

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO



Escuela Superior de Apan

TESIS DE LICENCIATURA

Caracterización de la función y localización de la proteína C19orf47 en cáncer cervicouterino

Para obtener el título de
Licenciada en Ingeniería en Biociencias

P R E S E N T A

Aranza Xareny Lozano Sánchez

Director

Dr. Sergio Manuel Encarnación Guevara

Codirector

Dr. Martín Peralta Gil

Apan, Hidalgo., México a 11 de Julio de 2024

Universidad Autónoma del Estado de Hidalgo

Escuela Superior de Apan

Caracterización de la función y localización de la proteína C19orf47 en cáncer
cervicouterino

Tesis

Para obtener el título de la Licenciada en Ingeniería en Biociencias

Presenta

Aranza Xareny Lozano Sánchez

Director

Dr. Sergio Manuel Encarnación Guevara

Codirector

Dr. Martín Peralta Gil

Apan, Hidalgo

Julio, 2024

El presente trabajo se realizó en el Laboratorio de Proteómica del Centro de Ciencias Genómicas perteneciente a la Universidad Nacional Autónoma de México, campus Morelos, bajo la dirección de Dr. Sergio Manuel Encarnación Guevara y codirección de Dr. Martín Peralta Gil.

Este trabajo fue financiado por el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT-UNAM), donativo IN-213522 a S. E-G. El cual me otorgó una beca de estudio de licenciatura, apoyo que agradezco ampliamente.



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
Escuela Superior de Apan
Campus Apan

Apan, Hgo., a 11 de julio de 2024
Número de control: ESAP-290/2024
Asunto: Autorización de impresión de tesis

Mtra. Ojuky del Rocío Islas Maldonado
Directora de Administración Escolar
PRESENTE

Por este medio le notifico que el jurado asignado a la pasante de la licenciatura en Ingeniería en Biociencias **Aranza Xareny Lozano Sánchez** (360275), a quienes les presentó el trabajo para su titulación en la modalidad de tesis "**Caracterización de la función y localización de la proteína C19orf47 en cáncer cervicouterino**", después de su revisión autorizó la impresión.

A continuación, se presentan las firmas de conformidad de los miembros de este jurado:

PRESIDENTE: Dr. Sergio Manuel Encarnación Guevara

SECRETARIO: Dr. Martín Peralta Gil

PRIMER VOCAL: Dr. Víctor Hugo Pérez España

SEGUNDO VOCAL: Dr. Jaime Alioscha Cuervo Parra

TERCER VOCAL: Dra. Teresa Romero Cortes

SUPLENTE 1: Dra. María del Pilar Gutiérrez Amador

SUPLENTE 2: Dra. Lis Tamayo Rivera

ATENTAMENTE
"Amor, Orden y Progreso"



Dr. Jorge Zuno Silva
Director de la Escuela Superior de Apan



Carretera Apan-Calpulalpan Km. 8,
Colonia Chimalpa Tlalayote, Apan, Hidalgo; C.P. 43900
Teléfono: 771 71 72000 ext. 5800, 5801
esap@uaeh.edu.mx



www.uaeh.edu.mx

DEDICATORIA

Esta tesis se la dedico con todo mi amor y gratitud a quienes han sido el ancla y los faros en mi vida.

Primero, a Dios, mi roca eterna, por guiarme y darme la fuerza para perseverar en cada paso.

A mi mamá, Erika, el pilar inquebrantable en el que siempre he encontrado refugio. Cada palabra de aliento, cada enseñanza y todo tu amor han sido las semillas de mi fortaleza. Sin tu apoyo y amor incondicional, que han sido mi ancla en las tormentas más feroces, simplemente no habría llegado hasta aquí.

A mi dulce hermana, Atziry, por ser mi eterna compañera y confidente. Desde el día en que llegaste al mundo, has traído una alegría y un color indescriptibles a mi vida. Tu risa, tu amor y tu apoyo constante han sido una fuente de luz y esperanza en mi vida.

A mi amado abuelo, Roberto, cada paso que he dado en este viaje lleva la huella de tu fe en mí, tus consejos y de las lecciones que me has enseñado con tu propio ejemplo de fortaleza y dedicación. Sin tu amor, tu guía y tu apoyo, este sueño no habría sido posible.

A mi querida Anene, por cada consejo lleno de sabiduría, por ser una fuente inagotable de fuerza y amor. Aunque ya no estés aquí físicamente, tu espíritu y tus enseñanzas siguen iluminando mi camino y dando forma a quien soy hoy.

A Fernando, tu amor, paciencia y aliento me han dado la fortaleza para seguir adelante incluso en los días más oscuros. Gracias por creer en mí y por estar siempre ahí para celebrar mis logros y ofrecerme consuelo en los momentos difíciles.

Y a mi fiel manada, por acompañarme en cada desvelo, victoria y obstáculo, por brindarme una pata amorosa en todo momento. Ustedes han sido mi constante compañía y fuente de alegría durante todo este proceso.

Gracias a cada uno de ustedes por ser las piedras angulares de mi vida y por hacer posible este logro. Esta tesis es un tributo a su amor, a su dedicación y a la huella imborrable que han dejado en mi corazón.

AGRADECIMIENTOS

Quiero agradecer a

Al Dr. Sergio Manuel Encarnación Guevara por abrirme las puertas de su laboratorio y brindarme la oportunidad de formar parte de su equipo de trabajo. Agradezco su apoyo, dedicación, orientación y paciencia. Así como la confianza y el respaldo que me ha otorgado.

Al Dr. Martín Peralta Gil por la confianza depositada en mí e impulsarme a crecer. Agradezco infinitamente la motivación y el apoyo que me ha brindado desde el primer día. Así como sus consejos y el valioso conocimiento que me ha compartido.

A la maestra Magdalena Hernández Ortiz. Su apoyo incondicional, y palabras de aliento, así como sus valiosos consejos, han sido de gran ayuda durante este proceso. Aprecio enormemente el conocimiento que compartió conmigo y su asistencia en el desarrollo del proyecto. Su experiencia y respaldo han sido esenciales para superar los desafíos y alcanzar los objetivos propuestos. Gracias por obsequiarme un espacio seguro en un lugar completamente desconocido para mí.

A la Dra. Nilda Sánchez, quien, a pesar de estar lejos, siempre estuvo dispuesta a compartir su conocimiento y a apoyarme con mis dudas. Su disposición y ayuda han sido invaluable.

A la maestra Gloria Angelina Herrera Quitero por recibirme, apoyarme y guiarme desde mi llegada al laboratorio. Agradezco su infinita paciencia, cada una de sus enseñanzas y el conocimiento compartido para el desarrollo del proyecto, así como la motivación brindada para superar todas las metas y los obstáculos a los que nos enfrentamos, su apoyo fue fundamental para este proyecto. Gracias por la confianza, el cariño, la compañía, gracias por todo.

A todos y cada uno de los miembros del laboratorio de proteómica, quienes me recibieron, enseñaron y acompañaron a lo largo de este proceso. Agradezco al técnico Gabriel por su disposición, el conocimiento compartido y la ayuda brindada para el desarrollo de este proyecto. A Ramiro por auxiliarme incluso en las dudas más simples, por su apertura y colaboración en este trabajo, así como sus consejos y motivación.

Agradezco haber tenido el placer de conocerlos, gracias por la experiencia, las risas y el crecimiento.

A todos los profesores que formaron parte de mi formación profesional, así como a la institución que por cuatro años me acogió.

A mi familia que siempre me respaldó y apoyó.

RESUMEN

El Proyecto del Proteoma Humano es una iniciativa internacional que tiene como objetivo realizar una lista del total de proteínas codificadas por el conjunto de aproximadamente 20,000 genes que conforman el genoma humano, mediante técnicas estrictas que fundamenten su evidencia a nivel de proteínas como lo es la espectrometría de masas.

En este sentido se creó el Consorcio C-HPP el cual complementa del Proyecto del Proteoma Humano y que tiene como finalidad el identificar las proteínas codificadas por cada uno de los cromosomas y el genoma mitocondrial de la célula humana. Para lograr lo anterior se utilizan distintas técnicas experimentales y bioinformáticas que sugieran de forma asertiva en que procesos biológicos se encuentra implicadas estas proteínas sin función asignada para mejorar la comprensión de la biología humana. Para facilitar y organizar el trabajo de esta iniciativa, el C-HPP se distribuyó en 25 consorcios, cada uno destinado al estudio de cada uno de los 24 cromosomas humanos y el genoma mitocondrial, incluyendo los cromosomas X y Y. México se incorporó al proyecto adoptando el estudio del total de las proteínas codificadas por el cromosoma 19. Más recientemente se estableció como reto dentro del consorcio general no solo el identificar las proteínas codificadas por los cromosomas, además asignarle una función a cada una de ellas y dar claridad a lo que se conoce como el “proteoma oscuro”, que incluye las proteínas identificadas de manea confiable, pero sin tener una función conocida dentro de la célula, estas proteínas fueron catalogadas como uPE1.

De este modo un estudio proteómico realizado en células de cáncer cervicouterino detectó mediante espectrometría de masas un grupo de 11 proteínas sin función asignada, entre las cuales, C19orf47 destacó por sus niveles de expresión en líneas celulares de cáncer cervicouterino. C19orf47 es una proteína con función desconocida, según se reporta en la base de datos Nextprot, se sugiere

que se encuentra localizada en el nucleoplasma y puede ser blanco de diferentes modificaciones postraduccionales.

Para caracterizar esta proteína con el objetivo de asignarle una función, en el presente trabajo se realizaron múltiples análisis bioinformáticos que destacan su importancia evolutiva y enfatizan su relevancia funcional, igualmente, se realizaron ensayos de inmunofluorescencia con el fin de determinar su localización subcelular. Así mismo se planteó una metodología basada en co-inmunoprecipitación y LC-MS/MS para revelar sus interacciones proteicas y sugerir la función biológica de C19orf47 y sus implicaciones en el cáncer cervicouterino y con el virus del papiloma humano, el cual es el agente causal que se ha observado estar presente en casi la totalidad de los casos de cáncer cervical.

Los resultados mostraron que C19orf47 se encuentra distribuida en el citoplasma y en mayor medida, en el núcleo celular, también se determinó que podría estar involucrada en la respuesta inmunitaria, procesos de señalización celular, adhesión celular, y además destaca su posible participación en la glucólisis por senescencia, sugiriendo una relación con la resistencia a la hipoxia. Además, resalta como C19orf47 podría participar activante en distintos procesos propios del cáncer como la proliferación y migración celular.

ABSTRACT

The Human Proteome Project is an international initiative aimed at creating a comprehensive list of all proteins encoded by the approximately 20,000 genes in the human genome, using rigorous techniques that provide protein-level evidence, such as mass spectrometry.

In this context, the C-HPP Consortium was established to complement the Human Proteome Project. Its goal is to identify the proteins encoded by each of the human chromosomes and the mitochondrial genome. To achieve this, various experimental and bioinformatic techniques are employed to assertively suggest the biological processes in which these uncharacterized proteins are involved, enhancing the understanding of human biology. To facilitate and organize the work of this initiative, the C-HPP is divided into 25 consortia, each focusing on studying one of the 24 human chromosomes and the mitochondrial genome, including the X and Y chromosomes. Mexico joined the project by adopting the study of all proteins encoded by chromosome 19. More recently, the general consortium's challenge has expanded to not only identifying the proteins encoded by chromosomes but also assigning a function to each of them and clarifying what is known as the "dark proteome," which includes proteins reliably identified but with unknown functions within the cell, categorized as uPE1.

In this regard, a proteomic study of cervical cancer cells detected, through mass spectrometry, a group of 11 proteins with unassigned functions. Among these, C19orf47 stood out due to its expression levels in cervical cancer cell lines. C19orf47 is a protein with an unknown function; according to the Nextprot database, it is suggested to be located in the nucleoplasm and may be subject to various post-translational modifications.

To characterize this protein and assign it a function, this study performed multiple bioinformatic analyses highlighting its evolutionary importance and emphasizing its functional relevance. Additionally, immunofluorescence assays were

conducted to determine its subcellular localization. A methodology based on co-immunoprecipitation and LC-MS/MS was proposed to reveal its protein interactions and suggest the biological function of C19orf47 and its implications in cervical cancer and human papillomavirus (HPV), the causative agent observed in nearly all cases of cervical cancer.

The results showed that C19orf47 is distributed in the cytoplasm and, to a greater extent, in the cell nucleus. It was also determined that it might be involved in immune response, cell signaling processes, cell adhesion, and potentially in glycolysis through senescence, suggesting a relationship with hypoxia resistance. Moreover, C19orf47 could actively participate in cancer-related processes such as cell proliferation and migration.

ÍNDICE

DEDICATORIA	i
AGRADECIMIENTOS	ii
RESUMEN	iv
ABSTRACT	vi
1. INTRODUCCIÓN	1
1.1. Generalidades de las proteínas	1
1.2. El Proteoma Humano.....	2
1.3. Proteómica.....	3
1.4. Oncoproteómica.....	4
1.5. Alteraciones de las células cancerosas.....	5
1.6. Cáncer Cervicouterino	6
1.6.1. Etiología y factores de riesgo	7
1.6.2. Tratamiento.....	8
1.7. Modelos de estudio de cáncer cervicouterino.	8
1.7.1. Líneas celulares con propiedades carcinogénicas: C33a (VPH-), HeLa (VPH 18) y SiHa (VPH 16).....	9
2. ANTECEDENTES	11
2.1. La Organización del Proteoma Humano.	11
2.2. El Proyecto del Proteoma Humano (HPP).	11
2.2.1. El proyecto del Proteoma Humano centrado en Cromosomas... 13	
2.3. El Proyecto del Proteoma Humano B/D-HPP impulsado por Biología y Enfermedades.....	15
2.3.1. Incorporación de México al Proyecto del Proteoma Humano 16	
2.4. Cromosoma 19.....	16
2.5. C19orf47	17
3. JUSTIFICACIÓN	18
4. HIPÓTESIS	19
5. OBJETIVOS	19
5.1. Objetivo general	19
5.2. Objetivos particulares.....	19
6. METODOLOGÍA	20
6.1. Análisis comparativo, filogenético de C19orf47.....	20
6.2. Análisis bioinformático de la expresión de C19orf47.....	20
6.3. Cultivo celular.....	21
6.4. Inmunofluorescencia.....	21
6.5. Extracción de proteínas totales.....	22
6.6. Extracción de la fracción citoplasmática y nuclear.	22
6.7. Western Blot.....	23
6.8. Co-inmunoprecipitación de C19orf47	24
6.9. Identificación de proteínas por LC-MS/MS.....	25
6.10. Análisis bioinformático y estadístico de datos generados a partir de LC- MS/MS	26
6.11. Construcción de redes de interacción proteína-proteína y enriquecimiento biológico.....	26

7. RESULTADOS Y DISCUSIÓN	28
7.1. Análisis Filogenético de la secuencia de aminoácidos de C19orf47	28
7.2. Predicción de dominios en C19orf47	29
7.3. Análisis de los niveles de expresión genética en tejidos tumorales y no tumorales	31
7.4. Expresión de C19orf47 en diferentes líneas celulares (HaCaT, C33a, HeLa y SiHa)	32
7.5. Localización subcelular de C19orf47	33
7.5.1. Predicción de la localización subcelular de C19orf47	33
7.5.2. Detección de la localización de C19orf47 mediante fraccionamiento citoplasmático y nuclear.	35
7.5.3. Inmunofluorescencia de C19orf47	35
7.6. Co inmunoprecipitación de C19orf47	38
7.6.1. Identificación de proteínas por LC-MS/MS.	38
7.6.1.1. Construcción del interactoma general de C19orf47 y su enriquecimiento biológico y molecular.	41
7.6.1.2. Redes Interactómicas de C19orf47, mediante MCODE en HaCaT, C33a, HeLa y SiHa	48
7.6.1.3. Clases funcionales del interactoma de C19orf47 en HaCaT, C33a, HeLa y SiHa	50
7.6.2. Análisis comparativo, de la expresión proteica, entre las líneas celulares cancerosas (HeLa y SiHa) contra no cancerosas (HaCaT)...	54
7.6.3. Análisis comparativo, de la expresión proteica, entre las líneas celulares cancerosas HeLa (VPH 16) y SiHa (VPH 18) contra C33a (VPH -)	58
8. CONCLUSIONES.....	61
REFERENCIAS.....	62
LISTA DE FIGURAS	viii
LISTA DE TABLAS	ix
LISTA DE ABREVIATURAS.	x

1. INTRODUCCIÓN

1.1. Generalidades de las proteínas

Las proteínas son moléculas vitales para el funcionamiento celular en cualquier organismo vivo, constituyen más de la mitad del peso seco de la célula y son las responsables de la realización de todos los procesos biológicos, y en consecuencia del correcto funcionamiento de los organismos. Las proteínas canónicas son consideradas la versión principal de una molécula que ha completado correctamente y de forma habitual el proceso de síntesis proteica, es decir, surgen como productos de un gen específico (Alberts *et al.*, 2017).

La síntesis proteica es un proceso complejo donde la información fluye a través del ADN, se transcribe en ARNm y finalmente se traduce en polipéptidos o bien, proteínas. Esta relación direccional conocida como el dogma central de la biología molecular a pesar de ser un proceso sumamente preciso, no está exento de errores, las variaciones genéticas y estructurales pueden tener numerosos efectos en la traducción de proteínas generando múltiples variantes proteicas que en algunos casos y según su origen podrían tener consecuencias biológicas (Forgrave *et al.*, 2022; Vihinen, 2021).

Las variaciones en los genes y las proteínas pueden tener múltiples efectos según el origen de su causa, en este sentido es posible distinguir dos grupos de variaciones proteicas altamente relacionados entre sí. Su estudio, además de facilitar la comprensión de la función biológica de los seres vivos, favorece el diagnóstico, tratamiento y comprensión de múltiples enfermedades.

Cuando se habla de variaciones proteicas, es común utilizar el término proteoforma, este no debe confundirse con las isoformas, las proteoformas abarcan todas las formas posibles de una proteína, no solo varían en secuencia, sino también a la recombinación del ADN y a las modificaciones postraduccionales. Por otro lado, las isoformas son variantes de la misma proteína, se generan a partir de un único gen, su secuencia de aminoácidos y estructura difieren de su forma

canónica debido a procesos como el empalme alternativo en el ARNm, polimorfismos de nucleótido simple (SNP, por sus siglas en inglés) o variantes de inicio de traducción. Su descubrimiento marcó un hito en el estudio de los proteomas, o bien, de la proteómica al comprobarse la capacidad genética para crear proteínas catalíticamente diferentes partiendo de un mismo gen. Su identificación se ha facilitado gracias a matrices de pequeños fragmentos de ADN o ARN (microarreglos) y bibliotecas de ADN complementario. Las isoformas pueden tener funciones específicas según el tejido o la etapa de desarrollo en la que se encuentren, y además en condiciones patológicas están fuertemente relacionados con muchas enfermedades (Liu *et al.*, 2022; Regnier & Kim, 2018; Tao *et al.*, 2024).

1.2. El Proteoma Humano

El genoma humano proporciona la base genética para la síntesis de proteínas, cada gen codifica una secuencia específica de aminoácidos que se traduce en una cadena de polipéptidos, que después de experimentar modificaciones se convierte en parte del proteoma. De tal forma que el proteoma humano, representa la colección de proteoformas codificadas por los 20 mil genes del genoma humano, y actúa como una maquinaria celular que impulsa diversos fenotipos biológicos (Liu *et al.*, 2016).

El término proteoma fue empleado por primera vez por Marc Wilkins para referirse al conjunto total de proteínas codificadas por el genoma de un organismo bajo condiciones específicas (Wilkins *et al.*, 1996). Sin embargo, un proteoma además de proteínas canónicas también incluye productos proteicos alternativos y modificaciones postraduccionales (Carbonara *et al.*, 2021).

A pesar de que se conoce el número de genes en el genoma humano, la diversidad de proteínas canónicas, proteoformas, y modificaciones postraduccionales en las proteínas amplían significativamente el proteoma complicando su estudio. Su tamaño y complejidad refleja naturaleza biológica de los seres humanos, el análisis del proteoma humano permite comprender cómo las proteínas regulan funciones esenciales.

1.3. Proteómica

La proteómica es una ciencia ómica encargada del estudio integral de las proteínas que conforman un proteoma (Digre & Lindskog, 2021; Monti *et al.*, 2019). Hoy en día, la proteómica tiene como objetivo principal realizar la caracterización funcional y estructural de las proteínas, así como identificar sus interacciones y modificaciones postraduccionales (PTMs, del inglés Post-Translational Modifications).

La proteómica hace uso de un conjunto de tecnologías especializadas que permiten extraer, separar y analizar las proteínas que constituyen una muestra para estudiar su composición proteica.

Entre las diferentes metodologías para la separación de proteínas, se incluye a la electroforesis SDS-PAGE (del inglés, Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis), y la electroforesis bidimensional; la primera separa a las proteínas en función de su peso molecular y la segunda considera propiedades como el punto isoeléctrico, la masa y el estado de la proteína. Igualmente, la proteómica puede hacer uso de técnicas estrictas como la cromatografía líquida, entre otras metodologías, las cuales permiten la separación necesaria para la posterior identificación de las proteínas mediante espectrometría de masas (MS por sus siglas en inglés). Esta tecnología permite catalogar detalladamente los componentes proteicos que forman parte de una muestra (Jannetto & Danso, 2020; Laemmli, 1970).

Debido a su alta precisión, la MS se ha convertido en la tecnología utilizada preferentemente para realizar análisis cuantitativos y cualitativos de las proteínas en función de la relación masa-carga (m/z). Según los requerimientos experimentales, la MS puede ser acoplada a distintas metodologías adicionales de separación que maximicen su precisión, por lo que comúnmente puede combinarse con diferentes tipos de cromatografías (Morley & Minceva, 2021), en este sentido, la cromatografía líquida acoplada a espectrometría de masas (LC-MS/MS, por sus siglas en inglés) es ampliamente utilizada en investigaciones proteómicas. Esta

técnica analítica permite separar los diferentes compuestos proteicos en una muestra mediante una columna cromatográfica, posteriormente los compuestos separados ingresan al espectrómetro de masas donde se ionizan y separan según sus valores m/z que permiten determinar los niveles relativos e inferir las identidades y la abundancia de las proteínas presentes en la muestra (Shuken, 2023; Tang *et al.*, 2023).

1.4. Oncoproteómica

Las proteínas, al participar activamente en los procesos celulares, tienen múltiples aplicaciones clínicas, incluso dentro de la oncología. La oncoproteómica, o bien, la aplicación de la proteómica en el estudio del cáncer se ha posicionado como una herramienta útil para la identificación y el estudio de los cambios en la expresión proteica en distintos tipos de cáncer. Especializándose en aquellas proteínas que se correlacionan directamente con la presencia de esta enfermedad, en conjunto con aquellas que puedan ser blancos terapéuticos para su tratamiento. Además, busca comprender la importancia funcional de las proteínas, así como las posibles relaciones entre compuestos proteicos específicos bajo la presencia de las propiedades carcinogénicas (Boys *et al.*, 2023; Rani *et al.*, 2023; Su *et al.*, 2021).

Particularmente, las células cancerosas se caracterizan por alterar los procesos celulares normales, evadiendo la muerte celular programada o apoptosis. Y como consecuencia, estas células desreguladas se reproducen sin control, produciendo nuevas células dañadas que a su vez terminan generando las conocidas masas tumorales. Los tumores formados por las células con anomalías en muchos casos resultan malignos para la salud y traen consigo un conjunto de padecimientos que pueden ocasionar la muerte (Fares *et al.*, 2020; Vaghari-Tabari *et al.*, 2021). Durante las últimas décadas el cáncer se ha posicionado como una enfermedad compleja y multifactorial importante con altas tasas de incidencia y de mortalidad que se han vuelto un problema de salud pública a nivel nacional y mundial (SSA, 2023; Sung *et al.*, 2021).

Diversos estudios enfocados en el inicio de la carcinogénesis, o de la progresión tumoral, han identificado que la presencia de ciertas proteínas puede indicar la desregulación de procesos metabólicos o de vías de señalización celular. Estas proteínas se han estudiado y caracterizado hasta el punto de considerarse en biomarcadores para el diagnóstico, predicción y el pronóstico del cáncer (Hristova & Chan, 2019).

La medición de los biomarcadores y supresores de tumores, o bien de las proteínas potencialmente relacionadas a estas funciones, se basa en gran medida en múltiples métodos basados en anticuerpos, especialmente la inmunohistoquímica (IHC, por sus siglas en inglés) y la inmunofluorescencia (IMF, por sus siglas en inglés) para una sola o un pequeño grupo de proteínas. Este tipo de técnicas basadas en la tinción de proteínas en muestras celulares o tejidos, son útiles para identificar patrones de tinción específicos para procesos cancerosos y orientar la elección de tratamientos contra el cáncer. Además, permiten comprender la biología de las proteínas al aperturar el entendimiento de su localización, comportamiento, función, expresión y sus potenciales implicaciones en células normales y malignas (Martínez-Rodríguez *et al.*, 2021; Sheng *et al.*, 2023; Tan *et al.*, 2020).

