



UNIVERSIDAD
GERARDO BARRIOS®
Líderes en Gestión del Conocimiento



Vicerrectoría de Investigación
y Proyección Social

Proteínas Multifuncionales en Enfermedades Humanas

Informe Técnico de Investigación

José Félix Rojas Cabeza

El Salvador, 2022

Editorial Universidad Gerardo Barrios

Proteínas Multifuncionales en Enfermedades Humanas.

I^a Edición

José Félix Rojas Cabeza

Informe Técnico de Investigación 2022

Vicerrectoría de Investigación y Proyección Social

ISBN

Todos los derechos reservados. Esta publicación no puede ser reproducida, ni parcial ni totalmente, ni registrada en/o transmitida por un sistema de recuperación de información, en ninguna forma ni formato, por ningún medio, sea mecánico, fotocopiado, electrónico o cualquier otro sin el permiso previo y por escrito de la editorial.

editorial@ugb.edu.sv

© Universidad Gerardo Barrios, 2022

Índice

INTRODUCCIÓN	3
JUSTIFICACIÓN	5
OBJETIVOS	6
Objetivo General	6
Objetivos Específicos	6
METODOS	6
Bases de datos, servidores y herramientas bioinformáticas utilizadas	7
Selección de un listado proteínas moonlighting asociadas a enfermedades humanas	8
Odds Ratio (Razón de Probabilidades)	9
RESULTADOS	12
Selección de proteínas moonlighting afiliadas a enfermedades humanas	15
Razón de Probabilidades (Odds ratio)	16
Códigos de Ontología de Genes de Proteínas Moonlighting	18
Análisis de funcionalidad de proteínas moonlighting	19
Redes de interacción	24
Docking con ClusPro 2.0	26
DISCUSIÓN Y CONCLUSIONES	27
Interacción entre proteínas moonlighting humanas	27
Interacción entre proteínas moonlighting asociadas a enfermedades humanas	28
Análisis de códigos GO, funcionalidad de proteínas moonlighting y redes de interacción	29
Análisis del Docking	36
Comentarios finales	37
REFERENCIAS	38
AGRADECIMIENTOS	44
ANEXOS	44

INTRODUCCIÓN

La finalidad de este trabajo es mejorar conocimientos actuales sobre los mecanismos moleculares por los cuales las proteínas multifuncionales intervienen en enfermedades humanas. Específicamente para identificar causas de efectos secundarios, o encontrar usos de medicamentos para fines no descritos anteriormente.

Esto requiere mejorar la comprensión actual de la función bioquímica y celular de las proteínas multifuncionales, mejorar los procedimientos para identificar la función a partir de la secuencia o la estructura y mejorar las anotaciones de bases de datos proteicas (secuencias y estructura), asimismo, comprender la amplia variedad de funciones de las proteínas(1).

Las proteínas multifuncionales son centrales como rutas bioquímicas de muchas enfermedades, por lo tanto, el desarrollo de tratamientos nuevos se verá favorecido al entender cómo alcanzan distintas ubicaciones celulares(1–3) o cómo afectar una sola función como diana terapéutica(3,4).

Si bien las proteínas multifuncionales pueden ser una de las causas de las reacciones adversas a medicamentos, también es posible especular que dichas proteínas también pueden ser dianas terapéuticas útiles(5). Especialmente considerando casos como peptidilprolil isomerasa de *Legionella pneumophila*, originalmente identificada como un potenciador de infectividad de macrófagos; es un factor de virulencia importante en ciertas bacterias intracelulares, y les ayuda a entrar al macrófago uniéndose a colágeno tipo IV(6). Posteriormente, se llegó al desarrollo de varios inhibidores sintéticos de la adhesividad de Mip, incluidos algunos compuestos de cicloheximida novedosos con actividad antimicrobiana contra *L. pneumophila*(7).

Luego, la ataxia telangiectasia mutada (ATM) quinasa se llama así porque las mutaciones en esta proteína causan ataxia telangiectasia, una enfermedad neurodegenerativa poco común. Esta quinasa es una proteína multifuncional y ahora se ve como un objetivo terapéutico(8).

A medida que aumenta el número de proteínas multifuncionales y se identifican diferentes sitios activos multifuncionales, es probable que las funciones moleculares de estas proteínas revelen algunas de ellas como objetivos relevantes para el desarrollo de nuevos agentes farmacéuticos(5).

La forma de demostrar que la proteína es multifuncional debe ser necesariamente experimental(3,9), lo que representa una limitación de este tipo de estudios. Además, existen reportes de que resultados experimentales inesperados suelen ser descartados por los investigadores, debido a que los datos experimentales de proteómica contienen muchos falsos positivos, estimados en alrededor del 20%(9,10).

El estudio de las proteínas multifuncionales se complica, debido a la existencia de muchos factores que afectan la función de una proteína: celulares, niveles de expresión, moduladores alostéricos, modificaciones postraduccionales, secuestro y otros factores pueden afectar el nivel de función de la proteína(2,11). En el caso de las proteínas multifuncionales, factores celulares como estos también pueden afectar el tipo de función de la proteína (Fig. 1)

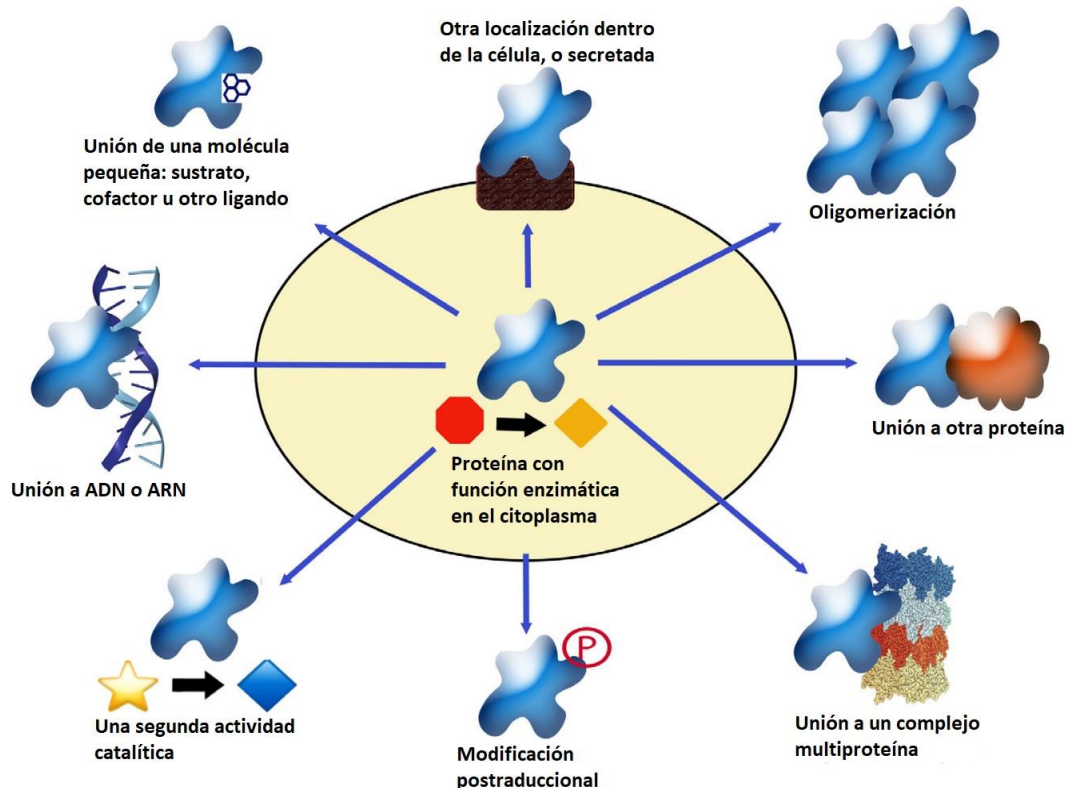


Figura 1. Factores que pueden afectar la función de una proteína multifuncional. Modificado de Jeffery, 2019(12).

Recientemente se han utilizado las siguientes aproximaciones que utilizan secuencias, estructuras, interactómica, algoritmos y programas para investigar multifuncionalidad(9): Búsquedas de homología remota utilizando Psi-Blast(13), Detección de motivos y dominios funcionales, Análisis de datos de bases de datos

de interacción proteína-proteína (PPI)(9,14), Análisis de coincidencia de la consulta de la secuencia con una base de datos 3D (ej. algoritmos como PiSITE(15) o Phyre2(16)), Análisis de correlación de mutaciones entre aminoácidos mediante algoritmos como MISTIC2(17). En el trabajo se pretende retomar algunas de estas técnicas para estudiar la interacción de proteínas moonlighting humanas asociadas a enfermedades con otras proteínas, buscando esclarecer los mecanismos moleculares de enfermedades(18–20).

JUSTIFICACIÓN

La proteómica ha cambiado de un enfoque en moléculas individuales a una perspectiva integrada, en la cual las proteínas emergen como actores dinámicos con múltiples funciones, en lugar de especialistas mono-funcionales(21).

Este trabajo pretende mejorar los conocimientos actuales sobre el involucramiento de las proteínas multifuncionales en los mecanismos moleculares de enfermedades humanas. Específicamente en términos de identificar causas de posibles efectos secundarios o reutilizar medicamentos para fines no descritos anteriormente.

Para esto se requiere mejorar la comprensión actual de funciones bioquímicas y roles celulares de proteínas multifuncionales, para mejorar los programas que identifican funciones a partir de secuencia o estructura, y mejorar las anotaciones de las bases de datos de secuencia y estructurales. Así como la comprensión de la amplia diversidad de funciones de las proteínas(22).

Al ser las proteínas multifuncionales centrales para el papel de muchas enfermedades, el desarrollo de nuevos tratamientos se vería favorecido al abordar preguntas como el mecanismo por el que alcanzan distintas ubicaciones celulares(22,23) o cómo afectar una sola función como diana terapéutica(2,23).

Si bien las proteínas multifuncionales pueden ser una de las causas de las reacciones adversas a medicamentos, también es posible especular que dichas proteínas también pueden ser dianas terapéuticas útiles(4).

A medida que aumenta el número de proteínas multifuncionales y comienza a identificarse el número de sitios activos multifuncionales, es más que probable que las importantes acciones en que participan estas proteínas conviertan a algunas de ellas en objetivos relevantes para el desarrollo de nuevos agentes farmacéuticos(4).

OBJETIVOS

Objetivo General

Examinar la relación entre proteínas multifuncionales y enfermedades humanas, tomando en cuenta los mecanismos moleculares de la enfermedad y su interacción con dianas farmacológicas.

Objetivos Específicos

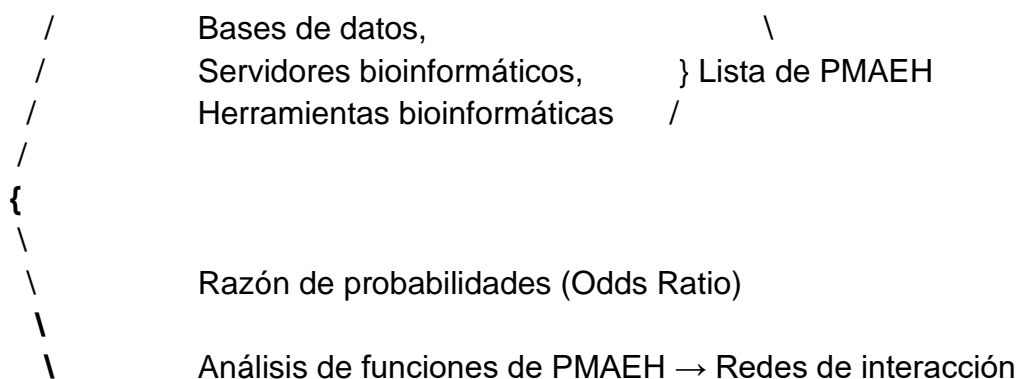
- Revisar si se pueden agrupar los mecanismos moleculares en que intervienen las proteínas multifuncionales asociadas con enfermedades.
- Revisar bases de datos de interacción de proteínas para hacer análisis de interacción local entre las proteínas asociadas encontradas en el mapeo.

METODOS

Generalidades:

Los métodos utilizados involucraron el uso de información previamente existente a través de Bases de datos, servidores, y herramientas bioinformáticas, para realizar un listado de las proteínas moonlighting humanas y un listado de las proteínas moonlighting humanas asociadas a enfermedades. Luego se determinó mediante una prueba t de dos muestras si las proteínas moonlighting humanas tienen más funciones que el promedio de las proteínas moonlighting de otros organismos. Además, se determinó si las proteínas moonlighting son más propensas a estar involucradas a enfermedades con una razón de probabilidades. Posteriormente se creó un repositorio de información sobre las funciones de los genes de proteínas moonlighting relacionados con enfermedades humanas utilizando GoTrapper. Finalmente, se hizo un análisis de enriquecimiento con GOnet.

