JC, Theil EC, et al. Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science.* 22 de diciembre de 2006;314(5807):1903-8.
23. Franco-Serrano L, Cedano J, Perez-Pons JA, Mozo-Villarias A, Piñol J, Amela I, et al. A hypothesis explaining why so many pathogen virulence proteins are moonlighting proteins. *Pathog Dis.* 1 de julio de 2018;76(5).
24. Pancholi V, Fischetti VA. A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity. *J Exp Med.* 1 de agosto de 1992;176(2):415-26.
25. Jeffery C. Intracellular proteins moonlighting as bacterial adhesion factors. *AIMS Microbiol.* 2018;4(2):362-76.
26. Tunio SA, Oldfield NJ, Berry A, Ala'Aldeen DAA, Wooldridge KG, Turner DPJ. The moonlighting protein fructose-1, 6-bisphosphate aldolase of *Neisseria meningitidis*: surface localization and role in host cell adhesion. *Mol Microbiol.* mayo de 2010;76(3):605-15.
27. Blau K, Portnoi M, Shagan M, Kaganovich A, Rom S, Kafka D, et al. Flamingo cadherin: a putative host receptor for *Streptococcus pneumoniae*. *J Infect Dis.* 15 de junio de 2007;195(12):1828-37.
28. Hennequin C, Porcheray F, Waligora-Dupriet A, Collignon A, Barc M, Bourlioux P, et al. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology (Reading, Engl).* enero de 2001;147(Pt 1):87-96.
29. Yamaguchi H, Osaki T, Kurihara N, Taguchi H, Hanawa T, Yamamoto T, et al. Heat-shock protein 60 homologue of *Helicobacter pylori* is associated with adhesion of *H. pylori* to human gastric epithelial cells. *J Med Microbiol.* octubre de 1997;46(10):825-31.
30. Wuppermann FN, Mölleken K, Julien M, Jantos CA, Hegemann JH. *Chlamydia pneumoniae* GroEL1 protein is cell surface associated and required for infection of HEp-2 cells. *J Bacteriol.* mayo de 2008;190(10):3757-67.
31. Garduño RA, Garduño E, Hoffman PS. Surface-associated hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. *Infect Immun.* octubre de 1998;66(10):4602-10.
32. Dallo SF, Kannan TR, Blaylock MW, Baseman JB. Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. *Mol Microbiol.* noviembre de 2002;46(4):1041-51.
33. Heilmann C, Hartleib J, Hussain MS, Peters G. The multifunctional *Staphylococcus aureus* autolysin aap mediates adherence to immobilized fibrinogen and fibronectin. *Infect Immun.* agosto de 2005;73(8):4793-802.

34. Kinhikar AG, Vargas D, Li H, Mahaffey SB, Hinds L, Belisle JT, et al. Mycobacterium tuberculosis malate synthase is a laminin-binding adhesin. *Mol Microbiol.* mayo de 2006;60(4):999-1013.
35. Jung C-J, Zheng Q-H, Shieh Y-H, Lin C-S, Chia J-S. Streptococcus mutans autolysin AtlA is a fibronectin-binding protein and contributes to bacterial survival in the bloodstream and virulence for infective endocarditis. *Molecular Microbiology.* 2009;74(4):888-902.
36. Allignet J, England P, Old I, El Solh N. Several regions of the repeat domain of the Staphylococcus caprae autolysin, AtlC, are involved in fibronectin binding. *FEMS Microbiol Lett.* 6 de agosto de 2002;213(2):193-7.
37. Sahoo S, Murugavel S, Devi IK, Vedamurthy GV, Gupta SC, Singh BP, et al. Glyceraldehyde-3-phosphate dehydrogenase of the parasitic nematode Haemonchus contortus binds to complement C3 and inhibits its activity. *Parasite Immunol.* diciembre de 2013;35(12):457-67.
38. Robinson MW, Buchtman KA, Jenkins C, Tacchi JL, Raymond BBA, To J, et al. MHJ0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of Mycoplasma hyopneumoniae. *Open Biol.* 17 de abril de 2013;3(4):130017.
39. Jarocki VM, Santos J, Tacchi JL, Raymond BBA, Deutscher AT, Jenkins C, et al. MHJ0461 is a multifunctional leucine aminopeptidase on the surface of Mycoplasma hyopneumoniae. *Open Biol.* enero de 2015;5(1):140175.