1.5. Alteraciones de las células cancerosas

Desde el inicio del estudio del cáncer, se determinó que la causa de este padecimiento se encontraba dentro de las lesiones celulares, o bien, daños o alteraciones en las células que pueden desencadenar cambios en su comportamiento estructura y función, estas alteraciones pueden promover el desarrollo del cáncer al desencadenar mutaciones genéticas, alterar el control del ciclo celular, evadir la muerte celular programada y modificar el microambiente celular, este descubrimiento motivó numerosos estudios posteriores para comprender los mecanismos de esta patología, sin embargo, no fue hasta el 2001 cuando Hanahan y Weinberg identificaron que el cáncer tiene una base genética y

epigenética capaz de causar mutaciones en genes responsables de iniciar la transformación de las células (Hanahan & Weinberg, 2000).

El proceso por el cuál una célula se transforma en una célula maligna, o bien, se torna cancerosa (carcinogénesis), consiste en un conjunto de eventos que además de dañar el ADN y el ciclo celular, les confieren a las células numerosas características genéticas que les permite mantener señales mediante la acumulación de mutaciones que promueven y mantienen las altas tasas descontroladas de proliferación formando tejidos anormales, otorgándoles la capacidad de evadir los supresores del crecimiento. Así mismo, estas mutaciones genéticas facilitan la desregulación del metabolismo y la evasión del sistema inmune, el cual, en condiciones normales, es capaz de eliminar a las células anormales que escapan del ciclo celular. Igualmente, estas mutaciones del ADN facilitan el desarrollo de mecanismos para resistir la apoptosis y aumentar la facultad que tienen las células malignas para invadir tejidos (metástasis) y posteriormente construir las vías necesarias para nutrir la masa tumoral generada (angiogénesis). En conjunto, estas características proporcionan un entorno favorable para el desarrollo y la expansión del cáncer, así mismo dificultan el éxito de los tratamientos cuando este se encuentra en un estado de progresión avanzada (Hanahan & Weinberg, 2000, 2011).

1.6. Cáncer Cervicouterino

El cáncer cervicouterino (CaCu) es un padecimiento ginecológico de lenta progresión que se torna más agresivo en sectores económicos vulnerables. Es el tercer cáncer con mayor índice de mortalidad entre las mujeres, con más de 348 mil defunciones por año y el cuarto más frecuente después del cáncer de mama, pulmón y colorrectal, con más de 662 mil casos nuevos (Bedell *et al.*, 2020; Cerqueira *et al.*, 2022; Sung *et al.*, 2021). Al igual que todos los tipos de cáncer, el CaCu es una alteración celular que se origina de la transformación maligna de las células que recubren el cuello uterino y se manifiesta inicialmente mediante lesiones precancerosas de lenta progresión (Polanco *et al.*, 2020.).

El cuello uterino se conforma de dos partes, el ectocérvix, que corresponde a la zona externa del cuello uterino y el endocérvix que se refiere a la parte interna del cuello uterino que forma la conexión entre la vagina con el útero. En este sentido, resulta importante señalar que el CaCu se diferencia en dos tipos, según las células en donde se origine este padecimiento, el carcinoma de células escamosas inicia en las células de ectocérvix, este es el tipo más recurrente de CaCu con alrededor del 80% del total de los casos; mientras que el adenocarcinoma se forma en las células que recubren el endocérvix y corresponde al 20% de los casos totales de este padecimiento. En la mayoría de los casos, esta patología está acompañada de lesiones definidas como Neoplasia Intraepitelial cervical (NIC) de primer a tercer grado, causadas por la persistencia de serotipos oncológicos del virus del papiloma humano (Baezconde-Garbanati *et al.*, 2019; Bedell *et al.*, 2020; Cerqueira *et al.*, 2022; NIH, 2021).

1.6.1. Etiología y factores de riesgo

El CaCu es el resultado de manifestaciones genéticas y epigenéticas, que en el 80% de los casos son causadas, por la infección del virus del papiloma humano (VPH) (Lei *et al.*, 2020). Los serotipos de alto riesgo, como el tipo 16 y 18, son responsables del 70% de los casos de CaCu atribuibles al VPH (Rerucha *et al.*, 2018; WHO, 2014). A pesar de que el VPH se coloca como el factor de riesgo principal para el desarrollo de esta enfermedad, existen diversos factores intrínsecos y extrínsecos relacionados con la progresión de células malignas, entre los cuales, se identifica la predisposición genética, la inmunodeficiencia y la inmunosupresión como factores clave que promueven tanto la infección por VPH como la progresión tumoral. Sin embargo, también se ha identificado que distintos hábitos aumentan la posibilidad de desencadenar el proceso de carcinogénesis como el consumo de tabaco que, debido a sus componentes tóxicos, daña la estabilidad de las células causando alteraciones en la señalización celular. Por otro lado, se ha demostrado que el uso prologado de anticonceptivos hormonales es otro factor que incentiva su evolución, sin embargo, se desconocen los mecanismos exactos que faciliten el desarrollo de CaCu (Li *et al.*, 2023; Polanco *et al.*, 2020.).

1.6.2. Tratamiento

La detección en etapas tempranas del CaCu, favorecen el pronóstico y la tasa de supervivencia, además apertura las opciones de tratamiento del paciente. Sin embargo, cabe aclarar que la elección del tratamiento depende completamente del estadio del cáncer (tamaño y diseminación del tumor) así como de la salud y el historial clínico del paciente.

Las opciones comprenden distintos tipos de cirugías que varían según el tamaño y la localización del tumor, por otra parte, la radioterapia y la quimioterapia suelen utilizarse tanto individualmente como combinadas para destruir e inhibir el desarrollo de las células malignas, sin embargo, los efectos secundarios de estos tratamientos son sumamente agresivos y requieren el uso adicional de medicamentos paliativos que disminuyan los efectos adversos. Finalmente existen tratamientos especializados como lo son la terapia dirigida y la inmunoterapia capaces de inhibir el crecimiento de los vasos sanguíneos tumorales y la interacción entre enzimas, proteínas y moléculas que favorecen la supervivencia de células tumorales, sin embargo, esto últimos son muy costosos por lo que difícilmente las personas pueden tener acceso a ellos (Enríquez-Aceves *et al.*, 2020; Mannucci *et al.*, 2020; Peters *et al.*, 2000).

1.7. Modelos de estudio de cáncer cervicouterino

Las regulaciones bioéticas y legales encargadas de limitar el uso de muestras biológicas humanas en investigaciones científicas tienen como objetivo proteger y garantizar los derechos de los pacientes y la integridad de los trabajos de investigación. Aunque estas regulaciones permiten el uso de muestras tumorales para la investigación, los procedimientos para su obtención y manejo son rigurosos para asegurar el consentimiento informado y la privacidad de los pacientes.

Debido a este tipo de limitaciones, es necesario adoptar modelos de estudio que se adapten a las necesidades experimentales y que al mismo tiempo permitan realizar la investigación pertinente que facilite la comprensión del cáncer. El uso de

líneas celulares y modelos murinos ha sido indispensable para el estudio de este tipo de padecimientos. Estos modelos, permiten un rango detallado de observación desde el inicio de la transformación de las células normales hasta convertirse en células malignas. También facilitan el estudio de las rutas metabólicas, y patológicas, minimizar la interferencia de variables externas al cáncer, así como identificar y probar posibles blancos terapéuticos (Delgadillo, 2014).

1.7.1. Líneas celulares con propiedades carcinogénicas: C33a (VPH-), HeLa (VPH 18) y SiHa (VPH 16)

Las líneas celulares son modelos *in vitro* bien establecidos que han aportado conocimientos valiosos en la investigación biomédica del cáncer. Estas líneas celulares tienen la capacidad de ser cultivada por un tiempo indefinido, manteniendo sus propiedades y características distintivas. Es por ello por lo que las líneas celulares provenientes de cáncer son capaces de ser propagadas rápidamente y por lo que, bajo las condiciones y el control adecuado, mantienen las propiedades carcinogénicas.

Las líneas celulares de cáncer presentan varias ventajas, con respecto al uso de muestras tumorales (biopsias), ya que pueden ser manipuladas bajo un ambiente controlado donde no existe interferencia de algún tipo con variables externas, manteniendo una población homogénea y un número ilimitado de muestra. Sin embargo, presenta desventajas relacionadas con la inestabilidad genómica derivada de la composición cromosómica aneuploide, que debido a la manipulación inadecuada pueden ocasionar la modificación de su composición genética (Dutil *et al.*, 2019; Masters, 2000).

Entre las líneas celulares más utilizadas en el ámbito científico se encuentra HeLa, siendo el linaje celular más antiguo y ha sido ampliamente caracterizada al ser empleada en numerosos estudios de distinta índole. Esta línea fue aislada en 1951 y proviene de una paciente de ascendencia africana de 31 años diagnosticada con adenocarcinoma cervical, contiene ADN del VPH del tipo 18 con un número modal de 82 cromosomas (Jordan, 2021).

La línea celular C33a, es una línea adherente que proviene de una paciente de 66 años caucásica diagnosticada con adenocarcinoma cervical. No tiene ADN de VPH y cuenta con un número modal de 46 cromosomas.

La línea celular SiHa fue aislada en 1970 a partir de un carcinoma de células escamosas de grado II obtenido de una paciente de 55 años de ascendencia asiática. Esta línea contiene ADN de VPH del tipo 16 con una o dos copias por célula (Skok *et al.*, 2021).

2. ANTECEDENTES

2.1. La Organización del Proteoma Humano

La secuenciación del genoma humano impulsó la creación de una organización científica internacional dedicada al estudio y la investigación del proteoma humano (Lander *et al.*, 2001). Es así como en el 2001 se consolidó la Organización del Proteoma Humano (HUPO del inglés, Human Proteome Organization), con la finalidad de promover el intercambio de datos y el desarrollo e innovación de tecnologías especializadas en el campo de la proteómica que fomenten la comprensión del proteoma humano (Hanash & Celis, 2002). Las primeras iniciativas de la HUPO incluyeron el estudio del proteoma presente en el plasma, hígado, cerebro, riñón y el sistema cardiovascular. Como resultado se incentivó la cooperación con otras instancias proteómicas como el Atlas de Proteínas Humanas (HTA, del inglés The Human Protein Atlas). Para la generación de anticuerpos contra todas las proteínas codificadas por el genoma humano (He, 2005; Omenn *et al.*, 2005; Ping *et al.*, 2005; Uhlén *et al.*, 2005; Yamamoto *et al.*, 2008).

2.2. El Proyecto del Proteoma Humano (HPP)

Se estima que el genoma humano comprende aproximadamente 20 mil genes, sin embargo, el proteoma considerado a partir de estos productos genéticos es enorme. Ya que aparte de las proteínas canónicas, se deben considerar un sin fin de proteoformas derivadas de la recombinación del ADN y del empalme alternativo de transcritos primarios, incluyendo numerosas PTMs, tales como: la acetilación, fosforilación o glicosilación, entre otras. Esto dificulta la definición de todas las proteínas que conforman el proteoma humano de una célula o tejido, en un momento dado.

Por esta razón el Proyecto del Proteoma Humano (HPP, por sus siglas en inglés), organizado por la HUPO, tiene como objetivo identificar, mapear, caracterizar y crear una confiable de las proteínas que conforman el proteoma humano. Apoyándose de las tecnologías más estrictas como MS, o ensayos

inmunológicos, e integrar la proteómica como un componente de la investigación multiómica con la finalidad de caracterizar la función primaria de las proteínas, sus funciones, modificaciones y sus redes de interacción, en la salud y en las enfermedades humanas (Legrain *et al.*, 2011; Omenn, 2021).

Como se describe en la Tabla 1, el total de proteínas que conforman el proteoma humano puede dividirse en cinco clases dependiendo de su evidencia proteica (PE). Sin embargo, cabe mencionar que estas pueden cambiar su clasificación según las actualizaciones reportadas.

Tabla 1. Clasificación de las proteínas según su evidencia proteica (PE, por sus siglas en inglés).

PE	Descripción.
PE1	Evidencia experimental a nivel de proteína confirmada mediante técnicas estrictas como Western blot o MS.
PE2	Evidencia a nivel de transcripción de ARNm.
PE3	Evidencia inferida por homología.
PE4	Predicción <i>in silico</i> .
PE5	Productos genéticos hipotéticos o pseudogenes.

Las proteínas de los grupos PE2 a PE4 son llamadas proteínas perdidas (MPs, del inglés Missing Proteins), los pseudogenes pertenecientes a PE5 no son contabilizadas en los reportes finales ya que carecen de evidencia suficiente de su existencia. Sin embargo, no basta con que una proteína sea clasificada como PE1 si no ha sido caracterizada funcionalmente, el grupo de proteínas PE1, sin función asignada es conocido como uPE1, (del inglés Uncharacterized Protein Existence 1), estas pertenecen al proteoma oscuro, es decir son aquellas proteínas de las que se tiene evidencia de su existencia mediante técnicas proteómicas, sin embargo, cuya función aún no ha sido determinada.

El HPP ha logrado grandes avances desde su lanzamiento, y está constituido por dos componentes fundamentales: El Proyecto del Proteoma Humano centrado de Cromosomas (C-HPP, del inglés Chromosome-Centric Human Proteome

Project) y El Proyecto del Proteoma Humano Impulsado por Biología y Enfermedades (B/D-HPP, del inglés Biology/Disease-driven Human Proteome Project) que en conjunto contribuyen activamente al entendimiento detallado del proteoma humano. Actualmente el C-HPP está constituido de 25 equipos, cada uno por cromosoma, que se encargan del estudio de cada cromosoma y del proteoma mitocondrial (Omenn *et al.*, 2024).

2.2.1. El proyecto del Proteoma Humano centrado en Cromosomas

Los cromosomas son estructuras subcelulares altamente organizadas dentro del núcleo celular, cada cromosoma consta de una sola molécula de ADN encargada de expresar las proteínas esenciales para el mantenimiento del correcto funcionamiento de los organismos, por ello resulta indispensable abordar el estudio individual de cada uno de los cromosomas del ser humano (Benbow, 1992; Hirano, 2015).

El Proyecto del Proteoma Humano centrado en Cromosomas (C-HPP) surge como una iniciativa del proyecto emblema de la HUPO y es el encargado de mapear las MPs expresadas por cada uno de los cromosomas, adicionalmente, incorporó a sus objetivos la caracterización funcional de aquellas proteínas sin función asignada (uPE1). El C-HPP tiene como objetivo principal mapear, anotar y caracterizar el proteoma humano cromosoma por cromosoma. Actualmente, los principales objetivos principales de este proyecto son mapear todas las proteínas faltantes y caracterizar las uPE1 de cada cromosoma (Paik *et al.*, 2018).

En la Tabla 2, se muestran los datos actualizados de las proteínas que conforman el proteoma humano en cada cromosoma según su clasificación basándose en su evidencia proteica, así mismo se muestran las proteínas faltantes (MPs) y las proteínas no caracterizadas (uPE1). El número de proteínas MPs ha reducido drásticamente en conjunto con las proteínas uPE1. Inicialmente se tenían reportadas 5511 MPs, reduciendo su número significativamente durante el 2023 a 1381 MPs y 1134 uPE1. Con el fin de analizar individualmente cada cromosoma, el C-HPP se distribuye en 25 equipos de 19 países (Figura 1).

Tabla 2. Datos actualizados de las proteínas clasificadas, según su evidencia proteica por cromosoma.

Chr	PE1	PE2	PE3	PE4	MPs	uPE1	%MP/PE1-PE4	% uPE1/PE1
1	1864	117	35	2	2018	141	7.63	7.56
2	1229	45	3	0	1277	71	3.76	5.77
3	1006	45	6	0	1157	56	4.81	5.54
4	707	27	12	0	746	45	5.22	6.37
5	838	34	2	0	874	48	4.17	5.76
6	938	52	4	1	995	56	5.71	5.96
7	878	90	7	3	978	48	10.26	5.54
8	627	32	10	1	670	34	6.49	5.43
9	703	51	8	2	764	54	7.97	7.66
10	690	38	1	1	730	42	5.50	6.12
11	1072	151	69	0	1292	61	17.04	5.68
12	962	40	8	0	1010	45	4.75	4.68
13	310	10	2	0	322	27	3.73	8.68
14	648	58	15	2	723	32	10.40	4.98
15	534	41	3	0	578	33	7.67	6.15
16	775	40	2	1	818	42	5.27	5.43
17	1080	60	4	0	1144	58	5.59	5.35
18	255	5	0	0	260	10	1.89	3.89
19	1318	77	12	1	1408	68	6.41	5.12
20	509	24	1	0	534	32	4.68	6.25
21	198	22	4	0	224	11	11.61	5.53
22	454	24	4	1	483	34	6.03	7.52
X	758	57	2	0	817	85	7.25	11.15
Y	33	7	1	0	41	1	20	3.13

(Omenn *et al.*, 2024)

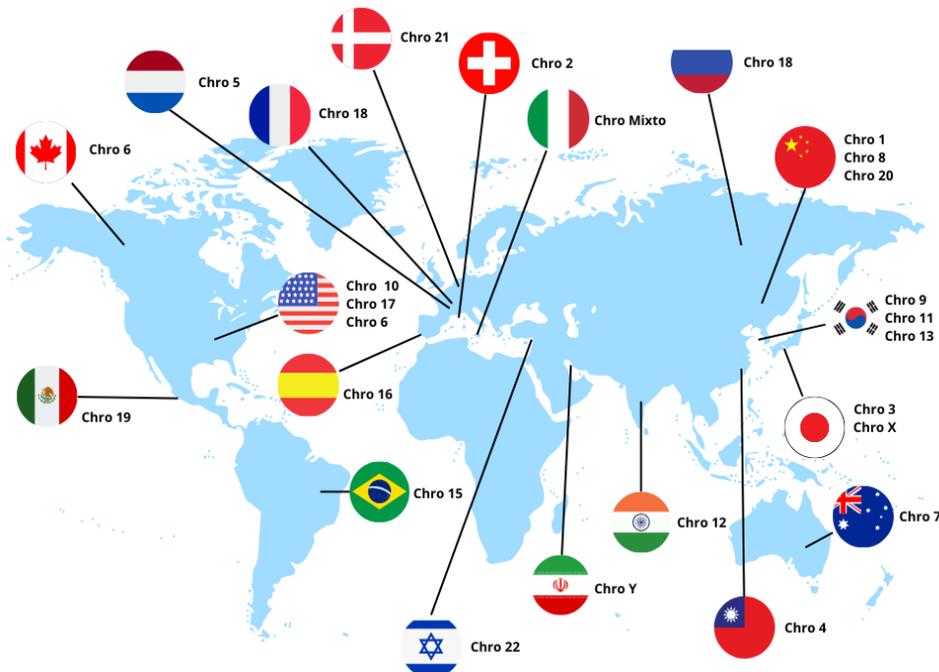


Figura 1. Colaboración internacional en el Proyecto del Proteoma Humano Centrado en Cromosomas (C-HPP). Los países participantes en el C-HPP para explorar las proteínas generadas en cada cromosoma se distribuye en 25 consorcios, cada uno dedicado al estudio de un cromosoma específico.

2.3. El Proyecto del Proteoma Humano B/D-HPP impulsado por Biología y Enfermedades

El Proyecto del Proteoma Humano impulsado para el entendimiento del papel de las proteínas en la Biología celular y en las enfermedades (B/D-HPP) es un complemento del HPP que se ocupa del mapeo y anotación del proteoma con el propósito de identificar y entender las funciones biológicas de las proteínas y su relación con la biología humana y las enfermedades.

Además de estudiar el proteoma, el B/D-HPP proporciona información valiosa sobre los mecanismos moleculares de las enfermedades en diferentes áreas de estudio que incluyen múltiples iniciativas especializadas, entre ellas, el cáncer. Dentro de esta iniciativa, se ha realizado un gran esfuerzo para entender la función

biológica del proteoma en los distintos tipos de cánceres. Con el fin de identificar biomarcadores que promuevan la detección temprana y a su vez desarrollar terapias personalizadas.

2.3.1. Incorporación de México al Proyecto del Proteoma Humano

Tras el lanzamiento del C-HPP, México fue uno de los países que se incorporaron a la iniciativa, fue así como adoptó el estudio del cromosoma 19 centrándose en identificar mediante técnicas proteómicas las 1408 MPs proteínas codificadas por genes pertenecientes a este cromosoma, y caracterizar funcionalmente 68 uPE1 codificadas por cromosoma 19. El consorcio mexicano adicionalmente realiza esfuerzos para analizar las redes de interacción proteica dentro del contexto biológico de las enfermedades, especialmente, en distintos tipos de cáncer (Gaceta, 2024).

Dentro de los trabajos realizados por el consorcio del laboratorio de proteómica del CCG, de la UNAM, el cual es el laboratorio central del consorcio mexicano del cromosoma 19. Un estudio proteómico logró identificar mediante MS a 11 uPE1 pertenecientes al cromosoma 19, entre las cuales un grupo destacó al incrementar su expresión en células tumorales con respecto a células no tumorales, este grupo incluye a LENG8, CCDC61, TMEM160, CCDC97 y C19orf48 (Gil *et al.*, 2017).

2.4. Cromosoma 19

El cromosoma 19 es uno de los 23 pares de cromosomas que conforman al ser humano, Este cromosoma es relativamente pequeño, con alrededor de 63 millones de pares de bases y contiene entre 2000 y 3000 genes. Entre sus funciones, los genes en el cromosoma 19 están implicados en diversos procesos biológicos esenciales, como el metabolismo, la función neurológica, Además, contiene genes que influyen en la regulación del sistema inmunológico, ayudando a coordinar la respuesta sistema inmune.(Grimwood *et al.*, 2004; Harris *et al.*, 2020).

Este cromosoma está relacionado con varios padecimientos como, la aterosclerosis, la distrofia muscular congénita, el síndrome de Peutz-Jeghers, el Alzheimer y tipos de cáncer. La relación entre el cáncer y el cromosoma 19 está determinada por la alta presencia de monosomías y genes supresores de tumores como GLTSCR2, SAFB y STK11 (Altalebi *et al.*, 2023; Cao *et al.*, 2020).

2.5. C19orf47

C19orf47, es una proteína codificada por el cromosoma 19, clasificada dentro del proteoma oscuro al no tener función asignada (uPE1). Según se reporta en bases de datos como InterproScan; una base de datos bioinformática que utiliza modelos predictivos para identificar patrones y dominios en las secuencias de proteínas (Paysan-Lafosse *et al.*, 2023), se predice que C19orf47 podría distribuirse con distintos grados de conservación en diversos organismos, incluyendo: mamíferos, reptiles, anfibios, peces, insectos y algunas plantas. Esta distribución sugiere que esta proteína pudiese tener relevancia biológica y funcional. Además, se prevé que tiene homólogos en diferentes especies con un alto grado de similitud e identidad en la secuencia de aminoácidos.

Con base con los datos de transcriptómica, tomados de Human Protein Atlas, el ARNm de C19orf47 se expresa en múltiples tejidos incluyendo: el cerebro, corazón, riñones, hígado, bazo, testículos, ovarios, piel y células sanguíneas. Igualmente se ha detectado una expresión diferencial, de C19orf47, en tejidos cancerosos en comparación con tejidos normales. También, se destaca la relación existente entre esta proteína y el tejido tumoral del cuello uterino.

C19orf47 tiene 10 variantes proteicas generadas por splicing alternativo, 189 genes ortólogos y su forma canónica está compuesta por 422 aminoácidos (aa). con un peso molecular de 45 KiloDaltons (kDa) y dos isoformas predichas de 38 y 30 kDa. Según se predice en la base de datos Nextprot (Zahn-Zabal *et al.*, 2019), que C19orf47 se localiza en el nucleoplasma y puede ser blanco de distintas modificaciones postraduccionales como fosforilación, ubiquitinación y sumoilación.

3. JUSTIFICACIÓN

El contenido del genoma humano incluye aproximadamente 20.000 genes codificantes, por el contrario, el espacio proteómico derivado de éstos es extenso debido a la cantidad de proteínas canónicas y proteoformas procedentes de una serie de modificaciones postraduccionales y variaciones en el ADN. El estudio del proteoma humano es esencial para comprender los mecanismos biológicos y funcionales del organismo en contextos homeostáticos y en la biología de las enfermedades. Sin embargo, la interpretación de la biología humana se ve obstaculizada por la falta de evidencia proteica y el desconocimiento funcional de algunos de las proteínas.

Por esta razón, es importante estudiar a las proteínas cuya función es desconocida (uPE1). La caracterización funcional de estas proteínas puede revelar funciones de estas en vías biológicas poco conocidas o no caracterizadas. Así como el descubrimiento de nuevos biomarcadores, o dianas terapéuticas, para contribuir al desarrollo de fármacos y terapias dirigidas. Se han realizado varios estudios proteómicos para mapear y caracterizar las proteínas y proteoformas que componen el proteoma humano. Concretamente en un análisis proteómico previo realizado en líneas de cáncer cervicouterino, se detectó a la proteína sin función asignada (uPE1) C19orf47, la cual se observó sobre expresada en células cancerosas respecto a las no cancerosas. En este sentido, este proyecto tiene como meta aportar el conocimiento necesario que a la caracterización de la función y localización de la proteína C19orf47 y su relación con el cáncer cervicouterino y la presencia del VPH.