Esquema



* Prot. Moonlighting Asociadas a Enf. Hum. (PMAEH)

Comentarios para aclarar esquema

Bases de datos, servidores y herramientas bioinformáticas: se utilizó información disponible para ser sistematizada y encontrar nuevos conocimientos de proteínas moonlighting involucradas en enfermedades humanas. Resulta en un **Listado de proteínas moonlighting asociadas a enfermedades humanas**

Cálculo de **Razón de Probabilidades (Odds Ratio):** Resalta que las proteínas moonlighting humanas tienen más probabilidades de estar involucradas en enfermedades.

Análisis de funciones de proteínas moonlighting asociadas a enfermedades: Se encontraron **Redes de interacción de proteínas moonlighting humanas** que deben interrogarse para generar más información. Queda claro que las proteínas multifuncionales interactúan entre sí frecuentemente, y que esas redes de regulación tienen información clave para entender cómo se mantiene la homeostasis y alterarla a través de medicamentos.

Bases de datos, servidores y herramientas bioinformáticas utilizadas

Se utilizaron o evaluaron las siguientes bases de datos:

Tabla 1. Bases de datos, servidores y herramientas bioinformáticas a evaluarse o utilizarse

Nombre	Tipo	Referencia	Enlace
APID	Base de datos	(24)	http://cicblade.dep.usal.es:8080/APID/
BioGRID	Base de datos	(25) (26)	https://thebiogrid.org/
Drugbank	Base de datos	(27) (28)	https://go.drugbank.com/
Gene Ontology	Base de datos	(29) (30)	http://geneontology.org/
HGMD	Base de datos	(31) (32)	http://www.hgmd.cf.ac.uk/ac/index.php
IMEx	Base de datos	(33)	https://www.imexconsortium.org/
InterPro	Base de datos	(34) (35)	https://www.ebi.ac.uk/interpro/
MultitaskProtDBII	Base de datos	(36)	http://wallace.uab.es/multitaskII/

OMIM	Base de datos	(37)	https://omim.org/
PDB	Base de datos	(38)	https://www.rcsb.org/
PiSITE	Base de datos	(15)	http://pisite.hgc.jp/cgi-bin/top.cgi
TTD	Base de datos	(39) (40)	http://db.idrblab.net/ttd/
STRING	Base de datos	(41)	https://string-db.org/
GOnet	Herramienta	(42)	https://tools.dice-database.org/GOnet/
GOTrapper	Aplicación	(43)	https://github.com/BioGeneTools/GOTrapper
ClusPro 2.0	Servidor Docking PPI	(44)	https://cluspro.bu.edu/

Selección de un listado proteínas moonlighting asociadas a enfermedades humanas

Se tomó una estrategia que consistió en seleccionar un conjunto de proteínas multifuncionales humanas, tomadas de la base de datos de proteínas multifuncionales MultitaskProtDB-II(36), y verificar la interacción de proteínas de interés relacionadas con enfermedades humanas con otras proteínas metabólicas de interés.

De los recursos evaluados para revisar la interacción proteína-proteína (PPI), se determinó que el más adecuado era STRING(41), basado en Cytoscape(45). STRING cuenta con más de 932,000,000 de interacciones(46), por lo que supera las otras opciones consultadas para realizar los análisis.

El análisis se realizó de la siguiente manera: Se tomó un listado de proteínas a las que se unen las proteínas moonlighting de MultitaskProtDB-II, según las interacciones descritas en BioGRID(25,26), para examinar las principales proteínas interactivas que presentan, y si comparten algunas de ellas.

Odds Ratio (Razón de Probabilidades)

Este tipo de test, también llamado “razón de probabilidades”, es una estadística que cuantifica la fuerza de la asociación entre dos eventos, A y B (47,48). El procedimiento seguido involucra definir una matriz de confusión, mostrada a continuación, y obtener la información asociada de la misma.

Tabla 2. Matriz de confusión

Test	Proteína Moonlighting	Proteína No Moonlighting
Grupo Asociada a Enfermedad	127	44
No Asociada a Enfermedad	5732	70045

Los resultados se pueden verificar en este [enlace](#) de la universidad de Hong Kong.

Extracción de códigos GO

Una vez que se obtuvieron las proteínas humanas asociadas a enfermedades, se incluyeron en GOTrapper(43) (Fig. 2), y se utilizó la opción de “within” para encontrar términos asociados en dentro de una sola lista de genes. También se utilizó threshold = 4, para obtener los códigos GO más generales. Nótese que la salida del archivo colocaba los códigos GO como únicos y unía los genes que tenían el código GO específico.

id	GO_id	GO_category	#_Background_Genes	Genes	Gene ID	Score	GO_term	
0	1	GO:0042803	MF	906	BST2, CDH1, CRYAB, TYMP, ENO1, HLCS, HES1, HSP...	684, 999, 1410, 1890, 2023, 3141, 3280, 3320, ...	8.8234	protein homodimerization activity
1	2	GO:0043312	BP	490	ALAD, ANPEP, ANXA2, ARG1, BST2, B4GALT1, GPI, ...	210, 290, 302, 383, 684, 2683, 2821, 2934, 314...	7.9366	neutrophil degranulation
2	3	GO:0005524	MF	1512	ACTB, ADK, ASS1, CFTR, EGFR, ERCC2, GK, GSK3B,...	60, 132, 445, 1080, 1956, 2068, 2710, 2932, 30...	9.5622	ATP binding
3	4	GO:0051301	BP	684	RCC1, ERCC2, FGF1, FGF2, PLK1, PPP1CC, RAN, RP...	1104, 2068, 2246, 2247, 5347, 5501, 5901, 6188...	8.4179	cell division
4	5	GO:0019901	MF	727	ACTB, CDKN1A, ATF2, CTNNB1, DNM2, EGFR, GSK3B,...	60, 1026, 1386, 1499, 1785, 1956, 2932, 3320, ...	8.5058	protein kinase binding

Figura 2. Encabezado de salida de GOTrapper con umbral de 4.

Posteriormente se utilizó un script en Python para generar un diccionario que le asignaba los códigos GO correspondientes a cada gen (Fig. 3).

```

GO_DIC = {}
for search_gene in tabla_1["Gene Symbol"]:
    temporary_list = []
    for i in range(len(GO_tool["Genes"])):
        if search_gene in GO_tool["Genes"][i]:
            temporary_list.append(GO_tool["GO_id"][i])
    GO_DIC[search_gene] = temporary_list

```

Figura 3. Código utilizado para asignar los códigos GO a cada gen

Luego las funciones biológicas se anotaron manualmente consultando QuickGO(49). Esto quedó registrado en un archivo adjunto en el anexo.

Análisis de funciones de proteínas moonlighting asociadas a enfermedades

Se asignaron categorías a cada enfermedad manualmente, consultando OMIM(50) y MedGen(51). Las categorías utilizadas fueron: "anemia", "atherosclerosis", "autoimmune", "blood", "cancer", "circulatory", "dermatitis", "genetic", "Hypertension", "immune", "Keloid", "metabolic", "myopathy", "necrosis", "neoplasia", "nervous", "bone", "retinal degradation", "virulence". Se decidió no tomar estos grupos para examinarlos precisamente, por razones de tiempo. Sin embargo, los genes de cada grupo quedan registrados en un archivo adjunto en el anexo.

También se identificaron en MultitaskProtDBII los genes humanos involucrados con enfermedades, se colocaron los 127 genes en string, tomando en cuenta textmining, databases e interacciones experimentales, y se identificaron 7 grupos de interacción ([enlace](#)). Luego los 7 grupos se examinaron con GONet(42), tomando la opción de "GOTerm annotation", sin los términos asociados a los genes, para interpretar más claramente la función de cada grupo.

Es interesante mencionar que GONet también puede utilizarse para análisis de enriquecimiento, en caso de que se desee revisar los niveles de expresión de los genes en cierto tipo específico de tejido.

Redes de interacción de proteínas moonlighting humanas

El análisis se realizó de la siguiente manera:

1. Se identificaron en MultitaskProtDBII los genes humanos involucrados con enfermedades,
2. Se colocaron los 127 genes en string, tomando en cuenta textmining, databases e interacciones experimentales,
3. Se identificaron 7 grupos de interacción (enlace),
4. Se descargaron de BioGRID los datos de las proteínas de interés de cada grupo,
5. Se convirtieron en .csv,
6. Se incorporaron a RStudio,
7. Se tomaron las columnas 8 y 9 del .csv, que contienen los genes que interactúan con la proteína de interés. Nótese que las columnas 8 y 9 corresponden al Gene Symbol de los interactores.
8. Las columnas 8 y 9 del .csv se colocaron en un vector de interacciones,
9. Se seleccionaron los valores únicos de ese vector de interacciones,
10. Finalmente, se exportaron que un archivo tipo .csv,
11. Se utilizó STRING para generar las redes de interacción. Las redes se crearon mostrando la red completa, solamente con interacción experimental, y con el nivel más alto de confianza.

```
##{r}
# HMGB1 109389
# MAPK1 111580

HMGB1 <- read.csv(file = './data/BIOGRID-GENE-109389-4.2.193.tab3.csv', sep
= ',', header = TRUE)
colnames(HMGB1)[1] <- "BioGRID.Interaction.ID"
MAPK1 <- read.csv(file = './data/BIOGRID-GENE-111580-4.2.193.tab3.csv', sep
= ',', header = TRUE)
colnames(MAPK1)[1] <- "BioGRID.Interaction.ID"

interacciones <- c(as.character(HMGB1[,8]),
  as.character(HMGB1[,9]),
  as.character(MAPK1[,8]),
  as.character(MAPK1[,9]))

int_unq_7 <- unique(interacciones)

write.csv(int_unq_7, file="./results/int_unq_7.csv")
##
```

Figura 4. Código utilizado para generar el archivo que se colocó en STRING para generar la red del grupo enf_7.

RESULTADOS

Una inspección inicial de las proteínas moonlighting humanas mostró una tendencia a interactuar entre sí. Se observaron 4 grupos principales: El grupo 1 se caracterizó por proteínas ribosomales, y factores de transcripción (MDM2, TP53, ...). El grupo 2 se encarga de la propagación de señales de fosforilación de Serina/Treonina quinasas. SMAD4 media la asociación diferencial con diferentes proteínas, mientras que RAS se encarga de GTPs. En resumen, el grupo 2 tiene que ver con enviar señales de fosforilación en el núcleo. El grupo 3 contiene principalmente proteínas relacionadas con el metabolismo de la glucosa, el grupo 4 tiene que ver con proteínas involucradas en la síntesis de RNA de transferencia, y otros genes que se muestran sin ser parte de un grupo específico. Posteriormente se realizó un agrupamiento de los genes de proteínas moonlighting relacionados con enfermedades humanas, y se observó mediante una prueba t de dos muestras que las cantidades de funciones de proteínas moonlighting no son mayores en humanos que en todo el conjunto. Sin embargo, un Odds ratio (Razón de Probabilidades) indicó que las proteínas moonlighting son más propensas a estar involucradas en enfermedades que otras proteínas. Posteriormente se obtuvieron los códigos GO ([ontología de genes](#)) para tener un repositorio de información sobre las funciones de los genes de proteínas moonlighting relacionados con enfermedades humanas con la(s) enfermedad(es) reportadas para dicho gen. Luego se realizó un agrupamiento de los genes moonlighting asociados a enfermedades humanas, y se obtuvieron grupos diferentes a los de proteínas moonlighting humanas. El grupo 1_enf era demasiado extenso para sistematizar la información, el grupo 2_enf contenía las proteínas ribosomales del grupo 2 de proteínas multifuncionales humanas, pero los factores de transcripción quedaron en otro grupo (4_enf), mientras que algunas de las proteínas involucradas en síntesis de RNA de transferencia quedaron en el grupo 3_enf. El grupo 5_enf tiene genes involucrados en adhesión celular, y metabolismo energético. El grupo 6_enf tiene genes involucrados en señalización celular al núcleo (RAN) y regulación del ciclo celular. Finalmente, el grupo 7_enf tiene genes involucrados en regulación de transcripción, y la vía MAP del ciclo celular (proliferación, diferenciación, regulación de la transcripción y desarrollo).

Se realizó una búsqueda en la base de datos MultitaskProtDB-II para seleccionar un listado de proteínas moonlighting que puedan estar implicadas en enfermedades humanas. Del total de proteínas moonlighting registradas en la base de datos, que cuenta con 694 proteínas moonlighting, se han escogido 171 para formar parte del listado, por corresponder a *Homo sapiens*.

