

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California**



**Maestría en Ciencias
Ciencias de la Vida
con orientación en Biomedicina y Biotecnología**

**Análisis proteómico comparativo entre tumores de mama
benignos y malignos en mujeres mexicanas**

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Vareska Lucero Zárate Córdova

Ensenada, Baja California, México
2019

Tesis defendida por
Vareska Lucero Zárate Córdoba

y aprobada por el siguiente Comité

Dr. Aldo Moreno Ulloa
Co-director de tesis

Dr. Israel Ramírez Sánchez
Co-director de tesis

Dra. Carolina Álvarez Delgado

Dr. Francisco Javier Estrada Mena

Dr. Carlos Alberto Brizuela Rodríguez



Dra. Patricia Juárez Camacho
Coordinadora del Posgrado en Ciencias de la Vida

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Vareska Lucero Zárate Córdoba © 2019

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis.

Resumen de la tesis que presenta **Vareska Lucero Zárate Córdova** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Vida con orientación en Biomedicina y Biotecnología.

Análisis proteómico comparativo entre tumores de mama benignos y malignos en mujeres mexicanas

Resumen aprobado por:

Dr. Aldo Moreno Ulloa

Co-director de tesis

Dr. Israel Ramírez Sánchez

Co-director de tesis

Cuando las células del seno proliferan sin control se da origen a la formación de una masa celular o tumor el cual se puede clasificar como benigno (TMB) o maligno (TMM), siendo que la presencia de este último en el seno diagnostica a un individuo con cáncer de mama. En México, el cáncer de mama es la neoplasia más prevalente y mortífera en mujeres. Complicando el escenario es el hecho de que un TMB puede transformarse en un TMM con el paso del tiempo. Sin embargo, los mecanismos involucrados en dicho proceso se desconocen, principalmente porque existe una limitada información acerca de sus diferencias a nivel molecular. Debido a que la función celular es directamente dictada por la acción de las proteínas, un estudio sobre el proteoma de los TMB y TMM puede proveer de información relevante que ayude a comprender sus diferencias. Este estudio evaluó y comparó el proteoma de los TMB (n = 10) y TMM (n = 6) de mujeres mexicanas. Para este fin, se utilizaron como herramientas de análisis la espectrometría de masas, softwares especializados y análisis estadísticos multivariantes para cuantificar la abundancia de proteínas entre ambos grupos. Mediante nuestro método, cuantificamos 2,168 proteínas con una confiabilidad mayor al 99% y tasa global de descubrimientos falsos menor al 1%. Un total de 523 proteínas mostraron una abundancia diferencial y significativa entre los TMM y TMB; 432 aumentaron y 86 disminuyeron en la cohorte de sujetos con TMM (± 1.5 fold change y $p < 0.05$). Los análisis multivariantes sugieren a cuatro proteínas como biomarcadores de malignidad incluyendo principalmente proteoglucanos moduladores de la matriz extracelular. La abundancia diferencial de dichas proteínas fue validada por Western blot. Además, el análisis de enriquecimiento por ontología de genes de las proteínas con un aumento significativo en los TMM sugiere una mayor activación principalmente de los procesos biológicos relacionados a la regulación de la expresión de SLIT y ROBO, enfermedades infecciosas, metabolismo de ARN, degranulación de neutrófilos, ciclo del ácido cítrico y transporte de electrones, y biogénesis y mantenimiento de orgánulos. Por otro lado, las proteínas con una disminución significativa sugieren una inhibición de procesos relacionados con el CHST6 defectuoso que causa MCDC1, organización de la matriz extracelular, regulación del transporte del factor de crecimiento similar a la insulina (IGF) y absorción por las proteínas de unión al factor de crecimiento similar a la insulina (IGFBPS) y remodelación de lipoproteínas plasmáticas. En conclusión, este es el primer estudio que evidencia diferencias a nivel molecular entre los TMB y TMM en mujeres mexicanas sugiriendo a ciertas proteínas y vías de señalización. No obstante, se requiere de un estudio con un mayor número de participantes para tomar en cuenta la heterogeneidad de la población mexicana.

Palabras clave: Benigno, maligno, mama, tumor, espectrometría de masas, swath-ms

Abstract of the thesis presented by **Vareska Lucero Zárate Córdoba** as a partial requirement to obtain the Master of Science degree in Life Science with orientation in Biomedicine and Bionanotechnology.

Comparative proteomic analysis of benign and malignant breast tumors of Mexican women

Abstract approved by:

Dr. Aldo Moreno Ulloa
Thesis co-advisor

Dr. Israel Ramírez Sánchez
Thesis co-advisor

Breast tumors are produced by an uncontrollable cell proliferation mechanism and can be classified as benign (TMB) or malignant (TMM). TMM or breast cancer is the neoplasia with the highest incidence and mortality in Mexican women. Over time, some types of TMB can transform into a TMM. However, the mechanisms involved in such process remain elusive and limited studies have examined the molecular differences between TMB and TMM. Hence, the aim of this study was to evaluate and compare the proteomic profile of TMB (n = 10) and TMM (n = 6) of Mexican women. For this purpose, liquid chromatography coupled to mass spectrometry (LC-MS) and multivariate statistical analysis were used to compare the breast proteome among both cohorts. Also, bioinformatic tools were used to investigate pathways associated with the proteins differentially abundant in TMM vs. TMB. Using our method, we quantified 2,168 proteins with a confidence greater than 99% and global false discovery rate lower than 1%. A total of 523 proteins showed a significant differential abundance (± 1.5 -fold change and p-value < 0.05) between TMM and TMB; 432 were up-regulated and 86 down-regulated in the TMM cohort. Multivariate analysis (PCA and PCVG) strongly suggested three proteins (i.e., extracellular matrix modulating proteoglycans) as markers of malignancy. These proteins were validated by Western blot analysis. In addition, the functional enrichment analysis of the up-regulated proteins in TMM using the REACTOME Database suggests a greater activation of the regulation of expression of SLIT and ROBO, signaling by ROBO receptors, and metabolism of RNA pathways (top-three). On the other hand, down-regulated proteins in TMM were associated with the defective CHST6 causes MCDC1, extracellular matrix organization, and regulation of Insulin-Like Growth Factor (IGF) and transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) pathways (top-three). In conclusion, our study revealed molecular differences at the protein level among TMB and TMM and strongly associated proteins of the extracellular matrix with the malignancy of breast tumors. However, a study with a larger number of participants is warranted to corroborate and extrapolate our findings.

Keywords: Benign, malignant, breast, tumor, proteomic, mass spectrometry, swath-ms

Dedicatoria

A todos los seres que amo.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico para realizar mis estudios de maestría con el No. de becario: 633638.

Al Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) y al Departamento de Innovación Biomédica por haberme dado la oportunidad de realizar mis estudios de maestría.

Al Dr. Aldo Moreno Ulloa por haberme aceptado en su laboratorio y haber contribuido con su guía y conocimientos a mi formación profesional.

Al Dr. Israel Ramírez, Dra. Carolina Álvarez, Dr. Carlos Brizuela y Dr. Francisco Estrada, mi comité de tesis, por sus valiosas observaciones y correcciones que enriquecieron y fortalecieron este trabajo.

A todas las personas que he conocido en la comunidad del CICESE que me han regalado invaluable anécdotas e increíbles experiencias.

Tabla de contenido

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatorias.....	iv
Agradecimientos.....	v
Lista de figuras.....	viii
Lista de tablas.....	ix
Capítulo 1. Introducción	
1.1 Antecedentes.....	1
1.1.1 Tumores de mama (TM).....	1
1.1.2. Tipos de TM.....	2
1.1.2.1 Tumores de mama benignos (TMB).....	2
1.1.2.2 Tumores de mama malignos (TMM).....	2
1.1.2.3 Clasificación de los TMM.....	3
1.1.2.4 Epidemiología de los TMM.....	4
1.1.2.5 Etiología de los TMM.....	5
1.1.2.6 Diagnóstico de los TMM.....	8
1.1.2.6.1 Proteínas como marcadores de malignidad de tumores de mama..	9
1.1.3 Proteómica.....	10
1.1.3.1 Proteómica tipo “shotgun”.....	11
1.1.3.2 Tipos de cuantificación proteómica.....	13
1.1.3.3 Tipos de adquisición de datos.....	13
1.1.3.4 Cuantificación SWATH – MS.....	15
1.1.3.5 Aplicación de la proteómica a la biología del cáncer.....	15
1.2 Justificación	17
1.3 Hipótesis	17
1.4 Objetivos.....	17
1.4.1 Objetivo general	17
1.4.2 Objetivos específico.....	17

Capítulo 2. Metodología

2.1 Diseño del estudio y obtención de muestras	18
2.2 Preparación de las muestras	19
2.2.1 Extracción de proteínas.....	19
2.2.2 Obtención de péptidos trípticos.....	20
2.3 Estandarización del método de LC-MS/MS.....	20
2.4 Cuantificación SWATH – MS.....	21
2.4.1 Análisis LC-MS/MS.....	21
2.4.2 Creación de una biblioteca espectral de proteínas.....	21
2.4.3 Análisis de enriquecimientos de ontología de genes GO de la biblioteca de proteínas.....	22
2.4.4 Cuantificación diferencial de proteínas de los TBM y TMM.....	23
2.5 Análisis estadístico.....	23
2.6 Identificación de vías de señalización asociadas a las proteínas diferencialmente abundantes en el TMM.....	24
2.7 Optimización de la biblioteca de proteínas.....	25
2.7.1 Modificación de los parámetros utilizados en la identificación y cuantificación de proteínas de la biblioteca espectral.....	25
2.7.2 Extensión de proteínas de la biblioteca espectral.....	25
2.7.3 Análisis de enriquecimientos de ontología de genes y cuantificación diferencial utilizando la biblioteca extendida	26
 Capítulo 3. Resultados	
3.1 Estandarización del método LC –MS/MS	27
3.2 Análisis de enriquecimientos de ontología de genes GO de la biblioteca de proteínas interna y extendida	29
3.3 Análisis SWATH-MS utilizando la biblioteca interna y extendida	32
3.4 Identificación de vías de señalización asociadas a las proteínas diferencialmente abundantes en el TMM.....	37
3.5 Creación de una biblioteca extendida y modificación de los parámetros para la optimización de la cuantificación de proteínas.	43
 Capítulo 4. Discusión	47
 Capítulo 5. Conclusiones	51
 Literatura citada	52
Anexos	61

Lista de figuras

Figura		Página
1	Tipos de tumor de mama. Creado con Biorender www.biorender.com	1
2	Clasificación de los TMM según (A) la localidad o ubicación (B) el grado (C) el estadio o (D) el subtipo molecular. (NIH, 2019a, 2019b; Johns Hopkins Medicine Pathology, 2019; National Breast Cancer Foundation, 2016).....	3
3	Distribución estandarizada para todas las edades de la tasa de incidencia de los TMM en el mundo (International Agency for Research on Cancer, 2018).....	4
4	Métodos de diagnóstico para los TMM (A) Mamografía, (B) Imagen por Resonancia Magnética, (C) Tomografía Óptica Difusa de Placa Paralela, (D) Histopatología (Lehman y Schnall, 2005; Choe <i>et al.</i> , 2009; Weigelt <i>et al.</i> , 2010).....	9
5	Diagrama de flujo de las dos metodologías asociadas a la proteómica LC-MS/MS: bottom up y top down. Modificado de Hixson <i>et al.</i> (2017).....	11
6	Esquema de los 3 pilares tecnológicos de la proteómica shotgun que incluyen 1) la separación de péptidos 2) espectrometría de masas y 3) bioinformática.....	12
7	Tipos de adquisición de datos de la proteómica cuantitativa (A) Adquisición dependiente de datos (DDA) (B) Adquisición independiente de datos (DIA) (C) La información del MS2 para el MS1 VLENTEIGDSIFDK ++ (D) Extracción de los MS2 en el tiempo de los datos de DIA. Modificados de (Hu <i>et al.</i> , 2016).....	14
8	Diagrama de flujo generalizado de la metodología empleada en este estudio.....	19
9	Comparación de los gradientes de elución de péptidos trípticos BCAEC analizados por LC-MS/MS. MS2 generados con los gradientes de elución de (A) 150 y (B) 120 min de tiempo de corrida, respectivamente. Cada círculo azul representa un MS2 proveniente de un péptido precursor.....	27
10	Comparación de las intensidades detectadas de 6 proteínas representativas al inyectar 0.5 ug y 1 ug de los extractos proteínicos de BCAEC en el LC-MS/MS. Las muestras fueron inyectadas por triplicado.....	28
11	Comparación de la señal de la abundancia del péptido VNFTVDQIR de la proteína Factor de Elongación 2 utilizando tamaños de ventana variables (intervalo de masa específico) de (A) 14 Y (B) 10 Da en la cuantificación SWATH-MS. Se muestra que la disminución del tamaño de ventana mejora considerablemente la señal de los péptidos.....	29
12	Representación del análisis de enriquecimiento de ontología de genes en función al término Componente Celular generado con la aplicación ClueGO del software Cytoescape para (A) la biblioteca interna en forma de redes de proteínas y (B) en forma de diagrama circular con sus respectivos porcentajes por grupo	31

13	Diagrama circular que muestra el porcentaje de proteínas de la biblioteca extendida asociadas a un Componte Celular.....	32
14	Gráfica de volcán que muestra las proteínas que modificaron su abundancia en los TMM (n=6) con relación a los TBM (n=10) en la cuantificación SWATH-MS utilizando la biblioteca interna (A) y extendida (B). Clasificación basada en análisis a nivel de proteína con un FC >1.5 y <0.6; y un pval <0.05.....	33
15	Análisis de componentes principales utilizando la biblioteca interna (A) y extendida (B). El paciente 14 fue excluido del análisis.....	33
16	Análisis de componentes principales por agrupación de variables PCVG para las proteínas cuantificadas con la biblioteca interna (A) y extendida (B). Las proteínas relacionadas con la sangre y el paciente 14 fueron excluidos del análisis.....	34
17	Comparación de las proteínas con abundancia disminuida en los TMM responsables de la separación de grupos sugeridas por el análisis PCVG para (A) la biblioteca interna y (B) extendida. Análisis no paramétrico con corrección T-test Welch. TMB, tumor de mama benigno. TMM, tumor de mama maligno. Los valores son media y rango percentil (10-90). * p <0.05.....	35
18	Comparación de las proteínas con abundancia aumentada en los TMM responsables de la separación de grupos sugeridas por el análisis PCVG para la (A) biblioteca interna y (B) extendida. Análisis no paramétrico con corrección T-test Welch. TMB, tumor de mama benigno. TMM, tumor de mamamaligno. Los valores son media y rango percentil (10-90). * p <0.05.....	36
19	Abundancia de las proteínas Mimencan, Decoria y Lumican analizadas por Western blot para las muestras de tumor maligno (TMB) y tumor benigno (TMM). Los valores son media y rango percentil (10-90). * p <0.05.....	37
20	Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca interna. (A) Representación de las vías de señalización relacionadas a las proteínas con menor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con menor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. p<0.05.....	39
21	Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca extendida. (A) Representación de las vías de señalización relacionadas a las proteínas con menor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con menor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. p<0.05.....	40
22	Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca interna. (A) Representación de las vías de señalización relacionadas a las proteínas con mayor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con mayor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. p<0.05.....	41

23	Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca extendida. (A) Representación de las vías de señalización relacionadas a las proteínas con mayor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con mayor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. $p < 0.05$	42
24	Esquema representativo que resumen las principales vías de señalización que estarían siendo alteradas en los TMM de acuerdo a los resultados de este estudio. Se sugiere una mayor activación en las vías en color rojo e inhibición en las de azul en función a la abundancia de las proteínas que integran cada vía.....	43
25	Creación de una biblioteca de proteínas extendida. (A) Evaluación de la compatibilidad entre la biblioteca interna y extendida. (B) Correlación del tiempo de retención, $R^2 = 0.95$. (C) Predicción del error promedio del tiempo de retención (RT) Correlación de la intensidad relativa de iones, (D) Correlación de Spearman, $R_{II} = 0.65$ Comparación de la cantidad de proteínas y péptidos obtenidos entre la biblioteca interna y la extendida.....	44
26	Resultados de la cuantificación SWATH con la biblioteca interna y extendida (A) Cantidad de proteínas y péptidos con diferentes FDR y (B) $FDR < 0.01$	45
27	Porcentaje de número de proteínas cuantificadas con diferentes FDR. El color gris oscuro indica el porcentaje de péptidos identificados con alta confianza ($FDR 0-0.1$)..	45
28	Evaluación de la cuantificación de proteínas utilizando la biblioteca interna y extendida. (A) Medición de la consistencia de la cuantificación entre el área bajo la curva (abundancia) de las proteínas con 3 FDR en función a su (B) coeficiente de variación, (C) coeficiente de correlación y Bland-Altman.....	46
29	Parámetros utilizados en la creación de un método con el algoritmo Paragon para la identificación de proteínas de TM con el software Protein Pilot.....	61
30	Código R utilizado con el paquete SwathXtend para extender la biblioteca interna con el proteoma de humano del repositorio SwathAtlas. El código permite la integración de las bibliotecas en formato .txt (Wu, 2019).....	62
31	Histograma de barras que representa el porcentaje de proteínas que se relacionaron con un término (vías de señalización REACTOME) en el análisis de enriquecimiento de genes de las proteínas con menor abundancia en los TMM utilizando la biblioteca extendida.....	63
32	Histograma de barras que representa el porcentaje de proteínas que se relacionaron con un término (vías de señalización REACTOME) en el análisis de enriquecimiento de genes de las proteínas con mayor abundancia en los TMM utilizando la biblioteca extendida.....	63

Lista de tablas

Tabla		Página
1	Tipos de factores de riesgo: los factores intrínsecos y extrínsecos. Los factores extrínsecos se subdividen en endógenos y exógenos (Wu <i>et al.</i> , 2018).....	5
2	Factores de riesgo extrínsecos de los TMM. Modificado de Momenimovahed (2019).....	6
3	Características histopatológicas y clínicas de los TBM y TMM.....	18
4	Lista de proteínas diferencialmente abundantes en los TMM que contribuyen en mayor medida a la separación de los grupos.....	34
5	Configuración utilizada en la cuantificación SWATH-MS con el software Peakview.	61
6	Configuración utilizada para el análisis multivariante PCVG en el software Markerview.....	61
7	Criterios utilizados en el análisis de enriquecimiento de ontología de genes con la aplicación ClueGO del software Cytoscape.....	62
8	Lista de ID Uniprot de las proteínas que modificaron su abundancia en los TMB y TMM utilizando la biblioteca interna y externa en la cuantificación SWATH-MS. El paciente 14 fue excluido del análisis.....	64

Capítulo 1. Introducción

1.1 Antecedentes

1.1.1 Tumores de mama (TM)

Los tumores, también conocidos como neoplasias, son una masa anormal de tejido que se origina por la proliferación incontrolada de las células de cualquiera de los tejidos del organismo. Las células tumorales actúan como anarquistas celulares desobedeciendo las reglas biológicas conocidas, como la capacidad de respuesta a los controles normales del crecimiento, lo que induce una desproporcionada y excesiva multiplicación celular (Cooper, 2000).

Los tumores se denominan de acuerdo al tipo de célula de la que provengan (originan). Los TM generalmente se originan en las glándulas que producen la leche o los conductos que la llevan al pezón. Según las características clínicas que presentan los tumores se pueden clasificar en benignos o malignos (Instituto Nacional del Cáncer, 2019.) (Figura 1).

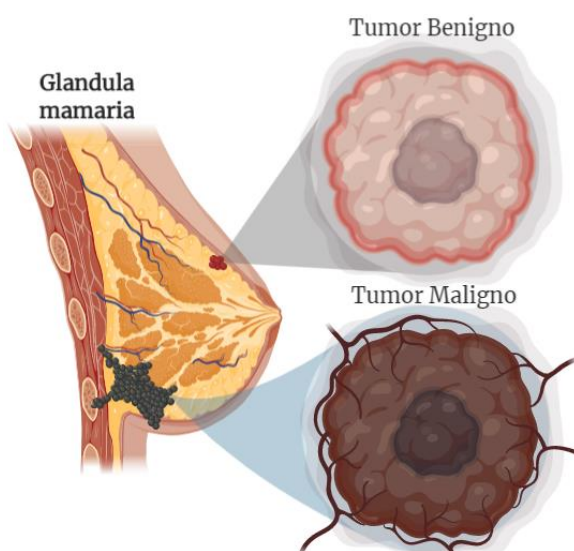


Figura 1. Tipos de tumor de mama. Creado con Biorender www.biorender.com

1.1.2 Tipos de TM

1.1.2.1 Tumores de mama benignos (TMB)

Anatomicamente, los TMB se encuentran en una zona bien localizada y su crecimiento permanece confinado en su ubicación original; en consecuencia, no invaden el tejido circundante ni migran a otros sitios del cuerpo. Los TMB conservan la integridad de la cápsula fibrosa o membrana basal (también llamada lámina basal) que rodea el seno y lo mantiene separado de otros tejidos (National Breast Cancer Foundation, 2019). La mayoría de los TMB no son potencialmente letales debido a que se pueden extirpar quirúrgicamente. Sin embargo, otros TMB pueden llegar a serlo si se presentan en órganos delicados como el cerebro o médula espinal. Los TMB no suelen ser recurrentes después de la cirugía, sin embargo, si esto llegara a ocurrir su aparición se manifestaría en su mismo lugar de origen (Sinha, 2018).

1.1.2.2 Tumores de mama malignos (TMM)

Los TMM son comúnmente conocidos como cáncer de mama y se caracterizan por su capacidad de propagarse por invasión y metástasis. Estas características se deben a la fragmentación y desaparición de la lámina basal producto de la secreción y actividad de enzimas proteolíticas (Snoj *et al.*, 2010). Los TMM crecen generalmente más rápido que los TBM y suelen reaparecer luego de una cirugía (Sinha, 2018). Se ha estimado entre un 15 y 16% de recurrencia de los TMM en mujeres después 5 a 10 años de haber recibido cirugía local como tratamiento (Holleczek *et al.*, 2019), (Colleoni *et al.*, 2016)

Estudios de oncología molecular comparada indican que la morfología de las células de los TMM consta de un núcleo grande e irregular, gran cantidad de ribosomas, poca cantidad de peroxisomas y disminución del volumen de las mitocondrias. Además, presenta cambios de composición de la membrana basal y en el número de receptores de superficie, microvellosidades atípicas en la superficie celular, anomalías en la glucólisis y en el ciclo del ácido tricarboxílico (TCA). Los TMM presentan la formación de vasos sanguíneos nuevos, proceso conocido como angiogénesis tumoral y pérdida de diferenciación que hace que las células tumorales se asemejen a las células embrionarias del tejido huésped (Baba y Căttoi, 2007).

1.1.2.3 Clasificación de los TMM

Los TMM se suelen clasificar de acuerdo a características anatómicas, celulares y moleculares. Esta clasificación se realiza en función a su localización, agresividad, capacidad de propagación y subtipo molecular. La clasificación basada en la localización se divide en dos tipos, carcinoma ductal y lobulillar; basada en la agresividad del tumor (grado de similitud del tumor al tejido normal en el microscopio) se clasifica en grados del 1 al 3; basada en la capacidad de propagación se clasifica en etapas del 0 al 4. Y finalmente de acuerdo al perfil molecular, es decir en función a los patrones de expresión génica se clasifican en cinco subtipos intrínsecos, los cuales son el tipo luminal A, luminal B, positivo para el receptor 2 de factor de crecimiento epidérmico humano (HER2), tumores basales (triple negativo) y de tipo normal (Dai *et al.*, 2015) (Figura 2).

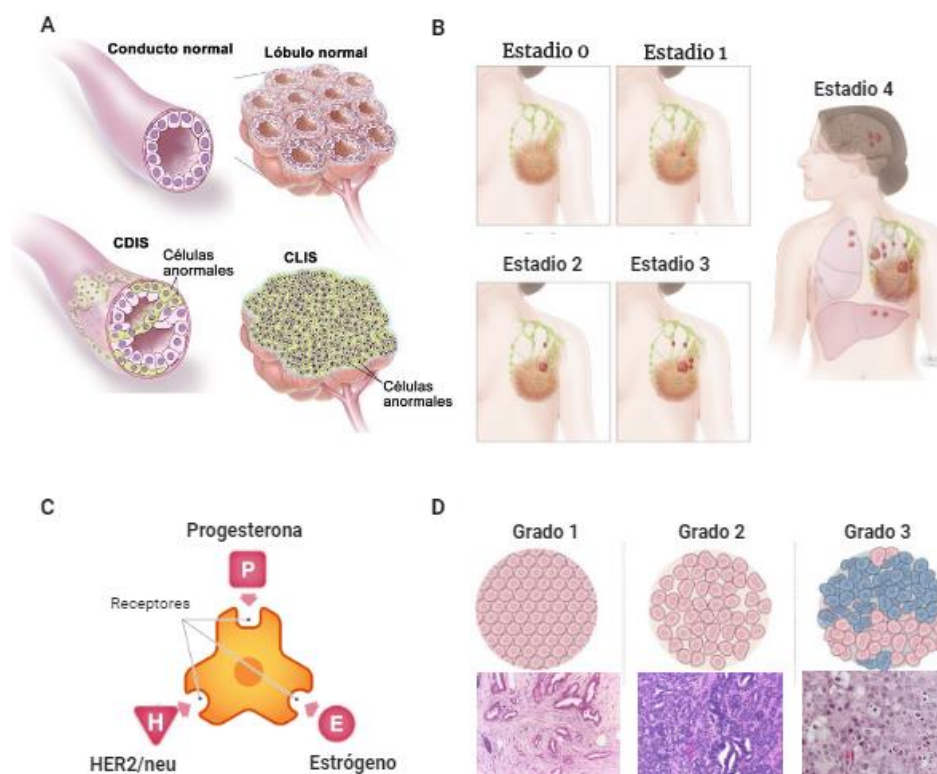


Figura 2. Clasificación de los TMM según (A) la localidad o ubicación (B) el grado (C) el estadio o (D) el subtipo molecular. (NIH, 2019a, 2019b; Johns Hopkins Medicine Pathology, 2019; National Breast Cancer Foundation, 2016)

1.1.2.4 Epidemiología de los TMM

Los TMM son una enfermedad que afecta principalmente a las mujeres, no obstante, los hombres también pueden ser afectados (18 - 90 y 0.16 - 1.24 por cada 100,000 mujeres y hombres a nivel mundial, respectivamente) (Ly *et al.*, 2013). Según la Organización Mundial de la Salud, los TMM son el cáncer con mayor incidencia en la población femenina a nivel global (World Health Organization, 2019; World Cancer Research Fund International, 2012). En el 2018, las tasas de incidencia más altas fueron calculadas para América del Norte y Europa Occidental, con valores de 91.6 y 91.1 por cada 100, 000 personas, respectivamente. Por otra parte, los valores de incidencia más bajos fueron para África media y Asia oriental (26.8 y 27 por cada 100, 000 personas, respectivamente) (Figura 3) (Ghonchek, 2016).

