

**UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA**

**Análisis de la expresión diferencial de proteínas totales de células de
cáncer de mama para la identificación de probables biomarcadores
útiles para el diagnóstico y/o pronóstico de esta enfermedad.**

TESIS

**Que para obtener el grado de
Doctor en Ciencias Biológicas y de la Salud**

PRESENTA

Biol. Exp. Karla Grisel Calderón González

Comité tutorial:

Directores de tesis

**Dr. Juan Pedro Luna Arias.
Dr. Humberto González Márquez.**



Ciudad de México, Abril, 2016.

El Doctorado en Ciencias Biológicas y de la Salud de la Universidad Autónoma Metropolitana pertenece al Padrón de Posgrados de Excelencia del CONACyT y cuenta con apoyo del mismo Consejo, con el convenio PFP-20-93.

Acta

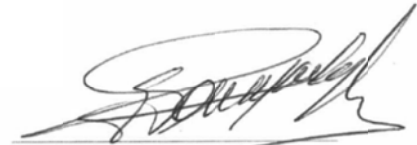
El jurado designado por las Divisiones de Ciencias Biológicas y de la Salud de las Unidades Cuajimalpa, Iztapalapa y Xochimilco aprobó la tesis que presentó:

la Biol. Exp. Karla Grisel Calderón González.

el día 15 de Abril del año 2016

Jurado

Presidente: Dr. Humberto González Márquez



Secretario: Dr. Juan Pedro Luna Arias



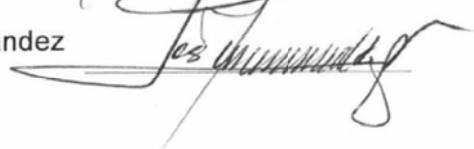
Vocal: Dra. Mina Konigsberg Fainstein



Vocal: Dr. Patricio Gariglio Vidal



Vocal: Dr. José Manuel Hernández Hernández



Dedicatoria

A mis padres

Beatriz y Sergio por ser un gran ejemplo de vida, por su presencia constante e invaluable, su aliento, enseñanza continúa, amor y apoyo incondicional.

A mis hermanos

Ana y Adrián, por darme la oportunidad de crecer a su lado y por su amor y apoyo incondicional.

A mi esposo

Jesús, por la dicha de su presencia en mi vida, por ser mi compañero y por su amor y apoyo incondicional.

Agradecimientos

Este trabajo fue realizado en el departamento de Ciencias Biológicas y de la Salud de la Universidad Autónoma Metropolitana, así como en el departamento de Biología Celular del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional. Durante la realización de este proyecto fui becaria CONACYT de Septiembre de 2011 a Septiembre de 2015.

Durante la realización de este proyecto se contó con el apoyo del Instituto de Ciencia y Tecnología del Distrito Federal (ICyTDF) con el proyecto ICyTDF-J.LA (CM-272/12-SECITI/033/2012), así como con el apoyo de CONACYT con el proyecto número 113674.

Agradezco al programa ECOS M10-A02 México-Francia, SEP-ANUIES-Conacyt por el apoyo financiero para la realización de una estancia académica corta en Lille, Francia.

Agradecimientos

Al Dr. Juan Pedro Luna Arias

Por aceptarme como parte de su equipo de trabajo, ha sido un gran privilegio trabajar a su lado. Gracias por la confianza que ha depositado en mí. También quiero agradecerle porque nunca escatimó esfuerzo ni tiempo para ayudarme. Gracias por su apoyo, consejos y enseñanzas.

Al Dr. Humberto González Márquez

Por la confianza que ha depositado en mí, porque desde que lo conocí me ha ofrecido su apoyo incondicional. Gracias por hacerme ver mis errores y gracias porque a pesar de ellos usted siempre me apoyo y defendió. Gracias por ser firme en los momentos en que se requería. Por sus múltiples consejos los cuales aprecio mucho.

A mis sinodales

Dra. Mina Konigsberg Fainstein, Dr. José Manuel Hernández Hernández, Dr. Patricio Gariglio Vidal por quienes siento un gran respeto y admiración. Gracias por aceptar ser sinodales en este trabajo y por las críticas, recomendaciones y valiosos comentarios para la consolidación de esta tesis.