Selección de proteínas moonlighting humanas

La primera fase del análisis involucró verificar si había alguna interacción entre proteínas moonlighting humanas. Esto se verificó utilizando STRING para determinar si había interacciones entre las proteínas obtenidas. La red obtenida tenía la siguiente configuración: Se mostró la red completa, se requirieron evidencias experimentales de interacción, y un nivel de interacción de alta confianza (score > 0.700). En el análisis se observaron 4 grupos con cinco interacciones o más entre proteínas de este tipo, y varios grupos con cuatro o menos interacciones (Fig. 5, Tabla 2).

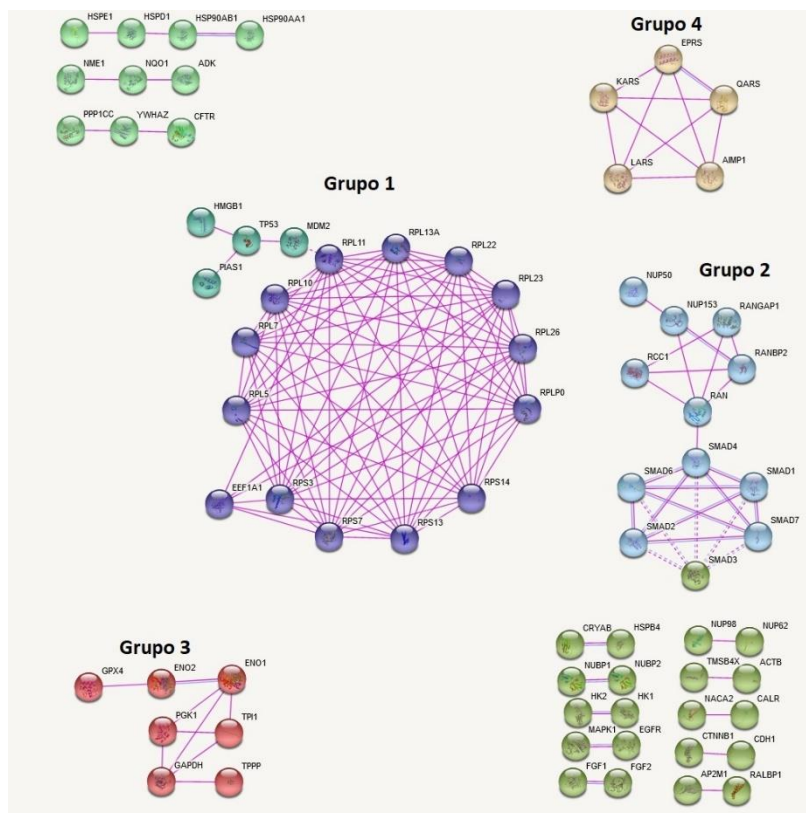


Figura 5. Grupos de interacción entre proteínas moonlighting humanas ([enlace](#)). Es importante destacar que la distancia entre nodos no está relacionada con semejanza entre las estructuras.

La cantidad de interacciones por proteínas es alta. En el Grupo 1 se encuentran TP53 (Tumor protein p53), MDM2 (un protooncogén), y RPL11 (Una proteína ribosomal); la cantidad de interacciones va de 1,379 a 129. En el grupo 2 se encuentran varias SMAD (Transductores de señal de TGF- β), RAN (un oncogén) y NUP153 (una nucleoporina); la cantidad de interacciones va de 380 a 51. En el grupo 3 Destacan GAPDH, y 2 Enolasas. Interesantemente, las interacciones reportadas incluyen interacción entre ENO1 y GAPDH.

Tabla 3. Grupos de interacción entre proteínas moonlighting humanas. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de [BioGrid](#).

G1		G2		G4		G8		G15	
HMGB1	312	NUP50	13	EPRS	28	CRYAB	127	NACA2	21
PIAS1	129	NUP153	14	QARS	29	HSPB4	70	CALR	23
TP53	1,379	RANBP2	9	AIMP1	5	G9		G16	
MDM2*	601	RANGA P1	21	LARS	14	HK1	168	CTNNB1	73
RPL11	370	RAN	20	KARS	24	Hk2	80	CDH1	65
EEF1A1	510	RCC1	22	G5		G10		G17	
RPS3	406	SMAD4	13	HSPE1	20	NUBP1	59	AP2M1	34
RPS7*	267	SMAD1	6	HSPD1	71	NUBP2	92	RALBP1	14
RPS13	321	SMAD7	8	HSP90A B1	9	G11			
RPS14	337	SMAD3	10	HSP90A A1	52	MAPK1	364		
RPL5	320	SMAD2	38	G6		EGFR	1,360		

RPL7	333	SMAD6	51	NME1	15 1	G12			
RPL10	458	G3		NQO1	65	FGF1	39		
RPL13 A	220	GPX4	52	ADK	26	FGF2	39		
RPL22	296	ENO2	11 0	G7		G13			
RPL23	359	ENO1	43 3	PPP1CC	66 5	NUP98	140		
RPL26	179	TPI1	18 8	YWHAZ	62 5	NUP62	220		
RPLP0	291	GAPDH	33 1	CFTR	85 9	G14			
		PGK1	18 2			TMSB4 X	40		
		TPPP	22			ACTB	497		

* Q00987 se cambió a A7UKY0 (uniprot_Id_1, en R). Ambas corresponden a la proteína MDM2. La primera está mejor descrita, pero la segunda es la que está reportada en MultitaskProtDBII.

** P62081 en la base de datos está como “40S ribosomal protein S7”, el nombre utilizado es RPS7.

Interaction

GAPDH [ENSP00000229239]

Glyceraldehyde-3-phosphate dehydrogenase. Has both glyceraldehyde-3-phosphate dehydrogenase and nitrosylase activities, thereby playing a role in glycolysis and nuclear functions, respectively. Participates in nuclear events including transcription, RNA transport, DNA replication and apoptosis. Nuclear functions are probably due to the nitrosylase activity that mediates cysteine S-nitrosylation of nuclear target proteins such as SIRT1, HDAC2 and PRKDC. Modulates the organization and assembly of the cytoskeleton. Facilitates the CHP1-dependent microtubule and membrane associations throu [...]

ENO1 [ENSP00000234590]

Alpha-enolase; Multifunctional en as its role in glycolysis, plays a pa processes such as growth control and allergic responses. May also i intravascular and pericellular fibrin to its ability to serve as a receptor plasminogen on the cell surface o such as leukocytes and neurons. i immunoglobulin production; Belor family

↔

Evidence suggesting a functional link:

Neighborhood in the Genome:	- disabled -
Gene Fusions:	- disabled -
Cooccurrence Across Genomes:	- disabled -
Co-Expression:	- disabled -
Experimental/Biochemical Data:	yes (score 0.811). In addition, putative homologs were found interact other organisms (score 0.186).
Association in Curated Databases:	- disabled -
Co-Mentioned in Pubmed Abstracts:	- disabled -
Combined Score:	0.839

Figura 6. Ejemplo de interacción entre ENO1 y GAPDH en la base de datos STRING.

Estas interacciones se calcularon tomando referencias de ensayos bioquímicos como entrecruzamiento (cross-linking), inmunoprecipitación y purificación por afinidad en tándem, etc. (fig. 6). Nótese que en “Show” se muestran las referencias que documentan las interacciones

Selección de proteínas moonlighting afiliadas a enfermedades humanas

Posteriormente se seleccionaron proteínas humanas con enfermedades asociadas de la base de datos utilizada(36), obteniéndose 127 proteínas.

Se hizo un análisis preliminar de las cantidades de funciones por proteína, mediante una prueba F, para confirmar que las varianzas son diferentes. Luego, una prueba t de dos muestras (two sample t test) indicó que las cantidades de funciones de

proteínas moonlighting no son mayores en *Homo sapiens* que en todo el conjunto. Estos datos están en el archivo de R utilizado para análisis de datos, adjunto como anexo.

```
{r}
var.test(fcn_go_count, fcnh_go_count, alternative="two.sided")
t.test(fcn_go_count, fcnh_go_count, var.equal = F)

F test to compare two variances

data: fcn_go_count and fcnh_go_count
F = 0.34379, num df = 2775, denom df = 683, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3047263 0.3862018
sample estimates:
ratio of variances
 0.3437924

Welch Two Sample t-test

data: fcn_go_count and fcnh_go_count
t = -5.5646, df = 802.2, p-value = 3.582e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.942732 -2.843410
sample estimates:
mean of x mean of y
 5.911023 10.304094
```

Figura 7. Prueba F y Prueba t de dos muestras de Welch.

Razón de Probabilidades (Odds ratio)

Ya ha sido descrito anteriormente que hay una gran proporción de proteínas moonlighting relacionadas con enfermedades(4). En este trabajo se realizó un Odds ratio con los datos más recientes. Se observó que 74% de las proteínas moonlighting humanas están relacionadas con enfermedades.

Interesantemente, las cifras difieren de la bibliografía consultada debido a que se han reportado nuevas proteínas moonlighting, y se han incluido muchas proteínas nuevas tanto a OMIM como a UniProt. Los cálculos fueron realizados con epiR(52) y a mano(48,53).

En la tabla 3 se muestra que las proteínas moonlighting son más propensas a estar involucradas en enfermedades humanas que el grupo general de proteínas.

Tabla 4. Resultados de análisis estadístico odds ratio, que muestra el grado de significación, con un intervalo de confianza de 95%.

	n	proporción
Número de proteínas moonlighting involucradas en enfermedades	127	74.269

Número de proteínas moonlighting No involucradas en enfermedades	44	25.731
Total	171	100.000
Número de proteínas humanas en UniProt y OMIM	5732	7.564
Número de proteínas humanas en UniProt y no OMIM	70045	92.436
Número de proteínas humanas en UniProt*	75777	100.000

		IC 95%	
Estimadores de Riesgo	Estimador	(Inferior,	Superior)
Riesgo absoluto en el grupo tratamiento	0.74	0.65	0.83
Riesgo absoluto en el grupo control	0.08	-0.02	0.17
Reducción absoluta de riesgo (ARR)	-0.67	-0.76	-0.58
Riesgo Relativo (RR)	9.82	8.96	10.76
Reducción del Riesgo Relativo (RRR)	-8.82	-9.76	-7.95
Odds ratio			
Probabilidades en el grupo tratamiento	2.11	1.73	2.48
Probabilidades en el grupo control	0.15	-0.12	0.42
Odds ratio	35.27	25.01	49.75

Códigos de Ontología de Genes de Proteínas Moonlighting

Gene Ontology (GO) es una importante iniciativa de bioinformática para unificar la representación de atributos de genes y productos de genes en todas las especies(30,54). Más específicamente, el proyecto tiene la intención de mantener y desarrollar un vocabulario controlado de atributos de genes y productos de genes; anotar genes y productos génicos, y asimilar y difundir datos de anotación; y proporcionar herramientas para un fácil acceso a todos los aspectos de los datos proporcionados por el proyecto, para permitir la interpretación funcional de los datos experimentales utilizando GO, por ejemplo, a través del análisis de enriquecimiento.