4. HIPÓTESIS

La proteína C19orf47, sin función asignada, codificada en el cromosoma 19 y detectada expresada en líneas celulares de cáncer cervical, se localiza tanto en el núcleo como en el citoplasma, y juega un importante papel en la regulación de procesos celulares clave como la proliferación, migración y adhesión en células cancerosas.

5. OBJETIVOS

5.1. Objetivo general

Determinar la función y localización de la proteína C19orf47 en células de cáncer cervicouterino y su relación con la presencia de VPH y el proceso de carcinogénesis.

5.2. Objetivos particulares

- Determinar los niveles de expresión de C19orf47 en líneas celulares no cancerosas (HaCaT) y de cáncer cervicouterino no infectadas e infectadas con VPH: C33a (VPH-), HeLa (VPH 18) y SiHa (VPH 16).
- Determinar la localización de C19orf47 mediante inmunofluorescencia en líneas celulares de cáncer cervicouterino.
- Mediante co-inmunoprecipitación seguida de análisis mediante LC-MS/MS, definir el interactoma proteico de C19orf47 en las líneas celulares de HaCaT, C33a (VPH-), HeLa (VPH 18) y SiHa (VPH 16).
- Proponer las potenciales funciones biológicas asociadas con C19orf47 en el cáncer cervical.

6. METODOLOGÍA

6.1. Análisis comparativo, filogenético de C19orf47

La secuencia de aminoácidos canónica de C19orf47 se obtuvo de la base de datos UniProt (del inglés, Universal Protein (Bateman *et al.*, 2023)). Para la búsqueda de secuencias homólogas de la secuencia canónica de C19orf47, se utilizó la herramienta básica de búsqueda de alineación local (BLAST, del inglés Basic Local Alignment Search Tool). De las secuencias resultantes, se seleccionaron las secuencias con un porcentaje de identidad entre 70 y 30%, y un e-value de $1e-5$ y $1e-10$, con la finalidad de identificar las secuencias altamente conservadas y diferenciarlas de aquellas con un alto grado de divergencia. Posteriormente se realizaron alineamientos múltiples, con las secuencias homólogas seleccionadas, con el programa Clustal (Sievers *et al.*, 2011). Finalmente se construyó un árbol filogenético, mediante el método de estimación de máxima verosimilitud (MLE, del inglés Maximum Likelihood Estimation), con apoyo del programa de Análisis Genético de Evolución Molecular (MEGA, del inglés Molecular Evolutionary Genetics Analysis) (Tamura *et al.*, 2021).

6.2. Análisis bioinformático de la expresión de C19orf47

Para explorar la expresión genética de C19orf47, se ingresó la secuencia canónica de aminoácidos en la plataforma que realiza Análisis Interactivo de Perfiles de Expresión Genética (GEPIA, del inglés Gene Expression Profiling Interactive Analysis) (Z. Tang *et al.*, 2017). En esta plataforma, se ingresó el nombre del gen correspondiente a C19orf47, y se exploró su expresión en tejidos normales y diversos tipos de cáncer. Posteriormente, se consultó la plataforma del Atlas de Proteínas Humanas (HPA, del inglés Human Protein Atlas) para obtener una visualización más detallada de la expresión de C19orf47 en diferentes modelos celulares. Finalmente se consultó las bases de datos InAct para explorar interacciones de la proteína con otras moléculas y proteínas (Orchard *et al.*, 2014).

6.3. Cultivo celular

Para la realización del proyecto, se utilizó la línea celular inmortalizada de queratinocitos transformados (HaCaT); las líneas celulares provenientes de cáncer cervicouterino sin VPH (C33a) e infectadas con VPH: HeLa (VPH 18) y SiHa (VPH 16). Las células se cultivaron en medio RPMI-1640, suplementado al 10% de Suero Fetal Bovino (SFB), con penicilina (100 U/ml) y estreptomycin (100 U/ml). a concentración constante de CO₂ de 5%, a una temperatura de 37°C y un ambiente saturado de humedad.

6.4. Inmunofluorescencia

1x10⁴ células de cada una de las cuatro líneas celulares utilizadas fueron cultivadas a 70% de confluencia por pozo en un Labteck de ocho pozos, las cuales se fijaron con paraformaldehído al 2% durante cinco minutos a temperatura ambiente (TA) Posteriormente se realizaron tres lavados continuos con PBS 1X filtrado, evitando la desecación de las células, y se añadieron 250 µl de solución de bloqueo (SFB 5%, tritón 0.05%, PBS 1X).

Para eliminar la solución de bloqueo, se lavaron las células por tres ciclos continuos con PBS 1X filtrado y se dejaron incubando con el anticuerpo primario diluido a una concentración de 1 µg/250 µl (*C19orf47 Polyclonal Antibody, invitrogen, PA5-59751*). El anticuerpo se diluyó en la solución de dilución (BSA 5%, tritón 0.05%, PBS 1X) y se dejó incubando durante 18 h a 4°C, sin agitación. Una vez más, se realizaron tres ciclos de lavado continuos con PBS 1X filtrado para retirar el exceso de anticuerpo primario y se incubó el anticuerpo secundario 1:250 (*Alexa FluorTM 488 goat anti-rabbit IgG (H+L), invitrogen, A11034*) durante una hora a TA protegido de la luz. Se realizaron tres ciclos de lavado continuos con PBS 1X filtrado, para posteriormente teñir el núcleo celular con DAPI bajo una concentración 1:500 durante quince minutos, con ligera agitación, cada cinco minutos, nuevamente protegido de la luz. Se realizó por última vez el ciclo de lavado con PBS 1X filtrado para después desmontar la estructura primaria del Labteck y sumergir la placa resultante cinco veces en agua mili Q filtrada para retirar los restos

de sales que pudieran precipitarse. Finalmente, se adicionaron a las muestras preparadas, 10 µl de medio de montaje (90% glicerol), y posteriormente, se colocó un cubreobjetos sobre ellas y se sellaron los bordes con acrílico para asegurar la integridad de las muestras durante el proceso de observación.

El análisis microscópico se realizó con un microscopio Confocal (*Olympus FV1000 Upright*), utilizando un objetivo de 60X y un aumento adicional de 2X. Las imágenes se adquirieron mediante la toma de diversos cortes para una observación detallada. Finalmente, las imágenes obtenidas fueron analizadas en el software libre FIJI (Schindelin *et al.*, 2012) siguiendo los procedimientos recomendados por el fabricante para garantizar la precisión y la fiabilidad de los resultados obtenidos en el análisis de las imágenes.

6.5. Extracción de proteínas totales

Se cultivaron a 90% de confluencia cada una de las líneas celulares (HaCaT, C33a, HeLa y SiHa) para ser lisadas con 200 µL de buffer de lisis [1mL Buffer lisis NP40 (NP40, Invitrogen, FNN 0021) + 10µL de inhibidores de proteasa (Halt Protease & Phosphatase Cocktail, Thermo Scientific 78440 (3X))]. La reacción se dejó incubar 30 min en hielo, homogenizando en vortex cada 5 minutos. Al concluir el tiempo, las muestras se centrifugaron a 13,000 rpm, a 4°C, durante 10 minutos. Una vez recuperado el sobrenadante y se determinó la concentración de proteína mediante la técnica colorimétrica de Bradford.

6.6. Extracción de la fracción citoplasmática y nuclear

Para la extracción de las fracciones citoplasmática y nuclear, la línea celular HeLa se cultivó a 90% de confluencia, posteriormente, las células se lavaron dos veces con PBS, enfriado en hielo, y luego se agregaron 200 µl de buffer de lisis (10mM KCl, 25nM HEPES a pH 7.8, 0.5% NP40, 1 nM PMSF, 10 µg/mL Leupetin, 25 µf/mL Aprotitin, 100 mM DTT, 5% glicerol, 0.1 nM EDTA, 20 Nm NaF, 1 nM Na₃VO₄ y H₂O destilada). Las células se rasparon sobre hielo y el sobrenadante resultante fue transferido a un tubo de centrifugación para ser incubado 30 min a

4°C con agitación constante, posteriormente, se centrifugó a 10,000 rpm, a 4°C durante 5 minutos, el sobrenadante fue recolectado como el extracto citoplasmático. Mientras que la fracción nuclear, contenida en el precipitado o “pellet”, fue lavada con 500 µl de buffer de lavado (10mM KCl, 25nM HEPES a pH 7.8, 1 nM PMSF, 10 µg/mL Leupetin, 25 µf/mL Aprotitin, 100 mM DTT, 5% glicerol, 0.1 nM EDTA, 20 Nm NaF, 1 nM Na₃VO₄ y H₂O destilada), dejándose incubar en hielo durante 3 minutos y luego se centrifugó se centrifugó a 10,000 rpm, a 4°C durante 5 minutos. El sobrenadante se descartó y se agregaron 150 µl de buffer de extracción (400mM NaCl, 25nM HEPES a pH 7.8, 1.5 nM MgCl₂ 1 nM PMSF, 10 µg/mL Leupetin, 25 µf/mL Aprotitin, 100 mM DTT, 10% glicerol, 0.1 nM EDTA, 20 Nm NaF, 1 nM Na₃VO₄, H₂O destilada); el “pellet” nuclear se resuspendió mediante suaves pipeteos y luego se incubó a 4°C durante 30 minutos con agitación constante. Posteriormente, la muestra se centrifugó a 13,000 rpm, a 4°C durante 5 minutos y se recolectó el sobrenadante como el extracto nuclear.

Las concentraciones de proteínas citoplasmáticas y proteínas nucleares se determinaron mediante cuantificación de proteínas con una curva estándar de Albúmina Sérica Bovina (BCA, por sus siglas en inglés) (*Pierce 660 nm Protein Assay Reagent, thermo scientific, 1861426*).

6.7. Western Blot

Los extractos proteicos se separaron mediante electroforesis en geles de poliacrilamida con dodecilsulfato sódico al 12% (SDS-PAGE, del inglés Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis). Posteriormente, el gel se transfirió a una membrana de Fluoruro de polivinilideno (PVDF, del inglés Polyvinylidene fluoride) con poro de 0.22-µM (*Inmmobilon-P, Milipore, MA, USA*). La membrana fue bloqueada con BSA al 5%, e incubada por toda la noche con el anticuerpo primario (*C19orf47 Polyclonal Antibody, invitrogen, PA5-59751*), a una concentración de 0.5 µg/ml a 4°C durante 16 horas con agitación constante. El anticuerpo secundario (mouse anti-rabbit IgG-HRP, Santa cruz, sc-2357) fue incubado una hora a temperatura ambiente a una concentración 1:10,000. La

visualización y análisis del resultado del Western Blot se realizó con el software de acceso libre Image Studio Digits.

En el caso del análisis, por Western Blot, de las fases nucleares y citoplasmáticas, se utilizó como control un anticuerpo primario anti-Histona (*Histone H2A X, Santa Cruz, sc-54606*), a una concentración 1:1000 y fue incubado toda la noche a 4°C.

6.8. Co-inmunoprecipitación de C19orf47

Para realizar el protocolo de Co-inmunoprecipitación, se utilizó un Kit de perlas magnéticas (*Dynabeads® Antibody Coupling Kit, invitrogen, 143.11D*), es importante destacar que las perlas no se centrifugaron en ninguna ocasión, todos los lavados se realizaron con ayuda con un imán.

Se utilizaron 5 mg de perlas/500 µl, las cuales se lavaron con 1 mL de la solución del Kit Dynabeads® Antibody Coupling, denominada C1. Después se incubaron con el anticuerpo primario de C19orf47 (C19orf47 Antibody, MybioSource, MBS7000536), a una concentración de 5 µg en 250 µl de la solución denominada C2 incluida en el Kit. La reacción se dejó incubando durante toda la noche, en agitación a 37°C. Al siguiente día se lavó el exceso de anticuerpo con 800 µl de la solución HB del Kit y después se realizó un lavado con 800 µl de la solución LB. Finalmente se realizaron tres ciclos de lavados con 800 µl la solución denominada SB.

Una vez acoplado el anticuerpo, las perlas magnéticas se lavaron dos veces con 200 µl de PBS 1X para ser transferidas a un tubo limpio y resuspendidas con 900 µl de buffer de extracción (Tris HCl 10 mM a pH8, KCl 200 mM, EDTA 2 mM, glicerol 40% y DTT 1mM). Se eliminó el sobrenadante, las perlas se resuspendieron en el extracto proteico y se dejaron incubando durante dos horas con agitación a 4°C. Se realizó una serie de cuatro lavados con 200 µl de buffer A/B (NaCl 100 mM, MgCl 2 Mm, DTT 1mM) y se transfirió a un tubo limpio. Las perlas se resuspendieron en 13 µl de buffer de elución (SDS 2%, Tris HCl 100 mM a pH 7.5, glicerol 10%,

EDTA 0.5 mM y DTT 100 mM) y se calentaron durante 10 minutos a 70°C, después se añadieron 117 µl de buffer de elución.

6.9. Identificación de proteínas por LC-MS/MS

Las proteínas se redujeron y alquilaron, luego se precipitaron con 9 volúmenes de etanol frío durante la noche a -20 °C. El sedimento se lavó tres veces con una solución de etanol al 90% y se solubilizó en bicarbonato de amonio 50 mM. Se agregó tripsina en una proporción de 1:50 (enzima:sustrato) y la muestra se incubó durante 16 ha 37 °C. Finalmente, la mezcla de péptidos se desalinizó con cromatografía de fase inversa C18, se filtró con unidades Millex GV 0.22 µm (Millipore), se secó y se almacenó a -80 °C hasta el análisis LC-MS/MS. Los péptidos se resuspendieron en las siguientes fases móviles: fase A (0,1 % de ácido fórmico en agua) y fase B (80:10 (v/v) acetonitrilo:agua, 0,1 % de ácido fórmico). y se separaron en un sistema Dionex Ultimate 3000, RSLCnano UHPLC en línea acoplado a un espectrómetro de masas de alta resolución Q-Exactive Plus (Thermo Fischer Scientific). Posteriormente, las muestras primero se capturaron en una precolumna (C18 PepMap 100, 5 µm, 100 M, 300 µm de diámetro interior x 5 mm) y luego se separaron siguiendo un gradiente de elución por 88 minutos, utilizando una columna capilar a 250 nL/min (columna EASY Spray, PepMap RSLC, C18, 2 µm, 100 A, 75 µm x 50 cm). Se utilizaron El espectrómetro de masas se operó en modo de adquisición dependiente de datos positivos y el rango completo de MS fue de 375 a 1200 m/z. Los 10 iones más intensos, se aislaron en el cuadrupolo y se fragmentaron bajo disociación colisional de alta energía con una energía de colisión normalizada del 27%. Los iones precursores se midieron a una resolución de 70.000 (a 200 m/z) y los fragmentos se midieron a 17.500. Sólo los iones con estados de carga de 2 y superiores se fragmentaron con una ventana de aislamiento de 2 m/z. La proteómica de la espectrometría de masas se analizó mediante el software Max Quant.

6.10. Análisis bioinformático y estadístico de datos generados a partir de LC-MS/MS

El análisis de los datos obtenidos de la espectrometría de masas se realizó utilizando el software de proteómica cuantitativa (MaxQuant), diseñado para analizar una gran cantidad de datos. Este programa utiliza una matriz de cuantificación, en donde las intensidades de proteínas analizadas, correspondientes a las 4 muestras experimentales, son transformadas a log₂, cada una con sus respectivas réplicas. El análisis estadístico se llevó a cabo mediante el uso del lenguaje de programación R (versión 4.4) y M-S stats (Kohler *et al.*, 2023). Para calcular los cambios de intensidad proteica. entre cada muestra experimental, se ajustó un modelo lineal organizando los errores con un modelo Bayesiano empírico. Se identificaron como proteínas diferencialmente expresadas (DEPs, del inglés Differentially Expressed Proteins) aquellas con un cambio mayor a 1.5 (fold change) y un valor p ajustado entre 0.5 y 1.5. Previo al análisis, se imputaron y normalizaron las intensidades por medias, asegurando la robustez y validez de las comparaciones entre muestras.

6.11. Construcción de redes de interacción proteína-proteína y enriquecimiento biológico

Una vez identificadas las DEPs, se realizó un análisis por medio de un diagrama de Venn, con la finalidad de identificar las proteínas presentes en cada línea celular y comenzar la construcción de los interactomas. Para predecir la función de las DEPs, se tomaron aquellas que se repitieron en los cuatro modelos para construir un interactoma general mediante la base de datos STRING (Szklarczyk *et al.*, 2023). Esta base de datos integra interacciones proteína-proteína y posibles funciones. Para identificar los procesos biológicos que se encuentran enriquecidos en el interactoma, se realizó un análisis mediante Metascape (Zhou *et al.*, 2019) donde se implementó un algoritmo de detección de complementos moleculares (MCODE) para identificar los componentes de red densamente conectados.

Para el cálculo de significación estadística se empleó el conjunto completo de genes del genoma como punto de partida. Los términos seleccionados fueron aquellos con un valor de $p < 0.01$, un mínimo de 3 genes asociados y un factor de crecimiento > 1.5 , esto indica una relación significativa entre los genes observados y esperados.

Los valores p fueron calculados utilizando la distribución hipergeométrica acumulativa, mientras que los valores q se calcularon utilizando el procedimiento de Benjamini-Hochberg para controlar el error en múltiples pruebas. Para agrupar los términos enriquecidos, se utilizó la puntuación Kappa como medida de similitud durante la agrupación jerárquica. Considerando como grupos, aquellos subconjuntos con una similitud superior al 0,3. El representante de cada grupo fue seleccionado con el término estadístico más significativo.

implica que además de mantener funciones críticas, no pueden tolerar mutaciones perjudiciales, también suelen involucrarse en procesos fundamentales como la replicación de ADN, la transducción de señales y el metabolismo celular (Piccinini & Milani, 2023).

Cabe mencionar que el porcentaje de homología, mostrada en los alineamientos de las secuencias, podría proporcionar semejanzas en la función proteica; sin embargo, no es suficiente ya que no es posible garantizar la similitud estructural en los niveles de pliegue o dominio para una adecuada comparación entre las funciones conocidas (Altschul, 1997; Dobson *et al.*, 2004). Por esta razón se requiere de estudios adicionales que permitan explorar y verificar sus funciones.

7.2. Predicción de dominios en C19orf47

Con el objetivo de identificar los dominios y regiones específicas de C19orf47, se realizó un análisis de su secuencia con el uso de herramientas bioinformáticas (Figura 3). Entre los hallazgos más relevantes, se logró identificar el dominio de la super familia SAM (del inglés, Sterile Alpha Motif), y su subdivisión SAM_CS047, así como otra familia adicional perteneciente al grupo de las proteínas no caracterizadas, DUF5577.

La identificación de dominios específicos proporcionó información valiosa sobre la posible función biológica de la proteína estudiada. El dominio SAM identificado en la región N-terminal sugiere que C19orf47 puede estar involucrado en la mediación de las interacciones proteína-proteína, jugando un papel importante en la señalización y adhesión celular, así como en la organización del citoesqueleto (Qiao & Bowie, 2005; Wang *et al.*, 2023).

Los dominios SAM tiene diversas funciones y pueden interactuar con proteínas, ARN y lípidos de membranas. Principalmente pueden desempeñar un papel en la oligomerización de proteínas y en procesos que van desde la traducción de señales hasta la regulación de la transcripción. Así mismo, las mutaciones en los dominios SAM se han relacionado con varias enfermedades. La subdivisión

SAM_CS047, detectada en la secuencia de C19orf47, es un dominio muy extendido en proteínas de señalización y regulación, aunque se desconoce la función exacta de las proteínas pertenecientes a este grupo (Kim, 2003).

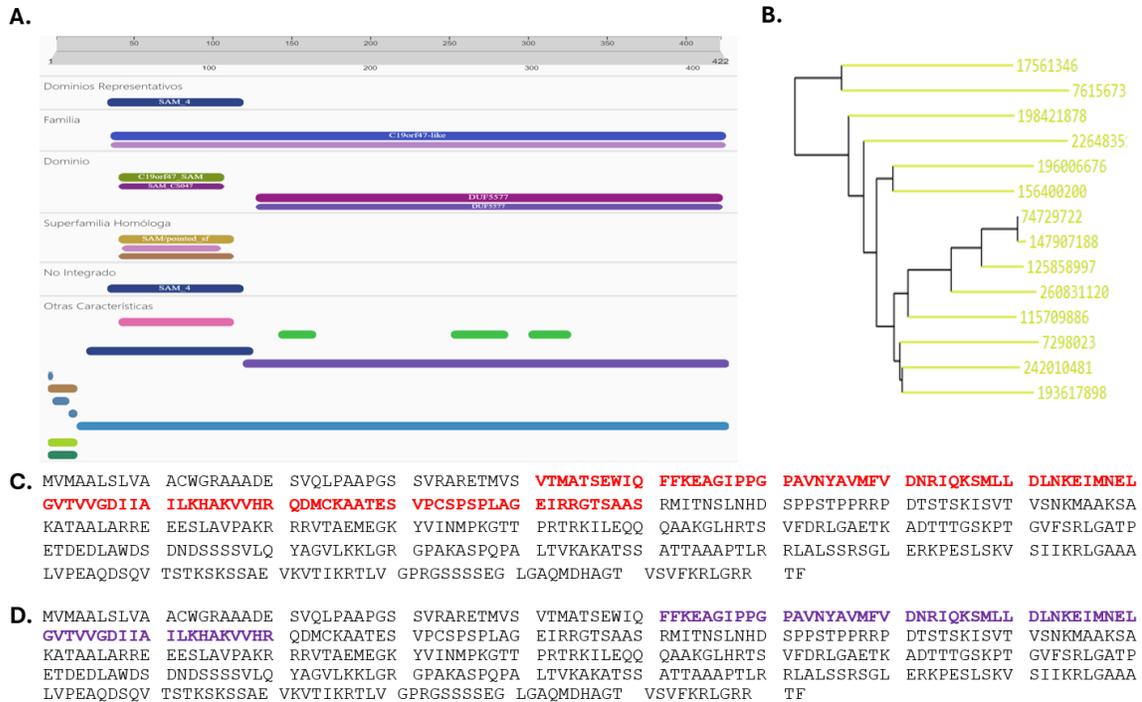


Figura 3. Dominios en la secuencia de aminoácidos de C19orf47. **A)** Se muestran diferentes dominios y regiones a lo largo de la secuencia, lo que proporciona una comprensión visual de su estructura y potenciales funciones, siendo el dominio SAM y su subdivisión CS047 los más destacables. **B) Taxonomía general de la subdivisión de SAM_CS047.** Árbol filogenético, con varias ramas, representando diferentes taxones dentro del metazoo, cada línea conecta a un nodo que representa la subdivisión del subdominio de SAM **C y D) Secuencia primaria de aminoácidos de C19orf47.** En color rojo se representa el dominio de la super familia SAM y en morado se muestran los aminoácidos del dominio SAM-CS047, en la secuencia de C19orf47.

El dominio SAM identificado en C19orf47, sugiere que esta proteína puede tener un papel en la mediación de interacciones proteína-proteína, la organización del citoesqueleto y la señalización y adhesión celular. Además, este dominio no solo está relacionado a las funciones ya mencionadas, sino que también ha sido asociado con diversas enfermedades, entre las cuales se incluyen al cáncer. Estas observaciones subrayan la importancia de investigar cómo las alteraciones en el dominio SAM, y el subgrupo reportado, podrían contribuir a la progresión tumoral y la resistencia a tratamientos.

7.3. Análisis de los niveles de expresión genética en tejidos tumorales y no tumorales

Para identificar el nivel de expresión de C19orf47, en distintos tejidos sanos y tumorales, se utilizaron las bases de datos de GEPIA y HPA (proteinatlas.org) (Uhlén *et al.*, 2015) (Figura 4).