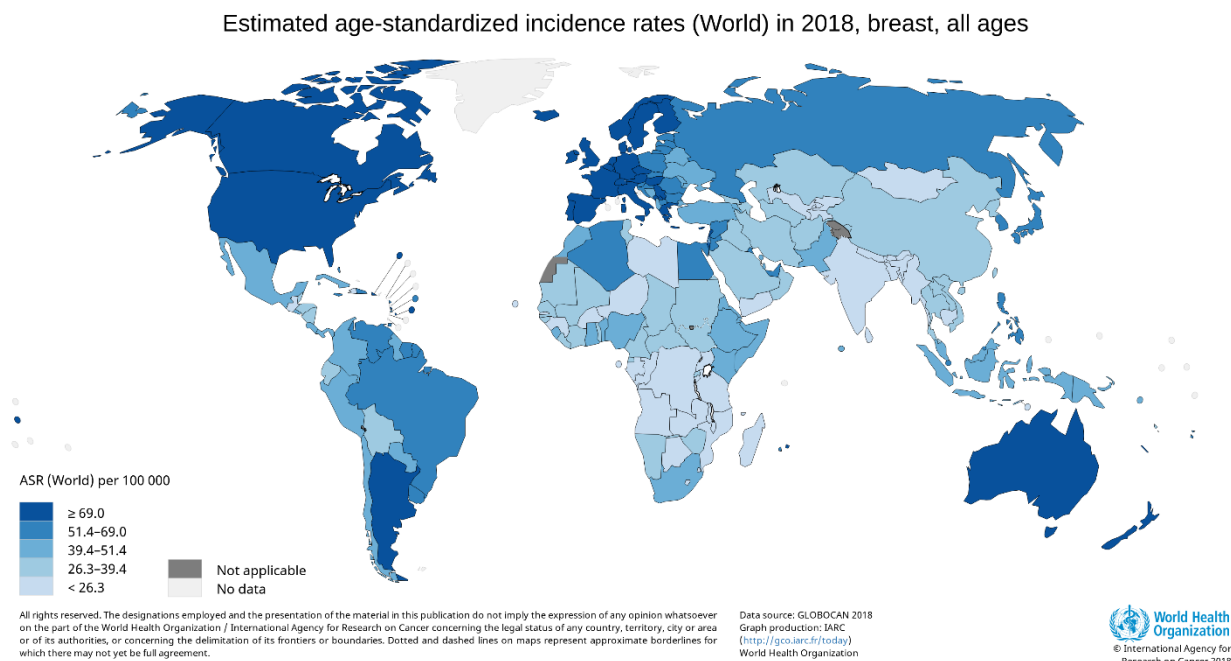


Figura 3. Distribución estandarizada para todas las edades de la tasa de incidencia de los TMM en el mundo. (International Agency for Research on Cancer, 2018)

En México, los TMM son el principal tipo de neoplasia maligna en mujeres (IMSS, 2015). De acuerdo con análisis internacionales, la tasa de incidencia de los TMM en México fluctúa entre 39.4 y 51.4 por cada 100,000 personas (Ghonchek, 2016). Por otro lado, estudios epidemiológicos en México indican 2,039 casos nuevos de TMM en el 2019 (9 primeras semanas). Siendo los estados con mayor incidencia: Baja California, Ciudad de México, Jalisco, Nuevo León y Veracruz (DGE, 2019). Las diferencias en las

incidencias podrían deberse a muchos factores entre ellos el estilo de vida, los factores genéticos y el medio ambiente de cada región.

1.1.2.5 Etiología de los TMM

La etiología de los TMM es poco conocida, no obstante, estudios recientes sugieren que los TMM son una enfermedad multifactorial, es decir, que su desarrollo está estrechamente vinculado con varios factores de riesgo (American Cancer Society, 2013). Estos factores se clasifican como extrínsecos e intrínsecos y tienen distinta influencia estadística en la incidencia de la enfermedad (Tabla 1).

Tabla 1. Tipos de factores de riesgo: los factores intrínsecos y no intrínsecos o extrínsecos. Los factores extrínsecos se subdividen en endógenos y exógenos (Wu *et al.*, 2018).

Factores de riesgo intrínsecos	Factores de riesgo Extrínsecos	
❖ Errores aleatorios en la replicación del ADN <i>No modificable</i>	Factores endógenos	Factores Exógenos
	❖ Edad biológica ❖ Susceptibilidad genética ❖ Hormonas ❖ Factores de crecimiento ❖ Inflamación ❖ Etnia <i>Parcialmente modificable</i>	❖ Radiación ❖ Carcinógenos químicos ❖ Virus ❖ Estilo de vida como fumar, vida sedentaria o mala alimentación <i>Modificable</i>

Los factores intrínsecos se relacionan a las mutaciones espontáneas que surgen como resultado de los errores aleatorios en la replicación del ADN. Los factores extrínsecos (Tabla 2) se subdividen en factores externos que son exógenos para el individuo como el estilo de vida; y los factores endógenos relacionados con las características internas que influyen en el control del crecimiento celular y la integridad del genoma como la edad y etnicidad (Tabla 2). Los factores extrínsecos son los que contribuyen con mayor magnitud al riesgo de desarrollar TMM (64.5%) (Wu *et al.*, 2016).

Tabla 2. Factores de riesgo extrínsecos de los TMM. Modificado de Momenimovahed (2019).

Tipo de factores de riesgo	Factores de riesgo individuales	Característica con mayor riesgo de desarrollar la enfermedad
Demográfico	Etnicidad	Blancas no hispanas
	Género	Femenino
	Edad	≥50 años
	Grupo sanguíneo	A y Rh ^s positivos
Reproductivo	Edad de la menarquia	Temprana
	Edad de la menopausia	Avanzada (≥ 50 años)
	Embarazo	Primer parto a edad avanzada o antes de las 33 semanas de gestación
	Aborto	Uno o más casos de aborto
Hormonales	Terapias hormonales	Uso de la píldora anticonceptiva oral, medicamentos estimulantes de la ovulación o terapia hormonal posmenopáusica
Hereditario	Factores genéticos	Mutaciones en los genes BRCA1 y BRCA2, polimorfismos de la metaloproteinasas de matriz, 3 o más casos de antecedentes familiares de TMM
Relacionados con los senos	Lactancia	Corta duración de la lactancia
	Senos	Aumento de densidad o trastornos benignos
Estilo de vida	Obesidad y sobrepeso	Niveles altos de insulina en respuesta a la obesidad
	Alcohol	Consumo de alcohol
	Fumar	Tabaquismo activo y pasivo
	Dieta	Consumo excesivo de carnes y grasas saturadas
	Actividad física	Poca o nula actividad física
	Vitamina D	Deficiencia de vitamina D
	Duración del sueño	Insomnio, poca duración y mala calidad del sueño
Otros	La contaminación del aire	Áreas urbanas con altos niveles de contaminación
	Trabajo nocturno	Aumento de los niveles de estrógeno a causa de la reducción de la melatonina por la exposición a la luz artificial durante la noche.
	Estatus socioeconómico	Estatus socioeconómico alto
	Diabetes	Diabetes en mujeres posmenopáusicas
	Radiación	Exposición a tratamientos con radiación

La edad y los antecedentes familiares han sido reportados como los factores de riesgos exógenos más importantes para las mujeres mexicanas (Calderón-Garcidueñas, 2000). Sin embargo, un estudio realizado en el 2017 afirmó que eran los contaminantes ambientales los principales responsables de la incidencia diferencial de TMM en distintas regiones del país (Castrezana, 2017). Asimismo, se ha reportado una relación entre la edad y ciertos tipos de tumor en el estudio realizado por Wu *et al.* (2018) donde los TMM con receptores de estrógenos parecen ser más frecuentes que los TMM triple negativo y HER2 positivo en sujetos de edad avanzada.

La etnicidad ha sido vinculada con la incidencia de TMM. Las mujeres blancas no latinas tienen una mayor incidencia en comparación con las latinas estadounidenses y las mujeres indígenas de América (141/89.9/54.8, respectivamente por cada 100, 000 personas) (Ci *et al.*, 2002). Fejerman *et al.* (2008) reportaron una asociación entre la ascendencia genética en latinas estadounidenses y el riesgo de los TMM y concluyeron que el riesgo de adquirir la enfermedad es mayor en las mujeres con una ascendencia europea (odd ratio=1.79) en comparación con las de ascendencia indígena americana. Sin embargo, en un estudio realizado a mujeres procedentes de países con bajo riesgo que migraron a E.E.U.U, se encontró que la segunda y tercera generación de los descendientes de estos migrantes, presentaron una tasa de incidencia de TMM más parecida a la de las mujeres de origen estadounidense que de su país de origen (Atkinson, 2018).

Asimismo, se ha reportado que las mujeres hispanas y afroamericanas tienen una mayor probabilidad de ser diagnosticadas con TMM en una etapa tardía (OR, 5.3) comparado con las mujeres blancas (Kamińska *et al.*, 2015). Existe evidencia de diferencias en la apariencia histológica, disminución de los niveles de expresión del receptor de esteroides y un incremento de la expresión de las proteínas reguladoras del ciclo celular en los TMM de mujeres afroamericanas en comparación con las mujeres blancas, lo que podría estar contribuyendo al diagnóstico tardío (Porter *et al.*, 2004). La expresión de ciertos genes en los TMM también puede estar asociado con el grupo étnico. Un estudio realizado por Lou *et al.* (2018) demostró que el 89% mujeres afroamericanas con TMM expresan el gen homeobox BP1, mientras que solo el 57% de las mujeres blancas caucásicas lo expresan ($p = 0.04$) (Chlebowski *et al.*, 2005).

Estudios recientes han corroborado que las diferencias por causas genéticas siguen siendo estadísticamente significativas incluso después del ajuste de variables socioeconómicas como el ingreso económico, la educación y el método de detección (Daly y Olopade, 2015). Sin embargo, la mayoría de los estudios realizados a nivel mundial involucran individuos de etnias distintas a las de nuestro país lo que restringe la interpretación y la extrapolación de los resultados a la población mexicana.

La población mexicana tiene una gran cantidad de variación genética que puede ser la base de fenotipos importantes en la investigación clínica. La ascendencia genética se ha asociado con numerosos criterios de valoración clínicos y riesgos de enfermedad como los recuentos de neutrófilos, los niveles de creatinina y la susceptibilidad al cáncer de mama lo cual es importante para el diseño e interpretación de los estudios médicos genéticos (Moreno-Estrada *et al.*, 2014).

1.1.2.6 Diagnóstico de los TMM

El diagnóstico clínico de los TMM y TMB actualmente se limita a técnicas de diagnóstico por contraste óptico (Figura 4). La mamografía de rayos X consiste en la medición de la densidad de la mama empleando rayos X y es considerada la técnica de diagnóstico estándar para la detección de TMM. Otros métodos de diagnóstico son la mamografía digital, que se diferencia de la tradicional por capturar de manera digital la imagen de rayos X de la mama; el ultrasonido, que utiliza ondas sonoras para identificar entre tejidos sólidos o quistes inofensivos; la imagen por resonancia magnética (MRI), que mide la respuesta del tejido de mama a un campo magnético; la tomografía óptica difusa de placa paralela (DOT), que mide la concentración de ciertas moléculas como oxi y desoxi hemoglobina, oxígeno, lípidos o la viscosidad de la matriz extracelular; la histopatología que realiza el análisis microscópico de los tejidos; y otras nuevas tecnologías como la gammagrafía y tomografía por emisión de positrones (PET), que analizan la temperatura, propiedades ópticas, eléctricas o elásticas de la mama (Paci, 2002; Nounou *et al.*, 2015).

Hasta la fecha, ninguna técnica óptica de diagnóstico puede detectar con precisión todos los tipos de anomalías en la mama, por lo que se recomienda el uso de múltiples técnicas de imagen para un diagnóstico más preciso. Además, se ha reportado que la mamografía de rayos x presenta una alta tasa de falsos positivos de TMM (~ 22% en mujeres menores de 50 años), baja sensibilidad a tumores pequeños (< 5mm) en el diagnóstico por ultrasonidos y baja especificidad entre TMM y TMB utilizando la MRI. Estas limitantes contribuyen a un sobret ratamiento de anomalías inofensivas como los TMB y representa un obstáculo para comenzar el tratamiento temprano en pacientes con tumores de mama (Choe *et al.*, 2009; Kukreti *et al.*, 2009; Nabavizadeh *et al.*, 2019; Yang *et al.*, 2015; Drukteinis *et al.*, 2013).

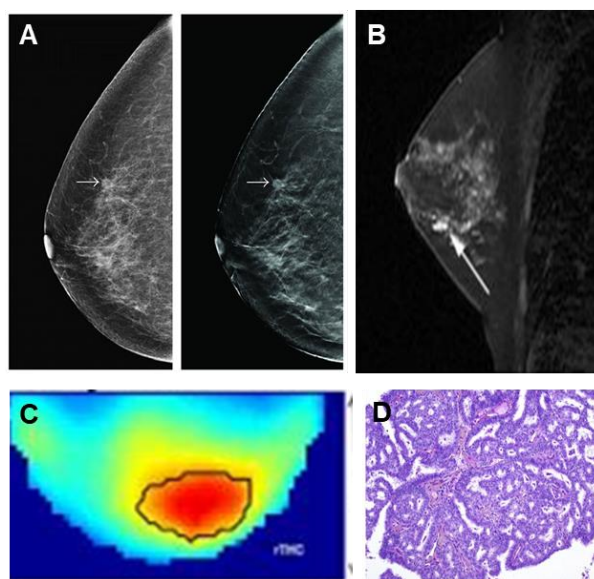


Figura 4. Métodos de diagnóstico para los TMM (A) Mamografía, (B) Imagen por Resonancia Magnética, (C) Tomografía Óptica Difusa de Placa Paralela, (D) Histopatología (Lehman y Schnall, 2005; Choe et al., 2009; Weigelt et al., 2010).

El mal diagnóstico podría atribuirse a la heterogeneidad tumoral incluyendo diversos componentes moleculares desconocidos, innatos de cada tumor que no están siendo tomados en cuenta a la hora del análisis. Por ello, la información adicional de otras características morfológicas y moleculares más precisas en ambos tipos de tumores podría mejorar el diagnóstico (Lehman y Schnall, 2005; Choe et al., 2009; Weigelt et al., 2010). Es así que, el descubrimiento de marcadores de malignidad confiables de los TMM se propone como una nueva estrategia que coadyuve a mejorar la sensibilidad, especificidad y precisión de los métodos convencionales, así como ayudar a comprender la biología del cáncer y el diagnóstico más preciso.

1.1.2.6.1 Proteínas como biomarcadores de malignidad de tumores de mama

Un biomarcador es una característica biológica cuantificable de naturaleza molecular indistinta (lipídeos, proteínas, genes, etc) que indica de forma reproducible procesos biológicos normales, patológicos o de respuesta a una exposición farmacológica. Los biomarcadores pueden ser de diagnóstico, monitoreo, predicción, pronóstico, riesgo, validación, respuesta a fármacos, entre otros (FDA-NIH Biomarker Working Group, 2017).

Los genes por sí solos no proporcionan un panorama completo de la biología ni los cambios relacionados al microentorno tumoral, porque el genoma de un organismo es estático (Dias *et al.*, 2016). Conocer la secuencia completa del genoma no es suficiente para aclarar la función biológica debido a que no existe una relación lineal determinante entre los genes y el proteoma de una célula.

El RNA mensajero es uno de los productos génicos involucrados en la dinámica celular (Pando-Robles y Lanz-Mendoza, 2009), sin embargo, no permiten comprender de manera exacta los mecanismos involucrados en la biología del cáncer debido a las modificaciones postraduccionales (PTM), empalmes alternativos o tasas diferenciales de renovación de proteínas. Además, la traducción de ARNm a proteínas es independiente a los niveles de abundancia de proteínas (Hixson *et al.*, 2017).

Por el contrario, las proteínas son las principales biomoléculas que ejercen la función celular y son capaces de dar una imagen panorámica de los procesos biológicos en función a su localización e interacción con otras proteínas (Liang, 2013). Los biomarcadores basados en proteínas son fundamentales para el diagnóstico fenotípico de los TMM. Sin embargo, no existe el conocimiento necesario de los perfiles proteómicos asociados a los TMM; conocimiento que sí existe en genómica (De Gruttola *et al.*, 2001; Cassidy y Méndez, 2015; Fan *et al.*, 2001; Nakagomi *et al.*, 2017; Uribe *et al.*, 2010; Sørli *et al.*, 2001).

Por tanto, para comprender integralmente un sistema biológico se debe estudiar al conjunto de proteínas o proteoma que compone a dicho sistema, ya que las proteínas dictan de manera directa la función celular.

1.1.3 Proteómica

El término proteómica se refiere al estudio a gran escala de las proteínas que constituyen una célula, tejido u organismo (proteoma) bajo condiciones definidas y específicas. En la proteómica, existen dos metodologías para la identificación de proteínas: el análisis de proteínas intactas (de arriba hacia abajo o *top down*) y el análisis de péptidos proteicos digeridos enzimáticamente (de abajo hacia arriba o *bottom up*) (Figura 5). La espectrometría de masas (MS) es una tecnología analítica esencial en el contexto de la proteómica debido a su alta capacidad de análisis y sensibilidad, capaz de detectar hasta la diezmilésima parte de la masa atómica (masa de un isótopo individual) de las moléculas analizadas (Pleil & Isaacs, 2016).

La MS puede emplear procedimientos experimentales adicionales como el análisis con marcaje isotópico (isotopic labelling) o sin marcaje (label free o basada en la intensidad de la MS) y la espectrometría de masas en tándem (MS/MS) para mejorar la sensibilidad y obtener más información del analito (Hixson *et al.*, 2017). El análisis sin marcaje es la metodología recomendada para el descubrimiento de biomarcadores (Levin *et al.*, 2007).

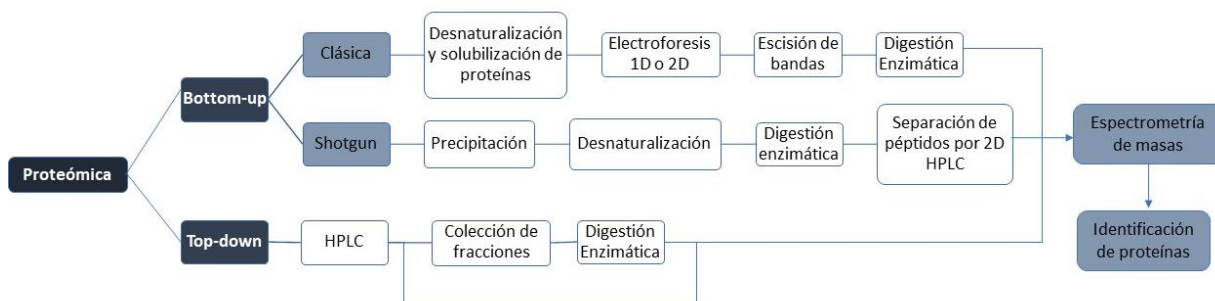


Figura 5. Diagrama de flujo de las dos metodologías asociadas a la proteómica LC-MS/MS: *bottom up* y *top down*. Modificado de Hixson *et al.* (2017).

1.1.3.1 Proteómica tipo “shotgun”

Debido a la complejidad de los sistemas biológicos, la proteómica moderna o *shotgun*, se basa en 3 pilares tecnológicos que incluyen 1) un método para separar proteínas/péptidos de una muestra compleja en función de sus propiedades químicas o físicas, 2) espectrometría de masas (MS, por sus siglas en inglés) para adquirir los datos necesarios para la identificación de proteínas individuales y 3) bioinformática para analizar y ensamblar los datos proporcionados por la MS (Figura 6) (Shah y Misra, 2011; Yu *et al.*, 2010).

La metodología de la proteómica *shotgun* comienza con la extracción de las proteínas de la muestra y su posterior digestión enzimática generalmente con tripsina. Generalmente se utiliza una cromatografía líquida de alta resolución (HPLC) para la separación de los péptidos digeridos que son posteriormente ionizados. Existe una gran variedad de fuentes de ionización disponibles para MS como la Ionización por Electrospray (ESI), Ionización Química a Presión Atmosférica (APCI), Fotoionización a Presión Atmosférica (APPI) e Ionización por Desorción Laser Asistida por Matriz (MALDI), de las cuales la técnica más usada para ionizar proteínas es la ESI. La ESI consiste en pasar el analito a través de un capilar al que se aplica un alto potencial eléctrico, luego la solución se dispersa en forma de spray formado por pequeñas

gotas cargadas, las cuales se evaporan rápidamente por un proceso de desorción o evaporación del solvente obteniendo péptidos o proteínas protonadas preferentemente en los aminoácidos histidina, arginina y lisina y en su extremo N-terminal (Siuzak, 2004).

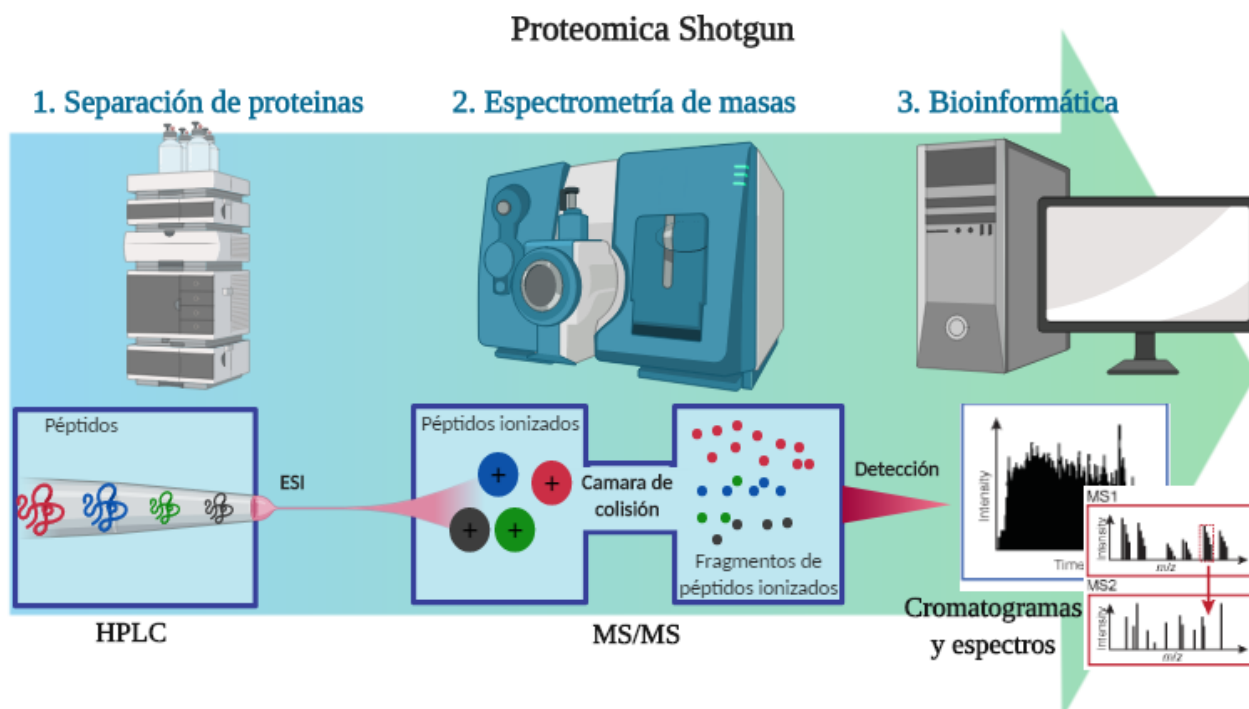


Figura 6. Esquema de los 3 pilares tecnológicos de la proteómica *shotgun* que incluyen 1) la separación de péptidos 2) espectrometría de masas y 3) bioinformática. Creado por Biorender. www.biorender.com

Los péptidos o proteínas ionizadas, también conocidos como iones parentales o precursores (MS1) son direccionados al analizador de masa. Existe una gran variedad de analizadores de masas, cada uno con su propio nicho de aplicación específico, que incluye el Cuadrupolo (Q), Tiempo de vuelo (TOF), Trampa de Iones, Trampa de Paul (3D), Trampa de Iones Linear (2D), Trampa Electrostática (Orbitrap), Trampa de Resonancia Ciclotrónica (ICR) e Híbridos en Tándem, como los Q-Trap, Q-TOF y Q-Orbitrap. Asimismo, los analizadores híbridos se caracterizan por presentar dos o más analizadores de masa diferentes separadas por una celda de colisión que genera iones de fragmentación (MS2) (Clarke, 2017). El espectrómetro de masas AB Sciex TripleTOF5600+ (SCIEX, Framingham, MA) presenta un analizador híbrido TOF unido a un triple Q que se destaca por su alto poder de resolución (30k) y una precisión de masa (<3 ppm) con altas tasas de adquisición espectral MS2 (20Hz) que permiten una potente plataforma de descubrimiento para mezclas complejas (Andrews *et al.*, 2011).

Los MS1 y MS2 individuales son seleccionados de acuerdo a la relación de su masa y carga (m/z) en el analizador de masas y finalmente detectados por el sistema de detección. Como resultado, se genera un cromatograma con información de la abundancia relativa de un ion específico en función al tiempo para realizar la cuantificación de proteínas. Finalmente, los datos proporcionados por el MS son analizados por herramientas bioinformáticas.

1.1.3.2 Tipos de cuantificación proteómica

Existen dos tipos de medición cuantitativa usando MS. La cuantificación absoluta, donde el objetivo es medir la concentración real de la proteína o péptido en la muestra; y la cuantificación relativa donde la concentración relativa de proteína/péptido es comparado entre dos o más muestras. La cuantificación absoluta se utiliza cuando se analizan pocos y conocidos péptidos o proteínas de interés, de lo contrario se utiliza la cuantificación relativa (Hixson *et al.*, 2017).

1.1.3.3 Tipos de adquisición de datos

Tradicionalmente, los métodos de adquisición de datos MS2 se han dividido en dos tipos: La adquisición dependiente de datos (DDA) o proteómica de descubrimiento, y la adquisición dirigida que incluye 3 su análisis, el monitoreo de reacción múltiple (MRM), la SRM (*Selected Reaction Monitoring*, Monitorización de Reacción Seleccionada) y la PRM (*Parallel Reaction Monitoring*, o Monitorización de Reacción Paralela) (Figura 7) (Fernández, 2016).

En la DDA los iones precursores o MS1 son elegidos estocásticamente en función de su abundancia. Los MS1 más abundantes se aíslan y fragmentan individualmente en iones de fragmentación o MS2 que luego son analizados con algoritmos de búsqueda. Este enfoque disminuye la reproducibilidad y evita la medición de péptidos de baja abundancia, en consecuencia, existe una pérdida de información. Sin embargo, la DDA tiene una gran amplitud de detección, es decir, permite identificar y cuantificar un enorme número de proteínas y no requiere ningún tipo de conocimiento previo acerca de la composición de la muestra (Aebersold & Mann, 2016). Por otra parte, el monitoreo de reacción múltiple (MRM) selecciona los iones precursores o MS1 que coinciden con un péptido de interés especificado previamente.

Los MS2 se enfocan en el pequeño conjunto de péptidos predeterminados. Este enfoque se utiliza cuando los objetivos principales del análisis es la cuantificación precisa, la reproducibilidad y la validación (García, 2017).