Una vez obtenido el listado, se observó que había una cantidad importante de códigos GO(29,30) asociados a la función canónica y a la principal de cada proteína moonlighting asociada a una enfermedad. Debido a que la cantidad de códigos GO asociados a cada proteína era bastante elevada (Fig. 8), se utilizó GOTrapper(43) para filtrar automáticamente los GO, con un umbral de 4, para obtener los más generales de la jerarquía, descartando los específicos. Esto que disminuyó la cantidad de códigos GO asociados a cada gen (ver anexos) de 1100 a 359. Sin embargo, hubo 5 genes que no fueron reconocidos: NACA2, XOR, SMN1 e IDH. En estos casos se colocaron todos los códigos GO de la base de datos de proteínas moonlighting. Como ejemplo: ANPEP, disminuyó de 25 códigos GO a 6 (Fig. 8)

Uniprot	Gene Symbol	Disease	GO_ID	Biological Process	GO_ID	Molecular function
P15144	ANPEP	[1] "Urinary Bladder Neoplasms, Dermatitis, Allergic Contact\N"	GO:0046718	viral entry into host cell	GO:0008233	peptidase activity
			GO:0043312	neutrophil degranulation	GO:0038023	signaling receptor activity
					GO:0008270	zinc ion binding
					GO:0042277	peptide binding
P18754	RCC1	[2] "Raynaud disease\N"	GO:0007088	regulation of mitotic nuclear division/viral entry into host cell	GO:0046982	protein heterodimerization activity
			GO:0007052	mitotic spindle organization	GO:0051020	GTPase binding
			GO:0051301	cell division	GO:0003682	chromatin binding
			GO:0051262	protein tetramerization		
			GO:0000082	G1/S transition of mitotic cell cycle		
P00558	PGK1	[3] "Phosphoglycerate kinase deficiency, hepatocellular carcinoma, Renal cell carcinoma\N"	GO:0061621	canonical glycolysis	GO:0016651	oxidoreductase activity, acting on NAD(P)H
			GO:0001666	response to hypoxia	GO:0005524	ATP binding
			GO:0043467	regulation of generation of precursor metabolites and energy		
			GO:0016525	negative regulation of angiogenesis		
			GO:0006094	gluconeogenesis		
			GO:0062013	positive regulation of small molecule metabolic process		
A1A528	ZW10	[4] "Osteosarcoma\N"	GO:0007088	regulation of mitotic nuclear division		
			GO:0007093	mitotic cell cycle checkpoint		
			GO:2001251	negative regulation of chromosome organization		
			GO:0010256	endomembrane system organization		
			GO:0033045	regulation of sister chromatid segregation		
			GO:1901991	negative regulation of mitotic cell cycle phase transition		
			GO:0051301	cell division		
			GO:0051656	establishment of organelle localization		
			GO:0000070	mitotic sister chromatid segregation		
			GO:0030010	establishment of cell polarity		
Q9H009	NACA2	[5] "Mutations and deletions are lethal in early development\N"	GO:0071840	cellular component organization or biogenesis	GO:0005634	nucleus
			GO:0022613	ribonucleoprotein complex biogenesis	GO:0005732	cytoplasm
			GO:0044085	cellular component biogenesis		
			GO:0008150	biological process		
			GO:0042254	ribosome biogenesis		

Figura 8. Ejemplo de los cinco primeros genes de la tabla. Para evitar incompatibilidad con GOTrapper se colocaron los códigos en inglés: Uniprot es el código de la proteína en Uniprot, Gene Symbol se refiere a la abreviatura tradicional de la proteína, Disease se refiere a la enfermedad a la que la proteína está asociada. GO_ID al código de referencia de ontología de gen de los procesos biológicos o funciones moleculares de cada proteína (ejemplo: RCC1 es una proteína que tiene 5 códigos GO asociados con procesos biológicos de división mitótica o celular, y

tres códigos GO de función molecular: unión a cromatina, a ATPasa, y a actividad de hetero dimerización de proteínas). Nótese que NACA2 tiene un color diferente porque fue uno de los 5 genes para los que se colocaron todos los GO. Los GO tachados corresponden a la categoría de compartimiento celular, que no se encontró en ningún otro caso.

El procedimiento empleado fue el siguiente:

1. Se importó la tabla donde se iban a incorporar los GO,
2. se importó la tabla resultante de aplicar GOTrapper con threshold de 4,
3. se realizó un diccionario que contenía el Gene Symbol de cada proteína moonlighting con sus funciones GO, que finalmente
4. se agregaron a la tabla con los GO.

Análisis de funcionalidad de proteínas moonlighting

Posteriormente se realizó un análisis de interacción utilizando STRING, para determinar las interacciones con las 127 proteínas moonlighting relacionadas con enfermedades humanas. En este caso el análisis se realizó con el nivel más alto de confianza (0.900), y tomando en cuenta sólo bases de datos (amarillo), datos minados de texto (azul) y experimentales (morado).

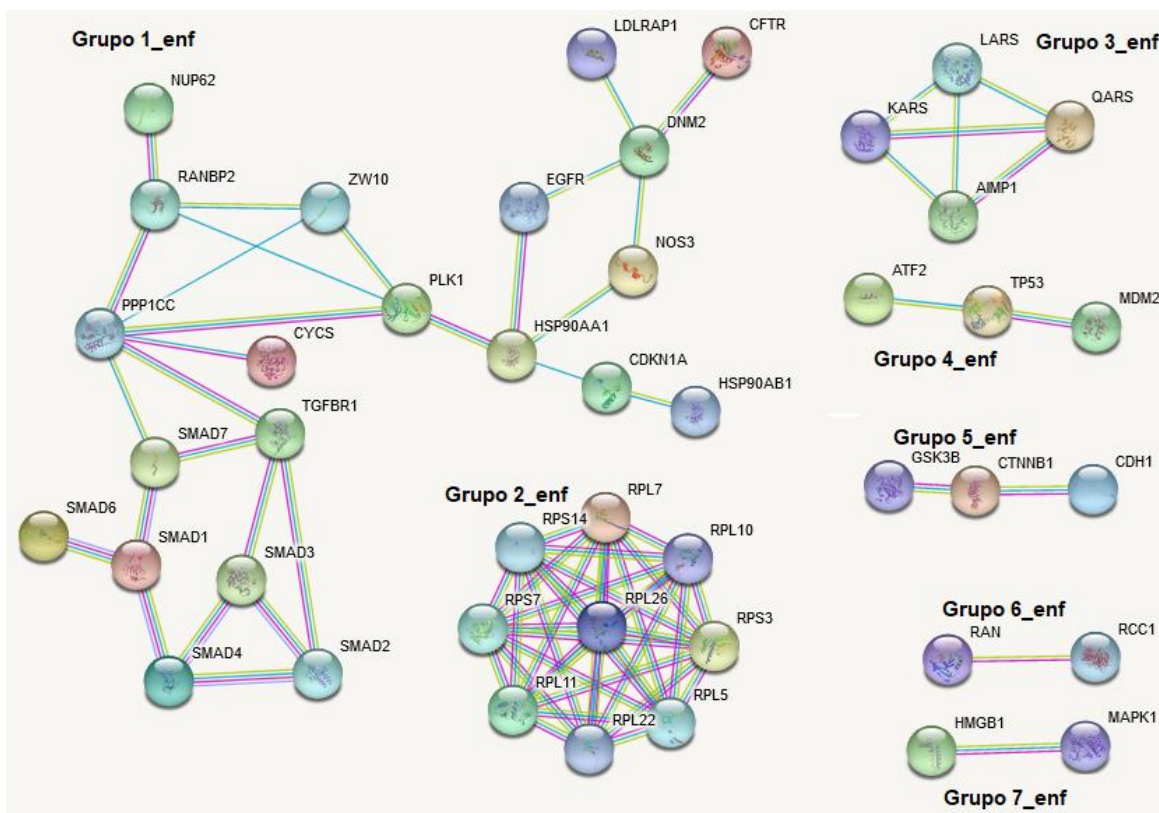


Figura 9. Red de interacción de proteínas moonlighting humanas ligadas a enfermedades ([enlace](#)). 123 proteínas

Tabla 5. Grupos de interacción entre proteínas moonlighting humanas relacionadas a enfermedades. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de [BioGrid](#).

G1_enf		G2_enf		G4_enf	
CDKN1A* [!]	335	RPL5	320	ATF2* ^{!!!}	227
CFTR	859	RPL7	333	MDM2	601
CYCS*	86	RPL10	458	TP53	1,379
DNM2*	173	RPL11	370		
EGFR	1,360	RPL22	296	G5_enf	
HSP90AA1	995	RPL26	179	CDH1	659
HSP90AB1	527	RPS3	406	CTNNB1	730
LDLRAP1*	30	RPS7	267	GSK3B*	577
NOS3* ^{!!}	88	RPS14	337		
NUP62	220			G6_enf	
PLK1	374	G3_enf		RAN	229
PPP1CC	665	AIMP1	147	RCC1	137
RANBP2	215	KARS	224		
SMAD1	168	LARS	245	G7_enf	
SMAD2	323	QARS	295	HMGB1	312
SMAD3	380			MAPK1	364
SMAD4	266				
SMAD6	51				
SMAD7	103				
TGFBR1*	211				
ZW10*	73				

* No aparecen en la red de interacción entre proteínas moonlighting humanas.

[!] Proteína sinónima con PAK3. ^{!!} Proteína sinónima con NANOS3. ^{!!!} Proteína sinónima con GDNF.

(izquierda) está involucrada con transporte mediado por vesículas, organización de membrana, morfogénesis celular, procesos metabólicos de lípidos y procesos metabólicos de moléculas pequeñas.

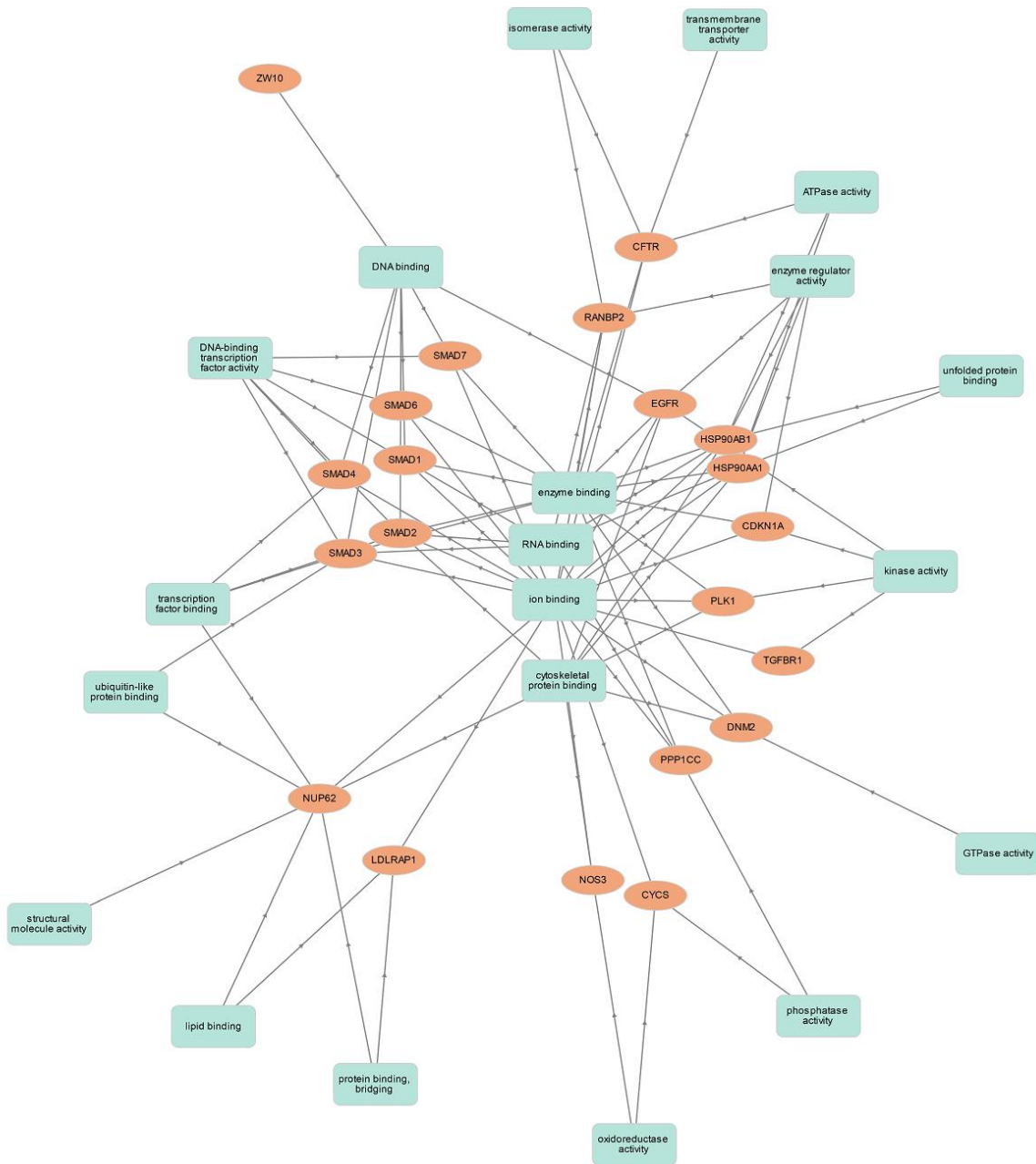


Figura 11. Mecanismos moleculares asociados a los genes del Grupo 1_enf. ([enlace](#))

Los procesos más comunes en cada gen se observan cerca del centro (Fig. 10): Procesos del sistema inmune, procesos homeostáticos, crecimiento, procesos biosintéticos, procesos metabólicos de compuestos de nitrógeno celular, desarrollo de estructuras anatómicas, diferenciación celular, respuesta al estrés, transducción

de señal, reproducción, procesos simbióticos, procesos de modificación celular de proteínas, organización de la mitocondria, ensamblado de complejos que contienen proteína. Otros procesos tienen menos conexiones, como locomoción (locomotion, arriba, centro) y envejecimiento (aging, abajo, centro), que se relacionan con solo dos genes. De igual manera se pueden revisar las funciones moleculares (Fig. 11), o las localizaciones celulares (Fig. 12).