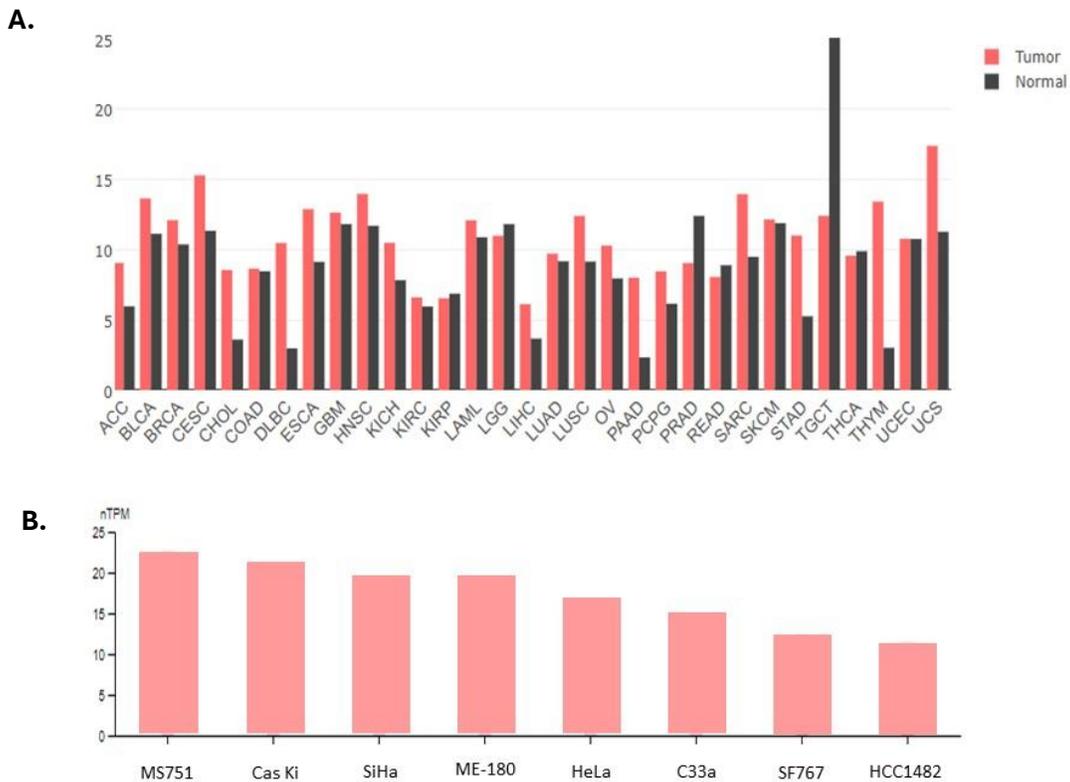


Figura 4. Expresión general de C19orf46 en tejidos no tumorales y tumorales. A) Presencia de la proteína C19orf47 en tejidos sanos contra tejidos tumorales. Tomado de GEPIA (Z. Tang *et al.*, 2017) B) Niveles existentes de ARNm de C19orf47, en diferentes líneas de cáncer cervicouterino. Tomado de The Human Protein Atlas (Uhlén *et al.*, 2015)

Los resultados demostraron que esta proteína se expresa significativamente en tejidos tumorales en comparación con tejidos no tumorales. En particular, se observó una alta expresión de C19orf47 en los siguientes tipos de cáncer: Carcinoma de células escamosas de cuello uterino (CESC, del inglés Cervical Squamous cell carcinoma and endocervical adenocarcinoma) y carcinosarcoma uterino (UCS, del inglés Uterine Carcinosarcoma) (Figura 4A). Mientras que los

niveles de ARNm, analizados en diferentes líneas de CaCu, presentan niveles más altos del mensajero de C19orf47 en las líneas tumorales de cérvix infectado con VPH en comparación con aquellas sin la presencia de la infección por VPH Figura 4B). Esto sugiere la implicación de C19orf47 en la progresión del CaCu.

Esto sugiere la implicación de C19orf47 en la progresión del CaCu. Diversos estudios han demostrado que la sobreexpresión de ciertas proteínas en cáncer está relacionada directamente con diversos procesos como la progresión tumoral y resistencia a la apoptosis, entre otros (Hozhabri *et al.*, 2022; Pessoa *et al.*, 2022).

7.4. Expresión de C19orf47 en diferentes líneas celulares (HaCaT, C33a, HeLa y SiHa)

Para determinar la expresión de la proteína C19orf47, en las líneas celulares HaCaT, C33a, HeLa y SiHa, se realizaron ensayos mediante Western Blot (WB) con extractos de proteínas totales (Figura 5).

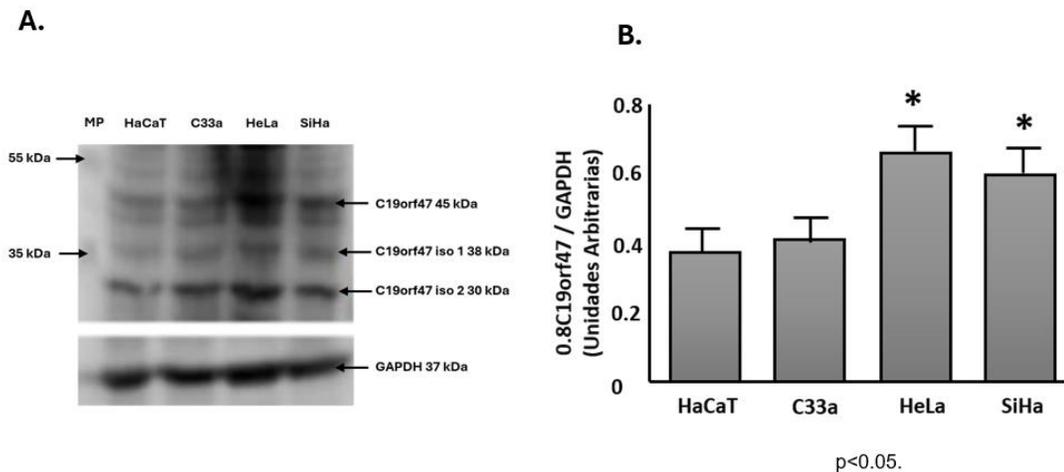


Figura 5. Expresión a nivel de proteína C19orf46 analizada mediante Western Blot (WB). **A.** Determinación de los niveles de proteína C19orf47, en líneas celulares CCU, por WB. Con su respectivo control de carga con la proteína GAPDH (del inglés Glyceraldehyde 3-phosphate dehydrogenase). **B.** Cuantificación de las bandas, por densitometría, del WB de C19orf47 en líneas celulares de CCU. Se muestra la media y la desviación estándar de la densitometría, de las bandas de C19orf47, respecto a actina en 3 experimentos independientes. Realizado con el software ImageJ y análisis de ANOVA, comparación múltiple Bonferroni (n=3). * p < 0.05.

En el WB se logró identificar la señal correspondiente a la proteína C19orf47 (45 kDa), con sus dos isoformas (38 y 30 kDa respectivamente), observándose diferencias significativas entre las líneas celulares. Específicamente, se encontró que en la línea celular HaCaT, no carcinogénica, la expresión de C19orf47 fue notablemente menor, en comparación con las líneas cancerosas. Esta diferencia sugiere que la expresión de C19orf47 podría estar relacionada con el estado maligno de las células ya que su expresión se observa incrementada su expresión en células tumorales.

Además, se observaron diferencias en la expresión de C19orf47 entre las líneas cancerosas no infectadas con VPH (C33a) y las infectadas con VPH (HeLa y SiHa). Las células infectadas, con VPH, mostraron una sobreexpresión notable de C19orf47 en comparación con las no infectadas ($p < 0.05$). Esto sugiere que la expresión de C19orf47 podría estar siendo regulada por la presencia del VPH, lo cual tendría implicaciones importantes en la progresión del cáncer y podría estar relacionada con el CaCu inducido por el VPH sugiriendo un papel potencial de esta proteína en la promoción de la proliferación celular o en la evasión de los mecanismos de control celular inducidos por el VPH.

En numerosos estudios se ha demostrado como los VPH de alto riesgo como el serotipo 16 y18, son la causa de varios cánceres humanos importantes tras la desregulación génica viral y la interacción entre proteínas y su capacidad de modular a supresores de tumores (Doorbar *et al.*, 2012). Por estas razones, se sugiere la existencia de un mecanismo por el cual el VPH, en conjunto con C19orf47, contribuyen a la progresión tumoral mediante la alteración de vías de señalización celular.

7.5. Localización subcelular de C19orf47

7.5.1. Predicción de la localización subcelular de C19orf47

Para predecir la localización subcelular de C19orf47 se emplearon múltiples herramientas bioinformáticas especializadas en el análisis de secuencias proteicas

como DeepLoc 2.1 (Odum *et al.*, 2024). Esta plataforma se basa en el análisis de la secuencia de aminoácidos de la proteína identificando la presencia de señales de localización nuclear (NLS, del inglés Nuclear localization signals), arrojando una localización principalmente nuclear definida por todas las herramientas utilizadas. De acuerdo con el análisis realizado con ProtT5-XL-Uniref50 (ProtT5), se predijo que C19orf47, además de distribuirse en el núcleo celular, podría tener presencia en el citoplasma (Figura 6).

Aunque la puntuación de localización citoplasmática fue menor en comparación con la señal nuclear, la presencia de ciertas señales sugiere que también puede desempeñar funciones tanto en el núcleo como en el citoplasma.

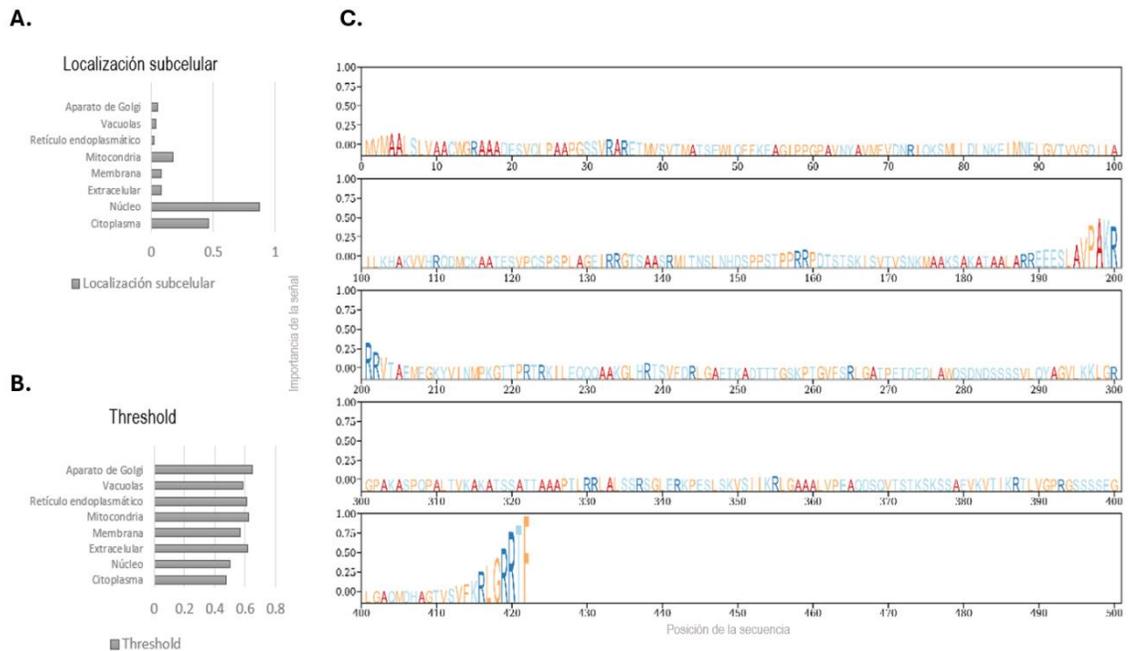


Figura 6. Predicción de la localización subcelular de C19orf47, mediante redes neuronales profundas, utilizando el método ProtT5. Este modelo proporciona una predicción más precisa a expensas de un tiempo de cálculo más largo debido al tamaño del modelo (3 mil millones de parámetros). **A.** Tabla de localizaciones subcelulares predichas. **B.** Umbrales de probabilidad para las localizaciones subcelulares. Se predice una localización si su probabilidad está por encima del umbral que se muestra. **C.** Representación de las señales de localización para C19orf47 para determinar su posible localización subcelular.

7.5.2. Detección de la localización de C19orf47 mediante fraccionamiento citoplasmático y nuclear

Como un método de validación, se utilizó la técnica WB con extractos de las fracciones citoplasmáticas y nucleares provenientes de la línea celular HeLa, puesto que ha demostrado ser la línea celular que mayormente expresa a C19orf47 (Figura 7). Los resultados confirmaron que C19orf47 se expresó tanto en la fracción citoplasmática como en la fracción nuclear. En la figura 7 se observa que la intensidad de la banda, correspondiente a C19orf47, fue similar en la fracción citoplasmática y nuclear. Vale la pena destacar la disminución de la isoforma 2 (30 kDa) en la fracción nuclear, misma que se ve sobre expresada en la fracción citoplasmática, lo que podría sugerir funciones diferentes en cada localización subcelular (Figura 7).

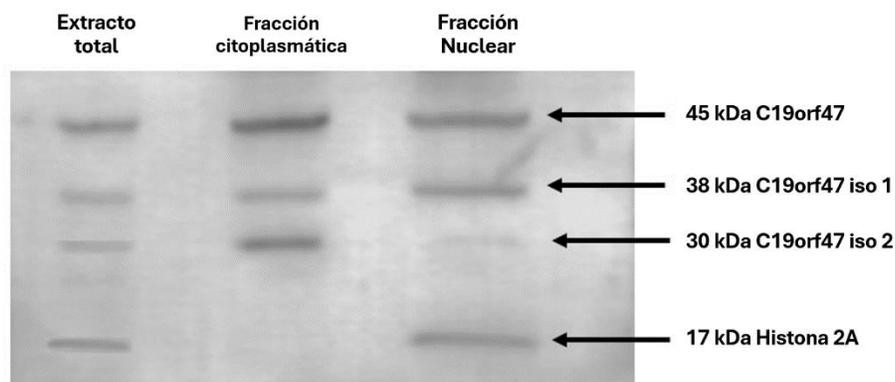


Figura 7. Validación de la localización subcelular de C19orf47 por WB en muestras de fracciones del citoplasma y del núcleo. En el ensayo se utilizó el anticuerpo primario, anti-C19orf47, y para el control nuclear se utilizó el anticuerpo primario, anti-Histona 2A. Como control positivo se utilizó un extracto de proteínas totales.

7.5.3. Inmunofluorescencia de C19orf47

Para confirmar la localización celular, de C19orf47, se realizaron ensayos de inmunofluorescencia en las diferentes líneas celulares. En la figura 8, se identifica la señal de C19orf47, misma que puede ser visualizada mediante el fluoróforo, representado en color verde, esta señal permite identificar la distribución y localización de C19orf47 en las líneas celulares HaCaT, C33a, HeLa y SiHa.

Adicionalmente, se tiñó con el colorante fluorescente DAPI el núcleo celular, este colorante se une al ADN y permite la visualización precisa de la estructura nuclear para proporcionar una referencia clara para identificar y analizar la ubicación de C19orf47 en relación con el núcleo celular. La superposición o “merge” de la señal verde de C19orf47 con la señal azul correspondiente a DAPI facilita la evaluación de la co-localización de C19orf47 en el núcleo y citoplasma, permitiendo una interpretación más compleja y precisa de la distribución subcelular de esta proteína.

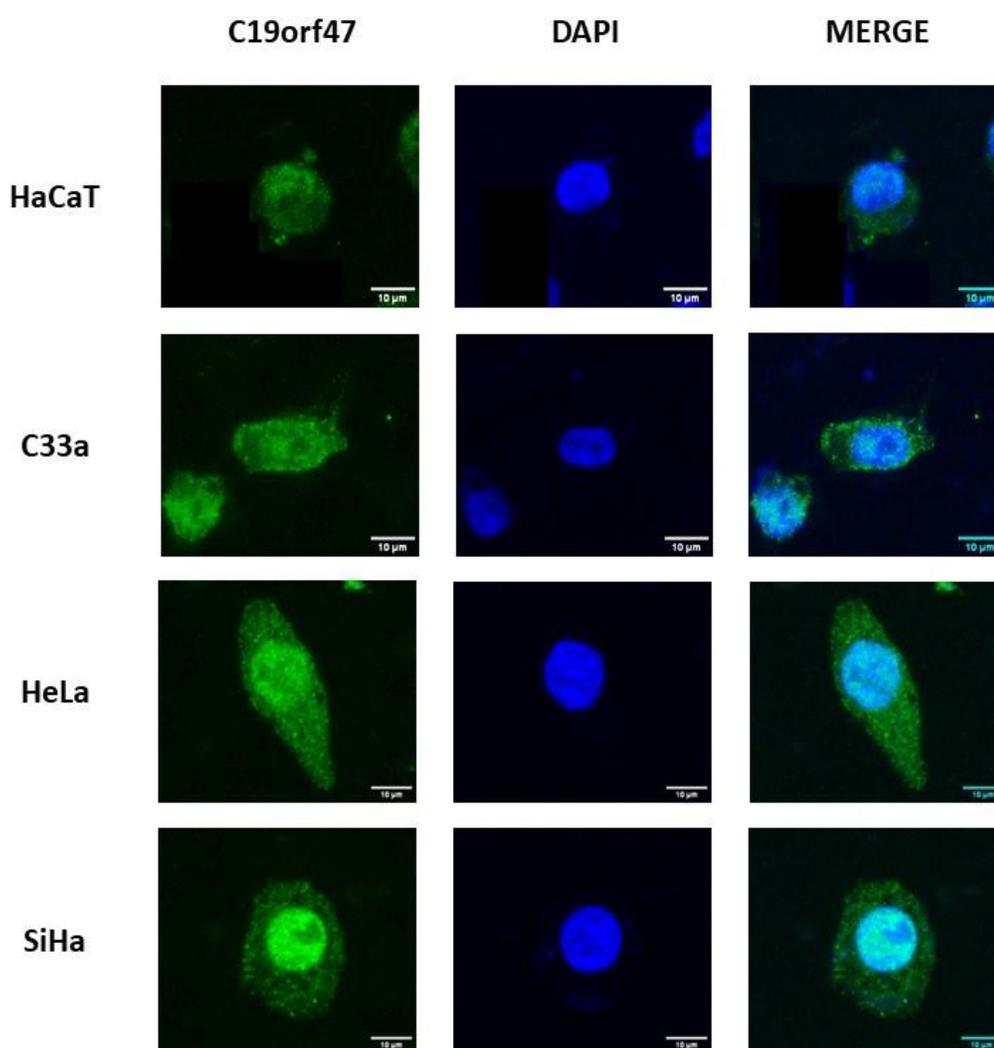


Figura 8. Microscopia confocal en líneas celulares de CaCu. Para identificar la señal de C19orf47 se utilizó un anticuerpo secundario acoplado al fluoróforo Alexa 488 (verde) y para la tinción nuclear se empleó el colorante fluorescente DAPI (del inglés 4',6-diamidino-2-phenylindole), mismo que se representa en color azul. Las imágenes se tomaron con el objetivo de 60X y un aumento de 2X a varios cortes.

En la línea celular, no cancerosa, HaCaT, C19orf47 se localizó predominantemente en el núcleo celular, y en mucho menor grado en el citoplasma. Mientras que, en las líneas celulares cancerosas, C33a, HeLa y SiHa, la señal se identificó con mayor intensidad en el núcleo, en comparación con el citoplasma. Particularmente la línea celular cancerosa, no infectada con VPH (C33a), presentó una señal más débil que las líneas de CaCu infectadas con VPH (HeLa y SiHa).

Aunque la literatura redice que C19orf47 se localiza en el nucleoplasma; los experimentos de fraccionamiento celular e inmunofluorescencias revelaron la presencia de C19orf47 en el núcleo y citoplasma en las líneas HaCaT, C33a, HeLa y SiHa.

En este sentido, tanto en líneas cancerosas, como no cancerosas, mantienen la misma tendencia de expresión celular. Y de acuerdo con los diferentes niveles de expresión de C19orf47, se puede identificar mayor expresión de la proteína en células cancerosas infectadas por VPH, en comparación con la menor expresión de C19orf47 de las células cancerosas, no infectadas con VPH. Mientras que las células no cancerosas, muestran la menor expresión de C19orf47.

La presencia de C19orf47 en más de un organelo subcelular resalta la complejidad de las funciones de la proteína y predice que su efecto en CaCu podría estar mediado por mecanismos distintos en cada localización. Sin olvidar que la isoforma 2 de C19orf47, disminuye drásticamente en la fracción nuclear de la línea celular HeLa, infectada con VPH 18, en comparación con la fracción citoplasmática. Por lo tanto, se podría sugerir que C19orf47 está involucrada en distintas funciones, de acuerdo con su localización.

7.6. Co inmunoprecipitación de C19orf47

Con el objetivo de identificar las proteínas que interactúan con C19orf47, se diseñó una estrategia que consta de tres pasos: 1) Co-inmunoprecipitación, 2) Análisis por LC-MS/MS con el método de adquisición dependiente de datos (DDA, por sus siglas en inglés) y 3) Análisis de los datos proteómicos. Esto proporcionó una visión detallada sobre las interacciones proteicas y sugiriendo una perspectiva funcional de esta uPE1.

La eficacia de la precipitación, del complejo de interacción de C19orf47, se validó mediante WB. Una vez confirmada la captura del complejo, el precipitado se analizó por LC-MS/MS por el método DDA, lo cual permitió la identificación y cuantificación de múltiples péptidos en cada inyección. Cabe mencionar que, para tener un análisis completo, cada muestra se analizó múltiples veces. Gracias a este análisis se obtuvieron las intensidades de cada péptido presente en las precipitaciones de C19orf47 en cada una de las líneas celulares utilizadas.

7.6.1. Identificación de proteínas por LC-MS/MS

Para garantizar la robustez y fiabilidad de los datos obtenidos del análisis por LC-MS/MS, las intensidades proteicas obtenidas se normalizaron por medias para ajustar los datos y asegurar comparaciones precisas entre las diferentes muestras realizadas (Figura 9A).

Como se observa en la Figura 9B, se elaboró un gráfico de doble entrada que representa de manera integral los péptidos identificados (morado) y cuantificados (azul). Este gráfico proporciona de forma detallada la relación entre la identidad y la abundancia de los péptidos en las diferentes muestras con sus respectivas repeticiones. Además, se observa que las líneas celulares tumorales infectadas con VPH (HeLa y SiHa), muestran una mayor diversidad y abundancia en comparación con las otras líneas celulares utilizadas.

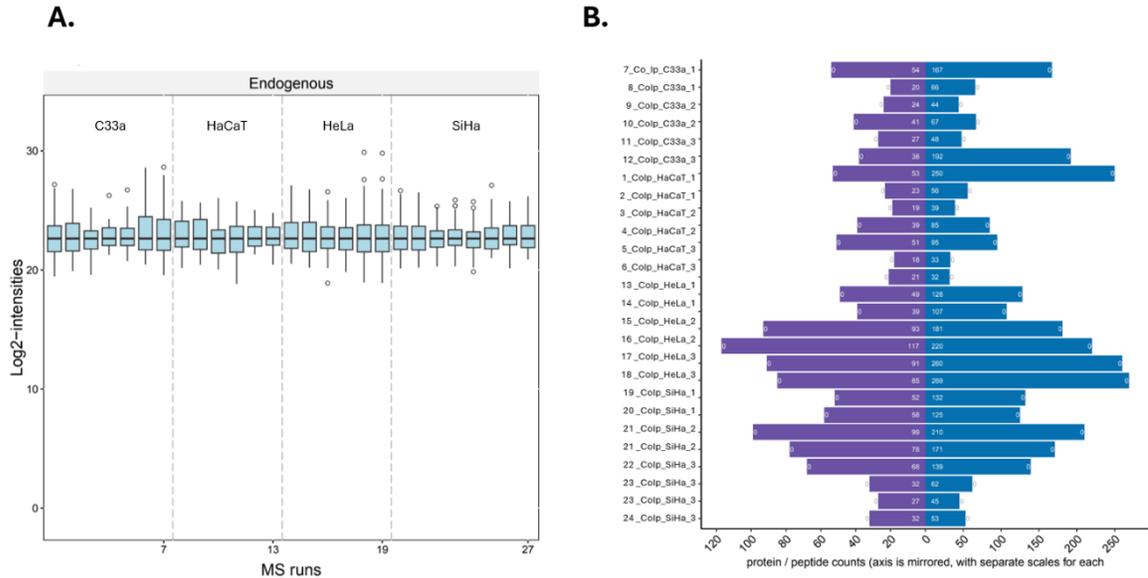


Figura 9. Espectrometría de masas de muestras co-inmunoprecipitadas de C19orf47. A. Normalización de las muestras inyectadas por medio de medias. **B.** Proteínas y péptidos identificados y cuantificados en cada inyección. En color morado se presentan los péptidos identificados y en color azul la abundancia de estos.

Para identificar mediante M-S stats las proteínas diferencialmente expresadas (DEPs) en cada una de las precipitaciones de cada línea celular analizada, se tomaron en cuenta aquellas con un cambio mayor a 1.5 (fold change) y un valor p ajustado entre 0.5 y 1.5.

Una vez identificadas las DEPs, se realizó un diagrama de Venn mediante Jvenn (Bardou *et al.*, 2014) para identificar las proteínas comunes entre las cuatro líneas celulares (Figura 10), en este diagrama se puede observar que HeLa continúa sobresaliendo, del resto de las otras líneas celulares, como la línea celular que presentó más interacciones de proteínas con C19orf47 (101 DEPs). Por otro lado, SiHa, otra línea de CaCu infectada con VPH, también mostró un número significativo de interacciones, aunque menor que HeLa con 93 DEPs identificadas (Figura 10B).