Al Dr. José Luis Gallegos Pérez

Por confiar en mí, por su motivación y por su apoyo incondicional en todo momento tanto académicamente como personalmente. Por ser una parte fundamental en mi desarrollo profesional y por enseñarme lo que sé de espectrometría de masas. Muchas gracias porque a pesar de la distancia siempre está al pendiente de mí (MM)

A la Bióloga María Luisa Labra

Por ser mi amiga y confidente. Gracias porque siempre que te necesite estuviste en el momento exacto. Gracias por tus consejos, apoyo y cariño. Por todos los momentos que hemos compartido. También quiero agradecerte por todo el apoyo que me diste durante la realización de este trabajo.

A Refugio Contreras

Gracias por ser una gran amiga, porque siempre me animabas cuando lo necesitada, porque esa alegría y optimismo que te caracteriza me contagia. Gracias por los buenos momentos que hemos pasado juntas. Gracias por todo el apoyo que me brindaste durante la realización de este trabajo.

A mis amigos de toda la vida.

Nelly, Lucy, Lysia, Hilda, Maribel, Edith, Robin. Muchas gracias por todo su apoyo, porque a pesar de la distancia, sé que cuento con ustedes. Gracias por que estuvieron conmigo en momentos muy difíciles, cuidándome y apoyándome. Gracias por todos los buenos momentos que hemos vivido. Jana gracias por tu amistad incondicional, las pláticas y los buenos momentos compartidos.

A mis amigos, hermanos y compañeros de laboratorio

Ely, Ravy, Fer, Erika, Jeny, por su amistad incondicional, por los buenos momentos, consejos, las pláticas, las risas, la confianza y todo su apoyo. A Cesar, Benito, Ale, Brenda, Faby y Bere por sus consejos y amistad. Porque todos ustedes hicieron mi estancia en el laboratorio muy amena y divertida.

A mis padres

Por ser unos excelentes padres. Porque han estado conmigo en todo mi proceso formativo, porque saben mejor que nadie lo que nos ha costado. Por todo su esfuerzo y sacrificio, por brindarme todo el amor, la comprensión, apoyo incondicional y confianza en cada momento de mi vida. Gracias por estar a mi lado en los buenos y malos momentos. Gracias por enseñarme a no rendirme y a levantarme cuando caigo.

A mis hermanos

Por ser los mejores hermanos. Gracias por todos los momentos que hemos vivido, por su amor y apoyo incondicional. Gracias porque se que siempre están conmigo.

A mis abuelos

Por todas las enseñanzas y el tiempo que compartimos.

A mi tía Rosa

Gracias por ser ese ángel guardián de mi familia. Gracias por que ha estado con nosotros en los momentos más difíciles. Gracias por todo su apoyo, comprensión y amor.

A mis suegros y cuñados.

Gracias por los buenos momentos.

A mi esposo

Gracias por tu amor y apoyo incondicional. Gracias porque siempre que te necesito estas ahí para apoyarme. Gracias por todos los momentos hermosos que hemos compartido, también te agradezco por no soltar mi mano y darme fuerza en los momentos más difíciles. Por ser mi esposo, amigo y compañero. Gracias por ser mi ejemplo de trabajo constante y porque me motivas a ser una mejor persona. Y muchas gracias por todo el apoyo académico que me has brindado.

RESUMEN

El cáncer de mama (CaMa) es el principal cáncer en mujeres de todo el mundo. En México es la primera causa de muerte por cáncer en la mujer. A pesar de que existen marcadores tumorales de suero como CEA (antígeno carcinoembrionario) y HER2, estos son detectados en estadios avanzados de la enfermedad y son usados como marcadores de progresión y recurrencia. La detección temprana de esta enfermedad ofrece un tratamiento oportuno con buenos resultados, por lo tanto, es necesario identificar nuevos marcadores que puedan conducir a una detección temprana, así como que puedan proveer evidencia de la efectividad de los tratamientos. El objetivo de este trabajo fue determinar los perfiles de expresión diferencial de proteínas de las líneas celulares MCF7 y T47D (Luminales), MDA-MB-231 (Baja Claudina) y SK-BR-3 (HER2+) en comparación con la línea celular control MCF 10A a través del marcaje con iTRAQ y espectrometría de masas en tándem, para identificar posibles biomarcadores de la enfermedad.