40. Min K-W, Lee S-H, Baek SJ. Moonlighting proteins in cancer. *Cancer Lett.* 1 de enero de 2016;370(1):108-16.
41. Wang X, Song X, Zhuo W, Fu Y, Shi H, Liang Y, et al. The regulatory mechanism of Hsp90alpha secretion and its function in tumor malignancy. *Proc Natl Acad Sci USA.* 15 de diciembre de 2009;106(50):21288-93.
42. Hance MW, Dole K, Gopal U, Bohonowych JE, Jezierska-Drutel A, Neumann CA, et al. Secreted Hsp90 is a novel regulator of the epithelial to mesenchymal transition (EMT) in prostate cancer. *J Biol Chem.* 2 de noviembre de 2012;287(45):37732-44.
43. Foo RS-Y, Nam Y-J, Ostreicher MJ, Metzl MD, Whelan RS, Peng C-F, et al. Regulation of p53 tetramerization and nuclear export by ARC. *Proc Natl Acad Sci USA.* 26 de diciembre de 2007;104(52):20826-31.
44. Jhaveri K, Taldone T, Modi S, Chiosis G. Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. *Biochim Biophys Acta.* marzo de 2012;1823(3):742-55.
45. Wang Y, Trepel JB, Neckers LM, Giaccone G. STA-9090, a small-molecule Hsp90 inhibitor for the potential treatment of cancer. *Curr Opin Investig Drugs.* diciembre de 2010;11(12):1466-76.
46. Cao L, Petrusca DN, Satpathy M, Nakshatri H, Petrache I, Matei D. Tissue transglutaminase protects epithelial ovarian cancer cells from cisplatin-induced apoptosis by promoting cell survival signaling. *Carcinogenesis.* octubre de 2008;29(10):1893-900.
47. RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. 2019. <http://www.rstudio.com/>.
48. Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R. rmarkdown: Dynamic Documents for R. R package version 2.2. 2020. <https://github.com/rstudio/rmarkdown>.
49. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller. dplyr: A Grammar of Data Manipulation. R package version 0.8.3. 2019. <https://CRAN.R-project.org/package=dplyr>.
50. Yihui Xie. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.26. 2019.

51. Hao Zhu. kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.1.0. 2019. <https://CRAN.R-project.org/package=kableExtra>.
52. Matt Dowle and Arun Srinivasan. data.table: Extension of 'data.frame'. R package version 1.12.8. 2019. <https://CRAN.R-project.org/package=data.table>.
53. Hadley Wickham. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. 2019. <https://CRAN.R-project.org/package=stringr>.
54. Hadley Wickham. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1), 1-29. 2011. <http://www.jstatsoft.org/v40/i01/>.
55. Hadley Wickham and Jennifer Bryan. readxl: Read Excel Files. R package version 1.3.1. 2019. <https://CRAN.R-project.org/package=readxl>.
56. Lucy D'Agostino McGowan and Jennifer Bryan. googledrive: An Interface to Google Drive. R package version 1.0.1. 2020. <https://CRAN.R-project.org/package=googledrive>.
57. s.r.o BS. GanttProject: free project management tool for Windows, macOS and Linux [Internet]. Gantt-Project. [citado 12 de junio de 2020]. Disponible en: <https://www.ganttproject.biz>.
58. HCMDB Human Cancer Metastasis Database [Internet]. [citado 13 de junio de 2020]. Disponible en: <https://hcmdb.i-sanger.com/>.
59. Retrieve/ID mapping [Internet]. [citado 13 de junio de 2020]. Disponible en: <https://www.uniprot.org/uploadlists/>.
60. Multitasking Protein DataBase [Internet]. [citado 15 de junio de 2020]. Disponible en: <http://wallace.uab.es/multitaskII/>.
61. Otto AM. Warburg effect(s)-a biographical sketch of Otto Warburg and his impacts on tumor metabolism. *Cancer Metab*. 2016;4:5.
62. Gelse K, Pöschl E, Aigner T. Collagens—structure, function, and biosynthesis. *Adv Drug Deliv Rev*. 28 de noviembre de 2003;55(12):1531-46.
63. Cretu A, Roth JM, Caunt M, Akalu A, Policarpio D, Formenti S, et al. Disruption of endothelial cell interactions with the novel HU177 cryptic collagen epitope inhibits angiogenesis. *Clin Cancer Res*. 15 de mayo de 2007;13(10):3068-78.
64. Caron JM, Ames JJ, Contois L, Liebes L, Friesel R, Muggia F, et al. Inhibition of Ovarian Tumor Growth by Targeting the HU177 Cryptic Collagen Epitope. *Am J Pathol*. 2016;186(6):1649-61.