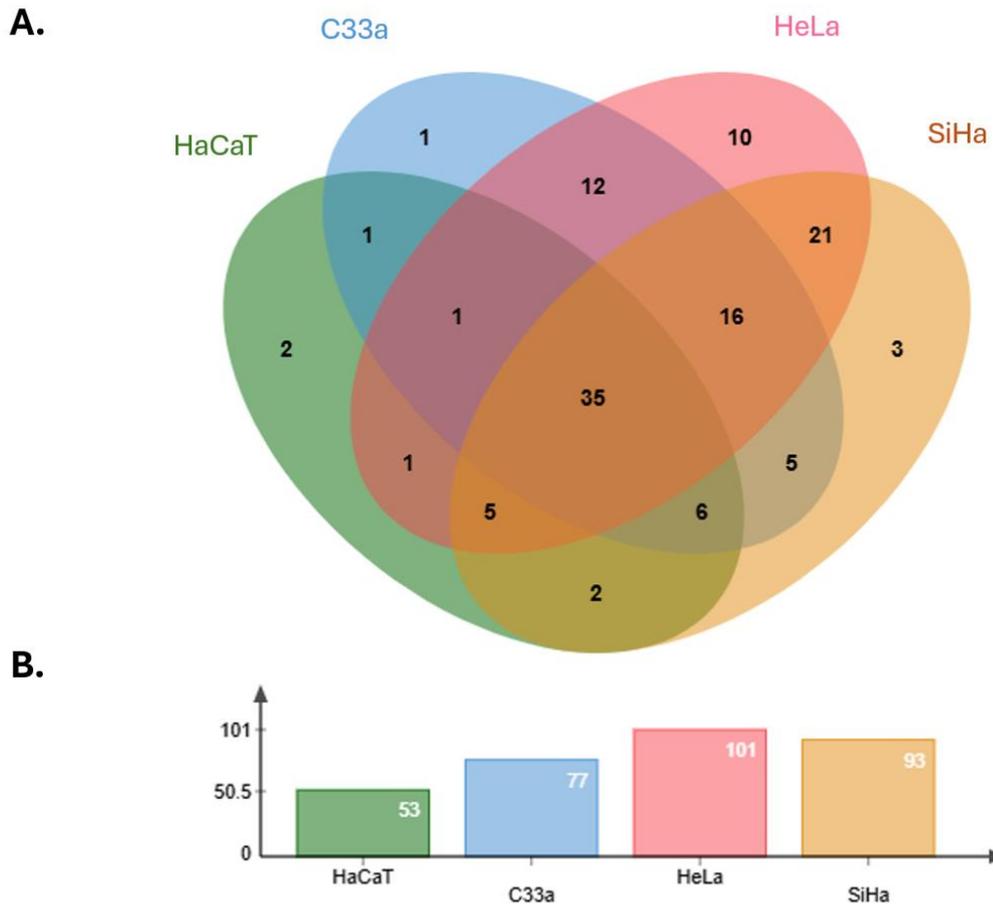


Figura 10. Análisis de interacciones Proteína-Proteína (P-P) en Líneas Celulares. **A.** El Diagrama de Venn representa a las proteínas DEPs compartidas entre las diferentes líneas celulares. HaCaT, C33A, HeLa y SiHa. Las áreas de intersección resaltan a las proteínas comunes entre las diferentes líneas celulares, brindando información sobre la superposición en la interacción proteica. **B.** El gráfico de barras representa el número de interacciones de cada línea celular, ofreciendo una visión comparativa de su contribución al conjunto de proteínas estudiadas.

La presencia del incremento de las interacciones en las líneas de CaCu, infectadas con VPH, sugieren una relación específica entre C19orf47 y la patología del cáncer cervical inducido por esta infección. En contraste, la línea celular no cancerosa, HaCaT, mostró un número significativamente menor de interacciones proteicas (53 DEPs), sugiriendo que las interacciones de C19orf47 podrían ser menos interactivas o diferentes en este tipo de células no cancerosas.

7.6.1.1. Construcción del interactoma general de C19orf47 y su enriquecimiento biológico y molecular

Para sugerir la posible función de C19orf47, resultó indispensable identificar la red de proteínas con las que interactúa. Ya que las proteínas forman una red dinámica de interacciones, con la finalidad de mecanizar sus funciones, a través de las relaciones entre los miembros que la conforman. De este modo, las redes interatómicas pueden brindar información útil sobre la función, los procesos biológicos y su implicación en la progresión de enfermedades (Barabási *et al.*, 2011; X. Zhao & Liu, 2019).

Para realizar un análisis detallado del interactoma de C19orf47, y poder sugerir su posible función se seleccionaron aquellas proteínas DEPs presentes en todas las líneas celulares, según los datos obtenidos del diagrama de Venn previamente descrito (Figura 10A). Posteriormente, estas DEPs se analizaron en la base de datos de STRING para generar el interactoma general de C19orf47 (Szklarczyk *et al.*, 2023). Esta base de datos es una herramienta útil en la interpretación funcional de las interacciones entre proteínas conocidas y predichas en base con datos experimentales y predicciones bioinformáticas, es por ello por lo que en ocasiones se omiten ciertas proteínas de una lista dada, o bien, no se reportan interacciones dentro del interactoma generado.

Como se muestra en la Figura 11, el interactoma resultante engloba un conjunto de 28 DEPs de las 35 que se identificaron en las líneas 4 celulares: HaCaT, C33a, HeLa y SiHa. La disminución en los miembros puede deberse a la falta de datos experimentales o bioinformáticos, o bien, a la repetición entre miembros.

Para obtener un análisis completo, se realizó una subdivisión en tres grupos (clusters, por su traducción al inglés) distintos. Con la finalidad de facilitar el entendimiento total de las interacciones, cada cluster destaca las potenciales interacciones y funciones asociadas con C19orf47 (Figura 11).

Los resultados arrojados por STRING, muestran que en el primer cluster (representado en color rojo) se encuentran proteínas que podrían estar involucradas en la regulación metabólica, respuesta inflamatoria, mantenimiento de la estructura celular y la estabilidad proteica. El segundo cluster, en color verde, predice que las proteínas que conforma este grupo corresponden a clases funcionales relacionadas con la regulación de la diferenciación celular, protección contra el estrés mecánico, adhesión y cohesión celular, mantenimiento del citoesqueleto e integridad en el epitelio. Finalmente, el tercer cluster, representado en color azul, podría estar implicado en el metabolismo lipídico, respuesta del sistema inmunitario, transporte intracelular y producción energética. Igualmente se identificó que este interactoma general participa en distintas vías metabólicas como la glucólisis por senescencia y el ciclo de Cori, debido a la abundancia de proteínas clave de estas vías como GAPDH y PKM.

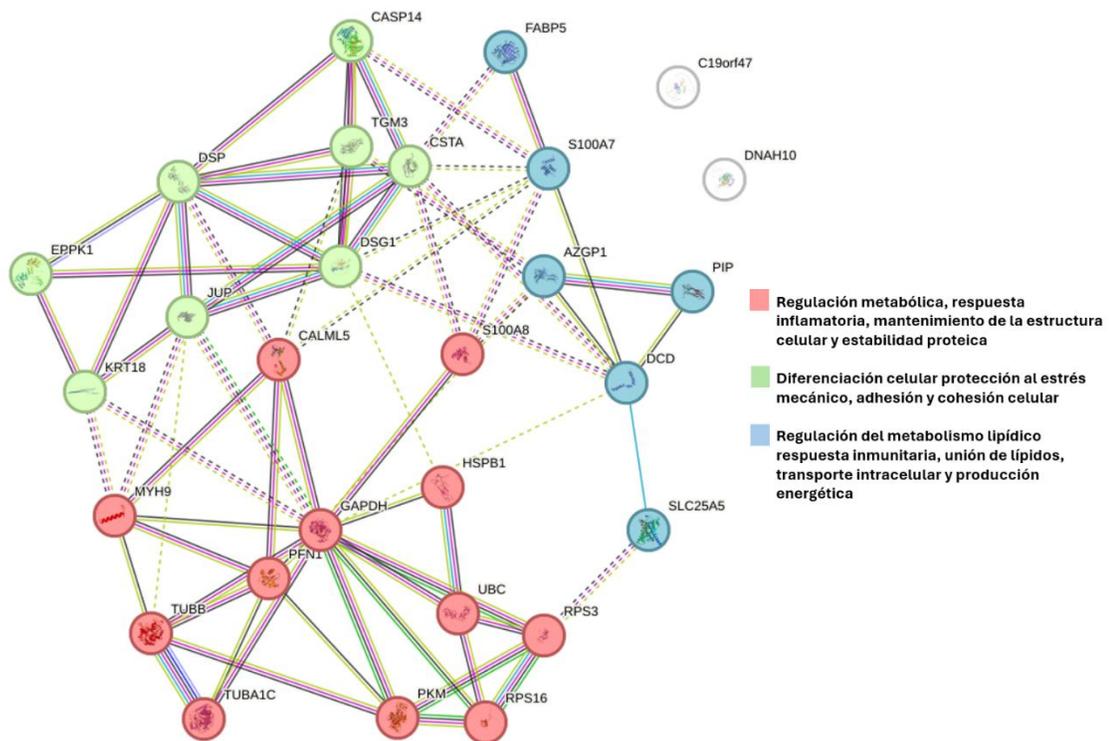


Figura 11. Red de interacciones P-P, obtenidas en el Interactoma de C19orf47. En la red se muestran tres clusters de las DEPs que interactúan con C19orf47, en las cuatro líneas celulares analizadas. Los nodos representan las proteínas que interactúan directamente con C19orf47 y las líneas indican las interacciones entre las proteínas. Los colores de los nodos indican los procesos biológicos en los que participan las proteínas, según la leyenda.

Posterior al análisis del interactoma de las DEPs comunes en las cuatro líneas celulares, se realizó un análisis mediante Metascape (Figura 12), esta es una plataforma bioinformática que permite integrar diversos datos ómicos de múltiples fuentes y es capaz de realizar análisis de enriquecimiento para asociar los genes diferencialmente expresados con términos funcionales específicos de GO (del inglés, Gene Ontology), rutas metabólicas o vías de señalización específicas usando bases de datos como KEGG, Reactome, o Pathway Commons. Además, busca genes que codifican factores de transcripción, que son proteínas que regulan la transcripción de otros genes. Estos factores suelen tener un papel central en la regulación de la expresión génica y pueden ser identificados por su participación en múltiples rutas y procesos funcionales relevantes.

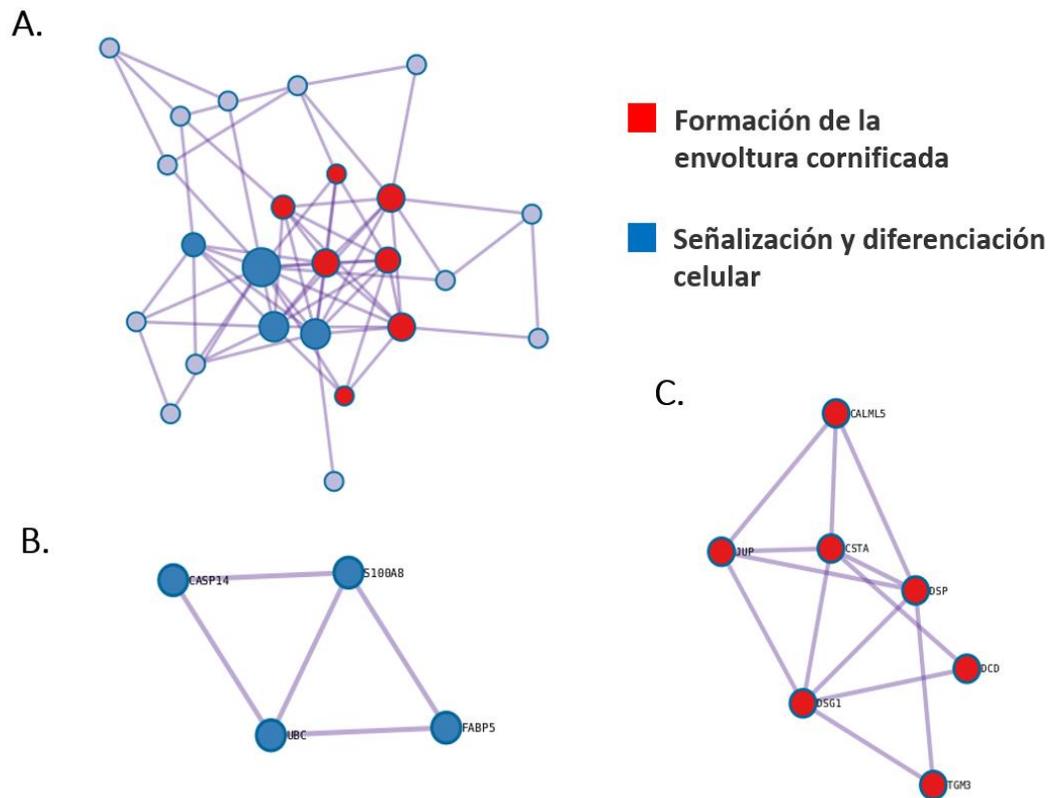


Figura 12. Interactomas proteína-proteína de líneas celulares identificados por MCODE (Proteínas compartidas). **A.** Los nodos representan proteínas y los bordes indican interacciones. Los componentes de red densamente conectados identificados por MCODE se resaltan con diferentes colores y se asignan funciones biológicas enriquecidas en cada módulo, como se indica en la leyenda. El tamaño de los nodos es proporcional al número de interacciones de cada proteína. **B.** Se presentan las proteínas relacionadas con la señalización, diferenciación y homeostasis celular. **C.** Se presentan las proteínas involucradas en la formación de la envoltura

Se identificaron dos clústeres principales que destacan por su relevancia funcional, el primero, representado en color azul, incluye una red de proteínas asociadas con la señalización, diferenciación y la homeostasis celular (Figura 12B), mientras que el segundo clúster, representado en color rojo está relacionado en la reticulación de péptidos (Figura 12C), esto lo que podría sugiriere una participación en procesos de PTM y estructurales dentro de la célula.

Para identificar la variedad de procesos biológicos, se realizó un enriquecimiento por ontología de genes (GOs, del inglés Gene Ontology), igualmente en Metascape con las DEPs identificadas en las cuatro líneas celulares utilizadas (HaCaT, C33a, HeLa y SiHa). Los resultados mostraron que el interactoma de C19orf47 está relacionado con la degranulación de neutrófilos, el desarrollo de la epidermis, la adhesión celular, la regularización de la migración celular y la estabilización de proteínas (Figura 13).

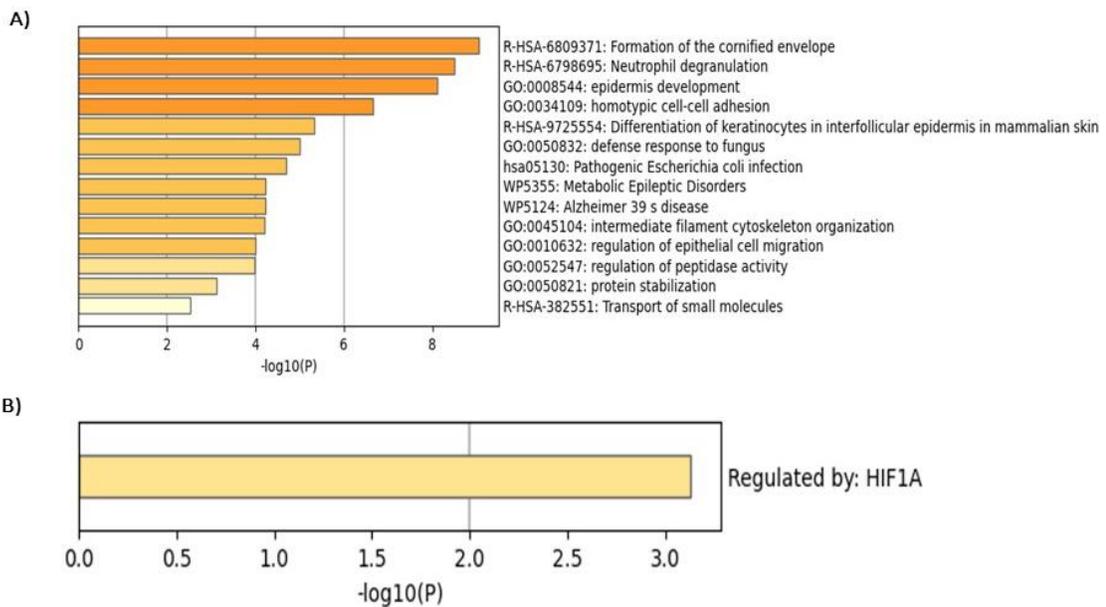


Figura 13 Enriquecimiento de funciones biológicas, en tejidos y de reguladores del interactoma proteína-proteína de C19orf47 en las líneas celulares HaCaT, C33a, HeLa y SiHa. A) Las barras representan las funciones sobrerrepresentadas, con sus respectivos valores de enriquecimiento y p-valores corregidos. B) Enriquecimiento de reguladores transcripcionales. Se representan al regulador transcripcional HIF1A, sobrerrepresentado en el interactoma. Los valores de enriquecimiento y significancia estadística se indican para cada categoría de regulador.

Se identificó a HIF1A como un posible factor de transcripción responsable de regular las interacciones y funciones del interactoma. Este factor de transcripción es conocido por su papel en la respuesta celular a la hipoxia, que se activa cuando hay bajas concentraciones de oxígeno, lo que provoca la transcripción de genes relacionados con la angiogénesis, el metabolismo energético y la supervivencia celular (Paredes *et al.*, 2021; Rashid *et al.*, 2021).

La detección del factor de transcripción HIF1A como un posible regulador, sugiere que C19orf47 tendría un papel en la regulación del ambiente celular en respuesta a la hipoxia, así como en la regulación de la expresión génica a niveles bajos de oxígeno, o bien, participar en la actividad transcripcional mediada por HIF1A.

Para comprender más a fondo el interactoma de C19orf47 y sus funciones biológicas, se construyó un interactoma individual a partir de cada una de las DEPs de cada una de las líneas celulares mediante STRING.

El análisis de los resultados, correspondientes a los interactomas de las cuatro líneas celulares, mostró que C19orf47 está involucrado en la regulación metabólica, las respuestas inflamatorias e inmunitarias, el mantenimiento de la estructura celular y la estabilidad proteica tanto en líneas cancerosas y no cancerosas.

Además, se esperaría que participará en la regulación de la diferenciación celular, la protección contra el estrés mecánico, la adhesión y la cohesión celular, el mantenimiento del citoesqueleto y la integridad del epitelio, así como en la regulación de la diferenciación celular, la respuesta del sistema inmune, el transporte intracelular y la producción energética (Figura 14).

En ausencia de propiedades carcinogénicas (HaCaT), C19orf47 mantiene interacciones mayoritariamente nucleares, esto podría indicar que podría influir en procesos biológicos fundamentales como mediador de las interacciones proteína-proteína, síntesis proteica y la regulación de la proliferación celular. Igualmente se

identifica que podría ser blanco de PTMs relacionadas con la acetilación, hidroxilación, mismas que podrías influir en su funcionalidad. Además, parece estar involucrada en funciones críticas para el mantenimiento de la función normal de las células epiteliales. Esto incluye la estructura celular, adherencia, integridad epitelial y el mantenimiento de la barrera epitelial funcional (Figura 14A).

En la línea celular C33a, se sugiere que C19orf47 estaría implicada en la síntesis y degradación de proteínas, proliferación y la unión celular. Además, podría estar asociada con la diferenciación y la señalización celular, así como la progresión tumoral (Figura 14B).

En contraste, en líneas celulares cancerosas infectadas por VPH (HeLa y SiHa), C19orf47 podría participar en la regulación de la estructura celular, respuesta al estrés, señalización celular, respuesta inflamatoria y la apoptosis. Además, podría interactuar en la síntesis de proteínas y el transporte intracelular, así como en la unión celular y el transporte celular. Igualmente se identificaron proteínas que están reguladas por la infección del VPH y ayudan a las células a adaptarse a condiciones adversas (Figura 14C y D).

Estos hallazgos sugieren que la función de C19orf47 podría estar mediada por la presencia de las propiedades carcinogénicas y al mismo tiempo estar involucrada en la progresión del cáncer bajo la presencia del VPH. En células no cancerosas como HaCaT, C19orf47 está más centrado en mantener la integridad y funcionalidad epitelial. Mientras que en células cancerosas y en presencia de VPH, C19orf47, podría involucrarse en la progresión tumoral, respuesta al estrés celular y la adaptación a condiciones adversas.

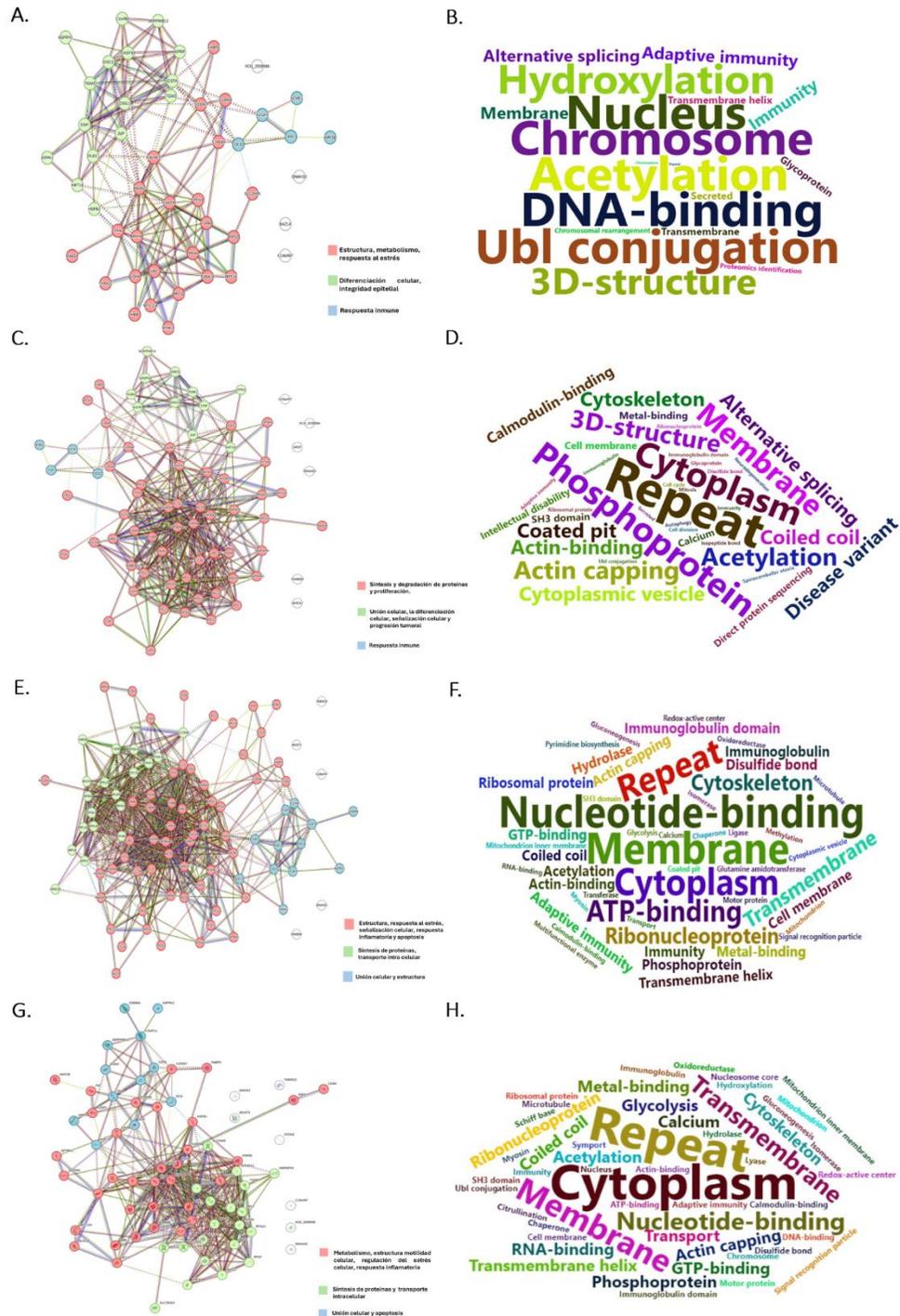


Figura 14. Interactomas de C19orf47, de cada una de las líneas celulares: HaCaT, C33a (VPH -), HeLa (VPH 16) y SiHa (VPH 18). Del lado izquierdo se observan las redes de los interactomas, de C19orf47, que correspondientes a las líneas celulares HaCaT (A), C33a (C), HeLa (E) y SiHa (G) respectivamente. Cada nodo representa las proteínas que interactúan directamente con C19orf47 y los bordes indican las interacciones entre ellas. Los colores de los nodos indican los procesos biológicos en los que participan las proteínas, según la leyenda. Del lado derecho se presentan los mapas de las clases funcionales, correspondientes a los interactomas de C19orf47, de cada línea celular: HaCaT (B), C33a (D), HeLa (F) y SiHa (H).

7.6.1.2. Redes Interatómicas de C19orf47, mediante MCODE en HaCaT, C33a, HeLa y SiHa

Para ampliar el entendimiento de los procesos biológicos o funciones en las que podría involucrarse C19orf47, se realizó un análisis por MCODE del interactoma de C19orf47 mediante Metascape a partir de las DEPs de cada una de las líneas celulares utilizadas (Figura 15). Este análisis individual ofrece variaciones entre los diferentes componentes moleculares con los que C19orf47 podría estar involucrado entre los diferentes contextos celulares y condiciones (como la infección por VPH).