Se identificaron 1,020 polipéptidos marcados con iTRAQ con al menos un péptido con 95% de confianza. Se utilizó una diferencia de 20% de expresión diferencial como valor de corte, obteniendo así 78 proteínas sobreexpresadas y 128 subexpresadas en común en todas las líneas celulares de cáncer de mama. Estas proteínas fueron clasificadas con el programa PANTHER en procesos biológicos, siendo las vías metabólicas las más afectadas. Por otro lado, las proteínas se analizaron con el programa STRING para determinar las redes de interacción proteína-proteína. El interactoma de las proteínas sobreexpresadas estuvo relacionado con topología del ADN, glicolisis, inicio de la traducción, corte y empalme del mRNA, la vía de las pentosas y la degradación por el proteosoma. La red de interacción de las proteínas subexpresadas incluyó varias proteínas mitocondriales, siendo el nodo más representativo el de la fosforilación oxidativa. También, se realizó una búsqueda en PubMed de las 206 proteínas identificadas. Con base en esta búsqueda se propusieron a las proteínas BAG6, DDX39, ANXA8 y COX4 como

probables biomarcadores en cáncer de mama, debido a que no han sido previamente reportadas y a que la información acerca de estas es escasa.

Por otro lado, también se identificaron proteínas exclusivas de cada una de las líneas celulares de cáncer de mama. Se obtuvieron 41 proteínas sobreexpresadas en MCF7, 59 en T47D, 55 en MDA-MB-231 y 74 en SK-BR-3. En el caso de las proteínas subexpresadas se identificaron 39, 32, 44 y 49 polipéptidos para MCF7, T47D, MDA-MB-231 y SK-BR-3, respectivamente. Las proteínas fueron analizadas de igual forma con PANTHER, siendo los procesos biológicos la categoría más afectada para las proteínas sobre y subexpresadas de cada línea celular. En cuanto al análisis con STRING, los interactomas de las proteínas sobreexpresadas tuvieron nodos relacionados con: Procesos metabólicos de mRNA (MCF7), biogénesis del ribosoma, transporte intracelular de proteínas (T47D), procesamiento de mRNA (MCF7 y T47D), ensamble de complejos de ribonucleoproteína, formación del complejo de preinicio de la traducción, regulación del inicio de la traducción (MDA-MB-231), procesos metabólicos de pequeñas moléculas (SK-BR-3), entre otros. Las redes de interacción proteína-proteína de los polipéptidos subexpresados fueron: Regulación de la polimerización y despolimerización de actina, activación de la respuesta inmune (T47D), transporte intracelular de proteínas, organización de la membrana (MDA-MB-231), biogénesis del ribosoma (SK-BR-3), entre otros.

ABSTRACT

Breast cancer is the principal cancer in women worldwide. In Mexico is the first cause of cancer death in women. Although there are serum tumor markers such as CEA (carcinoembryonic antigen) and HER2, they are detected in advanced stages of the disease and used as progression and recurrence markers. The early detection of this disease provides an appropriate treatment with satisfactory results, therefore is necessary identify new markers that can lead to early detection as well as provide evidence of the effectiveness of treatments. The aim of this work was to determine the differential protein expression profiles of breast cancer cell lines MCF7 and T47D (Luminals A), MDA-MB-231 (Claudin low) and SK-BR-3 (HER2+) in comparison to a normal control cell line MCF 10A by iTRAQ labelling and tandem mass spectrometry, in order to identify putative biomarkers of the disease.

We identified 1,020 iTRAQ-labelled polypeptides with at least one peptide identified with more than 95% in confidence. We determined a cutoff of 20% in differential expression, obtained 78 overexpressed and 128 subexpressed proteins in common in all breast cancer cell lines. These proteins were classified with PANTHER into biological processes, being the metabolic pathways the most affected. Also, the proteins were analysed with STRING for the generation of protein-protein networks. Interactome of overexpressed proteins was involved in DNA topology, glycolysis, translation initiation, splicing, pentose pathway, and proteasome degradation. The protein subexpressed network has several mitochondrial proteins; the main node was involved in oxidative phosphorylation. In addition, we performed a search for all the 206 differentially expressed polypeptides found in common in all breast cancer cell lines in the PubMed database. We propose BAG6, DDX39, ANXA8 and COX4 as putative biomarkers in breast cancer, because they not have been previously reported in breast cancer disease.