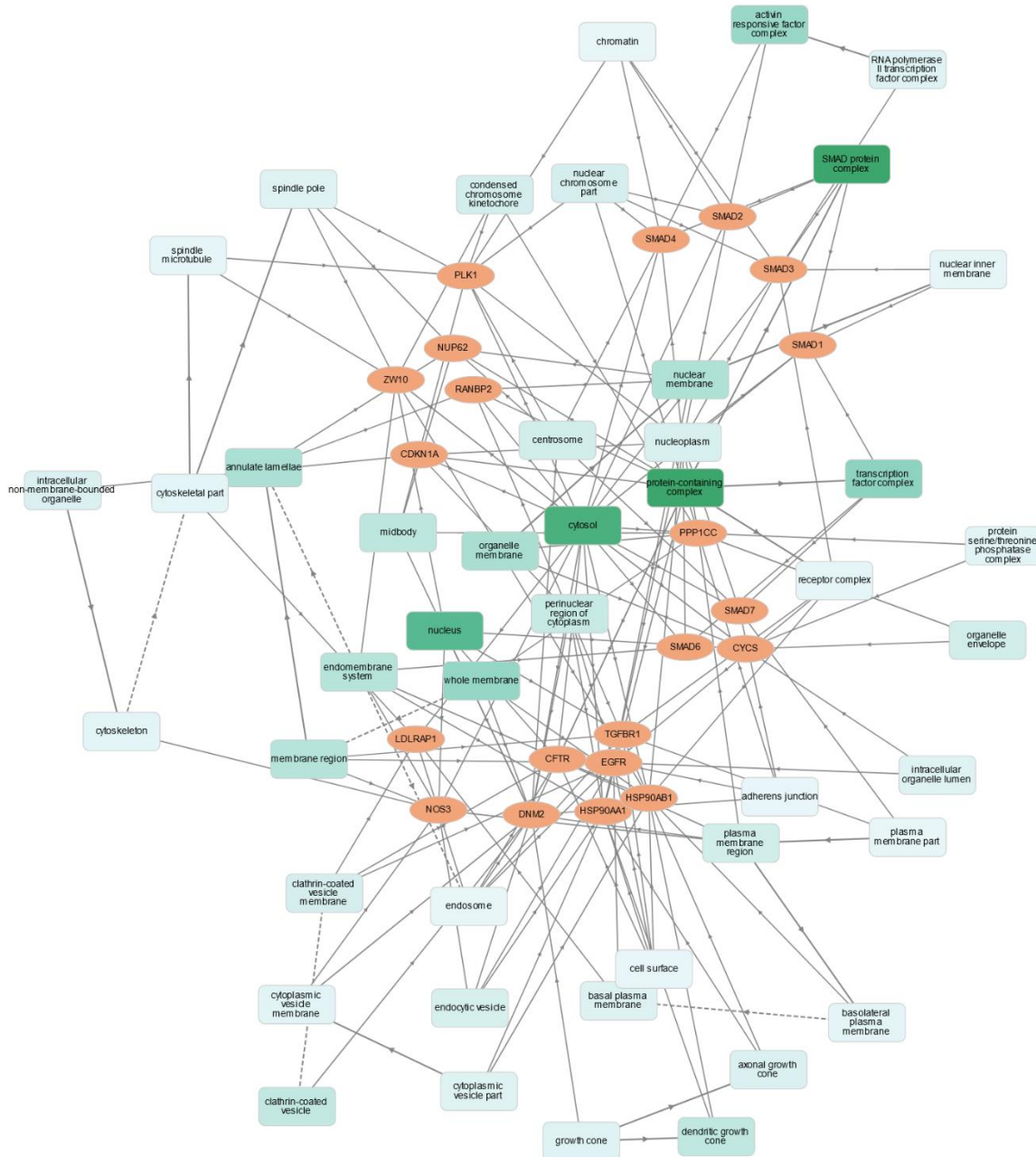


Figura 12. Compartimientos celulares asociados a los genes del grupo 1_enf. ([enlace](#))

Los mecanismos moleculares (Fig. 11) más comunes son de unión (i. e. regulación de la transcripción): a enzimas, a RNA, a iones y a proteínas citoesqueléticas. También es interesante notar que el grupo de SMAD (izquierda, arriba) muestra actividad de factor de transcripción de unión a DNA, y unión a DNA, junto con ZW10. Las otras actividades enzimáticas tienen que ver con metabolismo energético: oxidorreductasa, fosfatasa, GTPasa, cinasas (derecha). La mayoría de los mecanismos restantes también son de unión a DNA, a factores de transcripción, a proteínas similares a ubiquitina, y a lípidos. El resto de las funciones son: actividad de molécula estructural, actividad isomerasa y transporte transmembrana.

Finalmente, se investigó la localización celular (Fig. 12) de las proteínas expresadas por los genes del grupo 1_enf. Mientras más oscuro es el color verde, menor el p valor: así podemos estar seguros que las proteínas SMAD se asocian a su complejo; y que la mayoría de las proteínas están en el citosol; NOS3, ZW10 y SMAD6 están en núcleo. También se observan muchas interacciones en zonas como nucleoplasma, membrana nuclear, centrosoma, membranas de orgánulos y nuclear.

Las tres imágenes anteriores nos permiten entender detalles de los productos genéticos del Grupo_1enf.

Redes de interacción

Luego de seguir el procedimiento indicado para los 7 grupos indicados en la ilustración 13 y la tabla 4, se llegó a las siguientes redes de interacción, mostradas a continuación (Tabla 5 y Fig. 13).

Tabla 6. Redes de interacciones de proteínas para los grupos de proteínas moonlighting humanas asociadas a enfermedades

Grupo	Enlace	Interacciones	Genes
G1_enf	*	-	4612
G2_enf	enlace	1029	1063
G3_enf	enlace	478	551
G4_enf	enlace	1735	1790
G5_enf	enlace	1612	1612
G6_enf	enlace	297	318
G7_enf	enlace	624	636

* Para el momento de la redacción del trabajo, STRING soportaba máximo 2000 interacciones, lo que impidió la realización de la red de interacciones para el grupo 1.

Este grupo correspondía a proteínas ribosómicas de tipo L y S. Se observan hacia el centro, arriba-medio (Fig. 13). Se pueden notar algunos nodos que concentran

muchas interacciones que comunican la red de proteínas ribosómicas con otros elementos: destaca CDC5L (Cell division cycle 5 like protein) en el centro de la imagen, que une al grupo con el resto de interactores de la parte inferior.

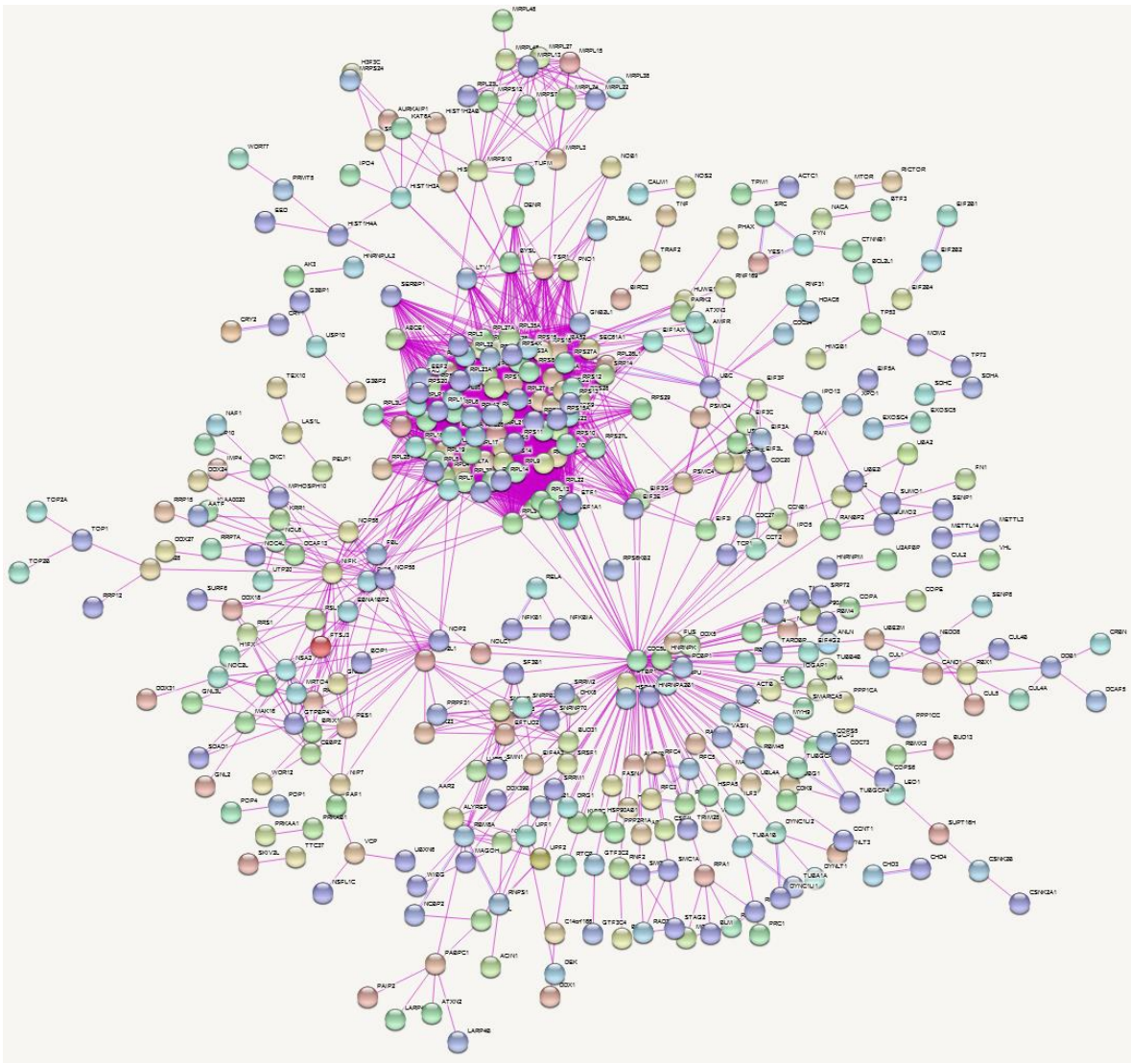


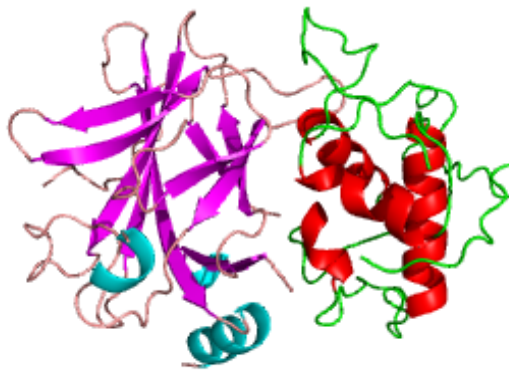
Figura 13. Red de interacción de proteínas del grupo 2_enf. ([enlace](#)), 1029 interacciones.

Arriba al centro, MRPL3 (Mitochondrial Ribosomal Protein L3) comunica al resto de las proteínas ribosomales mitocondriales con el grupo principal. Las otras interacciones ocurren por cúmulos de varios interactores, como el formado por FBL (Fibrilarina) y dos proteínas nucleolares (56 y 58), en la parte inferior izquierda, y cuatro ubiquitina-ligasas (AMFR, ATXN3, PARK2, HUWE1), que junto con UBC unen al grupo principal con el resto de las proteínas ubicadas en la parte inferior.

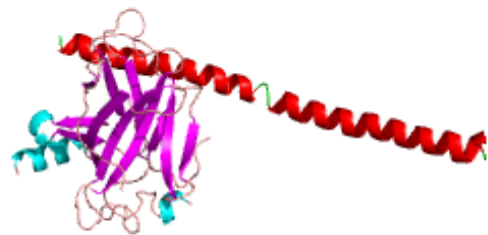
Es importante notar que los procesos biológicos en los que están involucrados más componente de este grupo son biogénesis ribosomal, traducción, targeting de proteínas, procesos simbióticos y procesos catabólicos que contienen bases nitrogenadas. Los compartimientos celulares en los que más componentes se encuentran involucrados son: núcleo, nucleoplasma, ribosoma, nucleolo, y citosol. Las funciones moleculares más relevantes son unión a RNA mensajero y ribosomal, y como constituyente estructural del ribosoma y unión a enzimas (RPL11, RPL5, RPS7 y RPS3).

Docking con ClusPro 2.0

ClusPro 2.0(55) generó varios modelos para cada pareja de receptor y ligando, según la formación de sitios de unión o interacción con baja energía. Las interacciones modeladas fueron de TP53 (Morado y azul) con ATF2 (Rojo y Verde) y TP53 (Morado y azul) con MDM2 (Rojo y Verde) (trabajo número 580167). El trabajo de TP53 y ATF2 dio 22 modelos de docking, mientras que el de TP53 y MDM2 dio 30 resultados. Los datos de cada interacción están como archivos adjuntos. Abajo se observan los 2 modelos que muestran la interacción con menor energía.