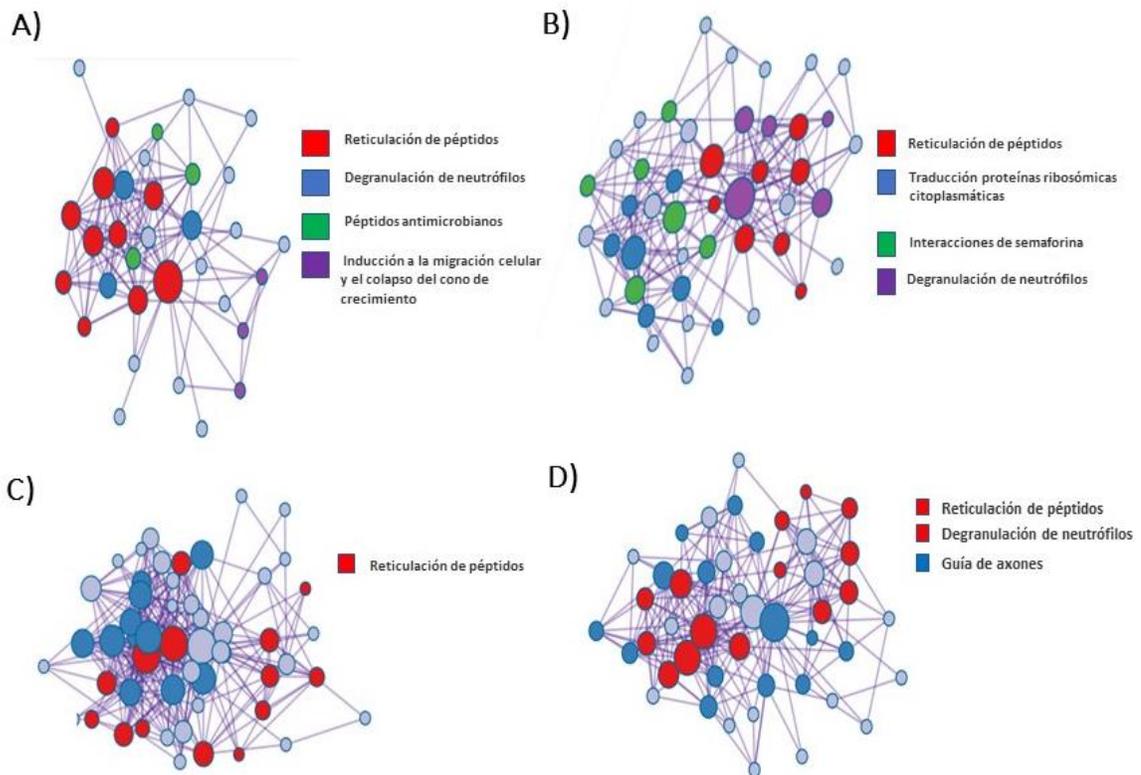


Figura 15. Red interatómica P-P de C19orf47, de 4 líneas celulares, identificados por MCODE. A-D) Redes Interatómicas en HaCaT, C33a, HeLa y SiHa, respectivamente. Los nodos representan proteínas y los bordes indican interacciones. Los componentes de red densamente conectados identificados por MCODE se resaltan con diferentes colores y se asignan funciones biológicas enriquecidas en cada módulo, como se indica en la leyenda. El tamaño de los nodos es proporcional al número de interacciones de cada proteína.

Al igual que en el análisis del interactoma general de C19orf47, se observó que en las cuatro líneas analizadas (HaCaT, C33a, HeLa y SiHa) procesos como la degranulación de neutrófilos y la reticulación de péptidos están presentes de manera consistente. Sin embargo, en comparación con el análisis general, además de los componentes ya mencionados, se observaron posibles implicaciones adicionales de C19orf47. En la línea celular no tumoral, HaCaT, se identificó que esta proteína podría estar involucrada en la dinámica celular. Además, se identificaron interacciones con péptidos antimicrobianos, así como la inducción a la migración celular y el colapso del cono de crecimiento.

Mientras que en la línea celular tumorales con ausencia de la infección por VPH (C33a), se observaron posibles implicaciones relacionadas con la interacción con la semaforina, mismas que no se habían identificado previamente. Esto sugiere que esta proteína podría tener un papel importante en la regulación del crecimiento celular y la señalización celular. Ya que las semaforinas, además de estar involucradas durante el desarrollo neuronal, están involucradas en la migración y la adhesión celular.

En contraste con las líneas tumorales infectadas por VPH (HeLa y SiHa) Se observan procesos similares a los obtenidos en el análisis general como la reticulación de péptidos. Aunque los procesos son consistentes con el análisis general, la infección por VPH podría modificar la intensidad o el contexto de estas interacciones, sugiriendo que C19orf47 podría desempeñar un papel diferente bajo condiciones de infección.

Esto sugiere que C19orf47 tendría un papel potencial en la regulación de procesos biológicos claves relacionados con la respuesta inmune y el comportamiento celular, dependiendo del contexto celular y la presencia de factores como la infección por VPH (HeLa y SiHa).

7.6.1.3. Clases funcionales del interactoma de C19orf47 en HaCaT, C33a, HeLa y SiHa

Con el fin complementar el enriquecimiento de las funciones se realizó un análisis a partir de genes de ontología (GOs) de los interactomas de C19orf47 de cada línea celular y encontrar diferencias significativas entre ellas, se repitió el análisis ontológico con cada uno de los modelos.

En el caso particular, en HaCaT, se encontró que C19orf47 podría estar involucrada en otros procesos además de la degranulación de neutrófilos, como la adhesión célula-célula y el desarrollo de la epidermis (Figura 16). A diferencia de las otras líneas celulares estudiadas, en HaCaT se observó que la interacción de C19orf47 podría estar regulada por factores de transcripción específicos como SP1, RELA y NFKB1. Estos factores de transcripción son conocidos por su participación en la respuesta inmune y la regulación de genes involucrados en la función de la barrera epitelial.

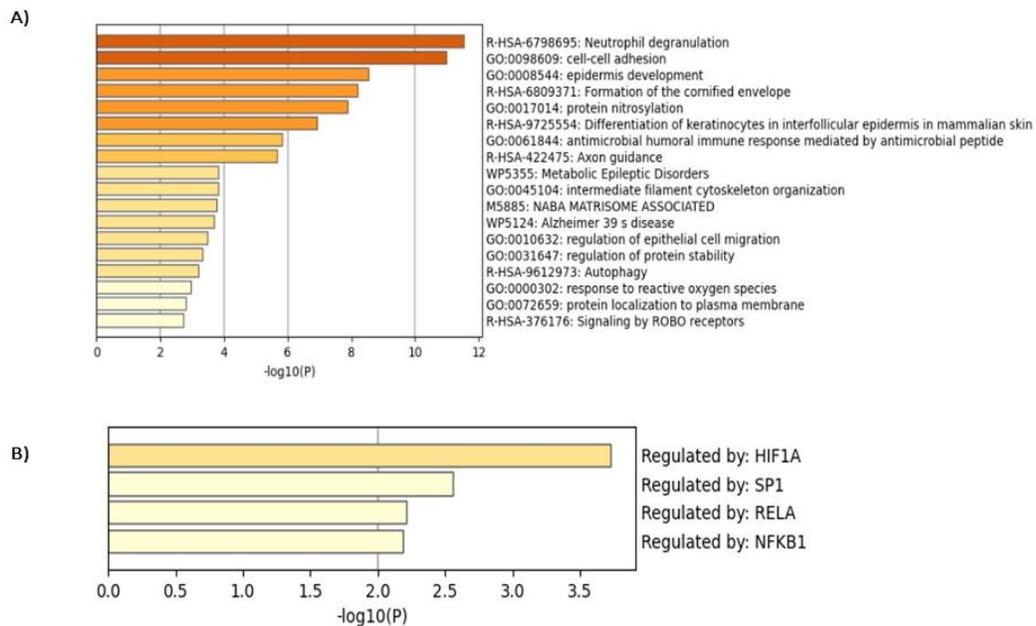


Figura 16. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular HaCaT. A) Enriquecimiento de funciones biológicas. Las barras representan las funciones más significativamente enriquecidas, con sus respectivos valores de enriquecimiento y p-valores corregidos. B) Enriquecimiento de reguladores. Se representan los tipos de reguladores sobrerrepresentados en el interactoma (HIF1A, SP1, RELA y NFKB1). Los valores de enriquecimiento y significancia estadística se indican para cada categoría de regulador.

La detección de HIF1A como posible regulador de C19orf47 en la línea celular no cancerosa, HaCaT, sugiere la participación de C19orf47 en distintos procesos de adaptación a distintas condiciones de estrés ambiental. La línea celular HaCaT como otros tipos celulares pueden enfrentar diversos factores de estrés que inducen una respuesta similar a la hipoxia activando HIF1A como parte del mecanismo de adaptación celular para manejar el estrés ambiental.

La presencia y la regulación específica de estos factores sugieren que C19orf47 puede estar integrada en redes de señalización que modulan la respuesta inmune local y la homeostasis del epitelio en condiciones no cancerosas. Lo que sugiere que en ausencia de CaCu, y de la infección por de VPH, C19orf47 podría tener influencia en procesos específicos relacionados con la función de barrera epitelial y la respuesta inmune local. Además. Podría participar en el mecanismo de adaptación celular para manejar el estrés ambiental.

Como se muestra en la Figura 17, en C33a, se observaron los patrones funcionales previamente identificados de C19orf47, Sin embargo, destaca que, en esta línea celular, C19orf47 podría estar involucrada en la señalización de VEGFA y VEGFR2, esto podría indicar que C19orf47 puede desempeñar un papel crucial en la regulación de la angiogénesis y la respuesta vascular. Así mismo estos resultados también sugieren que puede tener implicaciones importantes para la cohesión y la integridad del tejido celular, así como para la migración celular y la formación de estructuras multicelulares.

Lo resultados de este análisis por términos GOs mostraron la presencia de los reguladores HIF1A y SP1 (Figura 17). Por su parte y como ya se ha mencionado, HIF1A es conocido por su papel en la respuesta celular a la hipoxia: mientras que SP1, está involucrado en la regulación de múltiples genes relacionados con la proliferación células, la diferenciación y la respuesta a señales externas.

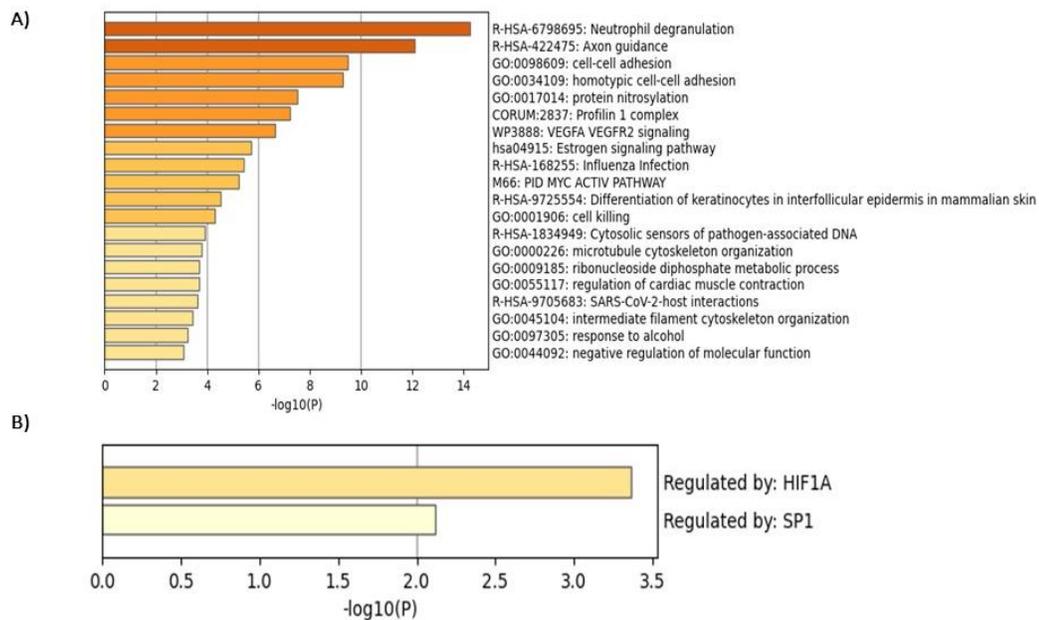


Figura 17. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular C33a. A) Enriquecimiento de funciones biológicas. Las barras representan las funciones más significativamente enriquecidas, con sus respectivos valores de enriquecimiento y p-valores corregidos. B) Enriquecimiento de reguladores. Se representan los tipos de reguladores identificado en el interactoma (HIF1A y SP1). Los valores de enriquecimiento y significancia estadística se indican para cada categoría de regulador.

En las líneas celulares HeLa y SiHa se observaron fuertes similitudes, se encontró que C19orf47 muy probablemente podría estar implicada, en procesos fundamentales de crecimiento, señalización, adhesión y cohesión celular, además posiblemente participa en la respuesta inflamatoria e inmune (Figura 18-19). A diferencia de las líneas HaCaT y C33a que presentaron más de un regulador, en líneas infectadas por VPH se observa que C19orf47 estaría siendo regulado únicamente por HIF1A lo que podría indicar que C19orf47 puede ser parte de los mecanismos adaptativos de las células ante condiciones de bajos niveles de oxígeno.

Además de su participación en las función previamente descritas, C19orf47 también mostró posibles implicaciones en respuestas inflamatorias e inmunes en ambas líneas celulares. Esto indica que podría estar involucrada en la modulación

de la respuesta del sistema inmunitario ante estímulos internos o externos, lo cual es crucial para la defensa del organismo y la regulación de la inflamación.

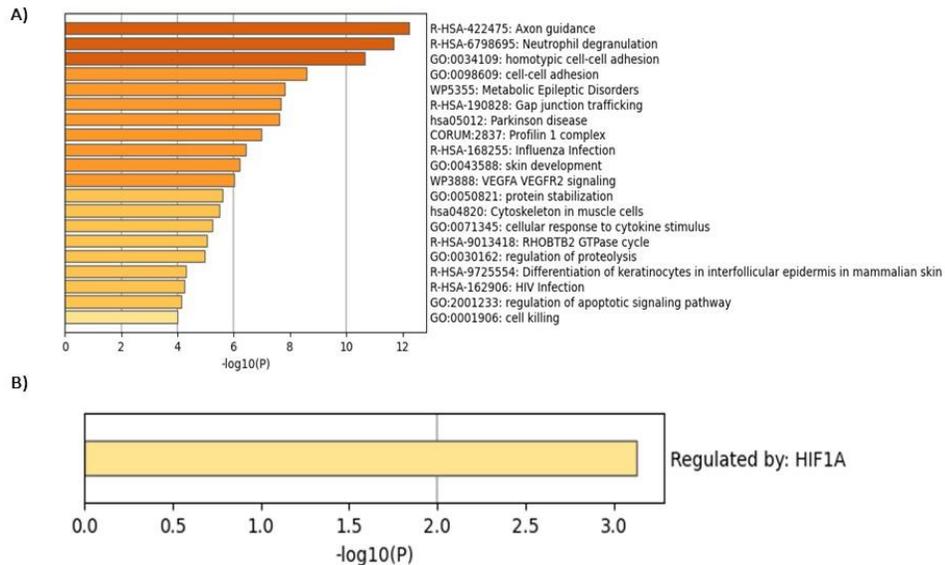


Figura 18. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular HeLa. A) Enriquecimiento de funciones biológicas. Las barras representan las funciones más significativamente enriquecidas, con sus respectivos valores de enriquecimiento y p-valores corregidos. B) Enriquecimiento de reguladores. Se representan el regulador sobrerrepresentado en el interactoma (HIF1A). Los valores de enriquecimiento y significancia estadística se indican para cada categoría de regulador.

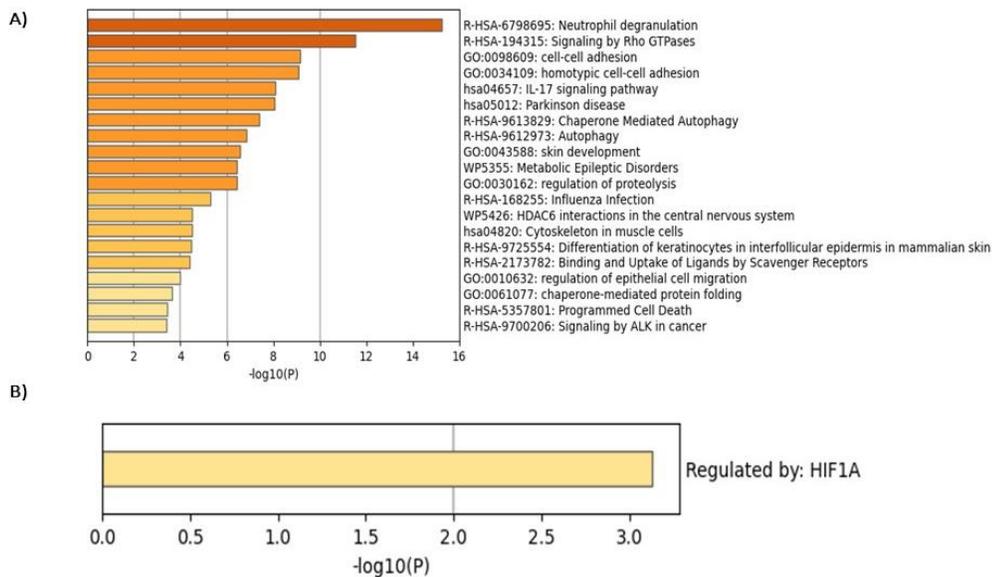


Figura 19. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular SiHa. A) Enriquecimiento de funciones biológicas. Las barras representan las funciones más significativamente enriquecidas, con sus respectivos valores de enriquecimiento y p-valores corregidos. B) Enriquecimiento de reguladores. Se representan la sobrerrepresentación del regulador HIF1A, en el interactoma de SiHa. Los valores de enriquecimiento y significancia estadística se indican para cada categoría de regulador.

7.6.2. Análisis comparativo, de la expresión proteica, entre las líneas celulares cancerosas (HeLa y SiHa) contra no cancerosas (HaCaT)

Para identificar a las proteínas sobre expresadas en las líneas celulares utilizadas y sus implicaciones biológicas, se realizó un análisis comparativo entre las intensidades proteicas obtenidas del análisis por LC-MS/MS de las precipitaciones de C19orf47 en cada línea celular.

Primero, se realizó la comparación de las líneas celulares cancerosas infectadas con VPH (HeLa y SiHa) contra la línea celular no cancerosa HaCaT (Figura 20). Para esto, se comparó la intensidad proteica de las DEPs (previamente identificadas) de las líneas celulares a comparar, mediante un volcano plot para identificar aquellas proteínas cuya intensidad proteica fuese significativa o única en cada una de las líneas comparadas y posteriormente, se construyó el interactoma de estas proteínas significativas. Esto permitió identificar patrones únicos de expresión proteica que podrían estar asociados con la transformación maligna y características específicas del cáncer. (Figura 20A).

Una vez identificado el interactoma de estas proteínas sobre expresadas, se logró identificar las funciones y procesos en los que están involucradas, en el caso de la comparación entre HaCaT y HeLa, se sugiere la intervención de C19orf47 en procesos relevantes como la regulación de la proliferación celular, adhesión invasión, migración y metástasis, además de la regulación del citoesqueleto y la respuesta al estrés celular. Igualmente, se identificó que el factor de transcripción HIF1A nuevamente podría ser el regulador de estas interacciones proteicas (Figura 20A-B).

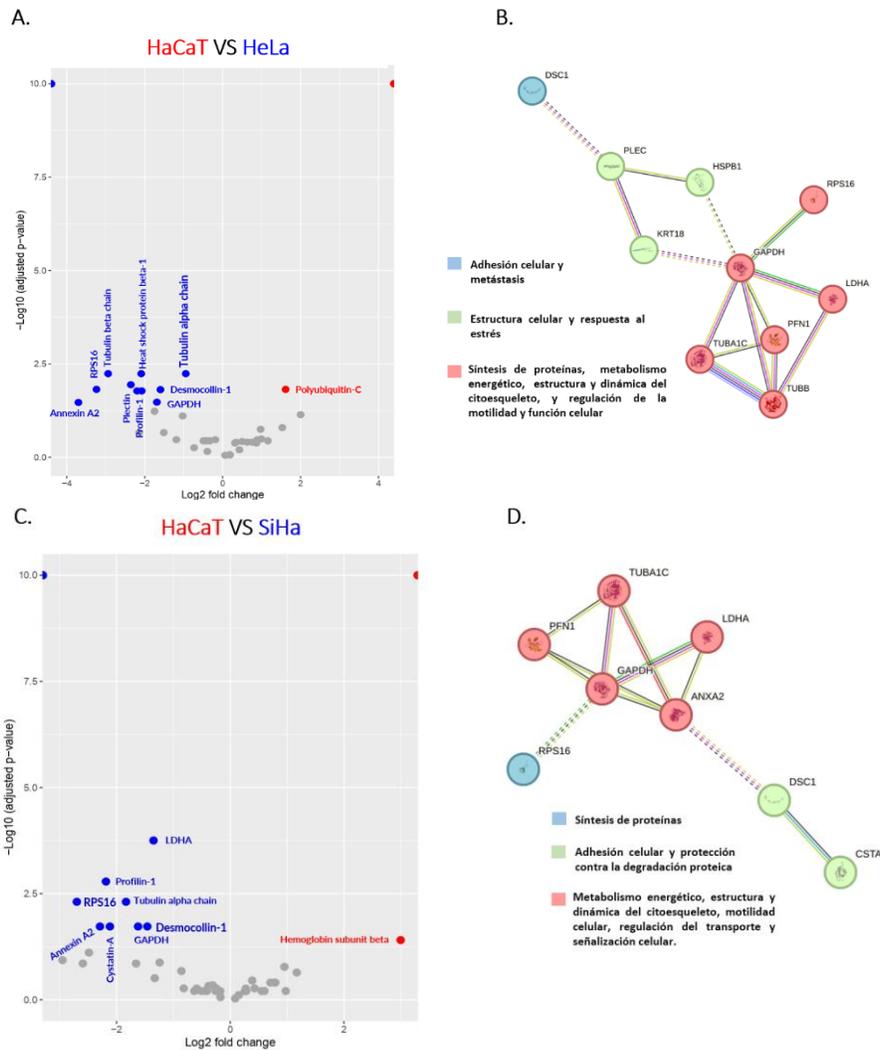


Figura 20. Análisis de la expresión proteica en líneas celulares cancerosas infectadas con VPH contra cancerosas no infectadas. Se realizó un análisis proteómico para comparar los niveles de expresión de diferentes proteínas entre las líneas celulares cancerosas infectadas con VPH contra la línea no infectada C33a **A)** Volcano plot mostrando las proteínas representativas entre C33a y HeLa, las proteínas mostradas en azul representan a aquellas que se sobre expresan en HeLa a diferencia de HaCaT. **B)** Interactoma resultante de las líneas sobre expresadas en HeLa. **C)** Volcano plot con las proteínas sobre expresadas en SiHa. **D)** Interactoma de las proteínas sobre expresadas en SiHa

Se sabe que, bajo condiciones celulares normales, HIF1A está inactivo y se degrada rápidamente en presencia de oxígeno. Sin embargo, bajo condiciones de hipoxia o en muchos tipos de cáncer, HIF1A se estabiliza y activa la transcripción de genes que favorecen la adaptación celular, incluyendo aquellos relacionados con la glucólisis por senescencia y la angiogénesis. La persistente presencia de este

factor en diferentes interactomas de C19orf47, especialmente en HeLa y SiHa, mismas que son células que presentan un crecimiento rápido y a menudo, una neovascularización ineficaz que conduce a zonas de baja oxigenación sugiere que HIF1A podría activar la expresión de genes que promueven la glucólisis, incluso en presencia de oxígeno, contribuyendo al efecto Warburg. Este efecto implica una preferencia por la glucólisis para la producción de ATP, a pesar de la disponibilidad de oxígeno, resultando en una acumulación de lactato.

Los resultados obtenidos de la comparación entre HaCaT y SiHa, sugiere que C19orf47 desempeña un papel en la proliferación celular, metástasis, la supervivencia de tumores bajo condiciones de estrés y la adhesión celular. Además, se sugiere que podría desempeñar un papel importante en la adaptación a entornos hipóxicos (Figura 20C-D). Igualmente se predice que C19orf47 podría tener una implicación glucolisis por senescencia y en el efecto Warburg.

En resumen, el análisis comparativo entre las líneas cancerosas, contra el control de líneas no cancerosas (HaCaT), reveló un conjunto de proteínas significativas en HeLa y en SiHa, conformado por ANAX2, CSTA, DSC1, GAPDH, HSPB1, PLEC, KRT18, LDHA, PFN1, RSP14, RSP16, TUBA1C y TUBB. Esto sugiere que C19orf47 podría desempeñar un papel crucial en el cáncer, participando en la regulación de la proliferación celular, la adhesión, invasión y migración celular, esenciales para la metástasis, y en la regulación del citoesqueleto. Además, podría estar implicada en la respuesta al estrés celular y la adaptación a entornos hipóxicos, Igualmente se sugiere una intervención entre C19orf47, el efecto Warburg y, por consiguiente, la glucólisis por senescencia.