Recientemente, se ha desarrollado otro método de muestreo llamado adquisición independiente de datos (DIA) donde todos los MS1 dentro de un intervalo de masa seleccionado se aíslan, fragmentan y analizan en MS2 (Hu *et al.*, 2016). La cuantificación se basa en la comparación de espectros DIA con bibliotecas espectrales que muestran patrones de fragmentación precisos y determinados empíricamente para cada péptido en la biblioteca (Amodei *et al.*, 2019). Un enfoque alternativo para la cuantificación del proteoma que combina un método DIA de alta especificidad con una nueva estrategia de extracción de datos dirigida para extraer los conjuntos de datos de fragmentos de iones resultantes es la cuantificación SWATH – MS (Gillet *et al.*, 2012)

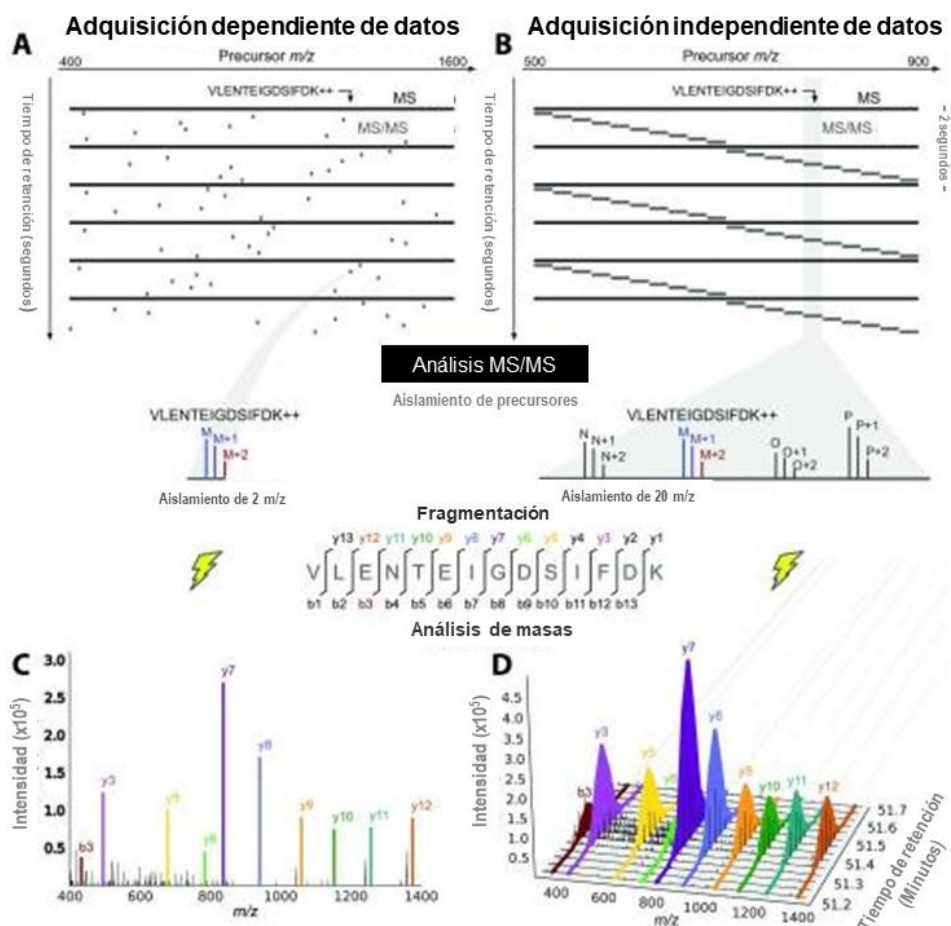


Figura 7. Tipos de adquisición de datos de la proteómica cuantitativa (A) Adquisición dependiente de datos (DDA) (B) Adquisición independiente de datos (DIA) (C) La información del MS2 para el MS1 VLENTEIGDSIFDK ++ (D) Extracción de los MS2 en el tiempo de los datos de DIA. Modificados de (Hu *et al.*, 2016).

1.1.3.4 Cuantificación SWATH - MS

La adquisición secuencial de ventanas de todos los espectros de masas teóricos (SWATH – MS) es una variante de los métodos de DIA. Este esquema de adquisición fragmenta sistemáticamente todos los MS1 en intervalos de masa específico consecutivos y que son denominadas como ventanas de aislamiento. Esto da como resultado un conjunto de MS2 multiplexados continuos en la intensidad como en las dimensiones del tiempo de retención que son utilizados en la cuantificación junto con una biblioteca de análisis generada previamente. La biblioteca de proteínas proporciona información cromatográfica y espectrométrica de masas de los péptidos de interés en forma de los llamados parámetros de consulta de péptidos (PQP, por sus siglas en inglés) (Ludwig *et al.*, 2018; Rosenberger *et al.*, 2014).

1.1.3.5 Aplicación de la proteómica a la biología del cáncer

En las últimas décadas, el desarrollo de tecnologías proteómicas ha permitido un gran avance en la caracterización molecular de los TMM mediante la identificación de nuevos biomarcadores que han ayudado a comprender las vías de señalización que gobiernan su desarrollo, progresión y diferenciación entre subtipos de TMM. Por ejemplo, se han identificado fenotipos de los TMM ER α + y ER α que indican un aumento de abundancia en proteínas relacionadas principalmente con el metabolismo de los aminoácidos, el proteasoma y el metabolismo de los ácidos grasos en los TMM ER α +; y un aumento en aquellas relacionadas con la ruta de la glucólisis en los TMM ER α - (Rezaul *et al.*, 2010). Del mismo modo, se han identificaron diferencias a nivel de abundancia de proteínas entre los subtipos de TMM luminal A, Her2 y triple negativo (Tyanova *et al.*, 2016).

Sanders *et al.* (2008) reportaron que la Ubiquitina, Calciclina y proteína de unión al Calcio S100 A8 (S100-A8) presentan una disminución estadísticamente significativa en el tejido tumoral (n=122) en comparación con el tejido normal (n=167). También se han sugerido diferencias a nivel de proteínas entre TMM de tipo ductal infiltrante y tejido de mama normal de mujeres tunecinas (n=10), donde se identificaron una serie de proteínas como la Enolasa 1, Fosfoglicerato quinasa 1, Desoxihemoglobina, Mn-superóxido dismutasa. La Cadena B de alfa-cristalina (CRYAB), Proteína de choque térmico 27 (HSP27), Proteína inhibidora de la quinasa Raf (RKIP), Ribonucleoproteína nuclear heterogénea A2/B1, Cofilina 1 y Peptidilprolil isomerasa con una abundancia aumentada en los TMM; y a las Proteínas inmunoglobulina

pesada gamma 1 (IGHG1) y Componente C3 del complemento (C3) con abundancia disminuida (Kabbage *et al.*, 2008).

Se ha estudiado el proteoma de TMM en otros grupos étnicos. Un estudio realizado en mujeres coreanas, reportó diferencias a nivel de proteínas entre tejido normal y los TMM ductal in situ (n=1) e invasivo (n=2). Las proteínas Calreticulina (CRT) y la proteína Disulfuro isomerasa A3 (PDIA3) aumentaron su abundancia en los TMM ductal in situ e invasivos con respecto al tejido de mama normal (Song *et al.*, 2012). Por otro lado, un estudio realizado a mujeres brasileñas comparó el proteoma de TMM ductal invasivo y el tejido no tumoral circundante; e identificó 161 proteínas desreguladas de las cuales 33 proteínas relacionadas al citoesqueleto y proteínas de unión a la actina disminuyeron su abundancia en los TMM respecto al tejido normal; mientras que 128 proteínas relacionadas a enzimas metabólicas aumentaron su abundancia (Da Costa *et al.*, 2015). En mujeres mexicanas, se ha reportado aumentada la abundancia de la proteína Gluoxalasa 1 (GLO1) en TMM (n=98) en comparación del tejido de mama normal circundantes del mismo paciente, una enzima involucrada en la desintoxicación del Metilglioxal, un producto Citotóxico de la glucólisis. (Fonseca-Sánchez *et al.*, 2012).

Además de análisis proteómicos en tejido tumoral, se han reportado estudios en líneas celulares humanas. Calderón-González *et al.* (2015) sugieren que las proteínas con mayor abundancia en los TMM incluyen proteínas mitocondriales involucradas en la fosforilación oxidativa. Otro estudio comparó 16 líneas celulares y 4 tejidos tumorales de tipo triple negativo, y reportó una disminución en la abundancia de la Proteína fosfatasa 1A (PPM1A) y aumento en Periostina (POSTN), Miosina cadena ligera quinasa (MYLK) y complejo de antígeno leucocitario humano (HLA-A) en los TMM (Lawrence *et al.*, 2015). En México se ha reportado un estudio en líneas celulares MCF-7 y biopsias de TMM de tipo triple negativo (n=6) que reveló la existencia de 16 proteínas con mayor abundancia en los TMM entre las cuales destacan el Inhibidor de disociación Rho GDP 2 (RhoGDI-2), la Proteína de choque térmico 27 (HSP27), el superóxido dismutasa 1 (SOD1), la Enzima conjugadora de ubiquitina E2 N (UBE2N), Complejo activador del proteasoma subunidad 1 (PSME1) y la Cadena ligera de ferritina (FTL) (Muñiz Lino, *et al.*, 2014).

Los estudios proteómicos relacionados al cáncer de mama han estado enfocados a la diferenciación de los subtipos de los TMM y comparación con el tejido de mama normal. Sin embargo, los análisis entre TMB y TMM son escasos. Además, el tipo de muestra analizada generalmente corresponde a líneas celulares de cáncer de mama, así como tejidos de mama de mujeres de etnias diferentes a las de México debido a lo cual existe la necesidad de estudiar TBM y TMM para conocer cuál es la diferencia a nivel de proteínas entre estos tipos de tumor en la población mexicana.

1.2 Justificación

Se ha demostrado que los factores genéticos y ambientales contribuyen a la incidencia de los TMM, lo cual sugiere heterogeneidad molecular de la patología en función a la etnia y localización geográfica. Sin embargo, en México, no hay estudios que describan un perfil proteómico de TMM y TMB a pesar de que el cáncer de mama ocupa el primer lugar en incidencia entre las neoplasias malignas en las mujeres.

Comprendiendo la biología de los TMM Y TMB se podría mejorar la sensibilidad y precisión de los métodos de diagnóstico cuya alta tasa de falsos positivos contribuye a que los tratamientos y procedimientos clínicos sean inadecuados e inespecíficos para el tipo de tumor de los pacientes.

1.3 Hipótesis

Existe una diferencia en la abundancia de proteínas entre los TBM y TMM de mujeres mexicanas.

1.4 Objetivos

1.4.1 Objetivo general

Evaluar cambios en la abundancia de proteínas entre los TMB y TMM de mujeres mexicanas.

1.4.2. Objetivos específicos

- 1.4.2.1 Estandarización de un método de detección y cuantificación de proteínas por LC-MS/MS para tejido tumoral de mama.
- 1.4.2.2 Creación de una biblioteca informática espectral de proteínas de tejido tumoral de mama.
- 1.4.2.3 Cuantificación de la abundancia de las proteínas de TBM y TMM.
- 1.4.2.4 Identificación de redes proteínicas y vías de señalización asociadas a las proteínas diferencialmente abundantes entre TBM y TMM.
- 1.4.2.5 Optimización de la biblioteca informática espectral de proteínas del tejido tumoral de mama.

Capítulo 2. Metodología

2.1 Diseño del estudio y obtención de muestras

En este estudio piloto, se analizaron dieciséis muestras de mujeres mexicanas diagnosticadas con tumor de mama (TMM n=6 y TMB n=10). Todas las muestras se obtuvieron del Hospital General Zona Norte de Puebla de la Secretaría de la Salud del Estado de Puebla. La aprobación ética para la recolección de muestras de las pacientes fue aceptada por la Junta de Revisión Institucional de acuerdo con la Ley de Portabilidad y Responsabilidad de Seguros Médicos (HIPAA, por sus siglas en inglés) y todos los participantes firmaron formularios de consentimiento informado. Las muestras se obtuvieron de un estudio previo (Perez-Ortiz *et al*, 2017). Las mujeres diagnosticadas con TMM recibieron un tratamiento previo de 80 mg/m² de paclitaxel por 12 semanas como quimioterapia neoadyuvante. Todas las muestras fueron obtenidas por cirugía conservadora de mama (tumorectomía) y posteriormente congeladas a -80 °C. Las muestras recolectadas se evaluaron y agruparon en el análisis de acuerdo con el informe histopatológico después del diagnóstico. Las características histopatológicas y clínicas se resumen en la Tabla 3.

Tabla 3. Características histopatológicas y clínicas de los TBM y TMM

Tipo del tumor	TBM	TMM
Edad	22-68	40-72
Diagnostico	Mastitis, fibroadenoma	HER2Neu, Triple Negativo y Luminal B
Terapia	Sin terapia	Paclitaxel
Diámetro del tumor	Mayor a 2 cm	Mayor a 2 cm

La metodología empleada en este estudio se ilustra en resumen en la **Figura 8**.

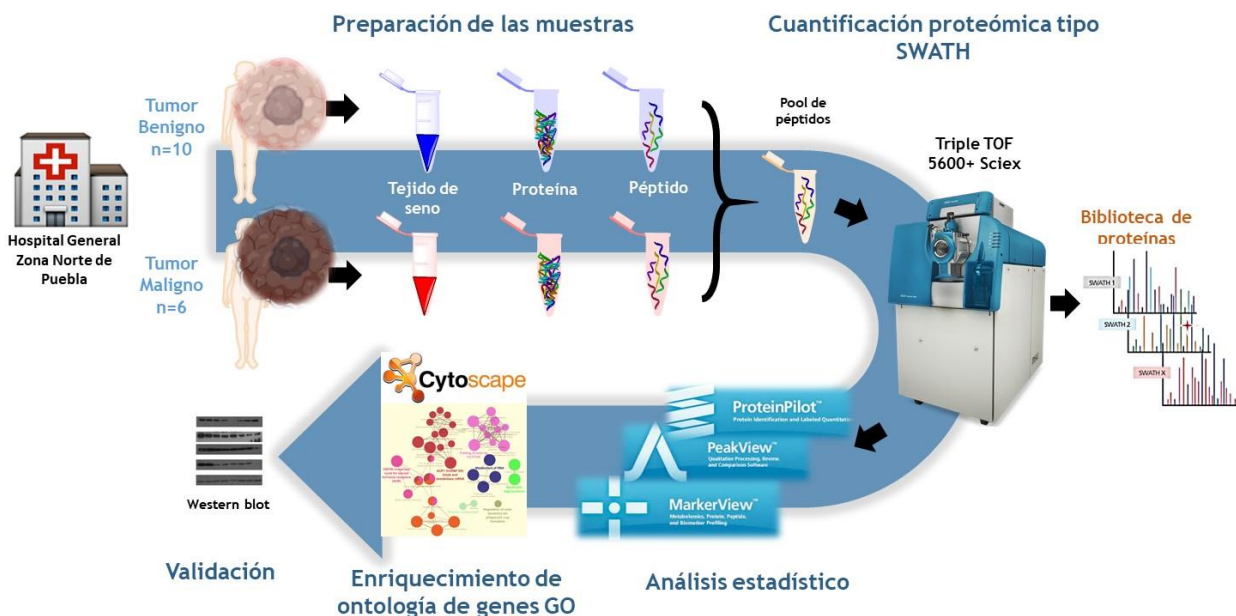


Figura 8. Diagrama de flujo generalizado de la metodología empleada en este estudio

2.2 Preparación de las muestras

2.2.1 Extracción de proteínas

Las dieciséis muestras de tejido mamario fueron maceradas con un homogenizador eléctrico (Polytron) en un tubo de vidrio y lisadas con 150-400 μl de buffer para lisis preparado con 0.5 M de NH_4HCO_3 a pH 8.0, 8 M de urea, 150 mM de NaCl, 3 mM MgCl_2 y 0.5% de Triton X-100 suplementado con inhibidores de proteasas (1mM-2mM Na_3VO_4 , 1mM NaF, 1mM PMSF, coctel inhibidor de proteasas [Sigma Aldrich] y coctel inhibidor de proteasas HALT [Sigma Aldrich]).

Posteriormente, se sonicaron por 5 minutos dos veces. Entre cada sonicación se realizó una agitación vigorosa con un vórtex por 30 segundos. Adicionalmente, las muestras se agitaron por inversión por 30 min a 4 $^{\circ}\text{C}$ y se centrifugaron a 14,000 rpm por 10 min. El sobrenadante (proteínas) fue recolectado y almacenado a -80 $^{\circ}\text{C}$. Las muestras fueron manejadas en baño de hielo durante todos los pasos. Finalmente, la determinación de la concentración proteínica se realizó mediante el ensayo de Bradford.

2.2.2 Obtención de péptidos trípticos

Los péptidos se obtuvieron con el kit comercial de digestión de proteínas Pierce™ Mass Spec Sample Prep Kit for Cultured Cells de la marca Thermo Fisher Scientific. Este procedimiento consistió en la dilución de las muestras en buffer de lisis celular a un volumen final de 100 ul y la posterior reducción de las proteínas con 2.1 ul de ditioneitol (DTT) incubándolas por 45 min a 50 °C. Las proteínas se alquilaron con 11.5 ul de iodoacetamida (IAA) incubándolas protegidas de la luz a temperatura ambiente por 20 min. Posteriormente, las proteínas se precipitaron con 460 ul de acetona al 100% durante toda la noche. Se agregó 50 ul de acetona al 90%, se centrifugó a 14,000 rpm a 4 °C por 5 min y se removió el sobrenadante. Se dejó secar a temperatura ambiente y después se le agregó 50 ul de buffer de digestión a los precipitados (proteínas). Finalmente, se agregaron 2 y 4 ul de las endoproteinasas Lys-C y tripsina y se incubaron por 2 y 3 horas, respectivamente. Las muestras fueron congeladas a -80 °C y secadas mediante centrifugación y alto vacío utilizando un Sistema SpeedVac (Savant, Thermo Fisher Scientific).

Después de la digestión de proteínas, se continuó con el proceso de extracción de detergente con acetato de etilo siguiendo el protocolo de Yeung Yee-Guide (Yee-Guide y Stanley, 2011). Luego, las muestras fueron secadas utilizando el sistema SpeedVac y resuspendidas en 95:5 agua:acetonitrilo con ácido fórmico al 0.1%. La concentración de los péptidos trípticos obtenidos se determinó mediante el empleo del kit Pierce™ Quantitative Colorimetric Peptide Assay de la marca Pierce de Thermo Fisher Scientific y un Espectrofotómetro UV-VIS-multiplaca 96 EPOCH (Biotek, Winooski, EE.UU) (absorbancia a 480 nm).

2.3 Estandarización del método de LC-MS/MS

Antes de analizar los péptidos resultantes de las muestras de TBM y TMM por LC-MS/MS, se estandarizó el método utilizando muestras de proteína provenientes de células endoteliales de la arteria coronaria bovina (BCAEC) y tejidos de retina de ratón. Los parámetros que se evaluaron se describen a continuación.

Para corroborar que la detección de la abundancia de proteínas estaba vinculada intrínsecamente a la cantidad de muestra inyectada al equipo y no a alguna falla o sesgo del equipo de detención, se inyectaron dos cantidades distintas de péptidos trípticos de BCAEC (0.5 y 1 ug) por triplicado al equipo LC-

MS/MS y se evaluó su respuesta de detección. Por otro lado, se evaluaron la capacidad de adquisición de datos con dos tiempos de elución diferentes (120 y 150 min) y la especificidad de cuantificación SWATH-MS con dos tamaños mínimos de ventana variable (10 y 14 Da) utilizando los péptidos trípticos de retina de ratón.

2.4 Cuantificación SWATH – MS

2.4.1 Análisis LC-MS/MS

Los péptidos trípticos de TBM y TMM fueron analizados por LC-MS/MS. La LC se llevó a cabo utilizando un sistema Eksigent nanoLC[®] 400 (Eksigent, Dublin, CA) que consiste en una columna analítica HALO Fused-Core C18 (0.3 × 150 mm de núcleo sólido, 2.7 μm y un tamaño de poro de 90 Å, Eksigent AB Sciex, Foster City, CA, EE. UU). La velocidad de flujo fue de 5 μL/min y la cantidad de muestra inyectada de 4 uL. La separación de péptidos se realizó utilizando un gradiente de elución de 150 min con 0.1% de ácido fórmico en agua (fase móvil A) y 0.1% ácido fórmico en acetonitrilo (fase móvil B). El gradiente comenzó con 95-5% A-B por 5 min, manteniéndose constante por 35 min, cambiando a 90-10% A-B por 60 min. Después, la composición se cambió a 70-30% A-B por 35 min y por último a 0-100% A-B por 15 min. Dos muestras blanco (solvente) fueron inyectadas entre muestras para minimizar el arrastre de remanentes (carryover). El eluyente fue dirigido directamente al MS AB Sciex TripleTOF5600+ (SCIEX, Framingham, MA) usando ionización por electrospray (ESI) con las siguientes condiciones IonSpray Voltage Floating (ISVF), 5500 V; Temperatura de fuente, 450 ° C; Cortina de gas (CUR), 30; Fuente de iones Gas 1 (GS1), 25; Fuente de iones Gas 2 (GS2), 35.

2.4.2 Creación de una biblioteca espectral de proteínas

La creación de la biblioteca espectral SWATH-MS se realizó mezclando todos los péptidos trípticos de las muestras de TBM y TMM en un único pool y sometidos a un análisis LC-MS/MS, como se mencionó anteriormente, operado en modo de DDA. El pool de proteínas se inyectó por triplicado al espectrómetro de masas TripleTOF5600+ cambiando automáticamente entre el modo de adquisición

completa MS y MS/MS. El tiempo de acumulación para el análisis MS fue de 0.25 s en el intervalo m/z de 350-1200 Da y para el análisis MS/MS fue de 0.05 s en el intervalo m/z 50-1200 Da (tiempo de ciclo 1.3 s). Las configuraciones de la IDA fueron: estado de carga +2 a +4, intensidad 150 cps, excluir isótopos dentro de 6 Da, tolerancia de masa 50 mDa y un número máximo de iones candidatos 20. Se seleccionó "excluir iones objetivo anteriores" 20 s después de dos eventos y también se seleccionó "restar fondo dinámico" (Dynamic Background Subtraction). Esto significa que los 20 iones más intensos (intensidad > 150) con estados de carga de +2 a +4 se aislaron y fragmentaron secuencialmente dentro de la trampa de iones lineal usando CID y finalmente se excluyeron dinámicamente durante 20s para evitar la selección repetitiva del mismo péptido. El instrumento se calibró automáticamente utilizando soluciones de calibración TOF MS y MS / MS positivas antes y después de las inyecciones (<3.5 horas) para garantizar una precisión de masa de <5 ppm para los datos MS y MS/MS.

La identificación de proteínas se llevó a cabo mediante el algoritmo Paragon™ incluido en el Software Protein Pilot™ (versión 4.2, AB Sciex, Foster City, CA, EE. UU.) que se basa en la alineación de los datos LC-MS/MS con las secuencias de proteínas de la base de datos del proteoma humano revisado en formato FASTA (20,360 entradas) obtenido de Uniprot (<http://www.uniprot.org>). Los parámetros de búsqueda fueron los siguientes: tipo de muestra: identificación; alquilación de Cys: yodoacetamida; digestión: tripsina; instrumento: TripleTOF 5600; factores especiales, ninguno; especie: *Homo sapiens*; sin modificaciones biológicas ni sustituciones de aminoácidos; un mínimo de un péptido único y una tasa de descubrimiento falso (FDR) de software con cut off >1.3 (95% de confianza). La biblioteca de proteínas generada fue exportada en formato group. Adicionalmente, la biblioteca fue cargada en el programa PeakView (V2.1 con el complemento SWATH Quantitation) y exportada en formato CSV.

2.4.3 Análisis de enriquecimientos de ontología de genes GO de la biblioteca de proteínas.

El análisis de enriquecimientos de ontología de genes GO es una forma de estructurar u organizar un conjunto de proteínas en términos de los procesos biológicos, componentes celulares o funciones moleculares asociados entre ellas; es decir, se clasifican las proteínas según su ontología genética. En este estudio, se utilizó el software ClueGO (version 2.5.1) de la plataforma Cytoscape para identificar la localización dentro de la célula de las proteínas de la biblioteca espectral y evaluar la cobertura celular del método de creación de la biblioteca. Porque a mayor cobertura celular, mayor probabilidad de encontrar biomarcadores y mejora sensibilidad de detección.

2.4.4 Cuantificación diferencial de proteínas de los TBM y TMM

La DIA se generó con 56 ventanas de aislamiento variables utilizando SWATH Variable Window Calculator (AB Sciex, Foster City, CA, USA) basada en las frecuencias de precursores m/z en las muestras de DDA, con un mínimo ancho de ventana de 5 m/z. El tiempo de acumulación para el análisis MS TOF fue de 0.15 s en el intervalo 350-1200 Da y los datos MS/MS se obtuvieron de 350-1200 Da con un tiempo de acumulación de 0.05 s (tiempo de ciclo 3 s). Los demás parámetros fueron los mismos que el usado para el análisis DDA.

Para realizar la cuantificación diferencial de proteínas, los datos obtenidos de la extracción de datos dirigida (DIA) junto con la biblioteca de proteínas generada por DDA se analizaron con la aplicación de cuantificación SWATH[®] Adquisición de MicroApp 2.0 en PeakView versión 1.2 (AB Sciex, Foster City, CA, USA). PeakView extrae el área de los picos SWATH MS y obtiene los valores de intensidad, m/z y tiempo de retención de todas las proteínas para realizar la cuantificación. La alineación de las proteínas fue realizada por una curva de calibración del tiempo de retención de las proteínas más intensas que fueron seleccionadas manualmente. La configuración usada para la cuantificación SWATH-MS fue un número máximo de 6 péptidos por proteína que se incluirán de la biblioteca de iones importados, 6 espectros de fragmentación o iones producto por péptido, un umbral de confianza de péptidos de 99%, una tasa de falsos descubrimientos (FDR) menor al 1%, una ventana de tiempo de retención XIC de 50 ppm, 10 min y se excluyeron los péptidos modificados. Finalmente, las áreas se exportaron a Markerview (AB Sciex, Foster City, CA, USA) para un análisis estadístico adicional.

2.5 Análisis estadístico

Las áreas de los MS₂ de los péptidos trípticos de las proteínas fueron exportadas de PeakView a formato del software MarkerView[™] (AB Sciex) para el análisis estadístico. Las intensidades de las áreas de los MS₂ de los péptidos se normalizaron frente a la intensidad total (factor de la escala logarítmica de la abundancia de la muestra). La comparación de la abundancia de las proteínas entre el grupo de TBM y TMM se realizó mediante la prueba t de Student no preada de dos colas (varianza desigual, Welch). Las proteínas que fueron consideradas con un cambio significativo en la abundancia fueron las que presentaron un valor de p menor a 0.05 y un fold-change (proporción de cambio) de la abundancia de proteínas de TMM y la abundancia de proteínas de TBM, mayor a 1.5 y menor a 0.66 (aumento o

disminución en la abundancia de proteínas, respectivamente). Para visualizar las proteínas diferencialmente abundantes entre ambos grupos, se realizó una gráfica de volcán utilizando el software comercial GraphPad Prism (GraphPad Software Inc.). Adicionalmente se realizaron dos análisis multivariados, el análisis de componentes principales (PCA) y análisis de componentes principales por agrupación de variable (PCVG) para reconocer la tendencia de los grupos dentro del conjunto de datos y el descubrimiento de potenciales biomarcadores.