On the other hand, we identified sets of unique polypeptides in each breast cancer cell lines. We obtained 41 overexpressed proteins for MCF7, 59 in T47D, 55 for MDA-MB-231 and 74 for SK-BR-3. In the case of subexpressed polypeptides we found 39, 32, 44, and 49 polypeptides for MCF7, T47D, MDA-MB-231, and SK-BR-3, respectively. The proteins were analysed with PANTHER, being the biological processes the category more affected for overexpressed and subexpressed proteins of each breast cancer cell line. In STRING analysis, the interaction network of overexpressed proteins has nodes involved in mRNA metabolic processes (MCF7), ribosome biogenesis and intracellular transport of proteins (T47D), mRNA processing (MCF7 and T47D), ribonucleoprotein complex assembly, formation of the translation preinitiation complex, and regulation of translational initiation (MDA-MB-231) as well as small molecule metabolic processes (SK-BR-3), among others. For subexpressed polypeptides interactomes the processes involved were: Regulation of actin polymerization/depolymerization, activation of immune process (T47D), intracellular protein transport and membrane organization (MDA-MB-231), as well as ribosome biogenesis (SK-BR-3), among others.

ÍNDICE GENERAL

	Pág.
RESUMEN	i
ABSTRACT	iii
ÍNDICE GENERAL	v
ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABLAS	xii
CAPÍTULO I	1
1.1 Cáncer	2
1.2 Epidemiología	3
1.3 Glándula mamaria	5
1.4 Cáncer de mama	7
1.5 Factores de riesgo en cáncer de mama.....	14
1.6 Métodos de detección en cáncer de mama	16
1.6.1 Mamografía	16
1.6.2 Mamografía digital.....	17
1.6.3 Ultrasonografía.....	17
1.6.4 Imagen de resonancia magnética (MRI)	18
1.6.5 Cintimamografía.....	18
1.6.6 Tomografía computarizada por emisión de fotones individuales (SPECT)	19
1.6.7 Tomografía por emisión de positrones	19

1.7 Biomarcadores en cáncer de mama	19
1.7.1 Marcadores tisulares	23
1.7.2 Marcadores en suero	24
1.7.3 Marcadores de expresión génica	26
1.8 Tratamiento en cáncer de mama	28
1.9 Proteómica y espectrometría de masas.....	28
1.9.1 Espectrómetro de masas	31
1.9.1.1 Fuentes de ionización	31
1.9.1.2 Analizadores de masas	35
1.9.1.3 Detectores de masas	41
1.10 Estrategias de análisis en Proteómica.....	41
1.10.1 Aproximación Top-down (TD)	42
1.10.2 Aproximación Bottom-up	42
1.11 Proteómica cuantitativa.....	44
1.11.1 SILAC.....	45
1.11.2 DIGE	46
1.11.3 ICAT	47
1.11.4 iTRAQ	47
CAPÍTULO II	51
Antecedentes.....	52
CAPÍTULO III	57
1. Justificación	58
2. Pregunta de investigación.....	59

3. Hipótesis	59
4. Objetivos	60
4.1 Objetivo general	60
4.2 Objetivos específicos	60
CAPÍTULO IV	61
1. Cultivo celular	62
2. Extracción de proteínas	63
3. Deslipidación	64
4. Precipitación con TCA en acetona.....	64
5. Electroforesis en geles de acrilamida	65
6. Tinción de geles de poliacrilamida con nitrato de plata	65
7. Digestión de proteínas.....	66
8. Marcaje de péptidos con iTRAQ.....	67
9. Prefraccionamiento de péptidos	67
10. Análisis por espectrometría de masas	68
11. Identificación y cuantificación de proteínas.....	69
12. Análisis bioinformático	69
13. Inmunodetección sobre membrana de nitrocelulosa (Western blot).....	72
CAPÍTULO V	73
1. Estandarización de protocolos de extracción de proteínas.....	74
2. Perfiles de expresión diferencial de las líneas celulares de CaMa	80
3. Clasificación funcional de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.	91