TP53 Cadena A y MDM2 Cadena A



TP53 Cadena A y ATF2 Cadena D

Figura 14. Ejemplos de Docking de TP53 Cadena A con MDM2 Cadena A, y ATF2 Cadena D modelado con ClusPro 2.0 en un solvente balanceado.

La interacción de TP53 y de MDM2 se ha documentado anteriormente en 11 ocasiones en Homo sapiens (PMIDs: [24207125](#), [21857681](#), [19255450](#), [19234109](#), [18316739](#), [18172499](#), [17470788](#), [17438265](#), [17080083](#), [16432196](#), [926727](#)) y en 2 rutas curadas. Se puede verificar haciendo clic en la línea morada entre TP53 y MDM2 en el siguiente [enlace](#), correspondiente a la fig. 9. En la misma consulta de STRING mencionada anteriormente no se encontraron interacciones de TP53 y

ATF2 en humanos, aunque hay numerosos reportajes en otras especies.

Una revisión manual de la interacción de TP53 y ATF2 da 8 resultados: Algunas fuentes no fueron relevantes(56)(62), pero otras aportan ciertas ideas sobre los mecanismos de acción en que se involucra la interacción de estas dos proteínas: Implicación en "Angiosarcoma cutáneo"(56); activación de RAS-GTPasa(57); proteínas que interactúan con DNA o mRNA(59); genes de "resistencia a los fármacos" en el carcinoma hepatocelular (60); papel de ATF2 en el aumento independiente de TP53 en la expresión de KIN (proteína de unión a RNA y DNA) después de la irradiación gamma(61); activación de p38 MAPK y el aumento de las actividades de unión al DNA de ATF-2 y p53 median la sobreexpresión de ciclooxigenasa-2 en la senescencia prematura en fibroblastos pulmonares IMR-90(63).

DISCUSIÓN Y CONCLUSIONES

Como se mencionó anteriormente, la proteómica cambió de un enfoque en moléculas específicas a una perspectiva integrada, donde las proteínas son actores dinámicos con varias funciones(21). Esto trae varios retos: obtener información utilizable sobre la gran cantidad de secuencias e interrelaciones accesibles actualmente, y manejar otros factores que limitan la usabilidad de la información disponible.

Interacción entre proteínas moonlighting humanas

Al hacer la revisión de literatura, se encontró que las proteínas moonlighting de *E. coli* tienen más interacciones físicas con un número mayor de clases funcionales distintas (64). Además, la misma fuente (64) menciona que tienden a interactuar entre sí, por lo tanto se decidió hacer una revisión de la interacción entre proteínas moonlighting humanas.

Aparentemente, hay una tendencia en las proteínas moonlighting de humanos a interactuar entre sí (Fig. 5, Tabla 2). Una posible explicación yace en que la inversión metabólica para producir una proteína es alta, lo que daría una ventaja selectiva a especies que puedan utilizar una enzima para más de una función.

Un caso bien conocido de proteínas moonlighting humanas es el siguiente: alrededor de una docena de enzimas ubicuas fueron adoptadas para jugar un segundo papel en las células especializadas de la lente del ojo (12). Además, hallazgos recientes (12) indican que factores celulares que afectan la función de proteínas moonlighting soportan esta hipótesis de “economía metabólica”.

Es decir, la reutilización de una estructura preexistente, con plegamiento, estructura y estabilidad seleccionados evolutivamente, puede cambiar de función con alguno de los cambios indicados en la ilustración 1 (unión a compuestos pequeños: sustratos, productos, cofactores u otros ligandos; otras proteínas, o complejos multiprotéicos; unión a RNA o DNA, modificaciones post-traduccionales; oligomerización) para permitir la selección de una proteína que cumpla más de una función.

La posibilidad de modular distintas funciones con cambios que requieren poca energía es una ventaja adaptativa considerable, lo que favorecería a un organismo capaz de utilizar proteínas multifuncionales. Otra posible ventaja de la multifuncionalidad, además de la reutilización de secuencias previamente seleccionadas, tiene que ver con un requerimiento de menor cantidad de genes (i. e. el organismo podría ejecutar más funciones con la misma cantidad de genes que otro que no presente multifuncionalidad de proteínas).

Otro elemento a considerar es que la tasa evolutiva de una proteína está dictada por su nivel de expresión (65), lo que deja a proteínas expresadas constitutivamente en una posición favorable en términos de ser seleccionada como proteína multifuncional.

Interacción entre proteínas moonlighting asociadas a enfermedades humanas

No se observaron suficientes evidencias para afirmar que el número de clases funcionales de proteínas moonlighting en humanos es distinto al del resto de los organismos (Fig. 7).

Dos eventos son independientes si y solo si, el odds ratio es igual a 1, es decir, las probabilidades de un evento son las mismas en presencia o ausencia del otro evento. Si el odds ratio es mayor que 1, entonces A (la proteína es moonlighting) y B (la proteína está vinculada a una enfermedad) están asociados (correlacionados) en el sentido de que, en comparación con la ausencia de B, la presencia de B aumenta las probabilidades de A, y simétricamente la presencia de A aumenta las probabilidades de B. Por el contrario, si el OR es menor que 1, entonces A y B están

correlacionados negativamente y la presencia de un evento reduce las probabilidades del otro evento (47,48).

Es posible que, una alteración que impida las funciones de una proteína multifuncional perturbe varios sistemas, y tenga consecuencias más serias que la pérdida de función de una proteína con una sola función. Aunado esto a que la mayoría de las proteínas moonlighting son parte de procesos constitutivos (lo que resalta su importancia fisiológica), una perturbación en este tipo de proceso puede ser más seria que una perturbación que involucre una sola función.

Observando la tabla 3, se puede notar que el riesgo absoluto en el grupo tratamiento (0.74) y control (0.08) es menor que 1, lo que implica que es significativo. Asimismo, se puede notar que la reducción absoluta de riesgo no contiene 1 en el intervalo de confianza, lo que implica que el valor es significativo. El riesgo relativo (9.82) corresponde al número de veces que ambos grupos difieren. La reducción del riesgo relativo (-8.82) corresponde a la cantidad de veces que una proteína normal está relacionada con una enfermedad, comparada con las proteínas moonlighting.

Los resultados obtenidos sugieren que las proteínas moonlighting son propensas a estar involucradas en enfermedades humanas (Tabla 3). Esto coincide con la sugerencia de que el fenómeno de multifuncionalidad no es un fenómeno extraño, y que su estudio detallado puede explicar mecanismos moleculares de enfermedades y efectos colaterales de algunos medicamentos(4).

Análisis de códigos GO, funcionalidad de proteínas moonlighting y redes de interacción

Los códigos GO obtenidos a través de la herramienta GONet(82) se utilizaron para analizar las anotaciones de los distintos grupos examinados (grupo 1_enf: Fig. 10, 11, 12. Resto de los grupos: [enlace](#)).

También se examinaron las proteínas multifuncionales presentes tanto en la red de proteínas multifuncionales humanas tenían ciertas características en común, que se presentan en la tabla a continuación:

Tabla 7. Características comunes entre proteínas multifuncionales humanas y proteínas multifuncionales humanas relacionadas a enfermedades

Grupo	Proteínas en común	Procesos biológicos compartidos	Funciones moleculares compartidas	Compartimientos celulares compartidos
1_enf	CFTR, EGFR, HSP90AA1, HSP90AB1, NUP62,	=	=	Complejos que contienen proteínas (GO:0032991)

	NUP98, PPP1CC, RANBP2, SMAD1, SMAD2, SMAD3, SMAD4, SMAD6, SMAD7			
2_enf	RPL10, RPL11, RPL22, RPL26, RPL5, RPL7,RPS14, RPS3, RPS7	Procesos catabólicos compuestos que contienen nucleobases (GO:0034655) Procesos simbióticos (GO:0044403) Targeting de proteínas (GO:0006605)	Constituyente estructural del ribosoma (GO:0003735)	Citosol (GO:0005829) Ribosoma (GO:0005840)
3_enf	AIMP1, KARS, LARS, QARS	Traducción (GO:0006412) Proceso metabólico de aminoácidos celulares (GO:0006520) Proceso metabólico del tRNA (GO:0006399)	=	Citosol (GO:0005829) Complejos que contienen proteínas (GO:0032991)
4_enf	MDM2, TP53	Proceso de modificación de proteínas celulares (GO:0006464) Ciclo celular mitótico (GO:0000278)	Unión a enzimas (GO:0019899) Unión a iones (GO:0043167)	Citosol (GO:0005829) Complejos que contienen proteínas (GO:0032991) Nucleolo (GO:0005730)

		<p>Autofagia (GO:0006914)</p> <p>Diferenciación celular (GO:0030154)</p> <p>Ensamblamiento de complejos que contienen proteínas (GO:0065003)</p> <p>Procesos simbióticos (GO:0044403)</p> <p>Respuesta al estrés (GO:0006950)</p>		<p>Nucleoplasma (GO:0005654)</p>
5_enf	CDH1, CTNNB1	<p>Adhesión celular (GO:0007155)</p> <p>morfogénesis celular (GO:0000902)</p> <p>Diferenciación celular (GO:0030154)</p> <p>Organización de uniones celulares (GO:0034330)</p> <p>Procesos simbióticos (GO:0044403)</p>	=	<p>Membrana plasmática (GO:0005886)</p> <p>Complejos que contienen proteínas (GO:0032991)</p> <p>Espacio extracelular (GO:0005615)</p>
6_enf	RAN, RCC1	<p>Segregación de cromosomas (GO:0007059)</p> <p>Organización del citoesqueleto (GO:0007010)</p> <p>Procesos simbióticos (GO:0044403)</p>	Unión a iones (GO:0043167)	<p>Complejos que contienen proteínas (GO:0032991)</p> <p>Envoltura nuclear (GO:0005635)</p> <p>Nucleoplasma (GO:0005654)</p>

		División celular (GO:0051301)		
7_enf	HMGB1, MAPK1	Diferenciación celular (GO:0030154) Muerte celular (GO:0008219) Procesos homeostáticos (GO:0042592) Procesos del sistema inmune (GO:0002376) Respuesta al estrés (GO:0006950) Transducción de señal (GO:0007165) Transporte mediado por vesículas (GO:0016192)	Unión a enzimas (GO:0019899) Unión a iones (GO:0043167) Unión a factores de transcripción (GO:0008134) Unión a DNA (GO:0003677)	Complejos que contienen proteínas (GO:0032991) Endosoma (GO:0005768) Membrana plasmática (GO:0005886) Nucleoplasma (GO:0005654)

Es importante notar que este cuadro solamente marca los elementos comunes de los genes comunes entre los dos grupos mencionados, buscando facilitar la comprensión de los procesos comunes entre las proteínas moonlighting humanas que interactúan entre sí, y las proteínas moonlighting humanas relacionadas con enfermedades.

Podemos ilustrar la limitación del cuadro considerando el grupo 4_enf (ATF2, MDM2, TP53): En la fig. 15 se puede notar que existen 4 procesos biológicos comunes a los tres genes del grupo: ciclo celular mitótico, procesos de modificación de proteínas celulares, diferenciación celular, respuesta al estrés. Además, hay muchos procesos biológicos en los que están involucrados dos de los tres genes.