65. Caron JM, Han X, Contois L, Vary CPH, Brooks PC. The HU177 Collagen Epitope Controls Melanoma Cell Migration and Experimental Metastasis by a CDK5/YAP-Dependent Mechanism. *Am J Pathol*. octubre de 2018;188(10):2356-68.
66. Freimark B, Clark D, Pernasetti F, Nickel J, Myszka D, Baeuerle PA, et al. Targeting of humanized antibody D93 to sites of angiogenesis and tumor growth by binding to multiple epitopes on denatured collagens. *Mol Immunol*. julio de 2007;44(15):3741-50.
67. Rasmussen M, Jacobsson M, Björck L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem*. 22 de agosto de 2003;278(34):32313-6.
68. Caswell CC, Han R, Hovis KM, Ciborowski P, Keene DR, Marconi RT, et al. The Scl1 protein of M6-type group A Streptococcus binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Mol Microbiol*. febrero de 2008;67(3):584-96.
69. Sylvestre P, Couture-Tosi E, Mock M. A collagen-like surface glycoprotein is a structural component of the Bacillus anthracis exosporium. *Mol Microbiol*. julio de 2002;45(1):169-78.

70. Hynes RO. Structure of Fibronectins. En: Hynes RO, editor. *Fibronectins* [Internet]. New York, NY: Springer; 1990 [citado 15 de junio de 2020]. p. 113-75. (Springer Series in Molecular Biology).
71. Speziale P, Arciola CR, Pietrocola G. Fibronectin and Its Role in Human Infective Diseases. *Cells*. 26 de 2019;8(12).
72. Foster TJ, Geoghegan JA, Ganesh VK, Hook M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nature Reviews Microbiology*. enero de 2014;12(1):49-62.
73. Foster TJ. The MSCRAMM Family of Cell-Wall-Anchored Surface Proteins of Gram-Positive Cocci. *Trends in Microbiology*. 1 de noviembre de 2019;27(11):927-41.
74. Brown CK, Gu Z-Y, Matsuka YV, Purushothaman SS, Winter LA, Cleary PP, et al. Structure of the streptococcal cell wall C5a peptidase. *PNAS*. 20 de diciembre de 2005;102(51):18391-6.
75. Somarajan SR, Rosa SLL, Singh KV, Roh JH, Höök M, Murray BE. The Fibronectin-Binding Protein Fnm Contributes to Adherence to Extracellular Matrix Components and Virulence of *Enterococcus faecium*. *Infection and Immunity*. 1 de diciembre de 2015;83(12):4653-61.
76. Singh KV, Rosa SLL, Somarajan SR, Roh JH, Murray BE. The Fibronectin-Binding Protein EfbA Contributes to Pathogenesis and Protects against Infective Endocarditis Caused by *Enterococcus faecalis*. *Infection and Immunity*. 1 de diciembre de 2015;83(12):4487-94.
77. Arciola CR, Bustanji Y, Conti M, Campoccia D, Baldassarri L, Samori B, et al. *Staphylococcus epidermidis*-fibronectin binding and its inhibition by heparin. *Biomaterials*. 1 de agosto de 2003;24(18):3013-9.
78. Arciola CR, Campoccia D, Montanaro L. Implant infections: adhesion, biofilm formation and immune evasion. *Nature Reviews Microbiology*. julio de 2018;16(7):397-409.
79. Chen J-L, Hong J, Lu J-L, Chen M-X, Chen W-X, Zhu J-S, et al. Effect of non-anticoagulant N-desulfated heparin on expression of vascular endothelial growth factor, angiogenesis and metastasis of orthotopic implantation of human gastric carcinoma. *World J Gastroenterol*. 21 de enero de 2007;13(3):457-61.
80. Pouliot N, Kusuma N. Laminin-511: a multi-functional adhesion protein regulating cell migration, tumor invasion and metastasis. *Cell Adh Migr*. febrero de 2013;7(1):142-9.
81. van Wijk XM, Dohrmann S, Hallstrom BM, Li S, Voldborg BG, Meng BX, et al. Whole-Genome Sequencing of Invasion-Resistant Cells Identifies Laminin a2 as a Host Factor for Bacterial Invasion. *mBio*. 10 de 2017;8(1).
82. Zotero | Your personal research assistant [Internet]. [citado 13 de junio de 2020]. Disponible en: <https://www.zotero.org/>.