2.6 Identificación de vías de señalización asociadas a las proteínas diferencialmente abundantes en el TMM

Para facilitar la interpretación biológica y visualizar las redes proteínicas, se realizó un análisis de enriquecimiento de genes utilizando el software ClueGO y la base de datos REACTOME (<https://reactome.org>). Las proteínas cuya abundancia presentó un cambio significativo en el grupo de los TMM con respecto a los TBM fueron analizadas por separado. Los parámetros utilizados para las proteínas que disminuyeron y aumentaron significativamente en los TMM fueron los siguientes: Prueba estadística utilizada = Enrichment/Depletion (prueba hipergeométrica de dos lados); Método de corrección utilizado = Reducción de Bonferroni; Nivel mínimo de GO = 3; Nivel máximo de GO = 8; Número de genes = 3; porcentaje mínimo = 2.0 Grupo de fusión = false; Umbral de puntaje Kappa = 0.4 y *Sharing Group Percentage* = 50.0. Donde el nivel mínimo y máximo determina el intervalo en la los niveles jerárquicos GO y el grupo de fusión que reduce la redundancia mediante la selección de las relaciones jerárquicas en la red de proteínas que comparten genes asociados más representativos.

2.7 Optimización de la biblioteca de proteínas

2.7.1 Modificación de los parámetros utilizados en la identificación y cuantificación de proteínas de la biblioteca espectral

La extensión de proteínas de la biblioteca es una limitante para la cuantificación SWATH-MS debido a lo cual se incrementó el número de proteínas y péptidos identificados en la biblioteca de proteínas tumorales obtenida en este estudio mediante la modificación de la tasa de descubrimiento falso (FDR), unos de los parámetros de búsqueda del Software Protein Pilot™. La identificación de las proteínas de la biblioteca fue evaluada con 95% de confianza y la cuantificación SWATH-MS con un número máximo de péptidos por proteína de 6 y 100; y un XIC de 50 y 75 ppm. Asimismo, la influencia del número de péptidos utilizados para la creación de la curva de calibración del tiempo de retención de las proteínas más intensas fue evaluado con 17, 18, 19 y 40 péptidos.

2.7.2 Extensión de proteínas de la biblioteca espectral

Se realizó un proceso de extensión de proteínas de la biblioteca espectral tomando como referencia el método descrito por Wu *et al.* (2016). En resumen, esta metodología consistió en la unión de la biblioteca de proteínas de este estudio (biblioteca interna) y las proteínas de una biblioteca externa del proteoma humano del repositorio SwathAtlas (10,316 proteínas) obtenida con el mismo instrumento y la misma condición de cromatografía que la biblioteca interna. La biblioteca interna y la extendida se sometieron a un proceso de limpieza que eliminó los péptidos de baja confianza (99%) y los iones de baja intensidad (5). Posteriormente, se realizó una evaluación de la correlación de la intensidad relativa de los fragmentos iónicos (correlación de Spearman, ρ) y el tiempo de retención de las proteínas de ambas bibliotecas. Finalmente, la construcción de la biblioteca extendida se obtuvo con la función `buildSpectraLibPair` del paquete `SwathXtend`. Asimismo, se modificaron y evaluaron el parámetro de corte para los iones de baja intensidad (5 y 2000) en el proceso de limpieza de bibliotecas.

2.7.3 Análisis de enriquecimientos de ontología de genes y cuantificación diferencial utilizando la biblioteca extendida

La cuantificación SWATH-MS y el análisis de enriquecimientos de ontología de genes utilizando la nueva biblioteca extendida fueron analizados con los mismos parámetros utilizados en el análisis de la biblioteca interna.

Capítulo 3. Resultados

3.1 Estandarización del método de LC –MS/MS

La cobertura de detección de datos se evaluó con dos tiempos de elución de LC diferentes. En el análisis a 120 min se observó un flujo de datos interrumpido. En contraste, en el análisis a 150 min de elución se observó una continua detección de datos con una disminución de péptidos detectados después de los 110 min hasta una completa ausencia de detección a los 150 min. Lo que sugiere que a 150 min de elución la cobertura de adquisición de datos es mayor que 120 min (Figura 9)

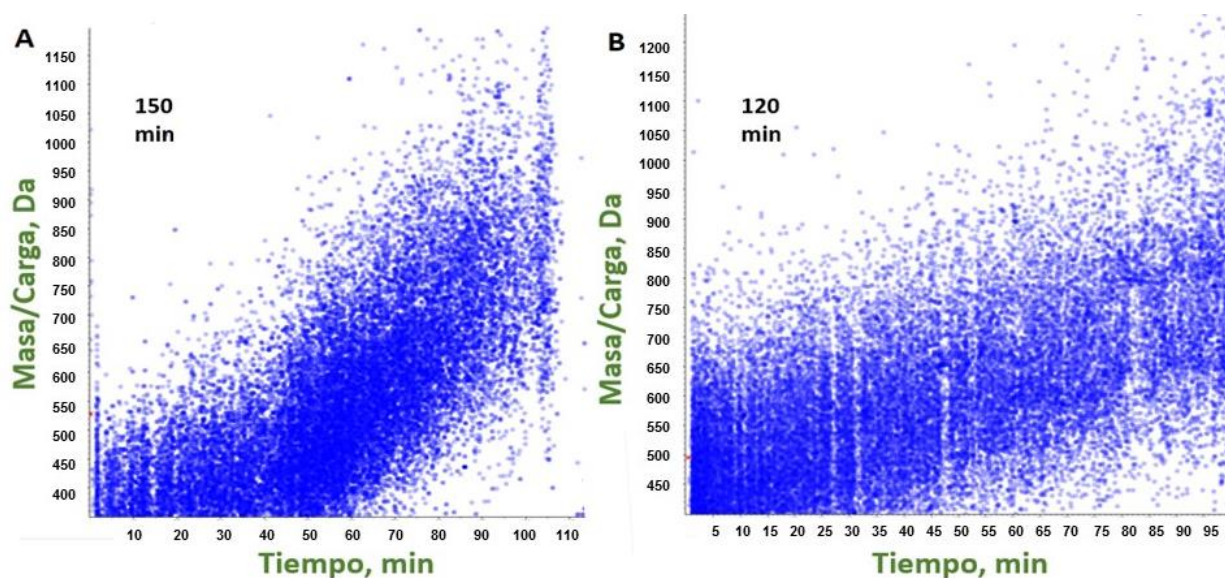


Figura 9. Comparación de los gradientes de elución de péptidos tripticos BCAEC analizados por LC-MS/MS. MS2 generados con los gradientes de elución de (A) 150 y (B) 120 min de tiempo de corrida, respectivamente. Cada círculo azul representa un MS2 proveniente de un péptido precursor

Con respecto a la etapa de cuantificación, se evaluaron dos cantidades de péptidos (0.5 y 1 ug). Las muestras fueron inyectadas por triplicado en el análisis LC-MS/MS. La evaluación de la respuesta de detección reveló que existe un cambio evidente en la magnitud de respuesta respecto a la concentración de las muestras (Figura 10).

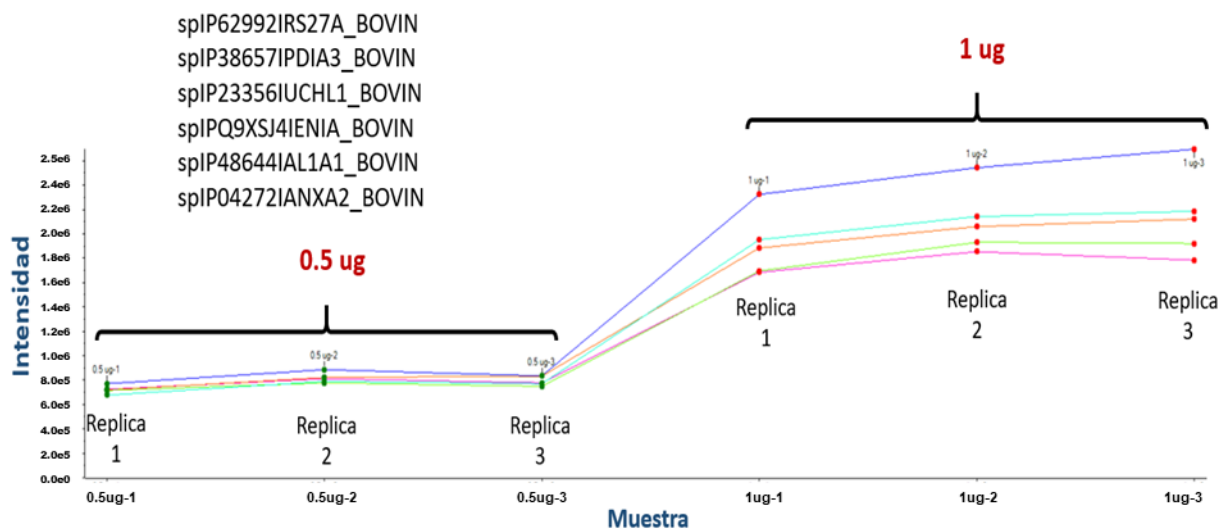


Figura 10. Comparación de las intensidades detectadas de 6 proteínas representativas al inyectar 0.5 ug y 1 ug de los extractos proteínicos de BCAEC en el LC-MS/MS. Las muestras fueron inyectadas por triplicado

Para adquirir datos de alta calidad y generar datos cuantitativos de mayor especificidad en la cuantificación proteómica tipo SWATH – MS se requiere el uso de ventanas de masa de tamaño variable. Trabajos previos han demostrado que la reducción del tamaño de la ventana puede proporcionar una mayor especificidad, menor interferencia y mejorar la cantidad de datos cuantitativos extraídos para aumentar finalmente la cobertura de detección. Las ventanas más pequeñas son deseables, especialmente en las regiones densas m/z donde se encuentran muchos MS1. En consecuencia, se evaluaron dos tamaños mínimos de ventana variable (10 y 14 Da) para el péptido VNFTVDQIR de la proteína Factor de Elongación 2 en muestras de retina de ratón. En la Figura 11 se muestra claramente que la disminución de la ventana variable de 10 a 14 Da mejora considerablemente la señal de los péptidos detectados.

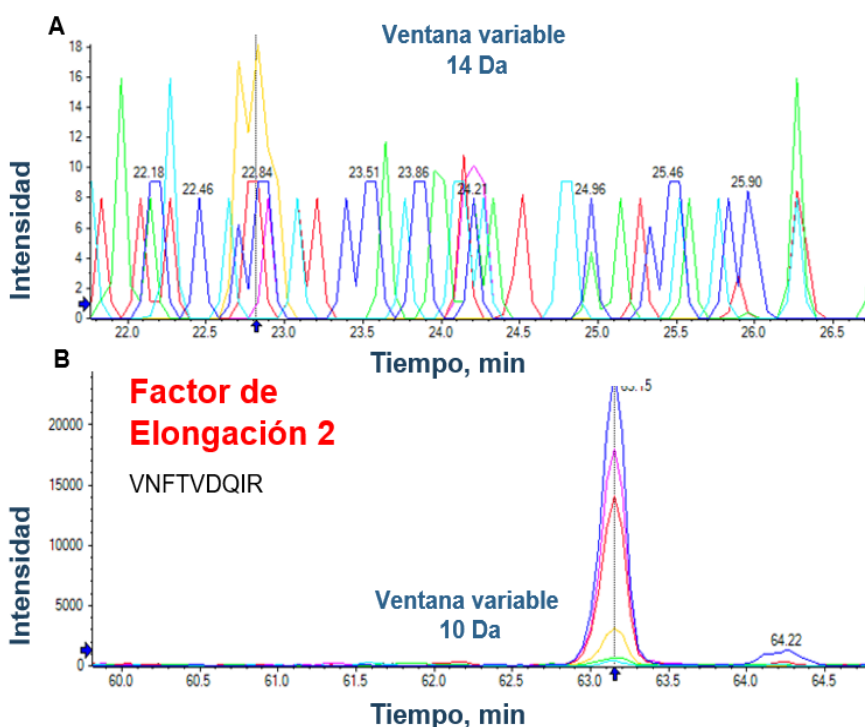


Figura 11. Comparación de la señal de la abundancia del péptido VNFTVDQIR de la proteína Factor de Elongación 2 utilizando tamaños de ventana variables (intervalo de masa específico) de (A) 14 Y (B) 10 Da en la cuantificación SWATH-MS. Se muestra que la disminución del tamaño de ventana mejora considerablemente la señal de los péptidos.

3.2 Análisis de enriquecimientos de ontología de genes GO de la biblioteca de proteínas interna y extendida

Se identificaron un total de 1,355 y 10,204 proteínas utilizando la biblioteca interna y extendida respectivamente con 95% de confianza. De las cuales se cuantificaron 1,221 y 2,168 proteínas (biblioteca interna y extendida, respectivamente) con un número máximo de péptidos por proteína 100, un XIC de 75 ppm y 13 péptidos en la curva de calibración del tiempo de retención. Se realizó un análisis de enriquecimiento de ontología de genes por componente celular de ambos análisis. En la gráfica se observa la agrupación de las proteínas por colores (clusters) en función a su localización en la estructura celular. El análisis ClueGO realiza automáticamente el cálculo de la significancia de los términos y grupos mediante la prueba Exacta Fisher y su corrección mediante Bonferroni. El tamaño de los nodos representa la significancia estadística de cada término y los círculos aislados representan los términos no asociados. La

red creada representa los términos como nodos que están vinculados en función de un nivel de puntaje kappa de 0.4.

Las proteínas identificadas con la biblioteca interna se clasifican según sus ubicaciones subcelulares en 12 grupos de los cuales el 21.43% de sus proteínas se encuentran asociados al lumen de vesículas citoplasmática, 14.29% a la mitocondria, 11.9% al lumen del retículo endoplasmático, 11.9% a la actina del citoesqueleto, 9.52% a núcleo, 9.52% a los ribosomas citosólicos, 7.14% a las vacuolas, entre otras (Figura 12). En cambio, las proteínas identificadas con la biblioteca extendida se clasificaron en 17 grupos de los cuales se destacaron el lumen de las vesículas secretoras (14.67%), la mitocondria (12%), el citoesqueleto (12%), el lumen del retículo endoplasmático (10.67%), los cromosomas (8%), vesículas citoplasmáticas (6.67%), complejo spliceosomal (1.33%), el aparato de Golgi (1.33) entre otras (Figura 13).

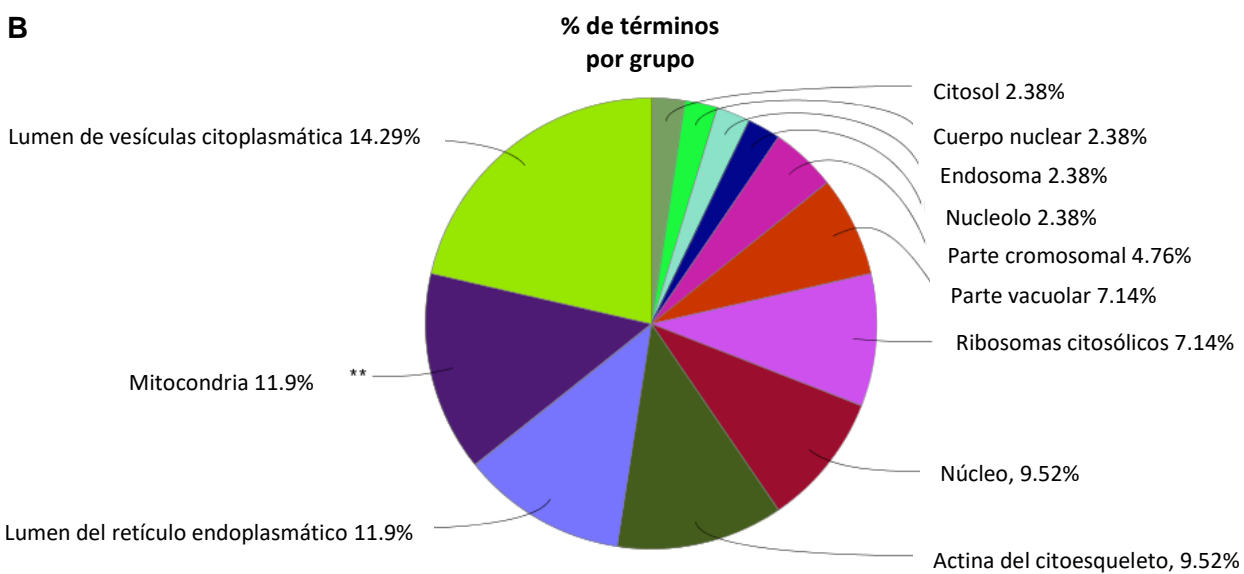
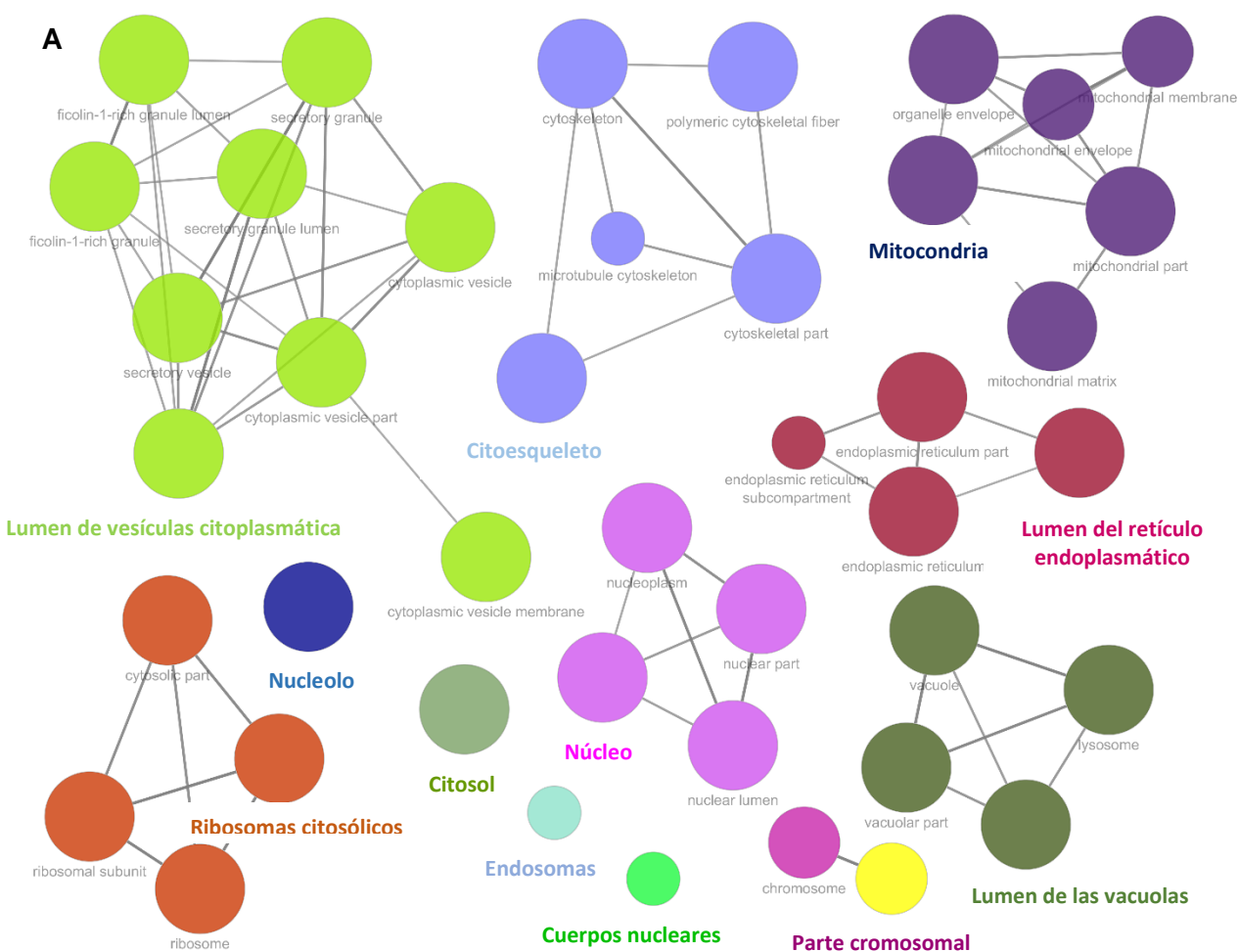


Figura 12. Representación del análisis de enriquecimiento de ontología de genes en función al término Componente Celular generado con la aplicación ClueGO del software Cytoescape para (A) la biblioteca interna en forma de redes de proteínas y (B) en forma de diagrama circular con sus respectivos porcentajes por grupo.

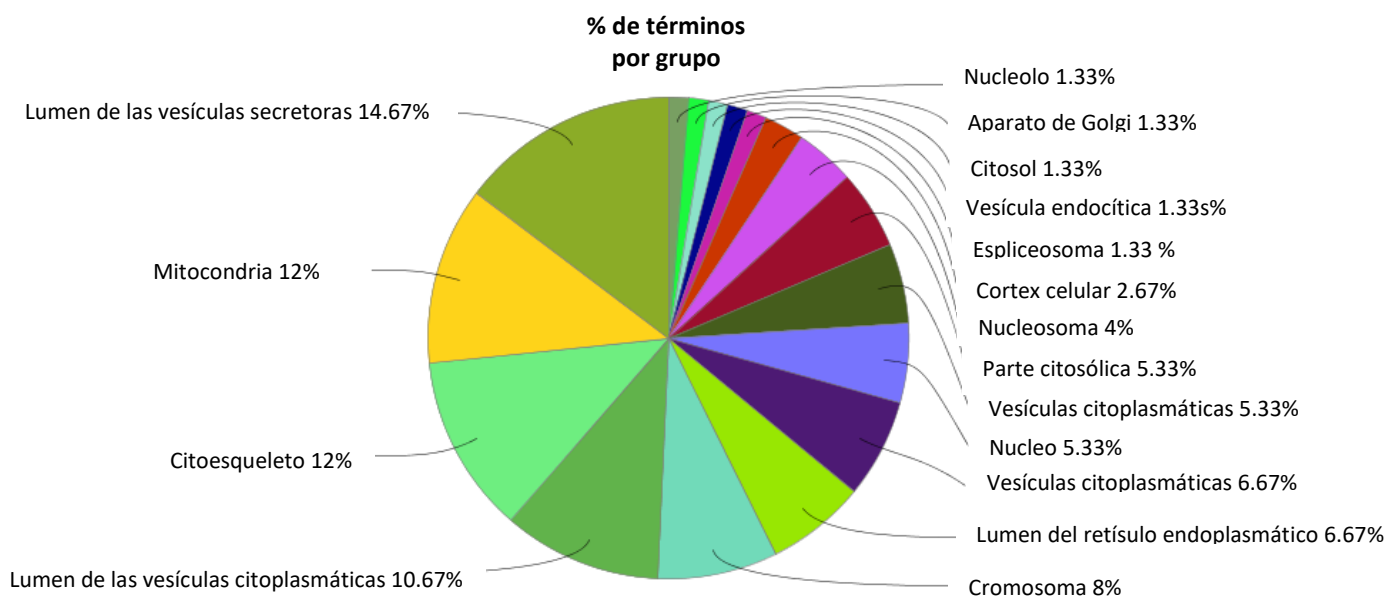


Figura 13. Diagrama circular que muestra el porcentaje de proteínas de la biblioteca extendida asociadas a un Componte Celular.

3.3 Análisis SWATH-MS utilizando la biblioteca interna y extendida

Los datos obtenidos de la cuantificación SWATH-MS se utilizaron para calcular la magnitud o proporción de cambio (FC, de sus siglas en inglés Fold Change) de la abundancia de las proteínas del TMM en relación a las del TBM (TMM/TBM). Solo las proteínas con un FC menor al 0.66 y mayor al 1.5 (es decir, una disminución del 60% o un aumento de 50% de abundancia) y un valor de significancia $p < 0.05$, fueron consideradas como aquellas que tuvieron un cambio significativo.

El análisis estadístico t de Student no pareado de dos colas (varianza desigual, Welch) reveló que 524 y 523 proteínas de TMM que cambiaron significativamente utilizando la biblioteca interna y extendida, respectivamente. De las cuales 34 proteínas disminuyeron y 490 aumentaron su abundancia en relación con los TBM con la biblioteca interna. Por otro lado, 86 proteínas disminuyeron y 432 aumentaron su abundancia utilizando la biblioteca extendida como se observa en las gráficas de volcán (Figura14).

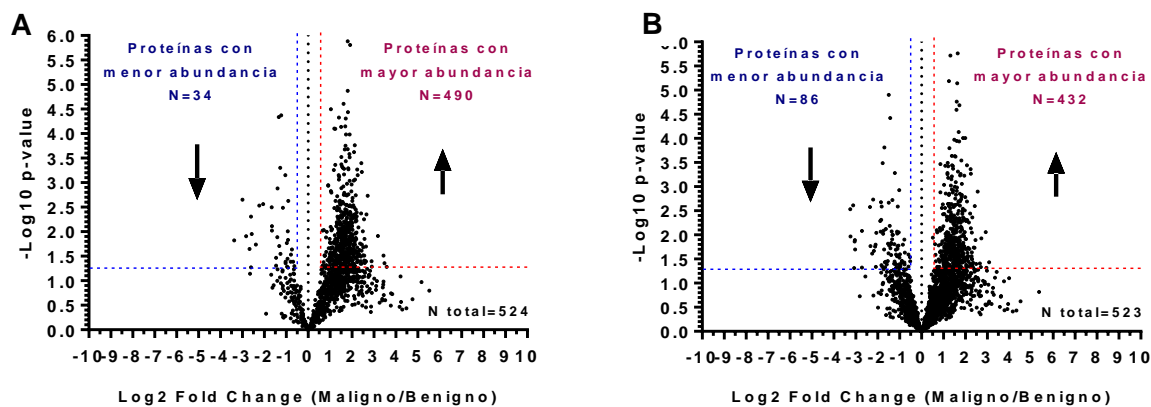


Figura 14. Gráfica de volcán que muestra las proteínas que modificaron su abundancia en los TMM (n=6) con relación a los TBM (n=10) en la cuantificación SWATH-MS utilizando la biblioteca interna (A) y extendida (B). Clasificación basada en análisis a nivel de proteína con un FC >1.5 y <0.6; y un pval <0.05

El PCA permitió la visualización bidimensional de la variabilidad o similitud entre las muestras de TBM y TMM en función a la abundancia de proteínas representado por la separación horizontal de las muestras. El PC1 y PC2 separan a los grupos en función a la abundancia de todas las proteínas de cada paciente. La variabilidad que muestra el PCA con los componentes principales PC1 y PC2 fue de 60.6% y 53.6% para la biblioteca interna y extendida, respectivamente. El análisis sugiere un patrón del comportamiento diferencial de las muestras de TBM y TMM, exceptuando al paciente 14 (TMM) (Figura 15).