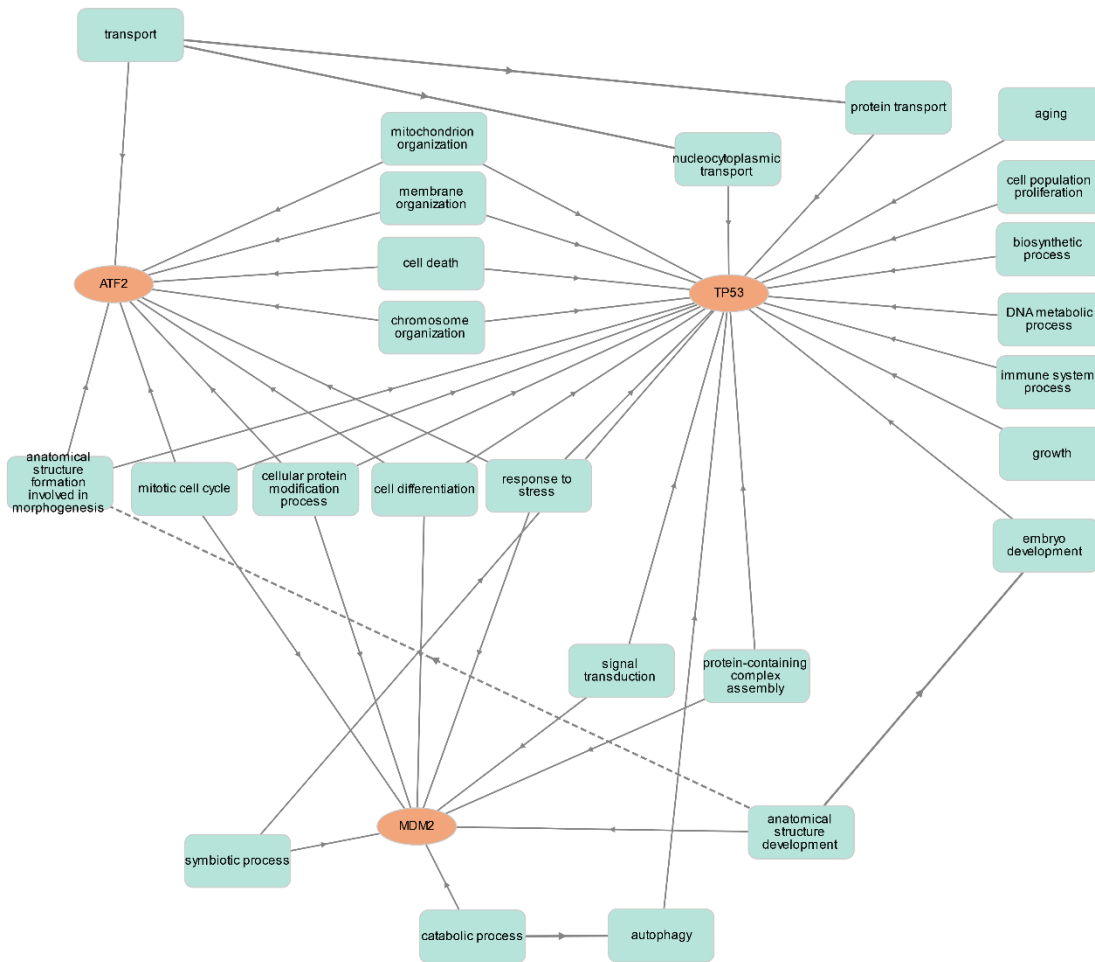


Figura 15. Procesos biológicos asociados a los genes del grupo G4_enf.

El mismo examen puede hacerse para las funciones moleculares (Fig. 16, donde se observan solamente dos funciones comunes: unión a iones y a enzimas) o compartimientos celulares (Fig. 17, donde se observa que sólo comparten localización en el nucleoplasma).

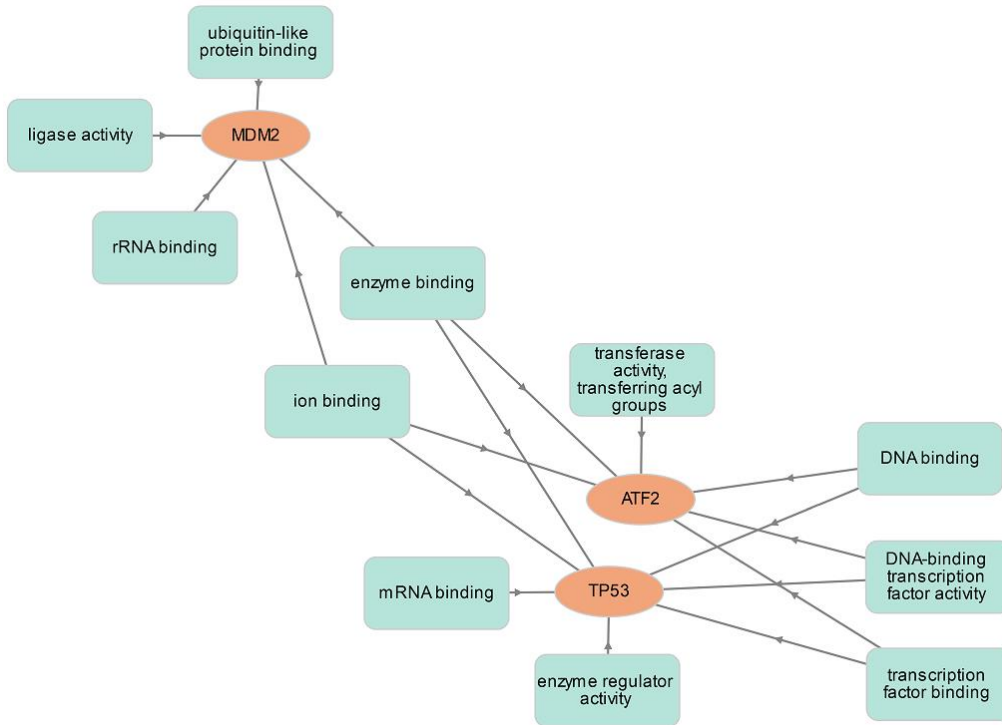


Figura 16. Mecanismos moleculares asociados a los genes del grupo G4_enf.

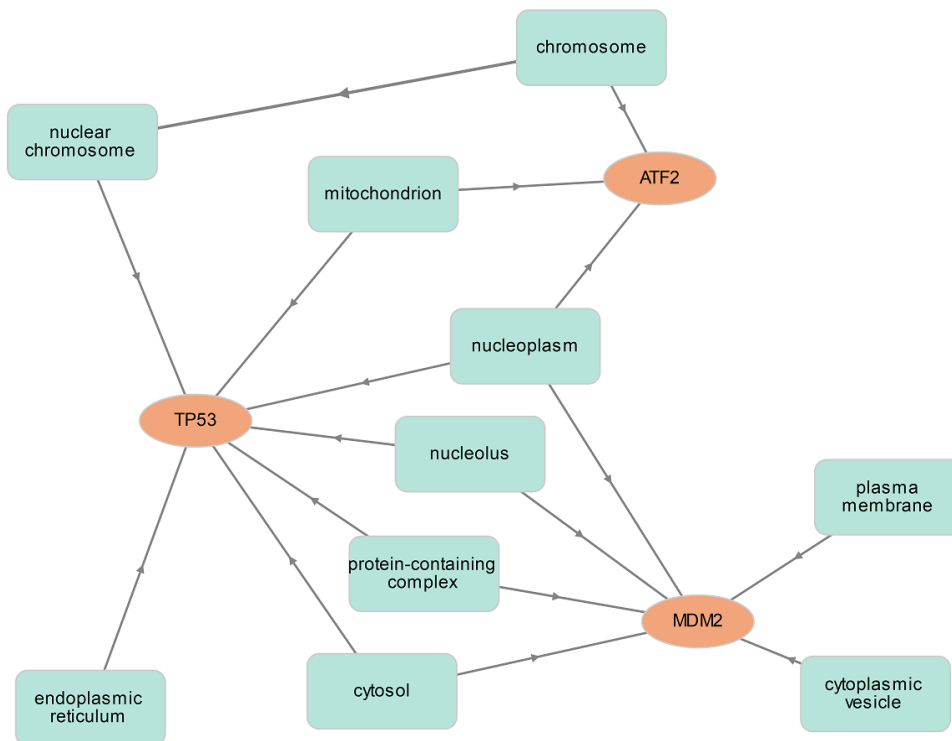


Figura 17. Compartimentos celulares asociados a los genes del grupo G4_enf.

Finalmente, se puede revisar la red de proteínas con la que interactúan las proteínas del grupo 4_enf (fig. 18). Se puede notar que esta red de interacción también llega a las proteínas estructurales ribosómicas del grupo 2_enf (Centro, hacia arriba), se ven las proteínas del grupo 5_enf (abajo, centro), y algunas SMAD del grupo 1_enf (abajo, derecha). Interesantemente, MAPK1 del grupo 7_enf (abajo, izquierda), aparece separada de HGMB1, que a su vez aparece interactuando con TP53 (centro, hacia la derecha). No se observan proteínas de grupo 3_enf, ni del grupo 6_enf.

A pesar de que se observan separaciones en la red de interacción, debe tomarse en cuenta que sólo se muestran las conexiones con score de interacción por encima de 0.9. Esta decisión se tomó para disminuir la cantidad de nodos que se mostraban en la figura y hacerla más comprensible. A pesar de eso, sin tener familiaridad con las proteínas que aparecen en la red, difícilmente puede hacerse una interpretación más detallada.

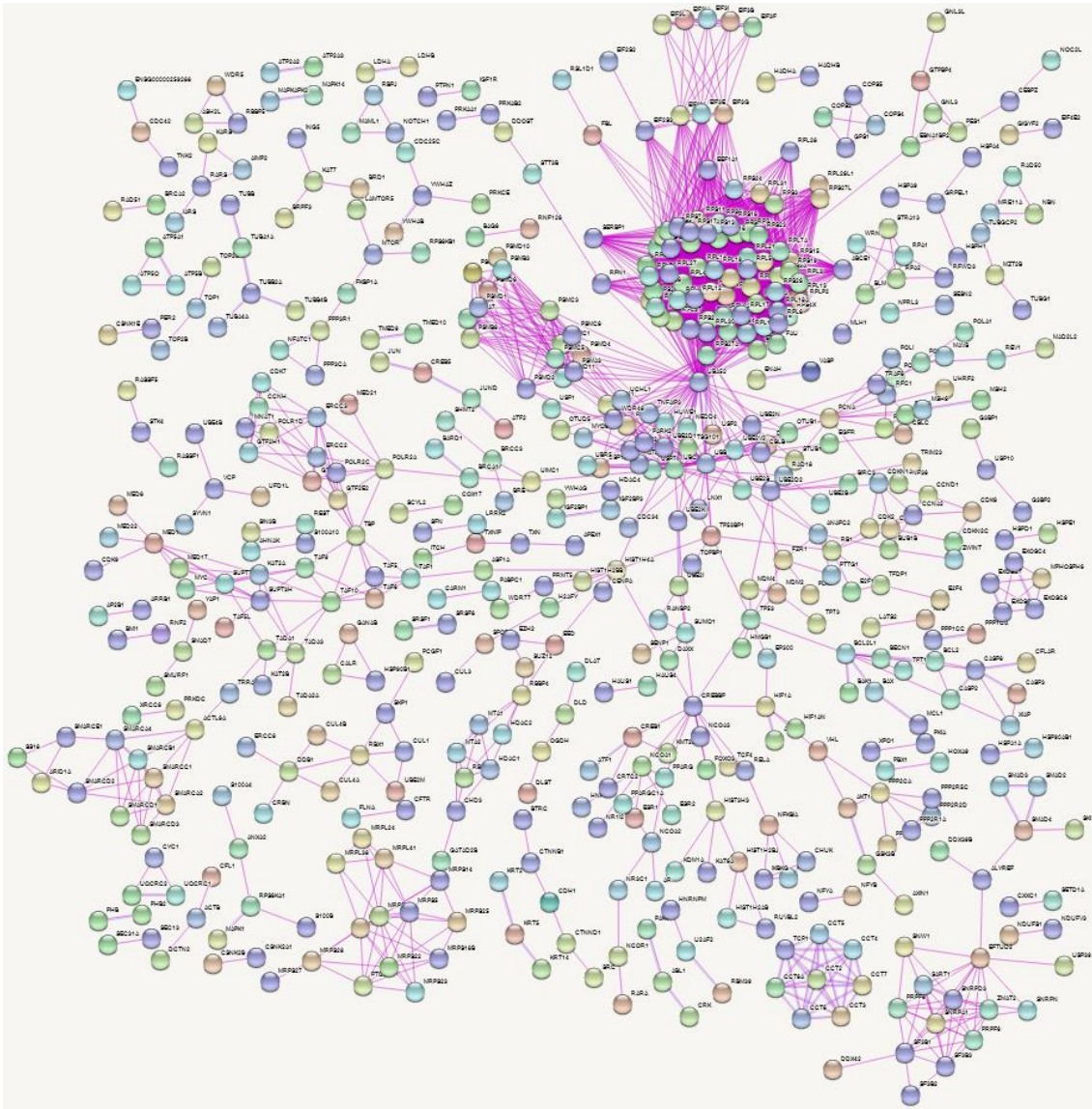


Figura 18. Red de interacción de proteínas del grupo G4_enf. ([enlace](#)) 1735 interacciones.

Parte de los temas que quedan por explorar serían análisis de enriquecimiento de los genes de interés en un caso específico de enfermedad, utilizando los datos de GONet o datos experimentales, o investigar cambios en el interactoma durante procesos patológicos. También pueden tomarse en cuenta análisis de enriquecimiento de tejidos específicos para observar perfiles de expresión más manejables en contextos fisiológicos más comprensibles.