Al tener una expresión significativa de LDHA y HSPB1 podemos sugerir que C19orf47 podría participar en el efecto de Warburg, ya que estas juegan un papel esencial en dicho efecto. En este caso, las células cancerosas metabolizan glucosa principalmente a través de la glucólisis aeróbica para apoyar la biosíntesis celular, este proceso es indispensable para el crecimiento tumoral rápido y para eludir la apoptosis. Así mismo estas proteínas están asociadas con respuestas adaptativas

a la hipoxia, lo que le confiere a una célula cancerosa la capacidad de sobrevivir a bajas concentraciones de oxígeno (Liu *et al.*, 2022; Sharma *et al.*, 2022).

Otro gen destacado es GAPDH, conocido principalmente por su función en la glucólisis, pero igualmente por funciones adicionales derivadas de PTMs, en la regulación de la muerte celular y la respuesta al estrés. También tiene implicaciones en la angiogénesis y la supervivencia celular bajo diferentes condiciones adversas, entre las que se incluye la hipoxia (Butera *et al.*, 2019; Paredes *et al.*, 2021; Sirover, 2018).

Por su parte la plectina (PLEC), actúa como un andamiaje celular relacionado con la adhesión, migración, invasión y metástasis. Debido a la expresión específica y la localización diferencial de la plectina en el cáncer, la mayoría de los investigadores la han convertido en un potente impulsor de las características malignas en muchos cánceres humanos (Gao *et al.*, 2023), Además. DSC1, KRT18, RSP14, RSP16, TUBA1C Y TUBB, son proteínas que ha sido identificadas en diversos modelos de estudio de cáncer como biomarcadores u oncogenes que participan en procesos metabólicos y de señalización celular, importantes para el cáncer, ya que actúan como reguladores del ciclo celular, potenciando directamente la capacidad de proliferación, invasión y migración de la células o procesos especializados como la alteración de la reparación del ADN y la senescencia (Chen *et al.*, 2021; Hu *et al.*, 2023; Hu *et al.*, 2022; Hurtado *et al.*, 2023; Kamal *et al.*, 2020; Zhang *et al.*, 2019; Zhao *et al.*, 2023). Mientras que estudios previos han mostrado que ANAX2, CSTA y PFN1 podrían desempeñar una función como supresores tumorales (Gau *et al.*, 2023; Lapcik *et al.*, 2023; Myklebust *et al.*, 2012; Wang *et al.*, 2019; Wang *et al.*, 2021; Yao *et al.*, 2014).

7.6.3. Análisis comparativo, de la expresión proteica, entre las líneas celulares cancerosas HeLa (VPH 16) y SiHa (VPH 18) contra C33a (VPH-)

Finalmente se realizó otra comparación entre las líneas cancerosas infectadas con VPH (HeLa y SiHa) contra la línea no infectada (C33a), profundizado en las diferencias de la expresión de proteínas que podrían estar relacionadas en la agresividad y capacidad viral del VPH en CaCu. (Figura 21).

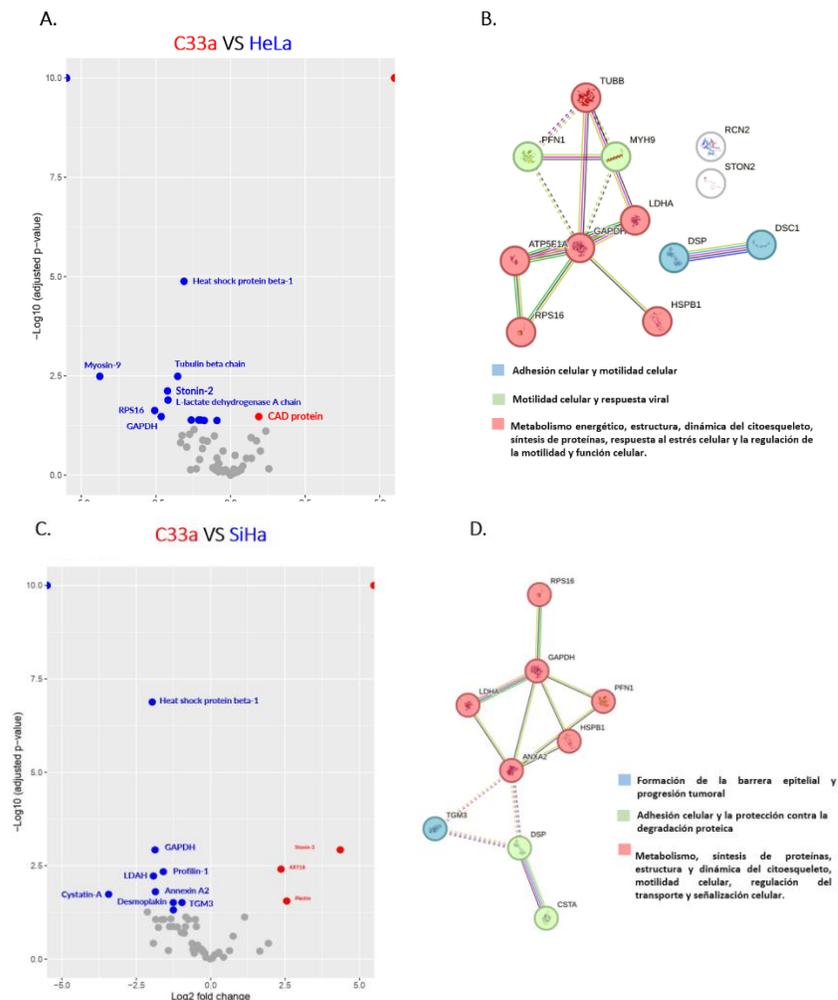


Figura 21. Análisis de la expresión proteica en líneas celulares cancerosas infectadas con VPH contra cancerosas no infectadas con VPH. Se realizó un análisis proteómico para comparar los niveles de expresión de diferentes proteínas entre las líneas celulares cancerosas infectadas con VPH contra la línea no infectada C33a **A)** Volcano plot mostrando las proteínas representativas entre C33a y HeLa, las proteínas mostradas en azul representan a aquellas que se sobre expresan en HeLa a diferencia de HaCaT. **B)** Interactoma resultante de las líneas sobre expresadas en HeLa. **C)** Volcano plot con las proteínas sobre expresadas en SiHa. **D)** Interactoma de las proteínas sobre expresadas en SiHa

Los resultados de la comparación entre C33a y HeLa, mostraron que C19orf47 podría influir en la capacidad del virus del VPH-18 para infectar y replicarse. Y sobrevivir en condiciones de estrés dentro de la célula huésped al regular el metabolismo energético y en la respuesta al estrés celular (Figura 21 A-B). Mientras que las diferencias entre las líneas celulares C33a y SiHa (infectada con VPH 16) sugieren que puede estar relacionada con la regulación del metabolismo energético y del estrés celular, así como en la supervivencia del virus dentro de las células huésped (Figuras 21C-D). Las proteínas identificadas abarcan una variedad de funciones clave que están estrechamente ligadas a la interacción entre el virus y las células hospederas, haciendo énfasis en el contexto del CaCu y el VPH.

En este sentido, ubicamos a ANXA2 como una proteína multifuncional que participa en procesos celulares vitales como la adhesión y metástasis del cáncer, además. Se ha demostrado que esta proteína desempeña un papel activo en infecciones, lo que sugiere un papel importante en la replicación viral y su patogénesis dentro de las células huésped (Park *et al.*, 2024). ATP5F1A es un componente del complejo V mitocondrial que no solo facilita el metabolismo energético celular, sino que también es capaz de interactuar con proteínas virales como el VPH en procesos relacionados con la respuesta celular a infecciones virales (You *et al.*, 2013). CSTA podría estar involucrada en la regulación de las respuestas inmunitarias ante la infección de este virus gracias a su función como inhibidora de proteasas de cisteína (Turk, 2008). El conjunto de DSC1 y DSP, involucradas en la adhesión celular y la integridad epitelial están asociadas con el cáncer y podrían ser relevantes en la interacción virus-huésped. Nuevamente GAPDH destaca por verse implicada en múltiples roles en los que se incluyen la replicación viral y la respuesta inmune gracias a sus múltiples funciones obtenidas de PTMs (Seidler, 2013).

A pesar de que genes como MYH9 y PFN1 desempeñan un papel crucial como supresores tumorales, se cree que podrían verse relacionados en la respuesta viral y su fitopatología (Liu *et al.*, 2024). Mientras que TGM3, conocida por su papel dentro del desarrollo de la piel y el cáncer, podría tener implicaciones en la

interacción viral debido a su capacidad para realizar modificaciones proteicas. Por último, TUBB, así como otras proteínas del citoesqueleto pueden ser cruciales para la entrada viral y la propagación intracelular del VPH (Chermnykh *et al.*, 2020).

A pesar de presentarse un panorama más despejado sobre la función de C19orf47 en la señalización celular y su desempeño en cáncer, aún no es posible asignarle con toda certeza una función exacta debido a la diversidad de interacciones y a su localización subcelular dual.

La relación evidente, entre C19orf47 y el VPH, en la progresión de CaCu, plantea preguntas sobre como exactamente interactúan e intervienen dentro de la biología celular y la patogénesis viral.

Cabe destacar que C19orf47 pertenece a una familia de proteínas no caracterizadas, esto representa un mayor desafío en términos de interpretación funcional y su respectiva comparación con las proteínas más estudiadas. Por lo que es necesario, realizar ensayos funcionales para confirmar los resultados, así como la importancia y relevancia clínica de las proteínas identificadas. Por esta razón es importante realizar estudios de sobreexpresión y silenciamiento génico, en modelos celulares y animales de experimentación, para confirmar su papel funcional en la patogénesis del cáncer cervicouterino. Además, investigaciones futuras podrían centrarse en comprender mejor las interacciones específicas de C19orf47 con otras proteínas y su impacto en vías biológicas específicas.

8. CONCLUSIONES

Con base con los resultados obtenidos en este trabajo, se determinó que C19orf47 es una proteína localizada en el citoplasma y núcleo celular, y se propone que, en esta localización dual, la proteína podría tener interacciones proteicas diferentes en cada complejo con otras implicaciones funcionales ya que demostró estar involucrada en procesos biológicos propios de ambos orgánulos.

Se sugiere que en condiciones no cancerosas (HaCaT), C19orf47 es una proteína involucrada en el mantenimiento de la señalización celular, la respuesta inmunitaria, la regulación del citoesqueleto y la respuesta al estrés celular. Bajo el contexto celular del CaCu, C19orf47 podría participar en procesos relevantes como la proliferación celular, adhesión, invasión, migración y metástasis. Además, se sugiere que C19orf47 podría influir en la capacidad del virus del VPH para infectar, replicarse y sobrevivir dentro de la célula huésped.

Se propone que el factor de transcripción HIF1A puede ser un regulador maestro del interactoma de C19orf47 en condiciones cancerosas, y que esta proteína sea clave en la resistencia a la hipoxia, glucólisis por senescencia y por consecuencia en el efecto Warburg. Mientras que, en condiciones celulares no tumorales, podría ser un factor clave para la resistencia al estrés ambiental.

El estudio de esta proteína proporciona una base para el entendimiento de los mecanismos de moleculares y biológicos del desarrollo y progresión de esta enfermedad. Igualmente contribuyen al estudio integral del proteoma humano, posicionando a C19orf47 como una pieza importante para el estudio del cáncer cervicouterino.

REFERENCIAS

- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2017). *Molecular Biology of the Cell* (J. Wilson & T. Hunt, Eds.). W.W. Norton & Company. <https://doi.org/10.1201/9781315735368>
- Altalebi, S. A. R., Haghi, M. & Hosseinpour Feizi, M. A. (2023). Expression study of microRNA cluster on chromosome 19 (C19MC) in tumor tissue and serum of breast cancer patient. *Molecular Biology Reports*, *50*(12), 9825–9831. <https://doi.org/10.1007/s11033-023-08801-x>
- Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Baezconde-Garbanati, L., Agurto, I., Gravitt, P. E., Luciani, S., Murphy, S., Ochoa, C., Gallegos, K., Barahona, R. & Rodríguez, Y. (2019). Barriers and innovative interventions for early detection of cervical cancer. *Salud Pública de México*, *61*(4), 456. <https://doi.org/10.21149/10425>
- Barabási, A.-L., Gulbahce, N. & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, *12*(1), 56–68. <https://doi.org/10.1038/nrg2918>
- Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, *15*(1), 293. <https://doi.org/10.1186/1471-2105-15-293>
- UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, *51*(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
- Bedell, S. L., Goldstein, L. S., Goldstein, A. R. & Goldstein, A. T. (2020). Cervical Cancer Screening: Past, Present, and Future. *Sexual Medicine Reviews*, *8*(1), 28–37. <https://doi.org/10.1016/j.sxmr.2019.09.005>

- Benbow, R. M. (1992). Chromosome structures. *Science Progress*, 76(301-302 Pt 3-4), 425–450.
- Boys, E. L., Liu, J., Robinson, P. J. & Reddel, R. R. (2023). Clinical applications of mass spectrometry-based proteomics in cancer: Where are we? *PROTEOMICS*, 23(7–8). <https://doi.org/10.1002/pmic.202200238>
- Butera, G., Mullappilly, N., Masetto, F., Palmieri, M., Scupoli, M. T., Pacchiana, R. & Donadelli, M. (2019). Regulation of Autophagy by Nuclear GAPDH and Its Aggregates in Cancer and Neurodegenerative Disorders. *International Journal of Molecular Sciences*, 20(9), 2062. <https://doi.org/10.3390/ijms20092062>
- Cao, C., Hong, P., Huang, X., Lin, D., Cao, G., Wang, L., Feng, B., Wu, P., Shen, H., Xu, Q., Ren, C., Meng, Y., Zhi, W., Yu, R., Wei, J., Ding, W., Tian, X., Zhang, Q., Li, W., Gao, Q., Chen, G., Li, K., Sung, W.-K., Hu, Z., Wang, H., Li, G. & Wu, P. (2020). HPV-CCDC106 integration alters local chromosome architecture and hijacks an enhancer by three-dimensional genome structure remodeling in cervical cancer. *Journal of Genetics and Genomics*, 47(8), 437–450. <https://doi.org/10.1016/j.jgg.2020.05.006>
- Carbonara, K., Andonovski, M. & Coorsen, J. R. (2021). Proteomes Are of Proteoforms: Embracing the Complexity. *Proteomes*, 9(3), 38. <https://doi.org/10.3390/proteomes9030038>
- Cerqueira, R. S., Santos, H. L. P. C. dos, Prado, N. M. de B. L., Bittencourt, R. G., Biscarde, D. G. dos S. & Santos, A. M. (2022). Controle do câncer do colo do útero na atenção primária à saúde em países sul-americanos: revisão sistemática. *Revista Panamericana de Salud Pública*, 46, 1. <https://doi.org/10.26633/RPSP.2022.107>
- Chen, B., Xu, X., Lin, D., Chen, X., Xu, Y., Liu, X. & Dong, W. (2021). KRT18 Modulates Alternative Splicing of Genes Involved in Proliferation and Apoptosis

- Processes in Both Gastric Cancer Cells and Clinical Samples. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.635429>
- Chermnykh, E. S., Alpeeva, E. V. & Vorotelyak, E. A. (2020). Transglutaminase 3: The Involvement in Epithelial Differentiation and Cancer. *Cells*, 9(9), 1996. <https://doi.org/10.3390/cells9091996>
- Delgadillo L. (2014). Dinámica Proteómica en Tumores de Cáncer Cérvico Uterino de líneas HeLa y SiHa [Tesis de maestría]. Universidad Nacional Autónoma de México.
- Digre, A. & Lindskog, C. (2021). The Human Protein Atlas—Spatial localization of the human proteome in health and disease. *Protein Science*, 30(1), 218–233. <https://doi.org/10.1002/pro.3987>
- Dobson, P., Cai, Y.D., Stapley, B. & Doig, A. (2004). Prediction of Protein Function in the Absence of Significant Sequence Similarity. *Current Medicinal Chemistry*, 11(16), 2135–2142. <https://doi.org/10.2174/0929867043364702>
- Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R. & Stanley, M. A. (2012). The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine*, 30, F55–F70. <https://doi.org/10.1016/j.vaccine.2012.06.083>
- Dutil, J., Chen, Z., Monteiro, A. N., Teer, J. K. & Eschrich, S. A. (2019). An Interactive Resource to Probe Genetic Diversity and Estimated Ancestry in Cancer Cell Lines. *Cancer Research*, 79(7), 1263–1273. <https://doi.org/10.1158/0008-5472.CAN-18-2747>
- Enríquez-Aceves, I., Galicia-Carmona, T., Coronel-Martínez, J. A., Espinosa-Romero, R., Calderillo-Ruíz, G., Cortés-Esteban, P. & Cetina-Pérez, L. (2020). Standard Treatment with Bevacizumab as Targeted Therapy in Cervical Cancer. *Revista de Investigación Clínica*, 72(4). <https://doi.org/10.24875/RIC.20000061>

- Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A. & Fares, Y. (2020). Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduction and Targeted Therapy*, 5(1), 28. <https://doi.org/10.1038/s41392-020-0134-x>
- Forgrave, L. M., Wang, M., Yang, D. & DeMarco, M. L. (2022). Proteoforms and their expanding role in laboratory medicine. *Practical Laboratory Medicine*, 28, e00260. <https://doi.org/10.1016/j.plabm.2021.e00260>
- Gaceta. (2024, January 29). ¿Qué es el Human Proteome Project en el que participa la UNAM? Gaceta UNAM.
- Gao, K., Gao, Z., Xia, M., Li, H. & Di, J. (2023). Role of plectin and its interacting molecules in cancer. *Medical Oncology*, 40(10), 280. <https://doi.org/10.1007/s12032-023-02132-4>
- Gau, D., Daoud, A., Allen, A., Joy, M., Sagan, A., Lee, S., Lucas, P. C., Duensing, S., Boone, D., Osmanbeyoglu, H. U. & Roy, P. (2023). Vascular endothelial profilin-1 drives a protumorigenic tumor microenvironment and tumor progression in renal cancer. *Journal of Biological Chemistry*, 299(8), 105044. <https://doi.org/10.1016/j.jbc.2023.105044>
- Gil, J., Ramírez-Torres, A., Chiappe, D., Luna-Peñaloza, J., Fernandez-Reyes, F. C., Arcos-Encarnación, B., Contreras, S. & Encarnación-Guevara, S. (2017). Lysine acetylation stoichiometry and proteomics analyses reveal pathways regulated by sirtuin 1 in human cells. *Journal of Biological Chemistry*, 292(44), 18129–18144. <https://doi.org/10.1074/jbc.M117.784546>
- Grimwood, J., Gordon, L. A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M., Aerts, A., Altherr, M., Ashworth, L., Bajorek, E., Black, S., Branscomb, E., Caenepeel, S., Carrano, A., Caoile, C., Chan, Y. M., Christensen, M., Cleland, C. A., Copeland, A., Dalin, E., Dehal, P., Denys, M., Detter, J. C., Escobar, J., Flowers, D., Fotopulos, D., Garcia, C., Georgescu, A. M., Glavina, T., Gomez, M., Gonzales, E., Groza, M.,

- Hammon, N., Hawkins, T., Haydu, L., Ho, I., Huang, W., Israni, S., Jett, J., Kadner, K., Kimball, H., Kobayashi, A., Larionov, V., Leem, S.-H., Lopez, F., Lou, Y., Lowry, S., Malfatti, S., Martinez, D., McCready, P., Medina, C., Morgan, J., Nelson, K., Nolan, M., Ovcharenko, I., Pitluck, S., Pollard, M., Popkie, A. P., Predki, P., Quan, G., Ramirez, L., Rash, S., Retterer, J., Rodriguez, A., Rogers, S., Salamov, A., Salazar, A., She, X., Smith, D., Slezak, T., Solovyev, V., Thayer, N., Tice, H., Tsai, M., Ustaszewska, A., Vo, N., Wagner, M., Wheeler, J., Wu, K., Xie, G., Yang, J., Dubchak, I., Furey, T. S., DeJong, P., Dickson, M., Gordon, D., Eichler, E. E., Pennacchio, L. A., Richardson, P., Stubbs, L., Rokhsar, D. S., Myers, R. M. & Rubin, E. M. Lucas, S. M. (2004). The DNA sequence and biology of human chromosome 19. *Nature*, 428(6982), 529–535. <https://doi.org/10.1038/nature02399>
- Hanahan, D. & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
- Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hanash, S. & Celis, J. E. (2002). The Human Proteome Organization: a mission to advance proteome knowledge. In *Molecular & cellular proteomics : MCP* (Vol. 1, Issue 6, pp. 413–414). <https://doi.org/10.1074/mcp.R200002-MCP200>
- Harris, R. A., Raveendran, M., Worley, K. C. & Rogers, J. (2020). Unusual sequence characteristics of human chromosome 19 are conserved across 11 nonhuman primates. *BMC Evolutionary Biology*, 20(1), 33. <https://doi.org/10.1186/s12862-020-1595-9>
- He, F. (2005). Human Liver Proteome Project. *Molecular & Cellular Proteomics*, 4(12), 1841–1848. <https://doi.org/10.1074/mcp.R500013-MCP200>

- Hirano, T. (2015). Chromosome Dynamics during Mitosis. *Cold Spring Harbor Perspectives in Biology*, 7(6), a015792. <https://doi.org/10.1101/cshperspect.a015792>
- Hozhabri, H., Ghasemi Dehkohne, R. S., Razavi, S. M., Razavi, S. M., Salarian, F., Rasouli, A., Azami, J., Ghasemi Shiran, M., Kardan, Z., Farrokhzad, N., Mikaeili Namini, A. & Salari, A. (2022). Comparative analysis of protein-protein interaction networks in metastatic breast cancer. *PLOS ONE*, 17(1), e0260584. <https://doi.org/10.1371/journal.pone.0260584>
- Hu, S., Cai, J., Fang, H., Chen, Z., Zhang, J. & Cai, R. (2023). RPS14 promotes the development and progression of glioma via p53 signaling pathway. *Experimental Cell Research*, 423(1), 113451. <https://doi.org/10.1016/j.yexcr.2022.113451>
- Hu, X., Zhu, H., Chen, B., He, X., Shen, Y., Zhang, X., Xu, Y. & Xu, X. (2022). The oncogenic role of tubulin alpha-1c chain in human tumours. *BMC Cancer*, 22(1), 498. <https://doi.org/10.1186/s12885-022-09595-0>
- Hurtado, R., Ramirez, A., Nabipur, L., Flores, J. & Tirado, C. A. (2023). The Key Role of the RPS14 Gene in Neoplasms and Solid Tumors. *Journal of the Association of Genetic Technologists*, 49(3), 121–126.
- Jordan, B. (2021). L'héritage d'Henrietta Lacks. *Médecine/Sciences*, 37(12), 1189–1193. <https://doi.org/10.1051/medsci/2021181>
- Kamal, M. A., Al-Zahrani, M. H., Khan, S. H., Khan, M. H., Al-Subhi, H. A., Kuerban, A., Aslam, M., Al-Abbasi, F. A. & Anwar, F. (2020). Tubulin Proteins in Cancer Resistance: A Review. *Current Drug Metabolism*, 21(3), 178–185. <https://doi.org/10.2174/1389200221666200226123638>
- Kim, C. (2003). SAM domains: uniform structure, diversity of function. *Trends in Biochemical Sciences*, 28(12), 625–628. <https://doi.org/10.1016/j.tibs.2003.11.001>

Kohler, D., Staniak, M., Tsai, T.-H., Huang, T., Shulman, N., Bernhardt, O. M., MacLean, B. X., Nesvizhskii, A. I., Reiter, L., Sabido, E., Choi, M. & Vitek, O. (2023). MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. *Journal of Proteome Research*, 22(5), 1466–1482. <https://doi.org/10.1021/acs.jproteome.2c00834>

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, H., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, F., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang,

J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J. E., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T. S., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, J., Thierry-Mieg, D., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J. & International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome International Human Genome Sequencing Consortium* The Sanger Centre: Beijing Genomics Institute/Human Genome Center. In *NATURE* (Vol. 409).

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), 680–685. <https://doi.org/10.1038/227680a0>

Lapcik, P., Sulc, P., Janacova, L., Jilkova, K., Potesil, D., Bouchalova, P., Müller, P. & Bouchal, P. (2023). Desmocollin-1 is associated with pro-metastatic phenotype of luminal A breast cancer cells and is modulated by parthenolide.