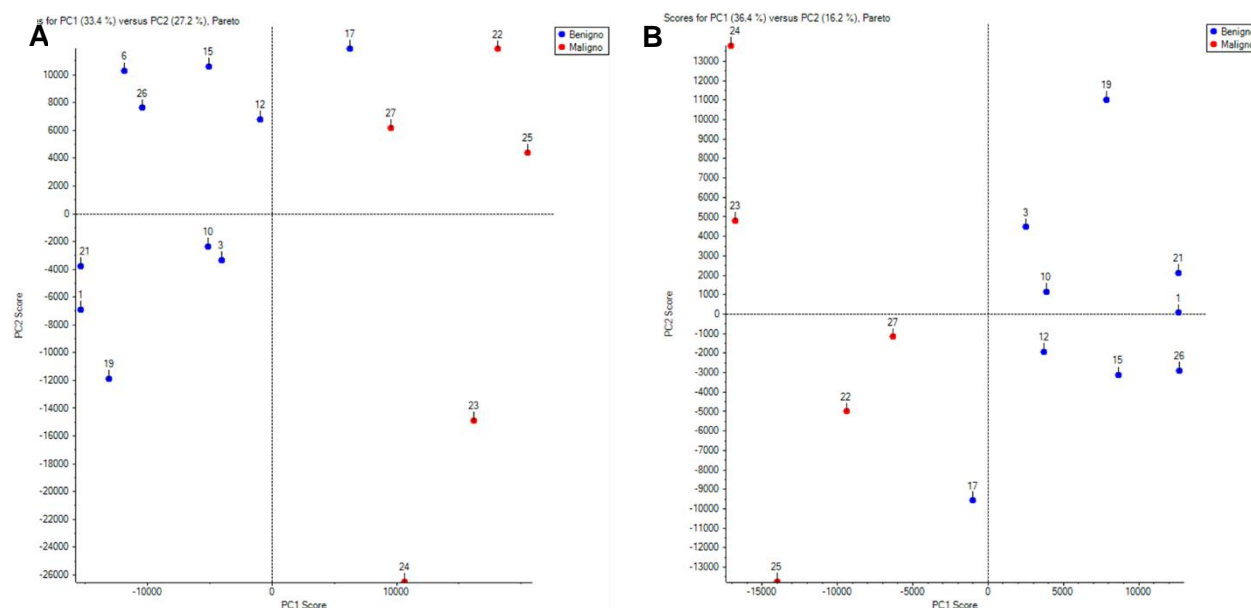


Figura 15. Análisis de componentes principales utilizando la biblioteca interna (A) y extendida (B). El paciente 14 fue excluido del análisis.

El PCVG es mostrado en la figura y muestra la correlación entre las abundancias relativas de las proteínas analizadas de las muestras de tumor benigno y maligno. El PCVG sugirió una disminución en la abundancia de las proteínas Mimecan, Decorina, Lumican, Serotransferrina y Cadena de colágeno alfa-1 (XIV) en los TMM con respecto a los TMB. Asimismo, sugirió un incremento en la abundancia de las proteínas Glicer aldehído-3-fosfato deshidrogenasa, Triosa fosfato isomerasa, Factor de elongación 1-alfa 1, Prosaposina y Piruvato quinasa isozimas M1 / M2 Tabla 1 (Figura 16) (Tabla 4).

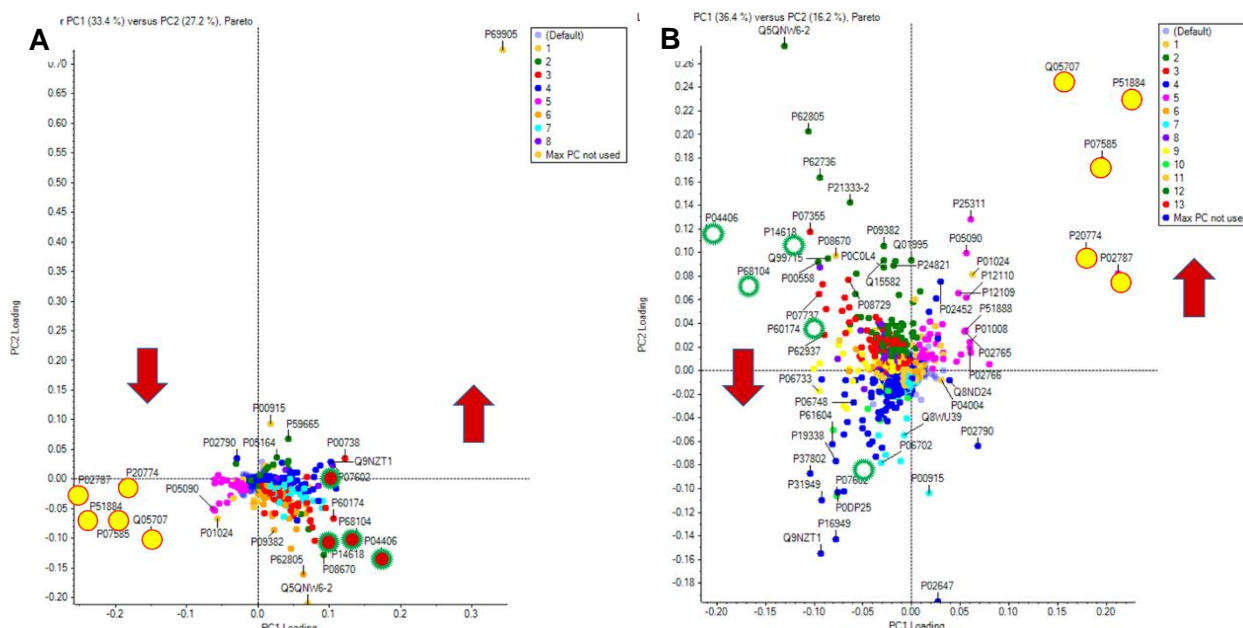


Figura 16. Análisis de componentes principales por agrupación de variables PCVG para las proteínas cuantificadas con la biblioteca interna (A) y extendida (B). Las proteínas relacionadas con la sangre y el paciente 14 fueron excluidos del análisis.

Tabla 4. Lista de proteínas diferencialmente abundantes en los TMM que contribuyen en mayor medida a la separación de los grupos

Proteínas	UniProt ID	TMM/TMB	
		Fold change(T/N)*	Abundancia
Mimecan	P20774	0.1154	↓
Decorina	P07585	0.2167	↓
Lumican	P51884	0.3230	↓
Serotransferrina	P02787	0.3555	↓
Cadena de colágeno alfa-1 (XIV)	Q05707	0.1214	↓
Glicer aldehído-3-fosfato deshidrogenasa	P04406	4.9047	↑
Triosa fosfato isomerasa	P60174	2.6716	↑
Factor de elongación 1-alfa 1	P68104	3.6964	↑
Isoenzimas M1/ M2 de piruvato cinasa	P14618	3.6311	↑
Prosaposina	P07602	2.7415	↑

*Fold change es la proporción de la abundancia media de los tumores maligno/benigno. Los péptidos correspondientes a cada proteína fueron cuantificados con el software SWAHT-MS.

Posteriormente, se graficó la abundancia diferencial de las proteínas responsables de la separación de los grupos, propuestas por el análisis PCVG tomando los valores de la cuantificación SWATH-MS (Figura 17 y 18). Y posteriormente las proteínas Mimecan, Decorina y Lumican fueron validadas por Western blot como se observa en la Figura 19.

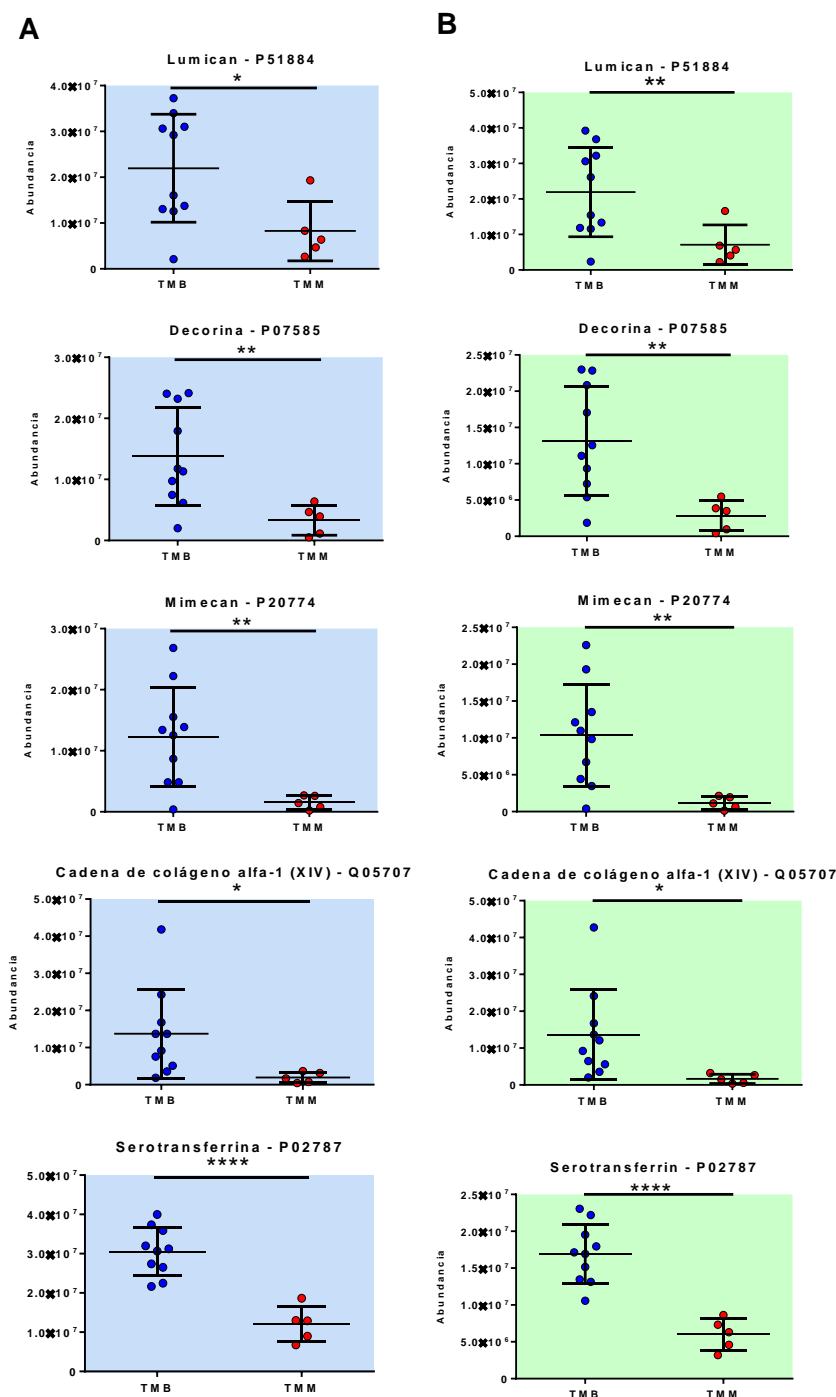


Figura 17. Comparación de las proteínas con abundancia disminuida en los TMM responsables de la separación de grupos sugeridas por el análisis PCVG para (A) la biblioteca interna y (B) extendida. Análisis no paramétrico con corrección T-test Welch. TMB, tumor de mama benigno. TMM, tumor de mama maligno. Los valores son media y rango percentil (10-90). * p < 0.05.

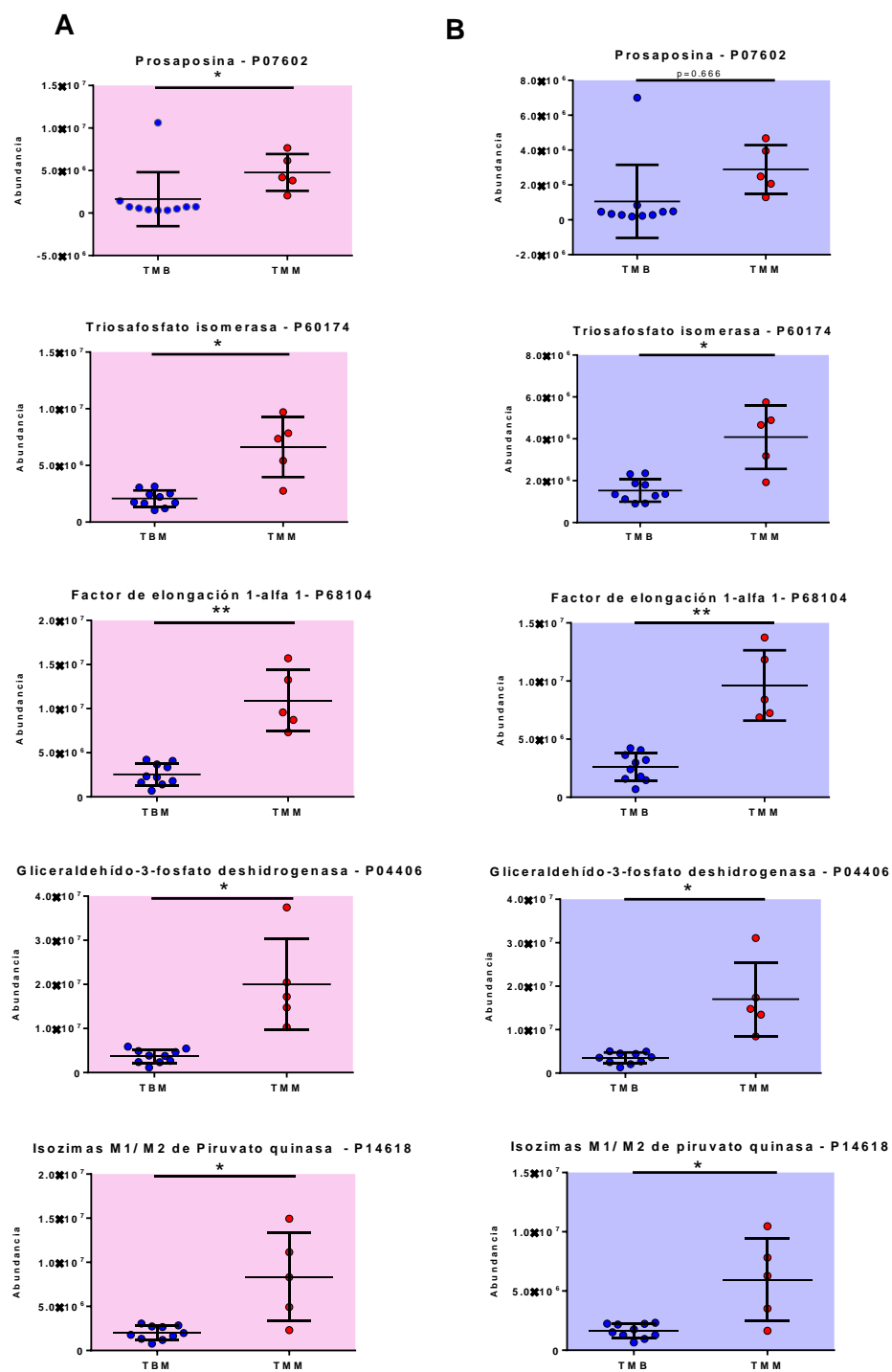


Figura 18. Comparación de las proteínas con abundancia aumentada en los TMM responsables de la separación de grupos sugeridas por el análisis PCVG para la (A) biblioteca interna y (B) extendida. Análisis no paramétrico con corrección T-test Welch. TMB, tumor de mama benigno. TMM, tumor de mama maligno. Los valores son media y rango percentil (10-90). * $p < 0.05$.

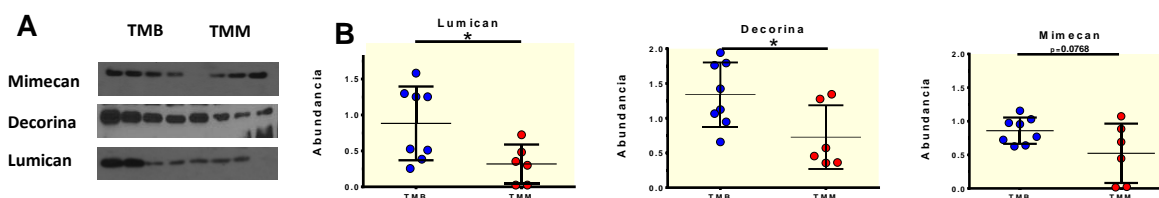


Figura 19. (A) Abundancia de las proteínas Mimecan, Decorina y Lumican analizadas por Western blot para las muestras de tumor maligno (TMB) y tumor benigno (TMM). Los valores son media y rango percentil (10-90). * $p < 0.05$

3.4 Identificación de vías de señalización asociadas a las proteínas diferencialmente abundantes en el TMM

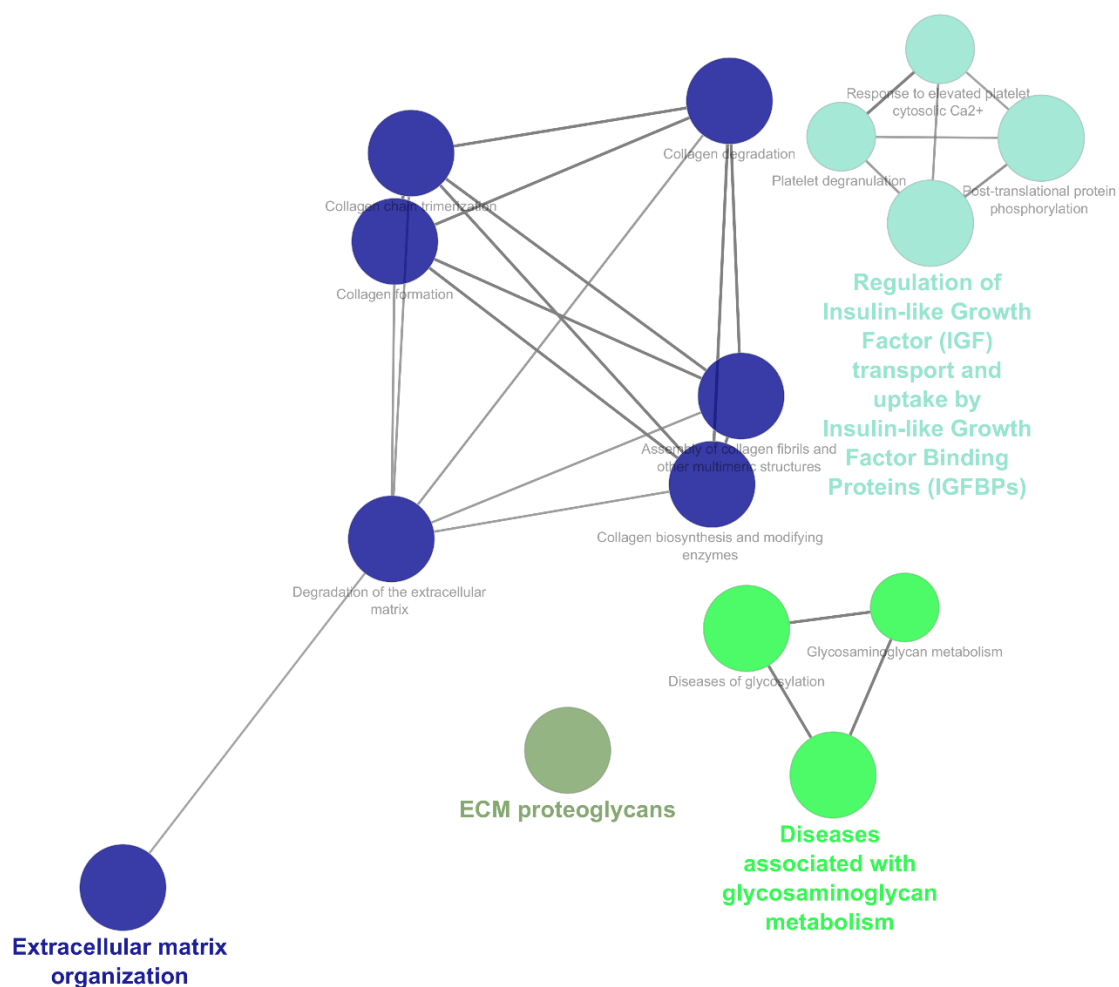
Para profundizar la interpretación de los datos, se realizó un análisis de enriquecimiento de genes a las proteínas cuya abundancia cambió significativamente en los TMM utilizando aplicación ClueGo del software Cytoscape. El análisis utilizando la biblioteca interna, indicó que el 48.67% de las 34 proteínas que disminuyeron en los TMM estaban asociadas con la organización de la matriz extracelular; el 26.67% con la regulación del transporte y captación del factor de crecimiento de tipo insulina (IGF) por las proteínas de unión al factor de crecimiento de tipo insulina (IGFBP); el 20% con desórdenes del metabolismo de los glicosaminoglicanos y el 6.67% con los proteoglicanos de la matriz extracelular (6.67%) (Figura 20).

Por otro lado, el análisis de enriquecimiento de ontología de genes utilizando la biblioteca extendida, correlacionó a las 86 proteínas con menor abundancia en los TMM con términos de REACTOME que incluyen la CHST6 defectuosa que causa la distrofia corneal macular tipo 1 (36.71%); organización de la matriz extracelular (32.14%); regulación del transporte y captación del factor de crecimiento de tipo insulina (IGF) por las proteínas de unión al factor de crecimiento de tipo insulina (IGFBP) (28.57 %) y metabolismo de las lipoproteínas (3.57%) (Figura 21).

Como se observa en la Figura 22, las proteínas con mayor abundancia en los TMM utilizando la biblioteca interna (490 proteínas) se asociaron principalmente con procesos biológicos como la regulación de la expresión de glicoproteína SLIT y roundabout ROBO (36.52%), señalización de receptores ROBO (36.52%), metabolismo de RNA (15.65%), traducción (3.48%), desgranulación de neutrófilos (3.48%), transporte mediado por vesículas (1.74%), entre otros. En cambio, las proteínas con mayor abundancia utilizando la biblioteca extendida se relacionaron a la regulación de la expresión de glicoproteína SLIT y

roundabout ROBO (49.33%), metabolismo de RNA (18.67%), enfermedades infecciosas (16%). Transición G2/M (6.67%), degranulación de neutrófilos (6.67%), el ciclo del ácido cítrico (TCA) y la cadena de transporte de electrones (1.33%), y el mantenimiento y biogénesis de los organelos (1.33%) (Figura 23).

A



B

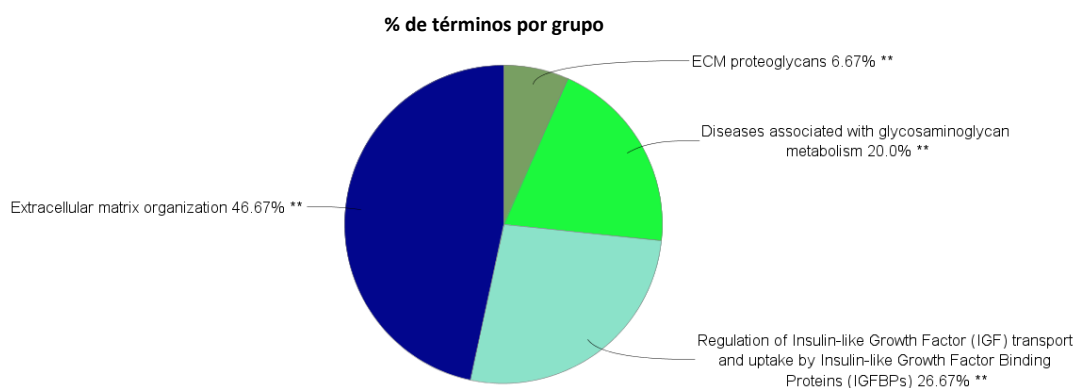


Figura 20. Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca interna. (A) Representación de las vías de señalización relacionadas a las proteínas con menor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con menor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. $p < 0.05$.