Análisis del Docking

En este caso no se contaba con motivos específicos de interacción, y se utilizó ClusPro2.0 como “prueba de concepto” para encontrar interacciones documentadas con anterioridad (MDM2-TP53 y ATF2-TP53). La parte valiosa del uso de este recurso fue el desarrollo de familiaridad y aprendizaje involucrados, que pueden utilizarse en contextos educativos para ilustrar ejemplos de interacción proteína-proteína con interpretaciones fisiológicas sencillas.

Uno de los temas por completar que sería de interés en cuanto a la interacción entre proteínas moonlighting es realizar el análisis de los modelos propuestos por ClusPro 2.0 para pares de proteínas de interés e identificar los aminoácidos de interacción, con la finalidad de revisar si coinciden con reportes de variantes asociadas a enfermedades, como la sección de Pathology/Biotech de UniProt.

Comentarios finales

Tomando en cuenta los resultados observados se pueden ver esbozos de una amplia red de interacciones, que empieza con proteínas moonlighting humanas, y de estas algunas se relacionan con enfermedades. Así, las proteínas moonlighting parecen parte de una gran red de interacciones de procesos fisiológicos de gran importancia, que suelen ser constitutivos.

La alta cantidad de interacciones que muestran las proteínas moonlighting humanas sugiere que son “encrucijadas” metabólicas, donde confluyen muchos procesos fisiológicos.

La anotación de un repertorio funcional de proteínas humanas y de otras especies ha impactado en los campos de la bioquímica y la genética, y posiblemente seguirá influyendo en preguntas científicas básicas y aplicadas, desde el problema del genotipo al fenotipo hasta nuestra comprensión de patologías humanas y el diseño de fármacos.

REFERENCIAS

1. Jeffery CJ. Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS J.* 2020;287(19):4141–9.
2. Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci* [Internet]. 1999 Jan;24(1):8–11. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0968000498013358>
3. Jeffery CJ. What is Protein Moonlighting and Why is it Important? *Moonlighting Proteins Nov Virulence Factors Bact Infect.* 2016;1–19.
4. Franco-Serrano L, Huerta M, Hernández S, Cedano J, Perez-Pons JA, Piñol J, et al. Multifunctional Proteins: Involvement in Human Diseases and Targets of Current Drugs. *Protein J* [Internet]. 2018;37(5):444–53. Available from: <http://dx.doi.org/10.1007/s10930-018-9790-x>
5. Henderson B, Fares MA, Martin ACR. Protein Moonlighting in Biology and Medicine. *Protein Moonlighting in Biology and Medicine.* 2016. 1–317 p.
6. Unal CM, Steinert M. Microbial Peptidyl-Prolyl cis/trans Isomerases (PPIases): Virulence Factors and Potential Alternative Drug Targets. *Microbiol Mol Biol Rev.* 2014;78(3):544–71.
7. Rasch J, Theuerkorn M, Ünal C, Heinsohn N, Tran S, Fischer G, et al. Novel cycloheximide derivatives targeting the moonlighting protein Mip exhibit specific antimicrobial activity against *Legionella pneumophila*. *Front Bioeng Biotechnol.* 2015;3(MAR):1–8.
8. Guleria A, Chandna S. ATM kinase: Much more than a DNA damage responsive protein. *DNA Repair (Amst)* [Internet]. 2016;39:1–20. Available from: <http://dx.doi.org/10.1016/j.dnarep.2015.12.009>
9. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I, et al. Bioinformatics and moonlighting proteins. *Front Bioeng Biotechnol.* 2015;3(JUN).
10. Lievens S, Eyckerman S, Lemmens I, Tavernier J. Large-scale protein interactome mapping: Strategies and opportunities. *Expert Rev Proteomics.* 2010;7(5):679–90.
11. Copley SD. Moonlighting is mainstream: Paradigm adjustment required. *BioEssays.* 2012;34(7):578–88.
12. Jeffery CJ. An enzyme in the test tube, and a transcription factor in the cell: Moonlighting proteins and cellular factors that affect their behavior. *Protein Sci.* 2019;28(7):1233–8.

13. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
14. Gómez A, Hernández S, Amela I, Piñol J, Cedano J, Querol E. Do protein-protein interaction databases identify moonlighting proteins? *Mol Biosyst.* 2011;7(8):2379–82.
15. Higurashi M, Ishida T, Kinoshita K. PiSite: A database of protein interaction sites using multiple binding states in the PDB. *Nucleic Acids Res.* 2009;37(SUPPL. 1):360–4.
16. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc [Internet].* 2015 Jun 7;10(6):845–58. Available from: <http://www.nature.com/articles/nprot.2015.053>
17. Colell EA, Iserle JA, Simonetti FL, Marino-Buslje C. MISTIC2: Comprehensive server to study coevolution in protein families. *Nucleic Acids Res.* 2018;46(W1):W323–8.
18. Caldera M, Buphamalai P, Müller F, Menche J. Interactome-based approaches to human disease. *Curr Opin Syst Biol [Internet].* 2017;3:88–94. Available from: <http://dx.doi.org/10.1016/j.coisb.2017.04.015>
19. Gonzalez MW, Kann MG. Chapter 4: Protein Interactions and Disease. *PLoS Comput Biol.* 2012;8(12).
20. Kuzmanov U, Emili A. Protein-protein interaction networks: Probing disease mechanisms using model systems. *Genome Med.* 2013;5(4):1–12.
21. Espinosa-Cantú A, Cruz-Bonilla E, Noda-Garcia L, DeLuna A. Multiple Forms of Multifunctional Proteins in Health and Disease. *Front Cell Dev Biol.* 2020;8(June):1–10.
22. Jeffery CJ. Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS J.* 2020;287(19):4141–9.
23. Jeffery CJ. Protein moonlighting: what is it, and why is it important? *Philos Trans R Soc B Biol Sci.* 2018 Jan 19;373(1738):20160523.
24. Alonso-López Di, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, et al. APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database.* 2019;2019(i):1–8.
25. Breitkreutz BJ, Stark C, Tyers M. The GRID: The General Repository for

- Interaction Datasets. *Genome Biol.* 2002;3(12).
26. Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D529–41.
 27. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36(SUPPL. 1):901–6.
 28. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074–82.
 29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology David. *Nat Genet.* 2000;25(1):25–9.
 30. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019;47(D1):D330–8.
 31. Cooper DN, Ball E V, Krawczak M. The human gene mutation database. *Nucleic Acids Res.* 1998;26(1):285–7.
 32. Stenson PD, Mort M, Ball E V., Chapman M, Evans K, Azevedo L, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet [Internet].* 2020;139(10):1197–207. Available from: <https://doi.org/10.1007/s00439-020-02199-3>
 33. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, et al. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat Methods.* 2012;9(4):345–50.
 34. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* 2009;37(SUPPL. 1):211–5.
 35. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 2019;47(D1):D351–60.
 36. Franco-Serrano L, Hernández S, Calvo A, Severi MA, Ferragut G, Pérez-Pons JA, et al. MultitaskProtDB-II: An update of a database of multitasking/moonlighting proteins. *Nucleic Acids Res.* 2018;46(D1):D645–8.
 37. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online

- Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33(DATABASE ISS.):514–7.
38. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallogr Sect D Biol Crystallogr.* 2002;58(6 I):899–907.
 39. Qin C, Zhang C, Zhu F, Xu F, Chen SY, Zhang P, et al. Therapeutic target database update 2014: A resource for targeted therapeutics. *Nucleic Acids Res.* 2014;42(D1):1118–23.
 40. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, et al. Therapeutic target database 2020: Enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 2020;48(D1):D1031–41.
 41. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res.* 2019;18(2):623–32.
 42. Pomaznoy M, Ha B, Peters B. GOnet: A tool for interactive Gene Ontology analysis. *BMC Bioinformatics.* 2018;19(1):1–8.
 43. Hassan H, Shanak S. GOTrapper: A tool to navigate through branches of gene ontology hierarchy. *BMC Bioinformatics.* 2019;20(1):1–6.
 44. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12(2):255–78.
 45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res [Internet].* 2003 Nov;13(11):2498–504. Available from: <http://ci.nii.ac.jp/naid/110001910481/>
 46. Ding Z, Kihara D. Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features. *Curr Protoc protein Sci.* 2018;93(1):e62.
 47. Stewart A. *Basic Statistics and Epidemiology: A Practical Guide: Health Education Research.* Oxford: Radcliffe Medical Press; 2002. 151 p.
 48. Pita-Fernández S, Vila-Alonso M, Carpena Montero J. Determinación de factores de riesgo. *Cad Aten Primaria.* 1997;4:75–8.
 49. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics.* 2009 Nov;25(22):3045–6.

50. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33(DATABASE ISS.):514–7.
51. National Center for Biotechnology Information. MedGen [Internet]. National Library of Medicine (US). 2004.
52. Stevenson M, Nunes T, Sanchez J, Thornton R, Reiczigel J, Robison-Cox J, et al. epiR: An R package for the analysis of epidemiological data. 2013.
53. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ.* 1986 Mar;292(6522):746–50.
54. The Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucleic Acids Res [Internet].* 2008 Jan 1;36(suppl_1):D440–4. Available from: https://academic.oup.com/nar/article/36/suppl_1/D440/2507489
55. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12(2):255–78.
56. Chen S-Y, Takeuchi S, Urabe K, Hayashida S, Kido M, Tomoeda H, et al. Overexpression of phosphorylated-ATF2 and STAT3 in cutaneous angiosarcoma and pyogenic granuloma. *J Cutan Pathol.* 2008 Aug;35(8):722–30.
57. von Elsner L, Hagemann S, Just I, Rohrbeck A. C3 exoenzyme impairs cell proliferation and apoptosis by altering the activity of transcription factors. *Naunyn Schmiedebergs Arch Pharmacol.* 2016;389(9):1021–31.
58. Elam MB, Majumdar G, Mozhui K, Gerling IC, Vera SR, Fish-Trotter H, et al. Patients experiencing statin-induced myalgia exhibit a unique program of skeletal muscle gene expression following statin re-challenge. *Artero R, editor. PLoS One.* 2017 Aug;12(8):e0181308.
59. Taguchi Y-H. Tensor decomposition-based unsupervised feature extraction identifies candidate genes that induce post-traumatic stress disorder-mediated heart diseases. *BMC Med Genomics.* 2017 Dec;10(S4):67.
60. Ding S. Methylation profile of the promoter CpG islands of 14 “drug-resistance” genes in hepatocellular carcinoma. *World J Gastroenterol.* 2004;10(23):3433.
61. Masson C, Menea F, Pinon-Lataillade G, Frobert Y, Radicella JP, Angulo JF. Identification of KIN (KIN17), a human gene encoding a nuclear DNA-binding protein, as a novel component of the TP53-independent response to ionizing radiation. *Radiat Res.* 2001;156(535):544.
62. Pratheeshkumar P, Sheeja K, Kuttan G. Andrographolide induces apoptosis

in B16F-10 melanoma cells by inhibiting NF- κ B-mediated bcl-2 activation and modulating p53-induced caspase-3 gene expression. *Immunopharmacol Immunotoxicol.* 2012 Feb;34(1):143–51.

63. Zdanov S, Toussaint O, Debacq-Chainiaux F. p53 and ATF-2 partly mediate the overexpression of COX-2 in H₂O₂-induced premature senescence of human fibroblasts. *Biogerontology.* 2009 Jun;10(3):291–8.
64. Khan I, Chen Y, Dong T, Hong X, Takeuchi R, Mori H, et al. Genome-scale identification and characterization of moonlighting proteins. *Biol Direct.* 2014;9(1).
65. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. *Nat Rev Genet.* 2015 Jul;16(7):409–20.

AGRADECIMIENTOS

Quiero mostrar mi gratitud a todas aquellas personas que estuvieron presentes en la realización de este proyecto, y agradecer su ayuda, palabras motivadoras, conocimientos, consejos y su dedicación.

A mis compañeros de Vicerrectoría, quienes a través de tiempo hemos fortalecido una amistad y creado un equipo de trabajo coordinado, muchas gracias por toda su colaboración, por convivir todo este tiempo conmigo, por compartir experiencias, alegrías, frustraciones, tristezas, retos, celebraciones y múltiples factores que ayudaron a que hoy seamos muy unidos. Por su confianza, y por crecer juntos día a día, mil gracias.

Finalmente, quiero agradecer especialmente a mi familia, en especial a mi esposa, y a mi madre.

ANEXOS

[Biogrid](#)

[Docking ClusPro2.0](#)

[GONet](#)

[GONet Grupos enf](#)

[Código R](#)

[Tabla enfermedades GO](#) (Fig. 8)

Vicerrectoría de Investigación y Proyección Social

Para conocer más
sobre esta iniciativa
puedes escribirnos o
llamarnos por

Whatsapp

2645-6500

WWW.UGB.EDU.SV

Campus San Miguel
Campus Usulután