4. Redes de interacción proteína-proteína de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama	93
5. Localización de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama en vías metabólicas.	96
6. IPA análisis de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.....	101
7. Selección de probables biomarcadores en cáncer de mama.	114
8. Validación de los probables biomarcadores para cáncer de mama.....	117
9. Perfil de expresión diferencial de las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+.	119
10. Clasificación funcional de las proteínas expresadas diferencialmente en las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+.....	119
11. Redes de interacción proteína-proteína de las proteínas encontradas expresadas diferencialmente en las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+	120
CAPÍTULO VI	129
Discusión	130
Conclusiones	147
Perspectivas	149
CAPÍTULO VII	151
Referencias.....	152
CAPÍTULO VIII	171
Anexo (Publicaciones)	171

ÍNDICE DE FIGURAS

	Pág.
Figura 1. Características de las células cancerosas.....	3
Figura 2. Incidencia y mortalidad de los principales tipos de cáncer	4
Figura 3. Anatomía (a) e histología (b) de la glándula mamaria.....	6
Figura 4. Desarrollo de la glándula mamaria.....	7
Figura 5. Principales clases de cáncer de mama..	9
Figura 6. Clasificación de los estadios en Cáncer de Mama.....	12
Figura 7. Análisis proteómico basado en espectrometría de masas..	30
Figura 8. Componentes de un espectrómetro de masas.....	31
Figura 9. Técnicas de ionización utilizadas en proteómica.....	35
Figura 10. Analizadores de masas utilizados en proteómica.....	40
Figura 11. Estrategias de análisis en proteómica.....	43
Figura 12. Tecnologías en proteómica cuantitativa..	49
Figura 13. Composición y mecanismo de acción de iTRAQ.....	50
Figura 14. Extractos totales de proteína obtenidos con Tritón-NP40, RIPA-Urea e hidrocloreuro de guanidina.....	74
Figura 15. Análisis del patrón electroforético de extractos totales obtenidos con Tritón-NP40, RIPA-Urea e hidrocloreuro de guanidina.	75
Figura 16. La benzonaza no elimina el ADN en extractos totales de proteína obtenidos con diferentes soluciones de lisis.	76
Figura 17. Eliminación de ADN del fago λ	77
Figura 18. Extractos totales de proteína obtenidos con GuHCl y GuHCl - RapiGest	78
Figura 19. Análisis del patrón electroforético de extractos totales obtenidos con GuHCl e GuHCl-RapiGest.....	79
Figura 20. Perfil electroforético de extractos totales de proteína de líneas celulares de cáncer de mama.	80

Figura 21. Identificación y cuantificación de péptidos de las líneas celulares de cáncer de mama a través de marcaje con iTRAQ y análisis por espectrometría de masas en tándem.....	81
Figura 22. Gráfica de volcán de las proteínas expresadas en las líneas celulares de cáncer de mama..	82
Figura 23. Diagramas de Venn-Euler de las proteínas diferencialmente expresadas en las líneas celulares de cáncer de mama.....	84
Figura 24. Clasificación funcional con PANTHER de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.....	92
Figura 25. Redes de interacción proteína-proteína de los polipéptidos sobreexpresados en común en todas las líneas celulares de cáncer de mama....	94
Figura 26. Redes de interacción proteína-proteína de los polipéptidos subexpresados en común en todas las líneas celulares de cáncer de mama.....	95
Figura 27. Visión global de las vías metabólicas afectadas en todas las líneas celulares de cáncer de mama..	99
Figura 28. Localización de los polipéptidos subexpresados en la cadena respiratoria obtenida con la base de datos KEGG.	100
Figura 29. Redes de interacción funcional del análisis Core I de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.....	111
Figura 30. Redes de interacción funcional del análisis Core II de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.....	112
Figura 31. Redes de interacción funcional del análisis Core III de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama..	113
Figura 32. Diagramas de Venn-Euler de probables paneles de biomarcadores obtenidos con el programa IPA y la base de datos de PubMed.....	116
Figura 33. Validación de proteínas candidatas a biomarcadores en cáncer de mama.	118