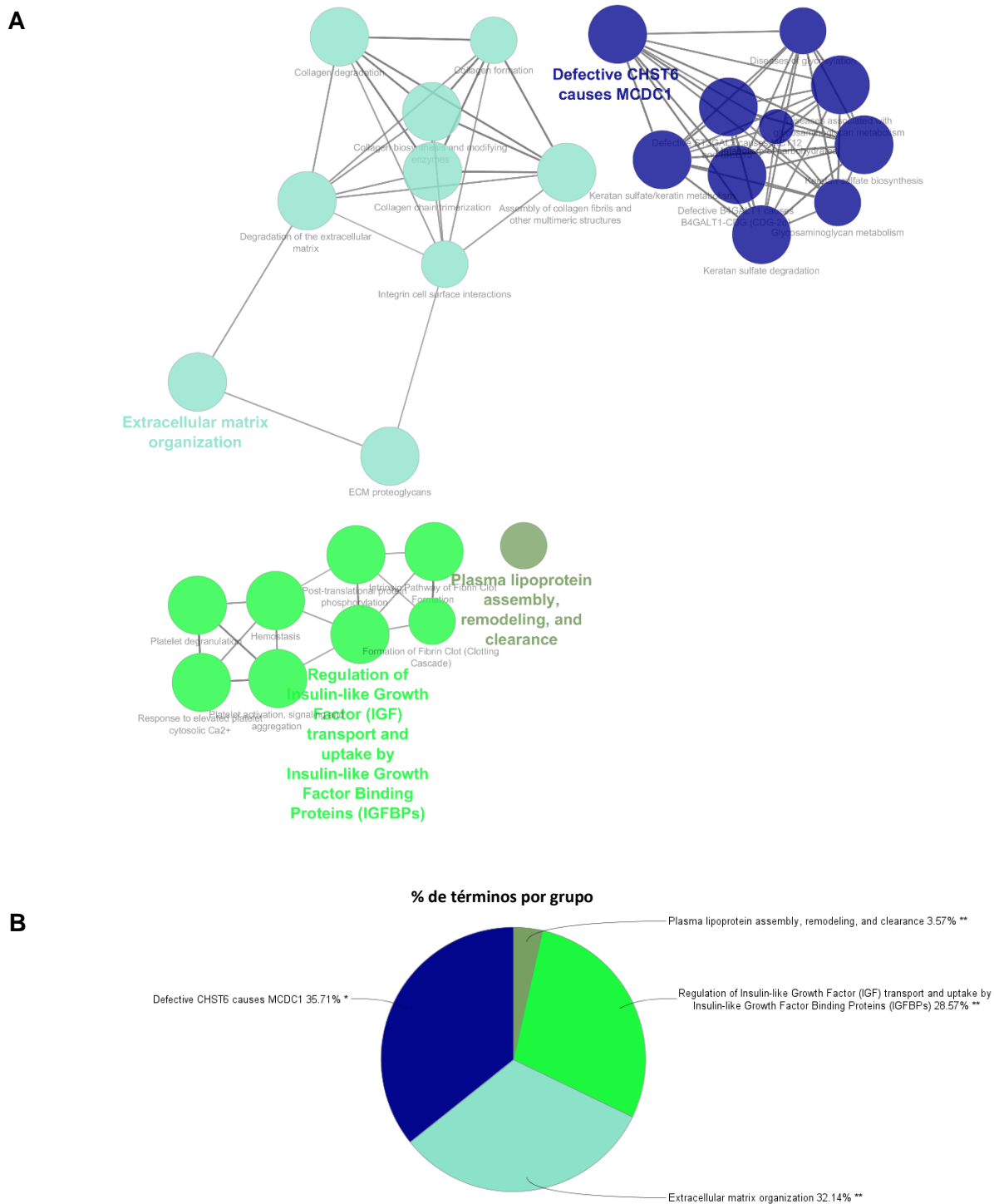
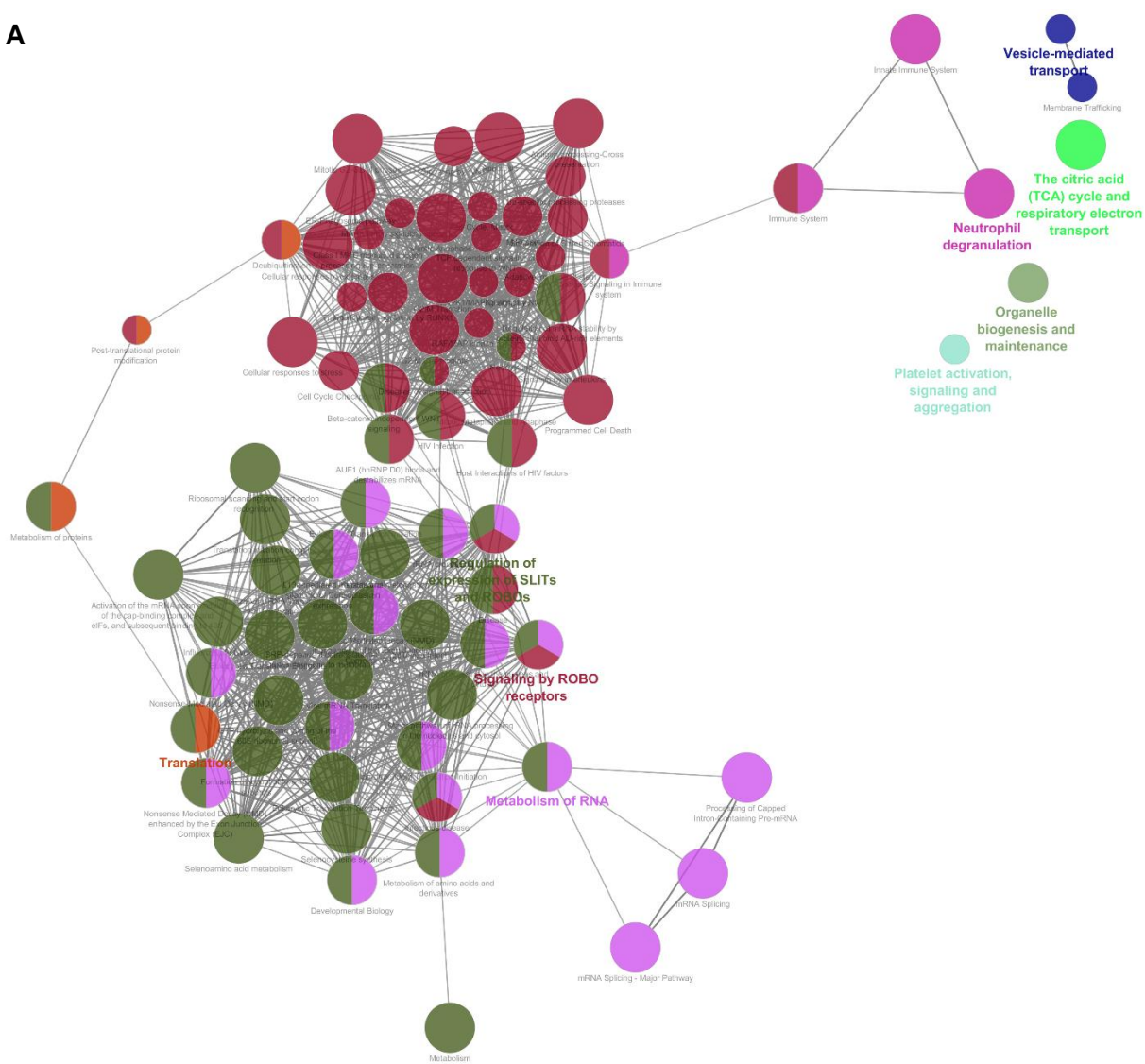


Figura 21. Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca extendida. (A) Representación de las vías de señalización relacionadas a las proteínas con menor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con menor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. $p < 0.05$.

A



B

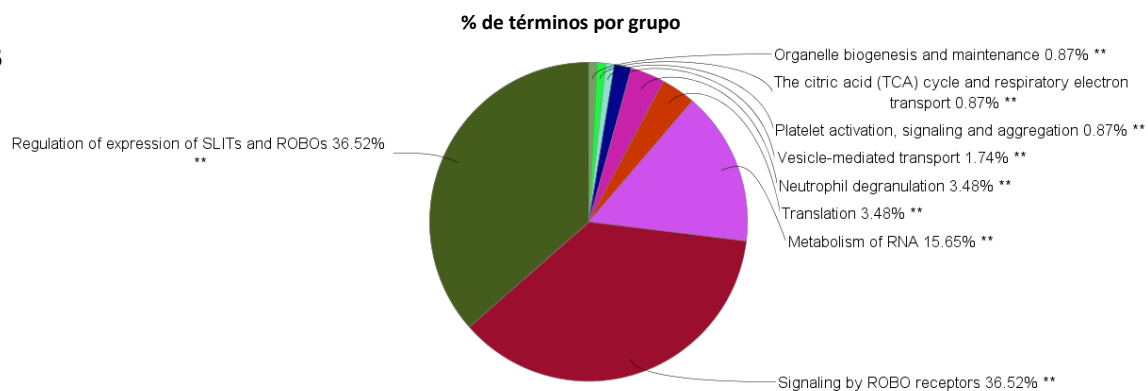
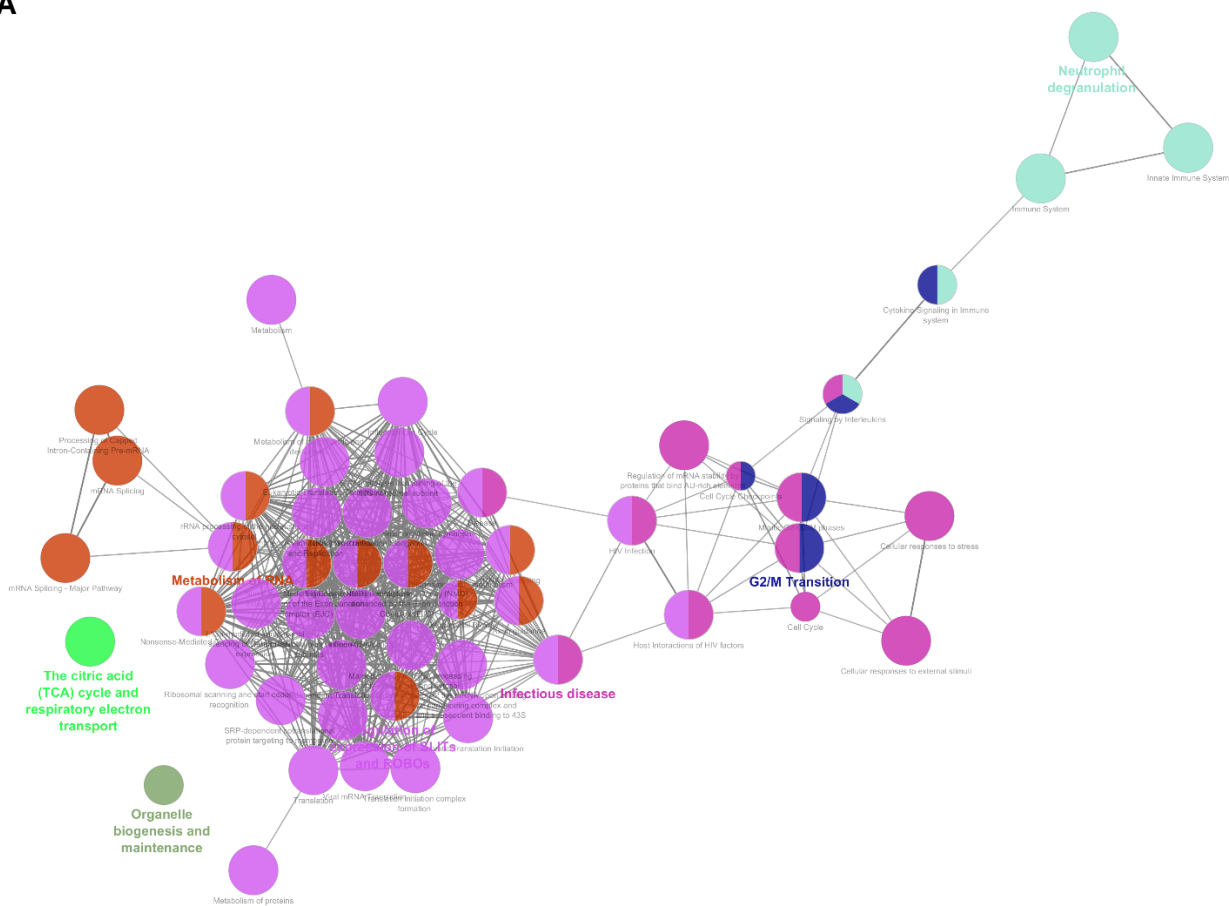


Figura 22. Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca interna. (A) Representación de las vías de señalización relacionadas a las proteínas con mayor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con mayor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. $p < 0.05$.

A



B



Figura 23. Análisis de enriquecimiento de ontología de genes utilizando las proteínas de la biblioteca extendida. (A) Representación de las vías de señalización relacionadas a las proteínas con mayor abundancia en los TMM (B) Diagrama pastel del porcentaje de las proteínas con mayor abundancia en los TMM asociadas a un término de REACTOME PATHWAY. $p < 0.05$.

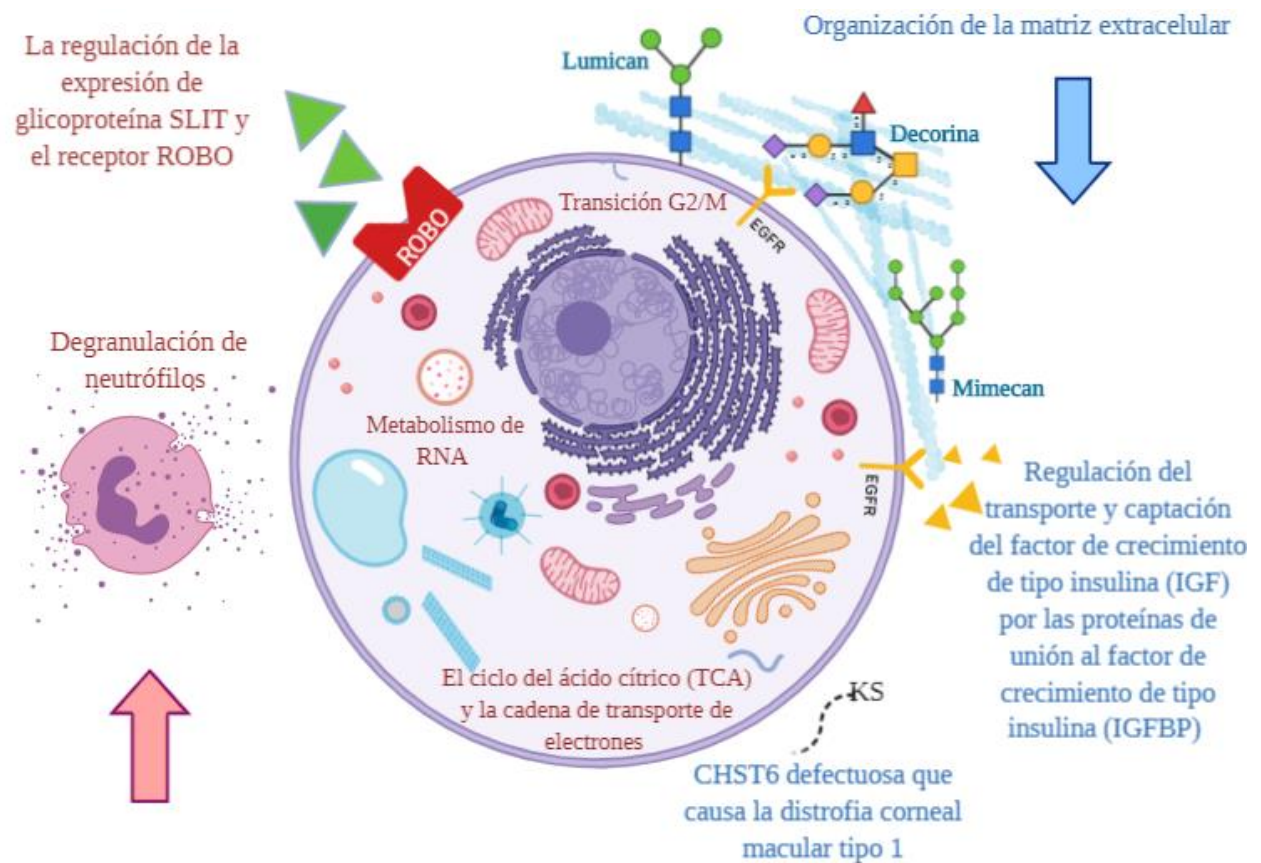


Figura 24. Esquema representativo que resume los hallazgos de este proyecto. En esta figura se muestran las vías de señalización alteradas en los TMM. Se sugiere una mayor activación en las vías en color rojo e inhibición en las de azul.

3.5 Creación de una biblioteca extendida y modificación de los parámetros para la optimización de la cuantificación de proteínas.

La biblioteca de proteínas interna de este estudio y una biblioteca externa del proteoma humano del repositorio SwathAtlas se sometieron a un proceso de limpieza antes de su combinación. Se eliminaron los péptidos de baja confianza (99%) y los iones de baja intensidad (5) de ambas bibliotecas y se evaluó su correlación. Para la correlación respecto al tiempo de retención (RT) se obtuvo un valor de 0.94 y para la correlación de la intensidad relativa de iones (RII), 0.94. Cabe destacar que dichos valores se encuentran dentro del intervalo de valores considerados para una buena correlación y calidad entre bibliotecas

($R^2 > 0.8$ y $RII > 0.6$). Se identificaron 10,204 proteínas de las cuales 413 pertenecían a la biblioteca interna (Figura 25).

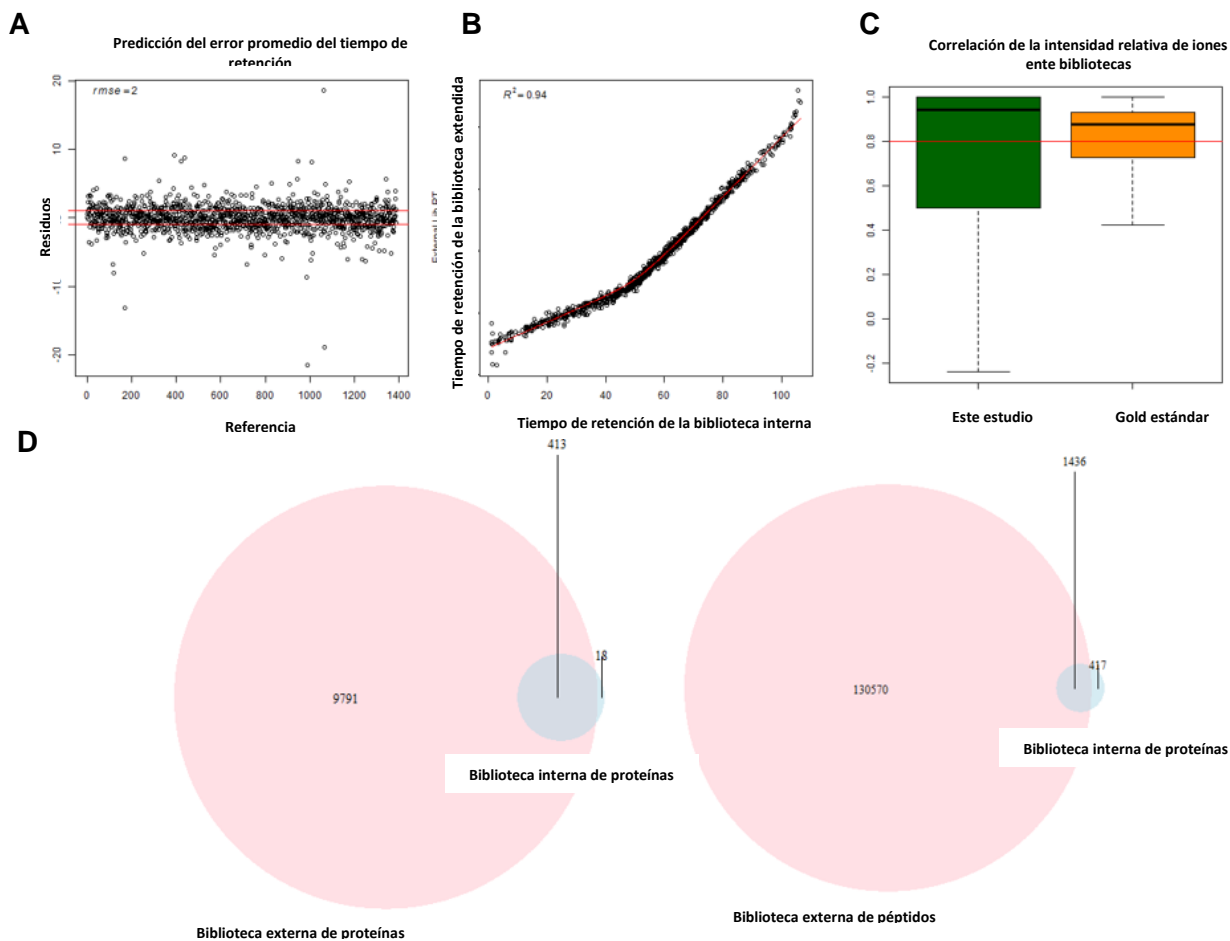


Figura 25. Creación de una biblioteca de proteínas extendida. (A) Evaluación de la compatibilidad entre la biblioteca interna y extendida. (B) Correlación del tiempo de retención, $R^2 = 0.95$. (C) Predicción del error promedio del tiempo de retención (RT) Correlación de la intensidad relativa de iones, (D) Correlación de Spearman, $RII = 0.65$ Comparación de la cantidad de proteínas y péptidos obtenidos entre la biblioteca interna y la extendida

Los resultados de la cuantificación SWATH usando la biblioteca extendida (2168 proteínas) duplicaron la cobertura de péptidos y proteínas cuantificadas con la biblioteca interna (1221 proteínas) con un $FDR < 0.01$. En la Figura 26A se observa que las proteínas cuantificadas disminuyen a medida que el FDR se vuelve más estricto debido a lo cual la confianza en las identificaciones de las proteínas cuantificadas aumenta a medida que aumenta el límite de pase del FDR. Asimismo, la gráfica de barras con las distribuciones FDR indican que las identificaciones de los péptidos de alta confianza (por ejemplo, $FDR < 0.01$, gris oscuro) aumentaron a medida que el filtrado FDR se vuelve más estricto (Figura 26 y 27).

La consistencia de la cuantificación se midió por el coeficiente de variación (CV) donde se observa que a medida que aumenta el valor de corte de FDR, el CV disminuye (Figura 28).

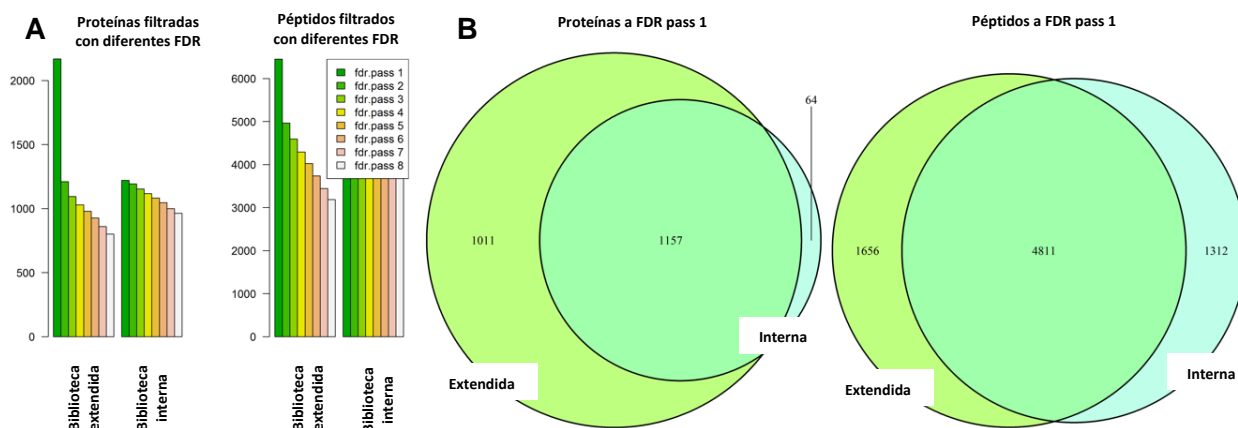


Figura 26. Resultados de la cuantificación SWATH con la biblioteca interna y extendida (A) Cantidad de proteínas y péptidos con diferentes FDR y (B) FDR<0.01.

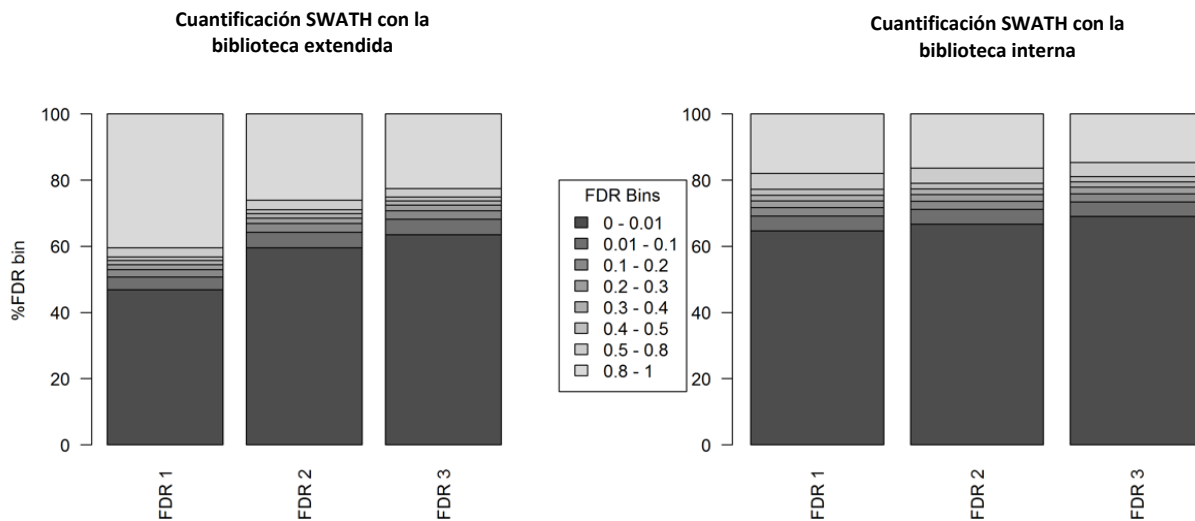


Figura 27. Porcentaje de número de proteínas cuantificadas con diferentes FDR. El color gris oscuro indica el porcentaje de péptidos identificados con alta confianza (FDR 0-0.1).

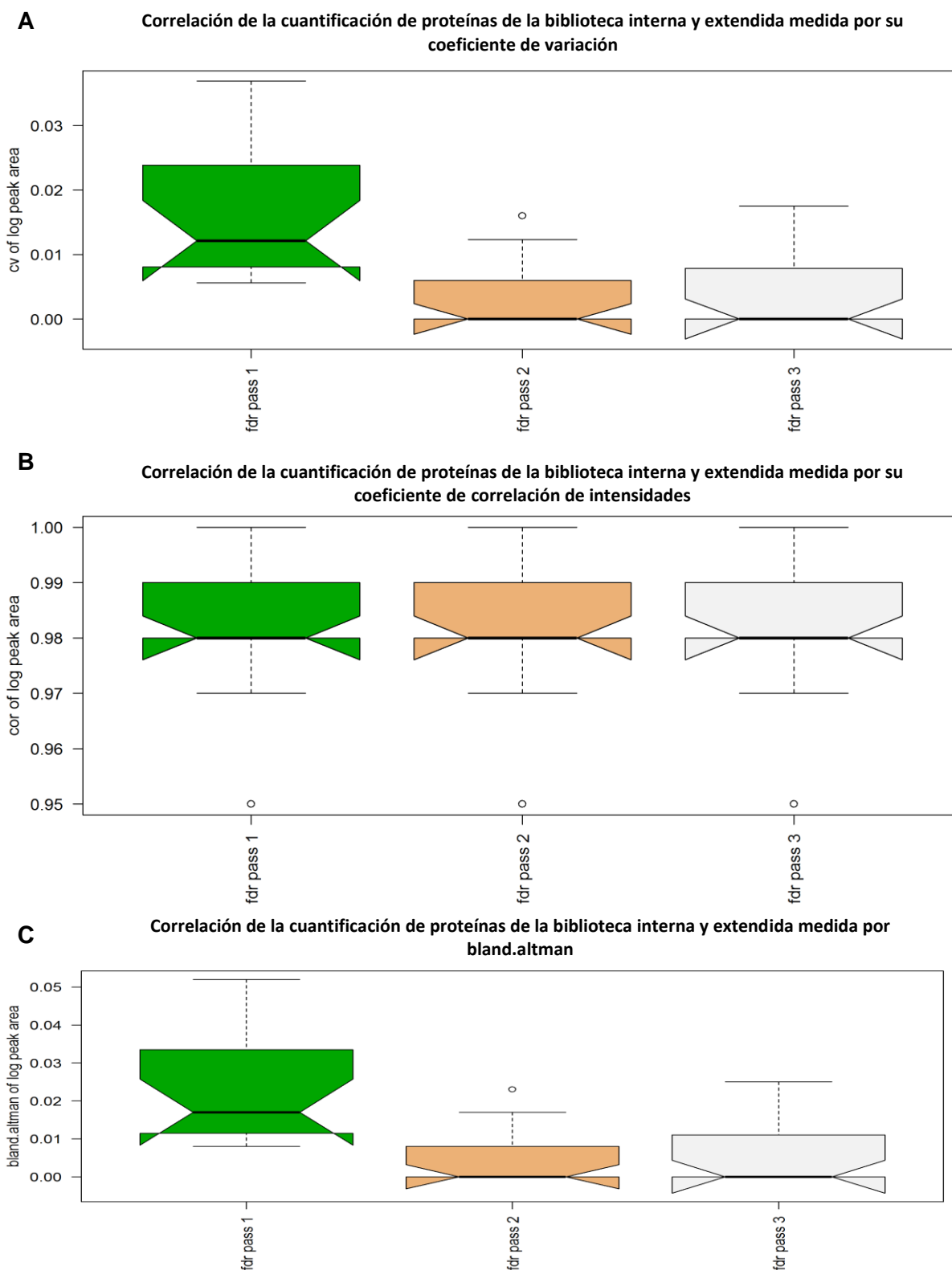


Figura 28. Evaluación de la cuantificación de proteínas utilizando la biblioteca interna y extendida. (A) Medición de la consistencia de la cuantificación entre el área bajo la curva (abundancia) de las proteínas con 3 FDR en función a su (B) coeficiente de variación, (C) coeficiente de correlación y Bland-Altman.