Cellular & Molecular Biology Letters, 28(1), 68. <https://doi.org/10.1186/s11658-023-00481-6>

Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C. H., Corthals, G. L., Costello, C. E., Deutsch, E. W., Domon, B., Hancock, W., He, F., Hochstrasser, D., Marko-Varga, G., Hosseini Salekdeh, G., Sechi, S., Snyder, M., Srivastava, S., Uhlén, M., Wu, C. H., Yamamoto, T., Paik, Y.-K. & Omenn, G. S. (2011). The Human Proteome Project: Current State and Future Direction. *Molecular & Cellular Proteomics*, 10(7), M111.009993. <https://doi.org/10.1074/mcp.M111.009993>

Lei, J., Ploner, A., Elfström, K. M., Wang, J., Roth, A., Fang, F., Sundström, K., Dillner, J. & Sparén, P. (2020). HPV Vaccination and the Risk of Invasive Cervical Cancer. *New England Journal of Medicine*, 383(14), 1340–1348. <https://doi.org/10.1056/NEJMoa1917338>

Li, X.-Y., Li, G., Gong, T.-T., Lv, J.-L., Gao, C., Liu, F.-H., Zhao, Y.-H. & Wu, Q.-J. (2023). Non-Genetic Factors and Risk of Cervical Cancer: An Umbrella Review of Systematic Reviews and Meta-Analyses of Observational Studies. *International Journal of Public Health*, 68. <https://doi.org/10.3389/ijph.2023.1605198>

Liu, J., Zhang, C., Zhang, T., Chang, C.-Y., Wang, J., Bazile, L., Zhang, L., Haffty, B. G., Hu, W. & Feng, Z. (2022). Metabolic enzyme LDHA activates Rac1 GTPase as a noncanonical mechanism to promote cancer. *Nature Metabolism*, 4(12), 1830–1846. <https://doi.org/10.1038/s42255-022-00708-4>

Liu, Q., Cheng, C., Huang, J., Yan, W., Wen, Y., Liu, Z., Zhou, B., Guo, S. & Fang, W. (2024). MYH9: A key protein involved in tumor progression and virus-related diseases. *Biomedicine & Pharmacotherapy*, 171, 116118. <https://doi.org/10.1016/j.biopha.2023.116118>

- Liu, Q., Fang, L. & Wu, C. (2022). Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes*, 13(3), 401. <https://doi.org/10.3390/genes13030401>
- Liu, Y., Beyer, A. & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, 165(3), 535–550. <https://doi.org/10.1016/j.cell.2016.03.014>
- Mannucci, A., Furniss, C. S., Ukaegbu, C., Horiguchi, M., Fehlmann, T., Uno, H., Yurgelun, M. B. & Syngal, S. (2020). Comparison of Colorectal and Endometrial Microsatellite Instability Tumor Analysis and Premm Risk Assessment for Predicting Pathogenic Germline Variants on Multigene Panel Testing. *Journal of Clinical Oncology*, 38(34), 4086–4094. <https://doi.org/10.1200/JCO.20.01470>
- Martínez-Rodríguez, F., Limones-González, J. E., Mendoza-Almanza, B., Esparza-Ibarra, E. L., Gallegos-Flores, P. I., Ayala-Luján, J. L., Godina-González, S., Salinas, E. & Mendoza-Almanza, G. (2021). Understanding Cervical Cancer through Proteomics. *Cells*, 10(8). <https://doi.org/10.3390/cells10081854>
- Masters, J. R. W. (2000). Human cancer cell lines: fact and fantasy. *Nature Reviews Molecular Cell Biology*, 1(3), 233–236. <https://doi.org/10.1038/35043102>
- Monti, C., Zilocchi, M., Colugnat, I. & Alberio, T. (2019). Proteomics turns functional. *Journal of Proteomics*, 198, 36–44. <https://doi.org/10.1016/j.jprot.2018.12.012>
- Morley, R. & Minceva, M. (2021). Liquid–Liquid Chromatography: Current Design Approaches and Future Pathways. *Annual Review of Chemical and Biomolecular Engineering*, 12(1), 495–518. <https://doi.org/10.1146/annurev-chembioeng-101420-033548>
- Myklebust, M. P., Fluge, O., Immervoll, H., Skarstein, A., Balteskard, L., Bruland, O. & Dahl, O. (2012). Expression of DSG1 and DSC1 are prognostic markers in

anal carcinoma patients. *British Journal of Cancer*, 106(4), 756–762.
<https://doi.org/10.1038/bjc.2011.548>

NIH. (2021, May). *What is cancer?* NIH.

Odum, M. T., Teufel, F., Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Winther, O. & Nielsen, H. (2024). DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Research*, 52(W1), W215–W220. <https://doi.org/10.1093/nar/gkae237>

Omenn, G. S. (2021). Reflections on the hupo human proteome project, the flagship project of the human proteome organization, at 10 years. In *Molecular and Cellular Proteomics* (20). American Society for Biochemistry and Molecular Biology Inc. <https://doi.org/10.1016/J.MCPRO.2021.100062>

Omenn, G. S., Lane, L., Overall, C. M., Corrales, F. J., Schwenk, J. M., Paik, Y. K., Van Eyk, J. E., Liu, S., Snyder, M., Baker, M. S. & Deutsch, E. W. (2018). Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. In *Journal of Proteome Research*, 17, (12), 4031–4041. American Chemical Society. <https://doi.org/10.1021/acs.jproteome.8b00441>

Omenn, G. S., Lane, L., Overall, C. M., Lindskog, C., Pineau, C., Packer, N. H., Cristea, I. M., Weintraub, S. T., Orchard, S., Roehrl, M. H. A., Nice, E., Guo, T., Van Eyk, J. E., Liu, S., Bandeira, N., Aebersold, R., Moritz, R. L. & Deutsch, E. W. (2024a). The 2023 Report on the Proteome from the HUPO Human Proteome Project. In *Journal of Proteome Research*, 23(2), 532–549. American Chemical Society. <https://doi.org/10.1021/acs.jproteome.3c00591>

Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y.-K., Yoo, J.-S., Ping, P., Pounds, J., Adkins,

- J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W. & Hanash, S. M. (2005). Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *PROTEOMICS*, 5(13), 3226–3245. <https://doi.org/10.1002/pmic.200500358>
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., ... Hermjakob, H. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42(D1), D358–D363. <https://doi.org/10.1093/nar/gkt1115>
- Paik, Y. K., Lane, L., Kawamura, T., Chen, Y. J., Cho, J. Y., Labaer, J., Yoo, J. S., Domont, G., Corrales, F., Omenn, G. S., Archakov, A., Encarnación-Guevara, S., Lui, S., Salekdeh, G. H., Cho, J. Y., Kim, C. Y. & Overall, C. M. (2018). Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. In *Journal of Proteome Research*, (17), (12), 4042–4050. American Chemical Society. <https://doi.org/10.1021/acs.jproteome.8b00383>
- Paredes, F., Williams, H. C. & San Martin, A. (2021). Metabolic adaptation in hypoxia and cancer. *Cancer Letters*, 502, 133–142. <https://doi.org/10.1016/j.canlet.2020.12.020>
- Park, I.-W., Fiadjoe, H. K. & Chaudhary, P. (2024). Impact of Annexin A2 on virus life cycles. *Virus Research*, 345, 199384. <https://doi.org/10.1016/j.virusres.2024.199384>

- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Lázaro Pinto, B., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H. & Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Pessoa, J., Martins, M., Casimiro, S., Pérez-Plasencia, C. & Shoshan-Barmatz, V. (2022). Editorial: Altered Expression of Proteins in Cancer: Function and Potential Therapeutic Targets. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.949139>
- Peters, W. A., Liu, P. Y., Barrett, R. J., Stock, R. J., Monk, B. J., Berek, J. S., Souhami, L., Grigsby, P., Gordon, W. & Alberts, D. S. (2000). Concurrent Chemotherapy and Pelvic Radiation Therapy Compared with Pelvic Radiation Therapy Alone as Adjuvant Therapy After Radical Surgery in High-Risk Early-Stage Cancer of the Cervix. *Journal of Clinical Oncology*, 18(8), 1606–1613. <https://doi.org/10.1200/JCO.2000.18.8.1606>
- Piccinini, G. & Milani, L. (2023). Germline-related molecular phenotype in Metazoa: conservation and innovation highlighted by comparative transcriptomics. *EvoDevo*, 14(1), 2. <https://doi.org/10.1186/s13227-022-00207-3>
- Ping, P., Vondriska, T. M., Creighton, C. J., Gandhi, T. K. B., Yang, Z., Menon, R., Kwon, M.-S., Cho, S. Y., Drwal, G., Kellmann, M., Peri, S., Suresh, S., Gronborg, M., Molina, H., Chaerkady, R., Rekha, B., Shet, A. S., Gerszten, R. E., Wu, H., Raftery, M., Wasinger, V., Schulz-Knappe, P., Hanash, S. M., Paik, Y.-K., Hancock, W. S., States, D. J., Omenn, G. S. & Pandey, A. (2005). A functional annotation of subproteomes in human plasma. *PROTEOMICS*, 5(13), 3506–3519. <https://doi.org/10.1002/pmic.200500140>

- Polanco, E. B., Águila Rodríguez, N., Guerra Villarpanda, D., Blanco Vázquez, Y., Rodríguez González, O., Santana, M. O., Aracelio, P. D. & Castellón, R. (2020). *Cáncer cérvico uterino: prevención y tratamiento Cervical uterine cancer: prevention and treatment*, 18(4), 685-693. <http://www.medisur.sld.cu/index.php/medisur/article/view/4648>
- Qiao, F. & Bowie, J. U. (2005). The Many Faces of SAM. *Science's STKE*, 2005(286). <https://doi.org/10.1126/stke.2862005re7>
- Rani, A., Devi Singh, V., Mazumder, R. & Dua, K. (2023). Cancer Proteomics for Cellular Dysfunction: Insights and Trends. *Current Pharmaceutical Design*, 29(9), 697–712. <https://doi.org/10.2174/1381612829666230316110932>
- Rashid, M., Zadeh, L. R., Baradaran, B., Molavi, O., Ghesmati, Z., Sabzichi, M. & Ramezani, F. (2021). Up-down regulation of HIF-1 α in cancer progression. *Gene*, 798, 145796. <https://doi.org/10.1016/j.gene.2021.145796>
- Regnier, F. E. & Kim, J. (2018). Proteins and Proteoforms: New Separation Challenges. *Analytical Chemistry*, 90(1), 361–373. <https://doi.org/10.1021/acs.analchem.7b05007>
- Rerucha, C. M., Caro, R. J. & Wheeler, V. L. (2018). Cervical Cancer Screening. *American Family Physician*, 97(7), 441–448.
- Secretaría de Salud. (2023, September 16). *México registra al año más de 195 mil casos de cáncer: Secretaría de Salud*. Gobierno de México.
- Seidler N. W. (2013). GAPDH, as a virulence factor. *Advances in experimental medicine and biology*, 985, 149–178. https://doi.org/10.1007/978-94-007-4716-6_5
- Sharma, D., Singh, M. & Rani, R. (2022). Role of LDH in tumor glycolysis: Regulation of LDHA by small molecules for cancer therapeutics. *Seminars in Cancer Biology*, 87, 184–195. <https://doi.org/10.1016/j.semcan.2022.11.007>

- Sheng, W., Zhang, C., Mohiuddin, T. M., Al-Rawe, M., Zeppernick, F., Falcone, F. H., Meinhold-Heerlein, I. & Hussain, A. F. (2023). Multiplex Immunofluorescence: A Powerful Tool in Cancer Immunotherapy. *International Journal of Molecular Sciences*, 24(4), 3086. <https://doi.org/10.3390/ijms24043086>
- Shuken, S. R. (2023). An Introduction to Mass Spectrometry-Based Proteomics. *Journal of Proteome Research*, 22(7), 2151–2171. <https://doi.org/10.1021/acs.jproteome.2c00838>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 1-6. <https://doi.org/10.1038/msb.2011.75>
- Sirover, M. A. (2018). Pleiotropic effects of moonlighting glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in cancer progression, invasiveness, and metastases. *Cancer and Metastasis Reviews*, 37(4), 665–676. <https://doi.org/10.1007/s10555-018-9764-7>
- Skok, K., Gradišnik, L., Maver, U., Kozar, N., Sobočan, M., Takač, I., Arko, D. & Kavalarič, R. (2021). Gynaecological cancers and their cell lines. *Journal of Cellular and Molecular Medicine*, 25(8), 3680–3698. <https://doi.org/10.1111/jcmm.16397>
- Su, M., Zhang, Z., Zhou, L., Han, C., Huang, C. & Nice, E. C. (2021). Proteomics, Personalized Medicine and Cancer. *Cancers*, 13(11), 1-31. <https://doi.org/10.3390/cancers13112512>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A*

- Cancer Journal for Clinicians*, 71(3), 209–249.
<https://doi.org/10.3322/caac.21660>
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J. & von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646.
<https://doi.org/10.1093/nar/gkac1000>
- Tamura, K., Stecher, G. & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>
- Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., Lim, J. C. T., Yeong, J. & Lim, T. K. H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Communications*, 40(4), 135–153.
<https://doi.org/10.1002/cac2.12023>
- Tang, M., Huang, P., Wu, L., Zhou, P., Gong, P., Liu, X., Wei, Q., Hou, X., Hu, H., Zhang, A., Shen, C., Gao, W., Tian, R. & Liu, C. (2023). Comprehensive Evaluation and Optimization of the Data-Dependent LC–MS/MS Workflow for Deep Proteome Profiling. *Analytical Chemistry*, 95(20), 7897–7905.
<https://doi.org/10.1021/acs.analchem.3c00338>
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C. & Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, 45(W1), W98–W102.
<https://doi.org/10.1093/nar/gkx247>
- Tao, Y., Zhang, Q., Wang, H., Yang, X. & Mu, H. (2024). Alternative splicing and related RNA binding proteins in human health and disease. *Signal Transduction*

and Targeted Therapy, 9(1), 1-33. <https://doi.org/10.1038/s41392-024-01734-2>

Turk, V., Stoka, V. & Turk, D. (2008). Cystatins: biochemical and structural properties, and medical relevance. *Frontiers in bioscience : a journal and virtual library*, 13, 5406–5420. <https://doi.org/10.2741/3089>

Uhlén, M., Björling, E., Agaton, C., Al-Khalili Szigyarto, C., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllerö, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S. & Pontén, F. (2005). A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Molecular & Cellular Proteomics*, 4(12), 1920–1932. <https://doi.org/10.1074/mcp.M500279-MCP200>

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Al-Khalili Szigyarto, C., Odeberg, J., Djureinovic, D., Ottosson Takanen, J., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. & Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220): 1260419. <https://doi.org/10.1126/science.1260419>

- Uhlén, M., Oksvold, P., Älgenäs, C., Hamsten, C., Fagerberg, L., Klevebring, D., Lundberg, E., Odeberg, J., Pontén, F., Kondo, T. & Sivertsson, Å. (2012). Antibody-based protein profiling of the human chromosome 21. *Molecular & cellular proteomics: MCP*, 11(3), M111.013458. <https://doi.org/10.1074/mcp.M111.013458>
- Paik, Y. K., Jeong, S. K., Omenn, G. S., Uhlen, M., Hanash, S., Cho, S. Y., Lee, H. J., Na, K., Choi, E. Y., Yan, F., Zhang, F., Zhang, Y., Snyder, M., Cheng, Y., Chen, R., Marko-Varga, G., Deutsch, E. W., Kim, H., Kwon, J. Y., Aebersold, R., ... Hancock, W. S. (2012). The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nature biotechnology*, 30(3), 221–223. <https://doi.org/10.1038/nbt.2152>
- Tabari, M., Ferns, G. A., Qujeq, D., Andevvari, A. N., Sabahi, Z. & Moein, S. (2021). Signaling, metabolism, and cancer: An important relationship for therapeutic intervention. *Journal of Cellular Physiology*, 236(8), 5512–5532. <https://doi.org/10.1002/jcp.30276>
- Vihinen, M. (2021). Functional effects of protein variants. *Biochimie*, 180, 104–120. <https://doi.org/10.1016/j.biochi.2020.10.009>
- Wang, T., Wang, Z., Niu, R. & Wang, L. (2019). Crucial role of Anxa2 in cancer progression: highlights on its novel regulatory mechanism. *Cancer Biology & Medicine*, 16(4), 671–687. <https://doi.org/10.20892/j.issn.2095-3941.2019.0228>
- Wang, Z., Jiang, C., Pang, L., Jia, W., Wang, C., Gao, X., Zhang, X., Dang, H. & Ren, Y. (2021). ANXA2 is a Potential Marker for the Diagnosis of Human Cervical Cancer. *Biomarkers in Medicine*, 15(1), 57–67. <https://doi.org/10.2217/bmm-2020-0629>
- WHO. (2014, January). *Cervical cancer*. World Health Organization: WHO.

- Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, Andrew. A., Hughes, G., Humphery-Smith, I., Williams, K. L. & Hochstrasser, D. F. (1996). From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Nature Biotechnology*, *14*(1), 61–65. <https://doi.org/10.1038/nbt0196-61>
- Yamamoto, T., Langham, R. G., Ronco, P., Knepper, M. A. & Thongboonkerd, V. (2008). Towards Standard Protocols and Guidelines for Urine Proteomics: A Report on the Human Kidney and Urine Proteome Project (HKUPP) Symposium and Workshop 6 October 2007, Seoul, Korea and 1 November 2007, San Francisco, CA, USA. *PROTEOMICS*, *8*(11), 2156–2159. <https://doi.org/10.1002/pmic.200800138>
- Yao, W., Ji, S., Qin, Y., Yang, J., Xu, J., Zhang, B., Xu, W., Liu, J., Shi, S., Liu, L., Liu, C., Long, J., Ni, Q., Li, M. & Yu, X. (2014). Profilin-1 suppresses tumorigenicity in pancreatic cancer through regulation of the SIRT3-HIF1 α axis. *Molecular Cancer*, *13*(1), 187. <https://doi.org/10.1186/1476-4598-13-187>
- You, W.-C., Chiou, S.-H., Huang, C.-Y., Chiang, S.-F., Yang, C.-L., Sudhakar, J. N., Lin, T.-Y., Chiang, I.-P., Shen, C.-C., Cheng, W.-Y., Lin, J.-C., Shieh, S.-H. & Chow, K.C. (2013). Mitochondrial protein ATPase family, AAA domain containing 3A correlates with radioresistance in glioblastoma. *Neuro-Oncology*, *15*(10), 1342–1352. <https://doi.org/10.1093/neuonc/not077>
- Zahn-Zabal, M., Michel, P.-A., Gateau, A., Nikitin, F., Schaeffer, M., Audot, E., Gaudet, P., Duek, P. D., Teixeira, D., Rech de Laval, V., Samarasinghe, K., Bairoch, A. & Lane, L. (2019). The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research*. *48*(D1):D328-D334. <https://doi.org/10.1093/nar/gkz995>
- Zhang, J., Hu, S. & Li, Y. (2019). KRT18 is correlated with the malignant status and acts as an oncogene in colorectal cancer. *Bioscience Reports*, *39*(8): BSR20190884. <https://doi.org/10.1042/BSR20190884>

- Zhao, X. & Liu, Z.-P. (2019). Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network. *Genes*, 10(2): 143, 143. <https://doi.org/10.3390/genes10020143>
- Zhao, Y., Wang, W., Li, J., Du, J., Xie, Q., Wang, M., Liu, Z., Huo, X., Zhao, F., Ren, D., Zhao, J. & Shen, G. (2023). Elevated expression of TUBA1C in breast cancer predicts poor prognosis. *PLOS ONE*, 18(11), e0263710. <https://doi.org/10.1371/journal.pone.0263710>
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1), 1523. <https://doi.org/10.1038/s41467-019-09234-6>

LISTA DE FIGURAS

Figura 1. Colaboración internacional en el Proyecto del Proteoma Humano Centrado en Cromosomas (C-HPP).....	15
Figura 2. Filogenia de C19orf47	28
Figura 3. Dominios en la secuencia de aminoácidos de C19orf47.	30
Figura 4. Expresión general de C19orf46 en tejidos no tumorales y tumorales....	31
Figura 5. Expresión a nivel de proteína C19orf46 analizada mediante Western Blot (WB)	32
Figura 6. Predicción de la localización subcelular mediante redes neuronales profundas, utilizando el método ProtT5.....	34
Figura 7. Validación de la localización subcelular de C19orf47 por WB en muestras de fracciones nuclear y citoplasmático	35
Figura 8. Microscopia confocal en células en líneas celulares CaCu	36
Figura 9. Espectrometría de masas de muestras co-inmunoprecipitadas de C19orf47.....	39
Figura 10. Análisis de interacciones Proteína-Proteína (P-P) en Líneas Celulares	40
Figura 11. Red de interacciones P-P, obtenidas en el Interactoma de C19orf47..	42
Figura 12. Interactomas proteína-proteína de líneas celulares identificados por MCODE (Proteínas compartidas).....	43
Figura 13. Enriquecimiento de funciones biológicas, en tejidos y de reguladores del interactoma proteína-proteína de C19orf47 en las líneas celulares HaCaT, C33a, HeLa y SiHa.	44
Figura 14. Interactomas de C19orf47, de cada una de las líneas celulares: HaCaT, C33a (VPH -), HeLa (VPH 16) y SiHa (VPH 18).	47
Figura 15. Red interatómica P-P de C19orf47, de 4 líneas celulares, identificados por MCODE	48
Figura 16. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular HaCaT.	50
Figura 17. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular C33a.....	52
Figura 18. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular HeLa.....	53
Figura 19. Enriquecimiento funcional en tejidos y reguladores del interactoma proteína-proteína de C19orf47 en la línea celular SiHa.	53
Figura 20. Análisis de la expresión proteica en líneas celulares cancerosas infectadas con VPH contra no cancerosas no infectadas.....	55

Figura 21. Análisis de la expresión proteica en líneas celulares cancerosas infectadas con VPH contra cancerosas no infectadas con VPH 58

LISTA DE TABLAS

Tabla 1. Clasificación de las proteínas según su evidencia proteica (PE, por sus siglas en inglés). 12

Tabla 2. Datos actualizados de las proteínas clasificadas, según su evidencia proteica por cromosoma... 14

LISTA DE ABREVIATURAS

Aa	Aminoácidos
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
ARNm	ARN mensajero
B/D-HPP	El Proyecto del Proteoma Humano Impulsado por Biología y Enfermedades
BLAST	Basic Local Alignment Search Tool.
BSA	Albúmina Sérica Bovina.
CaCu	Cáncer cervicouterino
CCDC61	Proteína centrosómica CCDC61
CCDC97	Proteína centrosómica CCDC97
CECSC	Carcinoma de células escamosas de cuello uterino
C-HPP	Proyecto del Proteoma Humano centrado en Cromosomas
Chr	Cromosoma
CO ₂	Dióxido de carbono.
DEPs	Proteínas diferencialmente expresadas
DTT	Ditiotreitol.
EDTA	Ácido etilendiaminotetraacético.
EyeOME	Proyecto del proteoma del ojo humano
GEPIA	Gene Expression Profiling
GLTSCR2	Gen de la Región Candidata de Supresor de Tumores de Glioma 2
GSC2	Homeobox protein goosecold-2

H ₂ O	Agua.
HCl	Ácido Clorhídrico.
HEPES	Ácido 4-(2-hidroxietil)-1-piperazina etanosulfónico.
HGI	Glicoproteómica humana.
HGP	Proyecto del Genoma Humano
HIPP	Inmuno-peptidoma.
HLPP	Proyecto del proteoma del plasma humano.
HPA	Human Protein Atlas
HPP	Proyecto del Proteoma Humano
HPPP	Proyecto del proteoma del plasma humano.
HUGO	Organización del Genoma Humano
HUPO	Organización del Proteoma Humano
IMF	Inmunofluorescencia
iMOP	Iniciativa para la proteómica de organismos modelo.
KCl	Cloruro de Potasio.
LC	Cromatografía Líquida
LC-MS/MS	Cromatografía Líquida acoplada a Espectrometría de Masas
LENG8	Miembro 8 del grupo de receptores de leucocitos
m/s	Relación masa-carga.
MgCl	Cloruro de Magnesio.
ML	Método de Máxima Verosimilitud.
MP	Proteínas perdidas
MS	Espectrometría de Masas

Na ₃ VO ₄	Orto-vanadato de Sodio.
NaCl	Cloruro de Sodio (sal común).
NaF	Fluoruro de Sodio.
NCBI	Centro Nacional de Información Biotecnológica
NIC	Lesiones neoplásicas intraepiteliales
NLS	Motivo de Localización Nuclear.
NTM	Neutrimina
ORF	Marco de Lectura Abierta (Open Reading Frame).
PBS	Solución Salina Tamponada con Fosfato.
PE	Evidencia proteica
pH	Potencial de Hidrógeno.
PMSF	Inhibidores de proteasas
PSMA6	Subunidad alfa tipo-6 del proteasoma
PTM	Modificaciones Postraduccionales
SAFB	Factor de Apego a Andamiaje B
SAM	Sterile Alpha Motif
SDS	Dodecil Sulfato de Sodio.
SDS-PAGE	Electroforesis en Gel de Poliacrilamida con Dodecil Sulfato de Sodio.
SFB	Suero Fetal Bovino
STK11	Serina/Proteína Quinasa de Treonina 11
TA	Temperatura Ambiente
TMEM160	Proteína transmembrana 160
UCS	Carcinosarcoma uterino

uPE1	Proteínas con función desconocida
VIH	Virus de Inmunodeficiencia Humana
VPH	Virus de Papiloma Humano