Figura 34. Diagramas de Venn-Euler de las proteínas sobreexpresadas en los diferentes tipos moléculas de líneas celulares de cáncer de mama.....	122
Figura 35. Diagramas de Venn-Euler de las proteínas subexpresadas en los diferentes tipos moléculas de líneas celulares de cáncer de mama.....	123
Figura 36. Clasificación funcional con PANTHER de las proteínas sobreexpresadas y subexpresadas exclusivamente encontradas en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama.....	124
Figura 37. Interactoma de los polipéptidos sobreexpresados exclusivos en cada una de las líneas celulares Luminal A de cáncer de mama.....	125
Figura 38. Interactoma de las proteínas sobreexpresadas exclusivas encontradas en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama.....	126
Figura 39. Interactoma de los polipéptidos subexpresados exclusivos en cada una de las líneas celulares Luminal A de cáncer de mama).	127
Figura 40. Interactoma de los polipéptidos subexpresados exclusivos encontrados en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama.....	128

ÍNDICE DE TABLAS

	Pág.
Tabla 1. Clasificación molecular de tumores en cáncer de mama.....	13
Tabla 2. Biomarcadores en cáncer aprobados por la FDA US.	22
Tabla 3. Proteínas sobreexpresadas encontradas en común en todas las líneas celulares de cáncer de mama por iTRAQ y espectrometría de masas en tándem....	85
Tabla 4. Proteínas subexpresadas encontradas en común en todas las líneas celulares de cáncer de mama por iTRAQ y espectrometría de masas en tándem....	87
Tabla 5. Clasificación de las proteínas sobre y subexpresadas con el programa IPA.	103
Tabla 6. Expresión diferencial de las proteínas comúnmente utilizadas para la caracterización de diferentes tipos de tumores	141
Tabla 7. Expresión diferencial de proteínas frecuentemente alteradas en cáncer..	142

CAPÍTULO I

INTRODUCCIÓN

1.1 Cáncer

La carcinogénesis es un proceso de múltiples pasos, que pueden ser el reflejo de alteraciones genéticas, epigenéticas o ambas, que conducen a una transformación progresiva de células normales hacia células con un alto grado de malignidad (Hanahan y Weinberg, 2011). Es importante mencionar que un solo cambio genético es raramente suficiente para el desarrollo de un tumor maligno. La mayoría de la evidencia apunta a un proceso de múltiples pasos y/o alteraciones secuenciales en diversos oncogenes, genes supresores de tumores y genes de microRNAs en células cancerosas (Croce, 2008).

Históricamente, los eventos de transformación en cáncer han sido definidos como eventos de iniciación (contribuyen en las primeras etapas de la transición neoplásica) y eventos de progresión (refiriéndose a procesos posteriores de transformación). Los oncogenes pueden ser activados por alteraciones estructurales derivadas de la mutación o fusión de un gen, por yuxtaposición de elementos potenciadores o por amplificación. Las translocaciones y mutaciones pueden ocurrir como eventos iniciadores o durante la progresión del tumor, mientras que la amplificación usualmente ocurre durante la progresión tumoral. Los productos de los oncogenes se clasifican en seis grandes grupos: factores de transcripción, remodeladores de la cromatina, factores de crecimiento, receptores de factores de crecimiento, transductores de señales y reguladores de la apoptosis (Croce, 2008). Alteraciones en estos elementos brindan a las células cancerosas ventajas sobre las células normales, como son señales de proliferación sostenida, insensibilidad a supresores de crecimiento, evasión a la destrucción inmune, inmortalidad replicativa, inflamación promotora de tumor, invasión y metástasis, angiogénesis sostenida, inestabilidad del genoma y mutaciones, evasión de la apoptosis y desregulación del metabolismo celular (Figura 1) (Hanahan y Weinberg, 2011).