Capítulo 4. Discusión

En México, los TMM ocupan el primer lugar de incidencia de las neoplasias malignas en las mujeres y hasta hoy, poco se sabe de su biología molecular. En consecuencia, existen dificultades en el diagnóstico de los distintos tipos de tumores y su tratamiento, ya que se entiende poco de los mecanismos que ocurren en las células malignas y que diferencian a estas de las células no malignas. Con el objetivo de profundizar en la biología del cáncer y comprender a nivel molecular los procesos que ocurren en las células malignas, este análisis prospectivo evaluó el proteoma de los TMM y los TMB en mujeres mexicanas utilizando la LC-MS/MS y softwares especializados.

Con la finalidad de reducir la dimensionalidad de los datos obtenidos de la LC-MS/MS, se realizó un análisis multivariante que incluyó un PCA y PCVG. Estos análisis son herramientas esenciales para la interpretación de los datos proteómicos y la identificación de biomarcadores (Long, 2013). Tomando en cuenta los resultados del PCA y PCVG de nuestro estudio, se sugieren a las proteínas Mimecan, Lumican y Decorina como potenciales biomarcadores de malignidad de los tumores de mama. Estas tres proteínas pertenecen a la familia de los pequeños proteoglicanos ricos en leucina (SRLP). Los SRLP son componentes de la matriz extracelular y participan en la organización estructural, proliferación, angiogénesis y migración de las células malignas en varios tipos de tumores (Appunni *et al.*, 2019).

Como se muestra en la Tabla 4, las proteínas Mimecan, Lumican y Decorina presentaron una abundancia disminuida en los TMM. Mimecan, también conocido como Osteoglicina o Factor Osteoinductivo, es un SRLP involucrado en la homeostasis de la matriz extracelular, proliferación y procesos inflamatorios en la mama. Se ha reportado una abundancia disminuida de Mimecan en TMM (Hu *et al.*, 2018a, 2018b; Moriggi *et al.*, 2018; Deckx *et al.*, 2016). Además, la disminución de su abundancia en los TMM, en mujeres de Países Bajos, se ha asociado a la recurrencia de varios tipos de TMM (Braakman, 2017). Estos datos coinciden con nuestros resultados a pesar de haber sido evaluados en mujeres de etnias diferentes a la de México, lo que sugiere un papel clave de esta proteína en células malignas.

La abundancia baja de Lumican en el estroma tumoral se correlaciona a la rápida progresión de los TM (Karamanou *et al.*, 2017), lo que coincide con los resultados de este estudio donde la disminución de la abundancia de Lumican demostró ser responsable de la diferenciación molecular entre los TMM y TMB. Sin embargo, se ha reportado que la abundancia de Lumican podría estar variando en función al tipo de TMM (Nikitovic *et al.*, 2014) como lo sugiere el estudio de Gámez-Pozo *et al.* (2017). En dicho estudio los autores demostraron que la abundancia de Lumican aumentó en los TMM ER+ en comparación con los

TMM triple negativo. Además, se ha demostrado que una mayor abundancia de Lumican en la matriz extracelular de los TMM se correlaciona con mayor agresividad tumoral y menor expresión de los receptores de estrógenos en los TMM (Ishiwata *et al.*, 2007). En nuestro estudio se analizaron tumores HER2Neu, Triple Negativo y Luminal B en el grupo de los TMM, de manera que se infiere una menor abundancia de Lumican para estos tipos de TMM.

Decorina compete con el factor de crecimiento epidérmico (EGF) por la unión al EGFR, siendo que ambas proteínas se sobreexpresan en diversos tipos de cáncer. Al unirse al EGFR, Decorina se dimeriza y posteriormente sufre una internalización y degradación. El interactoma de Decorina abarca procesos relacionados a la unión a la matriz, señalización protumorigénicas y proangiogénicas y procesos catabólicos intracelulares lo que sugiere su potencial función como agente oncosupresor (Neill & Iozzo, 2016). Sin embargo, los datos publicados en la literatura respecto a la abundancia de Decorina en los TMM son contradictorios. Una mayor abundancia de Decorina se ha relacionado con la resistencia a la quimioterapia de los TMM, pero también a la reducción del desarrollo y progresión del tumor (Bi & Yang, 2013). Se ha demostrado que la expresión del gen que codifica a Decorina aumenta en la matriz extracelular, pero disminuye dentro de las células de mama malignas. Además, las células MCF7 transducidas con Decorina exhiben una cohesión celular y tasa de mitosis disminuida, así como características apoptóticas aumentadas (Boström *et al.*, 2013). Nuestros resultados infieren que la abundancia disminuida de Decorina favorece la malignidad celular, por consiguiente, apoya su papel como agente oncosupresor en los TMM.

La variación de las abundancias de Mimecan, Lumican y Decorina entre los TMB y TMM, que se reportaron en este estudio, demostraron que las alteraciones en la estructura y de la matriz extracelular desempeñan un rol importante en la malignidad tumoral. Para extrapolar nuestros resultados, realizamos un análisis de enriquecimientos de ontología de genes. Este análisis consiste en relacionar un conjunto de proteínas en función a un término que incluye procesos biológicos, componentes celulares o funciones moleculares.

Las tres principales vías de señalización que se relacionaron con las proteínas que disminuyeron en los TMM, en el análisis de enriquecimiento de ontología de genes, fueron la CHST6 defectuosa que causa la distrofia corneal macular tipo1, organización de la matriz extracelular, y regulación del transporte y captación del IGF por las IGFBP.

En efecto, se ha reportado que las mutaciones en el gen que codifica la proteína CHST6 ocasionan que el queratán sulfato (KS) no sea sulfatado. Debido a que la sulfatación de los carbohidratos influye en la hidratación corneal, el KS no sulfatado pierde su hidrofilia, lo que contribuye a la opacidad corneal y a la pérdida de la agudeza visual (Wang *et al.*, 2017). Por otro lado, se ha reportado que los TMM son la fuente más común de metástasis en el globo ocular (Antosz *et al.*, 2014). La baja abundancia de las proteínas que están relacionadas a la vía CHST6 defectuosa que causa la distrofia corneal macular tipo 1 infieren que este proceso está vinculado a la malignidad tumoral de las mujeres diagnosticadas con TMM en este estudio.

Se ha reportado que la matriz extracelular cumple un papel crucial en la progresión, invasión y metástasis de los TMM (Jena & Janjanam, 2018; Walker *et al.*, 2018; Insua & Oskarsson, 2016; Oskarsson, 2013). Los resultados del análisis de enriquecimiento de ontología de genes demostraron que el 32.14 % de proteínas que disminuyeron su abundancia en los TMM, están relacionadas a la matriz extracelular lo que apoya los resultados obtenidos por el análisis multivariante. Asimismo, se ha demostrado en este estudio que el 28.57% de proteínas que disminuyeron su abundancia en los TMM están vinculadas a la regulación del transporte y captación del IGF por las IGFBP, lo que sugiere que hay una disminución en la actividad de esta vía de señalización. El IGF desempeña un papel importante en la función y el mantenimiento de la glándula mamaria y es un punto de convergencia para las principales vías de señalización implicadas en el crecimiento del cáncer de mama. La unión del IGF al receptor IGF desencadena una cascada de señalización que conduce a eventos proliferativos y antiapoptóticos (Christopoulos *et al.*, 2015). Por lo tanto, este evento sería una de los principales procesos celulares alterados en los TMM.

Por otro lado, los principales términos de la base de datos REACTOME que se relacionaron con las proteínas que aumentaron en los TMM, fueron la regulación de la expresión de la glicoproteína SLIT y ROBO, enfermedades infecciosas, metabolismo de RNA, degranulación de neutrófilos, el TCA y la cadena de transporte de electrones.

Varios estudios sugieren que la vía de señalización SLIT y ROBO está relacionada a procesos antitumorales debido a que se ha encontrado una menor abundancia de estas proteínas en los TMM. Además, una mayor presencia de SLIT y ROBO se ha relacionado con una mayor tasa de supervivencia en pacientes con varios tipos de cáncer, mientras que la baja presencia de estas proteínas se asocia a un aumento de la metástasis (Gara *et al.*, 2015). En contraste, en nuestro estudio se reveló una mayor abundancia de estas proteínas en los TMM.

El 22.67% de proteínas que aumentaron su abundancia en los TMM se relacionaron a los términos enfermedades infecciosas y la degranulación de neutrófilos de la base de datos REACTOME. En respuesta a una infección, los neutrófilos abandonan la circulación y migran hacia el foco inflamatorio. Los neutrófilos contienen varios subconjuntos de gránulos que se movilizan para fusionarse con la membrana celular lo que resulta en la exocitosis. Se ha descrito en la literatura que la liberación de estos gránulos ocasiona la activación celular asociada con la progresión tumoral. Los neutrófilos pueden promover la proliferación de células tumorales, estimulando la angiogénesis y la remodelación de la matriz y desactivando la inmunidad antitumoral dependiente de las células T (Mollinedo,2019). Además, se ha reportado un incremento de los neutrófilos en sangre de pacientes con TMM avanzadas o agresivos. Por lo tanto, se infiere que las vías de señalización asociadas a la degranulación de neutrófilos estarían participando activamente en procesos relacionados a los tumores HER2Neu, Triple Negativo y Luminal B de mujeres mexicanas.

En el mismo sentido, el 16% de proteínas que aumentaron su abundancia en los TMM sugieren un incremento en la activación del metabolismo del RNA y solo un 1.33% sugiere una mayor activación en el TCA. Nuestros resultados sugieren un mayor proceso metabólico oxidativo en células malignas. A pesar de que se ha reportado que las células cancerosas eluden el ciclo TCA y utilizan principalmente la glucólisis anaeróbica (Efecto Warburg), la evidencia emergente demuestra que ciertas células cancerosas como los TMM si estarían utilizando esta vía aerobia (Anderson *et al.*, 2018). Se ha investigado el metabolismo de la glutamina a través del ciclo TCA en los TMM triple negativos y se ha demostrado que la anaplerosis de glutamina es crítica para la supervivencia y el crecimiento de este tipo de tumores (Febg, 2018).

En resumen, con base en los resultados de este estudio, se observa que existe una diferencia a nivel proteómico entre los TBM y TMM de mujeres mexicanas. No obstante, nuestro estudio presenta algunas limitaciones que incluyen el análisis de un pequeño número de muestras en ambos grupos (TBM y TMM), lo que limita la extrapolación de nuestros resultados en la población general. Además, la terapia neoadyuvante de paclitaxel que fue administrada a los pacientes con TMM es un factor que podría estar interfiriendo en los resultados obtenidos.

Capítulo 5. Conclusiones

Los resultados de este estudio sugieren que existe una diferencia molecular —a nivel de proteínas— entre los TBM y TMM en mujeres mexicanas. Se propone por primera vez a un grupo de proteínas (mimecan, decorina y lumican) que en conjunto podrían ser indicadores de malignidad en tumores de mama. La función que resulta de la interacción entre estas proteínas podría explicar el proceso fisiológico o biológico de los TMM y su diferenciación con los TBM.

Basados en nuestros análisis de enriquecimiento de ontología de genes se infiere la presencia de una mayor activación en los procesos biológicos principalmente relacionados a la regulación de la expresión de la glicoproteína SLIT y ROBO, desgranulación de neutrófilos, del metabolismo de RNA y TCA y el transporte respiratorio de electrones, entre otros. Por otra parte, se sugiere una menor actividad de los procesos celulares relacionados al CHST6 defectuosa que causa la distrofia corneal macular tipo 1, organización de la matriz extracelular y regulación del transporte y captación del factor de crecimiento de tipo insulínico (IGF) por las proteínas de unión al factor de crecimiento de tipo insulínico (IGFBP) en los TMM.

Por ello, concluimos que la LC-MS/MS sin etiquetado junto con la cuantificación SWATH-MS es una herramienta útil para determinar diferencias entre los proteomas de los TMB y TMM.

Perspectivas futuras: Conociendo los mecanismos moleculares involucrados en cada tipo de tumor se podría desarrollar nuevos tratamientos que no se enfoquen en la supresión o activación de una sola molécula de interés sino de un conjunto de biomoléculas cuyo cambio en abundancia reflejaría el panorama global de la patología. Se requiere de un estudio con un mayor número de participantes para tomar en cuenta la heterogeneidad de la población mexicana.

Literatura citada

- Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347–355. doi:10.1038/nature19949
- American Cancer Society. (2013). Breast Cancer Facts & Figures 2012-2014. Breast Cancer Facts & Figures, 1–44. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2013-2014.pdf>.
- Amodei, D., Egertson, J., MacLean, B. X., Johnson, R., Merrihew, G. E., Keller, A., ... & MacCoss, M. J. (2019). Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *Journal of The American Society for Mass Spectrometry*, 30(4), 669-684.
- Anderson, N. M., Mucka, P., Kern, J. G., & Feng, H. (2018). The emerging role and targetability of the TCA cycle in cancer metabolism. *Protein & cell*, 9(2), 216-237.
- Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., & Muddiman, D. C. (2011). Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Analytical chemistry*, 83(13), 5442-5446.
- Antosz, Z. S., Walocha, J., Poręba, R., & Sioma-Markowska, U. (2014). Sudden loss of vision due to breast cancer metastasis to the eyeball. *Neuroendocrinology Letters*, 35(4).
- Appunni, S., Anand, V., Khandelwal, M., Gupta, N., Rubens, M., & Sharma, A. (2019). Small Leucine Rich Proteoglycans (decorin, biglycan and lumican) in cancer. *Clinica Chimica Acta*
- Atkinson, W. L. (2018). Epidemiology and Prevention of Measles. *Dermatologic Clinics*, 13(3), 553–559. [https://doi.org/10.1016/s0733-8635\(18\)30061-5](https://doi.org/10.1016/s0733-8635(18)30061-5)
- Baba, A. I., & Cătoi, C. (2007). Comparative oncology. Bucharest: Publishing House of the Romanian Academy.
- Bi, X. L., & Yang, W. (2013). Biological functions of decorin in cancer. *Chinese journal of cancer*, 32(5), 266.
- Boström, P., Sainio, A., Kakko, T., Savontaus, M., Söderström, M., & Järveläinen, H. (2013). Localization of decorin gene expression in normal human breast tissue and in benign and malignant tumors of the human breast. *Histochemistry and cell biology*, 139(1), 161-171.
- Braakman, R. B., Stingl, C., Tilanus-Linthorst, M. M., van Deurzen, C. H., Timmermans, M. A., Smid, M., ... & Umar, A. (2017). Proteomic characterization of microdissected breast tissue environment provides a protein-level overview of malignant transformation. *Proteomics*, 17(5), 1600213.

- Calderón-Garcidueñas, A. L., Parás-Barrientos, F. U., Cárdenas-Ibarra, L., González-Guerrero, J. F., Villarreal-Ríos, E., Staines-Boone, T., & Barrera-Saldaña, H. A. (2000). Risk factors of breast cancer in Mexican women. *Salud pública de México*, 42, 26-33.
- Calderón-González, K. G., Rustarazo, M. L. V., Labra-Barrios, M. L., Bazán-Méndez, C. I., Tavera-Tapia, A., Herrera-Aguirre, M. E., ... & León-Ávila, G. (2015). Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry. *Journal of proteomics*, 124, 50-78.
- Cassidy, M. R., & Méndez, J. E. (2015). Brca1 and brca2 in breast cancer and ovarian cancer. *Surgical Oncology a Practical and Comprehensive Approach*, 141–161. https://doi.org/10.1007/978-1-4939-1423-4_6
- Castrezana Campos, M. D. R. (2017). Geografía del cáncer de mama en México. *Investigaciones geográficas*, (93), 0-0.
- Chlebowski, R. T., Chen, Z., Anderson, G. L., Rohan, T., Aragaki, A., Lane, D., ... Prentice, R. (2005). Ethnicity and breast cancer: Factors influencing differences in incidence and outcome. *Journal of the National Cancer Institute*, 97(6), 439–447. <https://doi.org/10.1093/jnci/dji064>
- Choe, R., Konecky, S. D., Corlu, A., Lee, K., Durduran, T., Busch, D. R., ... & DeMichele, A. (2009). Differentiation of benign and malignant breast tumors by in-vivo three-dimensional parallel-plate diffuse optical tomography. *Journal of biomedical optics*, 14(2), 024020.
- Christopoulos, P. F., Msaouel, P., & Koutsilieris, M. (2015). The role of the insulin-like growth factor-1 system in breast cancer. *Molecular cancer*, 14(1), 43.
- Colleoni, M., Sun, Z., Price, K. N., Karlsson, P., Forbes, J. F., Thürlimann, B., ... & Goldhirsch, A. (2016). Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the international breast cancer study group trials I to V. *Journal of Clinical Oncology*, 34(9), 927.
- Cooper G.M. (2000). *The Cell: A Molecular Approach*. 2nd edition. Sunderland (MA): Sinauer Associates. The Development and Causes of Cancer. Recuperado de: <https://www.ncbi.nlm.nih.gov/books/NBK9963/>
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research*, 5(10), 2929.
- Daly, B., & Olopade, O. I. (2015). Race, ethnicity, and the diagnosis of breast cancer. *Jama*, 313(2), 141-142.
- De Gruttola, V. G., Clax, P., DeMets, D. L., Downing, G. J., Ellenberg, S. S., Friedman, L., ... & Zeger, S. L. (2001). Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. *Controlled clinical trials*, 22(5), 485-502. Deckx, S., Heymans, S., & Papageorgiou, A. P. (2016). The diverse functions of osteoglycin: a deceitful dwarf, or a master regulator of disease?. *The FASEB Journal*, 30(8), 2651-2661.

- Dias, M. H., Kitano, E. S., Zelanis, A., & Iwai, L. K. (2016). Proteomics and drug discovery in cancer. *Drug Discovery Today*, 21(2), 264–277. <https://doi.org/10.1016/j.drudis.2015.10.004>
- Drukteinis, J. S., Mooney, B. P., Flowers, C. I., & Gatenby, R. A. (2013). Beyond mammography: new frontiers in breast cancer screening. *The American journal of medicine*, 126(6), 472-479.
- Fan, S., Yuan, R. Q., Ma, Y. X., Meng, Q., Goldberg, I. D., & Rosen, E. M. (2001). Mutant BRCA1 genes antagonize phenotype of wild-type BRCA1. *Oncogene*, 20(57), 8215–8235. <https://doi.org/10.1038/sj.onc.1205033>
- FDA-NIH Biomarker Working Group (2017). BEST (Biomarkers, Endpoints, and other Tools). September 25, (Md), 55. <https://doi.org/10.1164/rccm.201301-0153OC>
- Fejerman, L., John, E. M., Huntsman, S., Beckman, K., Choudhry, S., Perez-Stable, E., ... & Ziv, E. (2008). Genetic ancestry and risk of breast cancer among US Latinas. *Cancer research*, 68(23), 9723-9728.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., ... & Liu, B. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, 5(2), 77-106.
- Fernández, V. (2016). Desarrollo de herramientas bioinformáticas para estudios de proteómica a gran escala de *Candida albicans*.
- Fonseca-Sánchez, M. A., Rodríguez Cuevas, S., Mendoza-Hernández, G., Bautista-Piña, V., Arechaga Ocampo, E., Hidalgo Miranda, A., ... & López-Camarillo, C. (2012). Breast cancer proteomics reveals a positive correlation between glyoxalase 1 expression and high tumor grade. *International journal of oncology*, 41(2), 670-680.
- Gámez-Pozo, A., Trilla-Fuertes, L., Berges-Soria, J., Selevsek, N., López-Vacas, R., Díaz-Almirón, M., ... & Moreno, F. G. (2017). Functional proteomics outlines the complexity of breast cancer molecular subtypes. *Scientific reports*, 7(1), 10100.
- Gara, R. K., Kumari, S., Ganju, A., Yallapu, M. M., Jaggi, M., & Chauhan, S. C. (2015). Slit/Robo pathway: a promising therapeutic target for cancer. *Drug discovery today*, 20(1), 156-164.
- García Marqués, F. J. (2017). Desarrollo de tecnologías proteómicas de última generación aplicadas al estudio del remodelado vascular. Tesis Doctoral. Universidad Autónoma de Madrid, España. 138 pp.
- Ghoncheh, M., Pournamdar, Z., & Salehiniya, H. (2016). Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pac J Cancer Prev*, 17(S3), 43-46.
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., ... & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics*, 11(6), O111-016717.

- Hixson, K. K., Lopez-Ferrer, D., Robinson, E. W., & Paša-Tolić, L. (2017). Proteomics. *Encyclopedia of Spectroscopy and Spectrometry*, 766–773. doi:10.1016/b978-0-12-803224-4.00061-3
- Holleccek, B., Stegmaier, C., Radosa, J. C., Solomayer, E. F., & Brenner, H. (2019). Risk of loco-regional recurrence and distant metastases of patients with invasive breast cancer up to ten years after diagnosis—results from a registry-based study from Germany. *BMC cancer*, 19(1), 520.
- Hu, A., Noble, W. S., & Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*, 5.
- Hu, X., Li, Y. Q., Li, Q. G., Ma, Y. L., Peng, J. J., & Cai, S. J. (2018). Osteoglycin (OGN) reverses epithelial to mesenchymal transition and invasiveness in colorectal cancer via EGFR/Akt pathway. *Journal of Experimental & Clinical Cancer Research*, 37(1), 41.
- Hu, X., Li, Y. Q., Li, Q. G., Ma, Y. L., Peng, J. J., & Cai, S. J. (2018). Osteoglycin-induced VEGF Inhibition Enhances T Lymphocytes Infiltrating in Colorectal Cancer. *EBioMedicine*, 34, 35-45.
- IMSS (2015). Cáncer de mama. Recuperado el 26 de Octubre de <http://www.imss.gob.mx/salud-en-linea/cancer-mama>
- Instituto Nacional de Cáncer (2019) Manual de enfermería oncológica. Recuperado el 25 de setiembre de http://www.msal.gob.ar/images/stories/bes/graficos/0000000011cnt-08-manual%20enfermeria_08-03.pdf
- Insua-Rodríguez, J., & Oskarsson, T. (2016). The extracellular matrix in breast cancer. *Advanced drug delivery reviews*, 97, 41-55.
- International Agency for Research on Cancer (2018). Estimated age-standardized incidence rate in 2018. Recuperado de <http://gco.iarc.fr>.
- Ishiwata, T., Cho, K., Kawahara, K., Yamamoto, T., Fujiwara, Y., Uchida, E., ... & Naito, Z. (2007). Role of lumican in cancer cells and adjacent stromal tissues in human pancreatic cancer. *Oncology reports*, 18(3), 537-543.
- Jena, M. K., & Janjanam, J. (2018). Role of extracellular matrix in breast cancer development: a brief update. *F1000Research*, 7.
- Johns Hopkins Medicine Pathology (2019). Breast Cancer & Breast Pathology. Recuperado el 08 de Agosto de: <https://pathology.jhu.edu/breast/>
- Kabbage, M., Chahed, K., Hamrita, B., Guillier, C. L., Trimeche, M., Remadi, S., ... & Chouchane, L. (2008). Protein alterations in infiltrating ductal carcinomas of the breast as detected by nonequilibrium pH gradient electrophoresis and mass spectrometry. *BioMed Research International*, 2008.
- Kamińska, M., Ciszewski, T., Łopacka-szatan, K., Miotła, P., & Starosławska, E. (2015). Review paper Breast cancer risk factors, 14(3), 196–202. <https://doi.org/10.5114/pm.2015.54346>

- Karamanou, K., Franchi, M., Piperigkou, Z., Perreau, C., Maquart, F. X., Vynios, D. H., & Brézillon, S. (2017). Lumican effectively regulates the estrogen receptors-associated functional properties of breast cancer cells, expression of matrix effectors and epithelial-to-mesenchymal transition. *Scientific reports*, 7, 45138.
- Kukreti, S., Cerussi, A. E., Tanamai, W., Hsiang, D., Tromberg, B. J., & Gratton, E. (2009). Characterization of metabolic differences between benign and malignant tumors: high-spectral-resolution diffuse optical spectroscopy. *Radiology*, 254(1), 277-284.
- Lawrence, R. T., Perez, E. M., Hernández, D., Miller, C. P., Haas, K. M., Irie, H. Y., ... & Villén, J. (2015). The proteomic landscape of triple-negative breast cancer. *Cell reports*, 11(4), 630-644.
- Lehman, C. D., & Schnall, M. D. (2005). Imaging in breast cancer: magnetic resonance imaging. *Breast Cancer Research*, 7(5), 215.
- Levin, Y., Schwarz, E., Wang, L., Leweke, F. M., & Bahn, S. (2007). Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *Journal of separation science*, 30(14), 2198-2203.
- Li, C. I., Malone, K. E., & Daling, J. R. (2002). Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. *Cancer Epidemiology and Prevention Biomarkers*, 11(7), 601-607.
- Clarke, W. (2017). Mass spectrometry in the clinical laboratory: Determining the need and avoiding pitfalls. In *Mass spectrometry for the clinical laboratory* (pp. 1-15). Academic Press.
- Liang, K.-H. (2013). Bioinformatics for biomedical science and clinical applications. *Proteomics*, 83–106. doi:10.1533/9781908818232.83
- Long, F. H. (2013). Multivariate analysis for metabolomics and proteomics data. In *Proteomic and metabolomic approaches to biomarker discovery* (pp. 299-311). Academic Press.
- Lou, Y., Fallah, Y., Yamane, K., & Berg, P. E. (2018). BP1, a potential biomarker for breast cancer prognosis. *Biomarkers in medicine*, 12(5), 535-545.
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular systems biology*, 14(8).
- Ly, D., Forman, D., Ferlay, J., Brinton, L. A., & Cook, M. B. (2013). An international comparison of male and female breast cancer incidence rates. *International journal of cancer*, 132(8), 1918-1926.
- Mollinedo, F. (2019). Neutrophil degranulation, plasticity, and cancer metastasis. *Trends in immunology*. 40(3):228-242 DOI:10.1016/j.it.2019.01.006
- Momenimovahed, Z. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world, 151–164.

- Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., ... & Ortiz-Tello, P. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*, 344(6189), 1280-1285.
- Moriggi, M., Giussani, M., Torretta, E., Capitanio, D., Sandri, M., Leone, R., ... & Gelfi, C. (2018). ECM Remodeling in Breast Cancer with Different Grade: Contribution of 2D-DIGE Proteomics. *Proteomics*, 18(24), 1800278.
- Muñiz Lino, M. A., Palacios-Rodríguez, Y., Rodríguez-Cuevas, S., Bautista-Piña, V., Marchat, L. A., Ruíz-García, E., ... López-Camarillo, C. (2014). Comparative proteomic profiling of triple-negative breast cancer reveals that up-regulation of RhoGDI-2 is associated to the inhibition of caspase 3 and caspase 9. *Journal of Proteomics*, 111, 198–211.
- Nabavizadeh, A., Bayat, M., Kumar, V., Gregory, A., Webb, J., Alizad, A., & Fatemi, M. (2019). Viscoelastic biomarker for differentiation of benign and malignant breast lesion in ultra-low frequency range. *Scientific reports*, 9(1), 5737.
- Nakagomi, H., Hirotsu, Y., Okimoto, K., Sakamoto, I., Amemiya, K., Nakagomi, S., ... Mochizuki, H. (2017). PALB2 mutation in a woman with bilateral breast cancer: A case report. *Molecular and Clinical Oncology*, 6(4), 556–560. <https://doi.org/http://dx.doi.org/10.3892/mco.2017.1189>
- National Breast Cancer Foundation (2019). Breast tumor. Recuperado el 26 de Setiembre de <https://www.nationalbreastcancer.org/breast-tumors>
- Neill, T., Schaefer, L., & Iozzo, R. V. (2016). Decorin as a multivalent therapeutic agent against cancer. *Advanced drug delivery reviews*, 97, 174-185.
- NIH (2019) Diccionario de cáncer. Estados Unidos de América. Recuperado de <https://www.cancer.gov/espanol/publicaciones/diccionario/def/carcinoma-ductal-in-situ>
- NIH (2019) Diccionario de cáncer. Estados Unidos de América. Recuperado de <https://www.cancer.gov/espanol/publicaciones/diccionario/def/carcinoma-lobulillar-in-situ>
- Nikitovic, D., Papoutsidakis, A., Karamanos, N. K., & Tzanakakis, G. N. (2014). Lumican affects tumor cell functions, tumor–ECM interactions, angiogenesis and inflammatory response. *Matrix Biology*, 35, 206-214.
- Nounou, M. I., Elamrawy, F., Ahmed, N., Abdelraouf, K., Goda, S., & Syed-Sha-Qhattal, H. (2015). Breast cancer: Conventional diagnosis and treatment modalities and recent patents and technologies supplementary issue: Targeted therapies in breast cancer treatment. *Breast Cancer: Basic and Clinical Research*, 9, 17–34. <https://doi.org/10.4137/BCBCR.S29420>
- Oskarsson, T. (2013). Extracellular matrix components in breast cancer progression and metastasis. *The Breast*, 22, S66-S72.

- Paci, E. (2002). Mammography and Beyond: Developing Technologies for Early Detection of Breast Cancer. *Breast Cancer Research* (Vol. 4). <https://doi.org/10.1186/bcr429>
- Pando-Robles, R. V., & Lanz-Mendoza, H. (2009). [The significance of proteomics in public health]. *Salud Pública de México*, 51 Suppl 3(1), S386-94. <http://doi.org/10.1590/S0036-36342009000900004>
- Perez-Ortiz, A. C., Ramírez, I., Cruz-López, J. C., Villarreal-Garza, C., Luna-Angulo, A., Lira-Romero, E., ... Estrada-Mena, F. J. (2017). Pharmacogenetics of response to neoadjuvant paclitaxel treatment for locally advanced breast cancer, 8(63), 106454–106467.
- Pleil, J. D., & Isaacs, K. K. (2016). High-resolution mass spectrometry: basic principles for using exact mass and mass defect for discovery analysis of organic molecules in blood, breath, urine and environmental media. *Journal of breath research*, 10(1), 012001.
- Porter, P. L., Lund, M. J., Lin, M. G., Yuan, X., Liff, J. M., Flagg, E. W., ... & Eley, J. W. (2004). Racial differences in the expression of cell cycle–regulatory proteins in breast carcinoma: Study of young African American and white women in Atlanta, Georgia. *Cancer*, 100(12), 2533-2542.
- Rezaul, K., Thumar, J. K., Lundgren, D. H., Eng, J. K., Claffey, K. P., Wilson, L., & Han, D. K. (2010). Differential protein expression profiles in estrogen receptor–positive and–negative breast cancer tissues using label-free quantitative proteomics. *Genes & cancer*, 1(3), 251-271..
- Rosenberger, G., Koh, C. C., Guo, T., Röst, H. L., Kouvonen, P., Collins, B. C., ... & Faini, M. (2014). A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Scientific data*, 1, 140031.
- Sanders, M. E., Dias, E. C., Xu, B. J., Mobley, J. A., Billheimer, D., Roder, H., ... & Caprioli, R. M. (2008). Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. *Journal of proteome research*, 7(4), 1500-1507.
- Shah, T. R., & Misra, A. (2011). Challenges in delivery of therapeutic genomics and proteomics. *Proteomics* 387–427.[doi:10.1016/b978-0-12-384964-9.00008-6](https://doi.org/10.1016/b978-0-12-384964-9.00008-6)
- Sinha, T. (2018). Tumors: benign and malignant. *Cancer. Therapy & Oncology International Journal*, 10(3), 1–3. <https://doi.org/10.19080/CTOIJ.2018.10.555790>
- Siuzdak, G. (2004). An introduction to mass spectrometry ionization: an excerpt from the expanding role of mass spectrometry in biotechnology, (2nd ed.); MCC Press: San Diego, 2005. *Journal of the Association for Laboratory Automation*, 9(2), 50–63.[doi:10.1016/j.jala.2004.01.004](https://doi.org/10.1016/j.jala.2004.01.004)
- Snoj, N., Dinh, P., Bedard, P., & Sotiriou, C. (2010). Molecular biology of breast cancer. *Essential Concepts in Molecular Pathology*, 341–349. <https://doi.org/10.1016/B978-0-12-374418-0.00026-8>
- Song, M. N., Moon, P. G., Lee, J. E., Na, M., Kang, W., Chae, Y. S., ... & Baek, M. C. (2012). Proteomic analysis of breast cancer tissues to identify biomarker candidates by gel-assisted digestion and label-free quantification methods using LC-MS/MS. *Archives of pharmacal research*, 35(10), 1839-1847.

- Sørbye, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., ... & Thorsen, T. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98(19), 10869-10874.
- Streckfus, C. F., Bigler, L. R., & Zwick, M. (2006). The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. *Journal of oral pathology & medicine*, 35(5), 292-300.
- Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., & Geiger, T. (2016). Proteomic maps of breast cancer subtypes. *Nature communications*, 7, 10259.
- Uribe, J. R., Hernández, C. A., Menolascino, F., Rodríguez, J. E., Istúriz, L. M., Márquez, M. E., ... Uribe, J. L. (2010). Clasificación molecular del cáncer de mama y su correlación clínica. *Revista Venezolana de Oncología*, 22(2), 109–116. <https://doi.org/10.4206/cuad.cir.2011.v25n1-10>
- Walker, C., Mojares, E., & del Río Hernández, A. (2018). Role of extracellular matrix in development and cancer progression. *International journal of molecular sciences*, 19(10), 3028.
- Wang, L., Tang, X., Lv, X., Sun, E., Wu, D., Wang, C., & Liu, P. (2017). CHST6 mutation screening and endoplasmic reticulum stress in macular corneal dystrophy. *Oncotarget*, 8(56), 96301.
- Weigelt, B., Geyer, F. C., & Reis-Filho, J. S. (2010). Histological types of breast cancer: how special are they?. *Molecular oncology*, 4(3), 192-208.
- World Cancer Research Fund International (2012). Breast cancer statistic. Recuperado el 10 de marzo de 2019 de: <https://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics> Recuperado el 30/05/18
- World Health Organization (2019). Breast cancer: Early diagnosis and screening. Recuperado en 10 de septiembre de 2019 de: <http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> (Recuperado el 21/05/18).
- WU J, Pascovici D (2019). SwathXtend: SWATH extended library generation and statistical data analysis. R package version 2.6.0.
- Wu, J. X., Pascovici, D., Ignjatovic, V., Song, X., Krisp, C., & Molloy, M. P. (2017). Improving protein detection confidence using SWATH-mass spectrometry with large peptide reference libraries. *PROTEOMICS*, 17(19), 1700174. doi:10.1002/pmic.201700174
- Wu, J. X., Song, X., Pascovici, D., Zaw, T., Care, N., Krisp, C., & Molloy, M. P. (2016). SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Molecular & Cellular Proteomics*, 15(7), 2501-2514.
- Wu, S., Powers, S., Zhu, W., & Hannun, Y. A. (2016). Substantial contribution of extrinsic risk factors to cancer development. *Nature*, 529(7584), 43.

- Wu, S., Zhu, W., Thompson, P., & Hannun, Y. A. (2018). Evaluating intrinsic and non-intrinsic cancer risk factors. *Nature communications*, 9(1), 3490.
- Yang, S. N., Li, F. J., Liao, Y. H., Chen, Y. S., Shen, W. C., & Huang, T. C. (2015). Identification of breast cancer using integrated information from MRI and mammography. *PloS one*, 10(6), e0128404
- Yu, Li-Rong & Stewart, Nicolas & Veenstra, Timothy. (2010). Proteomics. The Deciphering of the Functional Genome. *Essentials of Genomic and Personalized Medicine*. 89-96. 10.1016/B978-0-12-374934-5.00008-8.1

Anexos

Paragon Method: BREAST_CANCER Delete

Describe Sample

Sample Type: Identification

Cys Alkylation: Iodoacetamide

Digestion: Trypsin

Instrument: TripleTOF 5600

Special Factors:

- Phosphorylation emphasis
- Methyl esterification
- Phos-Tyr affinity column
- Gel-based ID

Species: Homo sapiens

Specify Processing

Quantitate Bias Correction Background Correction

ID Focus:

- Biological modifications
- Amino acid substitutions

Database: Human_Reviewed_20,341 entradas

Search Effort:

Rapid ID

Thorough ID

Results Quality:

Detected Protein Threshold [Unused ProtScore (Conf)] >: 0.05 (10.0%)

Run False Discovery Rate Analysis

Figura 29. Parámetros utilizados en la creación de un método con el algoritmo Paragon para la identificación de proteínas de TM con el software Protein Pilot.

Tabla 5. Configuración utilizada en la cuantificación SWATH-MS con el software Peakview.

Filtración de péptidos	
Número de péptidos por proteína	100
Número de transiciones (iones de fragmentación) por péptido	6
% de confianza	99
FDR	1.0
Exclusión de péptidos modificados	sí
Cromatogramas de iones extraídos (XIC)	
Tamaño de la ventana de extracción (min)	10
Anchura del XIC (ppm)	75

Tabla 6. Configuración utilizada para el análisis multivariante PCVG en el software Markerview

Número de PCs	4
Angulo delta (grados)	40
Distancia mínima desde el origen	0.2
Asignación de grupos automáticamente	sí
Distancia mínima del origen para comenzar un nuevo grupo	0.05
Ordenar por	Magnitud

```

> rm(list=ls())
> library(SwathXtend)
> filenames <- c("Lib2.txt", "Lib3.txt")
> libfiles <- paste(system.file("files",package="SwathXtend"),
+ filenames,sep="/")

> Lib2 <- readLibFile(libfiles[1], clean=TRUE)
> Lib3 <- readLibFile(libfiles[2], clean=TRUE)
> Lib2 <- cleanLib(Lib2, intensity.cutoff = 2000, conf.cutoff = 0.99,
+ nomod = FALSE, nomc = FALSE)
> Lib3 <- cleanLib(Lib3, intensity.cutoff = 2000, conf.cutoff = 0.99,
+ nomod = FALSE, nomc = FALSE)
> checkQuality(Lib2, Lib3)
> plotAll(Lib2, Lib3, file="allplots.xlsx")

> Lib2_3 <- buildSpectraLibPair(libfiles[1], libfiles[2], clean=T,
+ nomc=T, nomod=T, plot=F,
+ outputFormat = "peakview",
+ outputFile = "Lib2_3.txt")
> outputLib(Lib2_3, filename="Lib2_3.txt", format="peakview")

```

Figura 30. Código R utilizado con el paquete SwathXtend para extender la biblioteca interna con el proteoma de humano del repositorio SwathAtlas. El código permite la integración de las bibliotecas en formato .txt (Wu, 2019).

Tabla 7. Criterios utilizados en el análisis de enriquecimiento de ontología de genes con la aplicación ClueGO del software Cytoscape.

Tipo de ontología	REACTOME Pathways
pV	≤0.05
Intervalo GO Tree	3-8
Número de genes mínimo por clúster	3 /20
Porcentaje de genes del número mínimo de genes por cluster (%)	2
Kappa Score	0.4
Tipo de análisis estadístico	Enrichment/Depletion (prueba hipergeométrica de dos lados)
Corrección	Bonferroni
Diseño de visualización	Orgánico

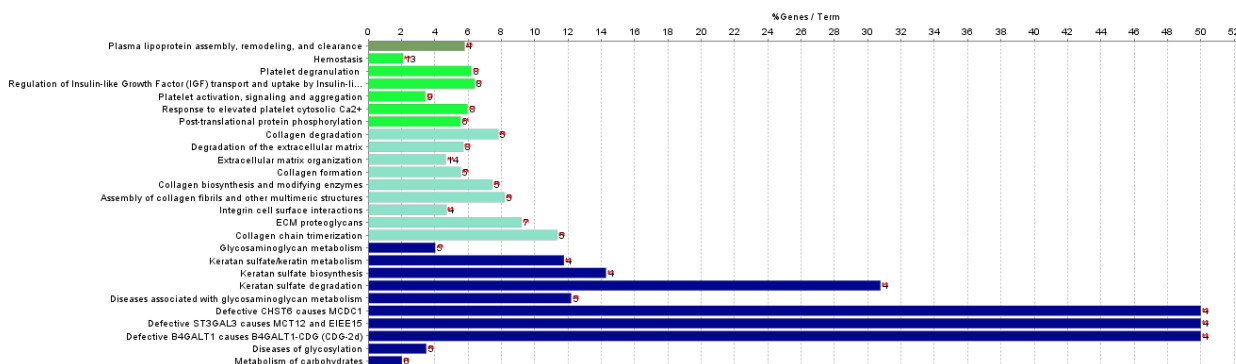


Figura 31. Histograma de barras que representa el porcentaje de proteínas que se relacionaron con un término (vías de señalización REACTOME) en el análisis de enriquecimiento de genes de las proteínas con menor abundancia en los TMM utilizando la biblioteca extendida.



Figura 32. Histograma de barras que representa el porcentaje de proteínas que se relacionaron con un término (vías de señalización REACTOME) en el análisis de enriquecimiento de genes de las proteínas con mayor abundancia en los TMM utilizando la biblioteca extendida.

Tabla 8. Lista de ID Uniprot de las proteínas que modificaron su abundancia en los TMB y TMM utilizando la biblioteca interna y externa en la cuantificación SWATH-MS. El paciente 14 fue excluido del análisis.

Proteínas con abundancia disminuida en los TMM con la biblioteca interna	Proteínas con abundancia disminuida en los TMM con la biblioteca extendida	Proteínas con abundancia aumentada en los TMM con la biblioteca interna			Proteínas con abundancia aumentada en los TMM con la biblioteca extendida		
AOA0C4DH72	P02652	P09497	P31943	P51659	Q9H930	P52907	P61254
P01042	P0DOX2	P30041	Q15046	Q8NFV4	Q13442	Q9UKM9	P31943
P35749	P0DOX7	Q99497	P62258	P62277	P12277	P25788	P55084
P39060	P04217	P00568	Q9UL46	P13667	P67809	Q96HJ9	P60903
P00734	P01619	O75323	Q12905	P12268	P04792	Q16576	P27348
P02652	AOA087WSZO	P29401	P17844	P43686	Q9H910	Q9Y2W1	Q15185
P00747	P43652	P05114	Q15435	Q9UIJ7	Q9H444	P25205	Q99829
P04433	P13716	P43034	P37108	P40227	Q16629	P11021	O00299
O15061	P02749	Q9HC38	P62714	Q9UKM9	O75150	P61970	P05455
Q9NRN5	P05546	P20042	P51149	P09234	Q16630	Q10567	Q9BS26
P51888	P01031	P49755	O75947	P62333	P46783	Q9UIJ7	P49321
P02774	Q06828	Q16836	Q12874	Q01518	P49903	P06744	P54578
P02766	P00747	Q9Y3C6	P17980	P30040	Q9Y2A7	P17174	P61981
P02768	P51948	P18206	P68036	P07910	P84103	P62277	P14866
P39059	P01700	Q09666	O60832	Q9UMS4	P45880	P14209	Q13510
P01008	P39060	O75531	P27816	P43304	P17096	P13667	P49458
Q15746	Q0VD83	Q9Y3D6	Q13561	Q07666	Q9BRP8	P10253	P29590
P02787	P01042	P07942	P48047	Q9NP97	Q00765	P62851	Q9NYL9
P02765	P60228	Q15942	P11279	P62316	P04406	Q9H299	O60749
P51884	Q96CT7	P14854	P23284	P29590	O95881	P04080	Q12874
P23946	P04433	P22061	Q15819	P20340	Q99623	P38919	Q92574
P12110	P00734	P30048	Q04637	Q08211	P19971	Q9NTK5	P24752
P01591	Q567U6	P11177	P55072	P52597	P36405	P08708	P12956
Q9UBX5	AOA0B4J1X5	P62873	Q14818	P06744	Q92882	P60953	Q1KMD3
P51911	Q9GZR2	O43639	P67870	P11940	P09914	Q8N5N7	P07686
P07585	P51888	P46459	P62750	P14314	P54577	P28066	P29966
Q8N474	Q92609	P27348	P40939	P28838	O75347	P43307	P30044
P22105	Q9H582	P62937	P25787	P55036	Q04917	Q9UNZ2	P43243
P08294	Q9UJV9	P00505	P07602	O95292	P07858	P05387	O14773
Q07507	P02774	Q9BRA2	P10606	P62308	P25774	Q15121	P17987
Q9BXN1	Q6UXG2	P30046	Q14126	P08579	Q08380	P09669	P61960
Q05707	Q9NWW8	P31937	P46109	Q00839	Q9UJ70	P06576	P09211
P20774	Q96RU2	P31946	Q13177	Q10567	P10809	P15880	P25705
P12273	O14578	P18669	P07203	Q15459	Q8TF72	Q13151	P54727
	Q9NRN5	Q14773	P50991	Q9HDC9	P21796	P62306	P27824
	P02766	P20700	Q9Y265	Q15233	P09651	Q9UJU6	Q7Z3U7
	P09001	P61026	P61960	Q14498	P23588	Q16718	Q13615
	Q13433	P27695	P13804	Q15084	P41250	Q15717	P04083
	O75529	Q96I99	P08865	P61978	P22392	Q99436	Q9NRV9
	P12109	Q13162	P05198	Q99873	P40121	P62847	Q99832
	Q96HS1	O75915	Q9Y678	E9PAV3	P62266	P63104	P23246
	P39059	Q9NQG5	P61019	P07737	P61956	Q13561	P68402
	P49792	P38117	P00491	P30626	Q13885	Q16543	P06733
	P02768	P11586	P62942	Q9Y524	Q15417	Q00688	P09417
	Q6UWY5	P24752	P61247	O95197	P15311	Q9Y3U8	Q96MM6
	P40818	P17858	P31942	Q32MZ4	P22307	P43304	P25398
	P02787	P07355	P09661	P10599	P01034	P37108	O75369
	P02765	Q9P258	P47914	P48444	O60506	Q92841	P52943
	Q7KZI7	O14979	P27797	P39019	P52272	P40926	P31040
	Q9UM54	P61586	O00233	O76003	Q96AG4	P46782	P62140
	Q6UWP2	P07741	Q04837	P08708	P04075	Q71UI9	Q9H1E3
	P15088	P78417	O75348	P24534	P29692	P13473	P00387

P51884	P48643	O75964	P20810	Q92688	P04843	Q92499
P16671	P09622	P17931	P22626	P07996	P35268	Q14103
P11171	Q9BS26	P36957	Q99584	P62081	Q9Y6N5	Q96KP4
P23946	Q9ULZ3	P10515	P26599	O15260	P31948	P08758
P61601	Q9BT78	Q02878	Q8NC51	Q99729	O75390	P63208
P12110	P40925	P49368	O60841	P19525	P61247	P14625
Q15746	Q15370	P39656	P07339	Q15102	P22314	Q9BVG4
Q8TET4	Q8NEV1	P52565	P22087	P09874	P26641	Q9NX63
P01008	P26440	P15880	Q96HJ9	Q15056	P05388	P07741
Q96JJ7	Q9UQ80	Q92945	Q9UHX1	P07437	O75964	P17980
Q8IX12	P37837	P51991	P38646	P11142	Q01844	P37837
Q07866	Q99832	O60888	P78371	Q96AE4	Q99714	P49368
Q8NBN7	P63173	Q92734	P23528	Q13247	P08579	Q14847
P51911	P51397	P50990	P46777	P98179	P62899	P09104
Q9UBX5	Q14847	P22314	Q15181	P60842	P60468	P51148
Q9Y570	P25398	O60664	P50213	P38159	P50502	Q9UQ80
P01591	Q14974	Q8IZQ5	P06748	P13073	Q9Y678	Q04637
Q96CW1	Q6IBS0	P62318	P20290	P41236	P21912	Q8NEV1
Q6WQC1	Q14103	P08571	Q01130	P08962	Q15365	P55769
Q9Y3E5	P30084	P49411	O15511	P68104	P62861	P63173
P35606	P51148	P36542	Q9Y2W1	P24539	Q00341	Q9Y265
P07585	P50570	Q96HC4	P62306	P09234	P62318	P40925
Q15654	O75874	Q9H3K6	P35244	O60869	P30101	P11177
P22105	P13284	Q12906	Q14444	P35232	P62258	Q12090
Q8N474	P07384	P46782	P62304	P43487	P62829	Q16563
P29762	P08758	Q13263	P61106	Q9NQC3	P39656	P02746
P08294	P62888	P50995	Q96QK1	Q15181	P36542	P41252
Q9BXN1	Q9Y490	Q5JTV8	Q9NQW7	P14618	Q13263	P78417
Q05707	P00387	Q13510	Q9P1F3	Q13435	Q5JTV8	Q14979
Q07507	Q96KP4	P13639	O60869	P26599	P51397	Q32MZ4
P05090	P68402	P58546	P11142	P14174	Q92945	P07954
P20774	P09417	Q07065	P14618	P10599	Q15843	P04424
P12273	Q92499	P45974	Q9NQC3	P62820	Q13409	P45974
Q13127	O60234	Q99714	P07900	P30050	P17931	P30084
	Q9NX63	P00492	P62269	Q96C19	Q8IZQ5	O75746
	O00231	Q15843	Q96AE4	P62304	Q12906	Q15370
	O43175	P62899	Q02818	P13693	Q96HC4	P62937
	Q8N163	P62861	P35659	Q01518	Q5HYI8	Q5GJ75
	Q14697	P62826	P43487	P50213	P61019	Q9BWH6
	P23246	O75131	P98179	P50454	P08567	P30046
	P63208	Q9UJU6	P49321	Q9Y6W5	P13639	P31937
	P04083	O75390	Q96C19	P62269	P62942	P31946
	P23396	P30101	P35232	Q02818	P60174	P22061
	P30044	P60174	P61923	P07900	P36957	P00505
	P14625	P13473	P13073	Q9Y5K6	Q02878	Q96I99
	P31040	P60468	P24539	P54920	O00629	P30048
	Q16563	Q9UNZ2	P68104	O15511	P31942	Q8IX01
	Q9UHD8	P50454	P41236	P07339	P55072	O43639
	P09104	P17174	P07437	P06748	Q15233	Q9Y3D6
	P09211	Q15121	Q96AG4	Q00839	P49411	O75531
	O75369	P05388	P04075	Q13011	P20618	Q09666
	P60866	Q16576	P38159	E9PAV3	P25787	Q14790
	P27824	P62314	P19525	P17844	P62750	
	P25705	P20618	Q9Y4Z0	P63241	P63220	
	O60749	Q71UI9	Q13247	O76003	P00491	
	P12956	Q9BVG4	O60506	Q9NQW7	P62495	
	P43243	P04843	P54577	P23528	P54819	
	P42126	Q9Y6N5	P41250	P20810	P13804	
	P20073	Q99436	Q99729	Q8NC51	P51991	
	P52943	P40926	Q15102	P11940	O75947	
	P54727	Q13011	P07996	P24534	P27797	
	P13010	O43242	P09874	P61978	P50995	
	Q1KMD3	O75396	P52272	Q9Y5Z4	O00233	
	Q9NRV9	Q9Y3U8	P29692	P07737	P46109	
	P06733	P54819	P01034	P78371	O75348	
	P61981	P50502	Q15056	Q16832	P50990	
	Q99829	Q00341	P60842	P07910	Q93050	
	P14866	Q96PK6	P08962	P22087	P10606	
	P00736	Q15717	P09651	P51659	P09661	

	P54578	Q16543	Q92688	P30626	Q15819
	P0DMV9	P06576	P22307	P46777	P48047
	Q13409	P35268	P62081	P62314	P08865
	P31483	O43390	P62820	P38646	P67870
	P60903	Q00688	P61956	P22626	P05198
	Q9NYL9	P62847	P62266	P40227	O43707
	Q16891	O00193	Q13885	P55036	P52565
	O43809	P09669	O00764	Q9NP97	P47914
	Q15907	P30050	P22392	P28838	P00492
	P49458	Q01844	P40121	P39019	O60664
	P05455	P09525	P23588	P80404	P07203
	P12830	P28066	Q08380	Q14498	P62888
	P63104	P60953	P15311	O43451	P23284
	P02747	P26641	P10809	P62308	P50991
	P29966	Q16718	P21796	Q15084	O43390
	Q9H1E3	P05387	Q04917	Q13242	Q15907
	P62495	P10253	Q14764	Q01130	P27816
	P35998	Q92841	Q99623	P52597	Q13838
	O00299	P04632	O75347	P04632	Q16698
	P28070	Q9H299	P25774	B2RTY4	O43143
	Q13838	P43307	Q9Y6W5	O95197	Q15435
	P07686	P61970	P04406	P62826	P61088
	P61254	P13693	Q13435	O95292	P55735
	P61916	P38919	Q16630	P20340	Q9Y490
	P62140	Q9NTK5	Q92882	P19878	P68036
	O43707	P14209	O95881	P00558	Q13148
	Q13185	Q9NR45	Q00765	P14314	P10515
	P63220	P62851	P49903	Q08211	O14818
	Q15185	P31948	P36405	P12268	O60832
	P55084	P52907	P19971	Q9UMS4	O60493
	Q13148	P63241	P84103	Q07666	P09525
	Q9BWD1	Q8WWX9	Q9UJ70	P20674	Q14108
	Q14108	P12236	P06703	P30040	P20073
	P34932	P04080	P17096	P30085	P28070
	O60493	P20674	P46783	Q9BRA2	P62714
	P67936	P05386	P45880	Q9Y224	P0DMV9
	Q16698	P54920	Q9BRP8	Q99426	P31483
	P17987	P11021	Q9H444	Q15061	P11279
	O43143	Q9Y224	Q9H910	Q99584	P67936
	O15260	Q99426	P67809	Q15459	P34932
	P07954	P00558	Q16629	P12236	P13010
	P62913	P30085		Q8WWX9	Q09028
	P25788	Q13442		Q13185	Q9Y3C6