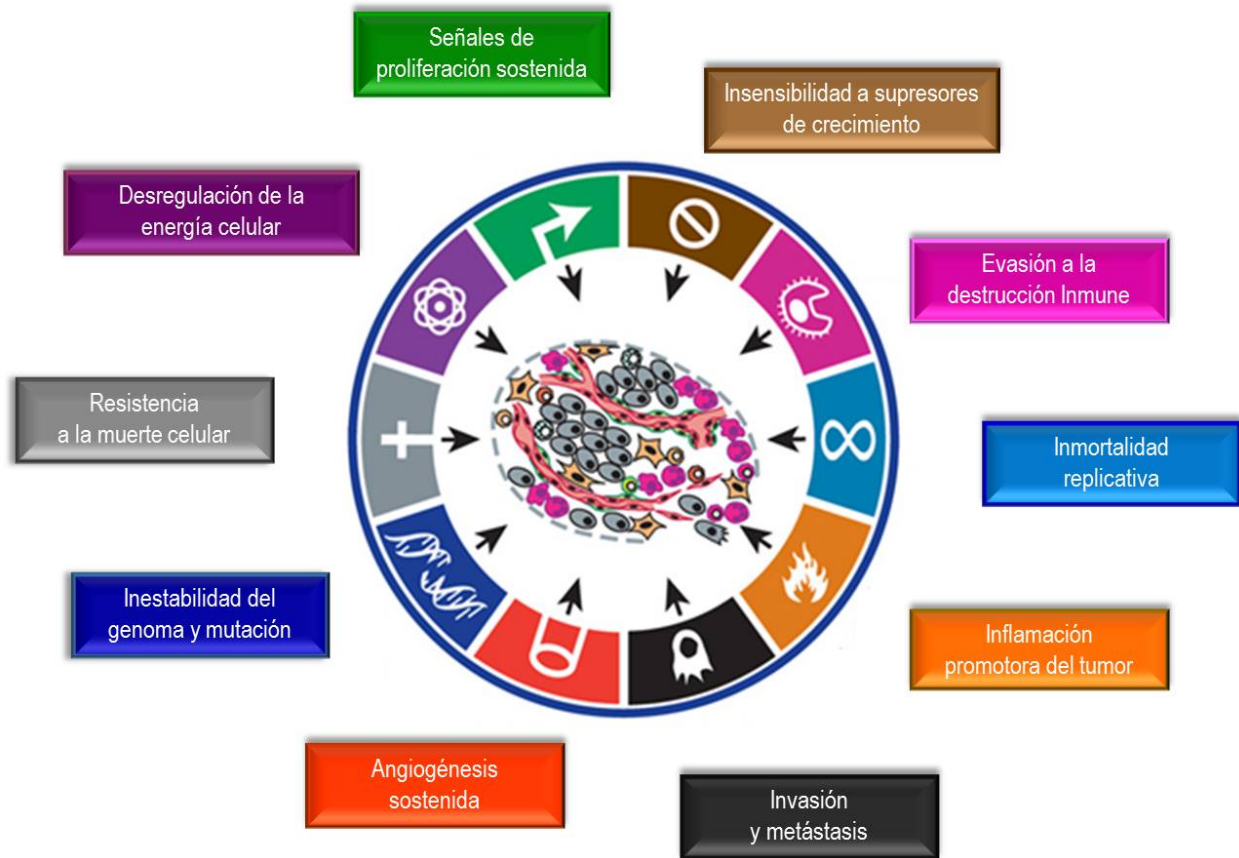


Figura 1. Características de las células cancerosas. La imagen muestra las características que las células normales adquieren durante el proceso de la carcinogénesis (Modificado de Hanahan y Weinberg, 2011).

1.2 Epidemiología

El cáncer es la quinta causa de muerte a nivel mundial (<http://www.who.int/mediacentre/factsheets/>). En el 2012 se registraron 14,1 millones de casos nuevos, 8,2 millones de muertes y 32,6 millones de personas que viven con cáncer (dentro de los 5 años del diagnóstico) en todo el mundo. El 57% (8 millones) de casos nuevos de cáncer, el 65% (5,3 millones) de las muertes y el 48% (15,6 millones) de los casos de cáncer prevalentes durante 5 años se produjeron en los países con menor desarrollo (http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx).

El cáncer de mama es el segundo cáncer más común en el mundo, de hecho, el tipo de cáncer más frecuente entre las mujeres con un estimado de 1,67 millones de casos nuevos de cáncer diagnosticados en el 2012 (25% de todos los cánceres) (Figura 2a). Es el cáncer más común en las mujeres, tanto en las regiones con mayor y menor desarrollo, contabilizándose 794,000 y 883,000 casos, respectivamente (http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx). En México, el cáncer es la tercera causa de muerte, registrándose 75,229 defunciones durante el 2013. El cáncer de mama es la primera causa de muerte por cáncer en la mujer con un registro de 5,548 defunciones en el 2013 (Figura 2b) (<http://www.inegi.org.mx/est>). El cáncer de mama masculino es poco frecuente. El número de casos de carcinomas de mama que se presentan en hombres es inferior a 1%. La mediana de edad en el momento del diagnóstico es de 60 a 70 años, aunque este cáncer puede afectar a hombres de todas las edades (<http://www.cancer.gov/espanol/tipos/>).

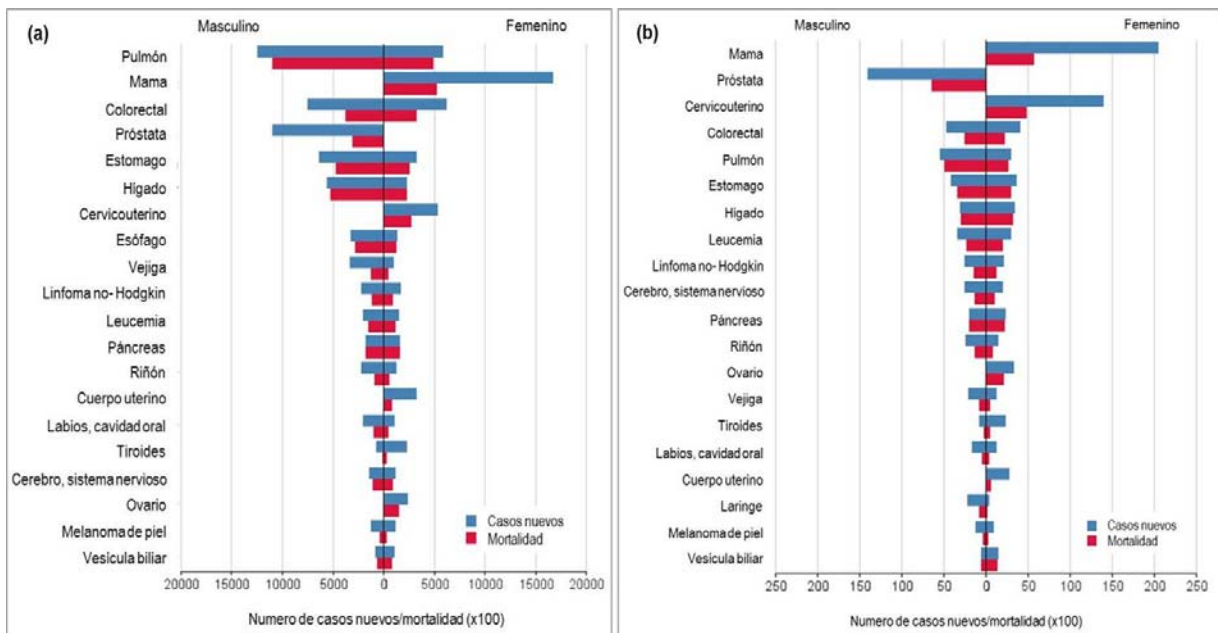


Figura 2. Incidencia y mortalidad de los principales tipos de cáncer. (a) En el mundo y en (b) México (http://globocan.iarc.fr/Pages/bar_sex_pop_sel.aspx).

1.3 Glándula mamaria

La glándula mamaria es un órgano secretor complejo que está formado por diferentes tipos celulares: células epiteliales (forman la red de conductos de la glándula), adipocitos (constituyen almohadillas de grasa en los que la red de conductos está embebida), células endoteliales vasculares (componen los vasos sanguíneos), células estromales, incluyendo fibroblastos y una variedad de células del sistema inmune (Watson y Khaled, 2008). La estructura de la glándula mamaria está compuesta por dos compartimentos principales: la porción glandular y el tejido conectivo y estromal. La parte glandular comprende de 15-20 lóbulos y cada uno de estos posee lóbulos más pequeños que finalizan en docenas de bulbos delgados que pueden producir leche (Figura 3). Los lóbulos están conectados por tubos finos llamados ductos que conducen al pezón. Las células que forman los ductos y lóbulos son epiteliales y su función principal es producir y secretar la mayoría de los constituyentes de la leche. Las células epiteliales están rodeadas por una capa de células mioepiteliales unidas a la membrana basal y su papel es mantener la estructura tubular de los ductos y lóbulos. Alrededor de los lóbulos y los ductos, se encuentra tejido graso y conectivo, compuesto de fibroblastos en presencia de abundante matriz extracelular y adipocitos. Por otro lado, vasos sanguíneos y linfáticos irrigan la glándula mamaria, además de estar inervada principalmente por nervios sensoriales simpáticos (Hondermarck, 2003).

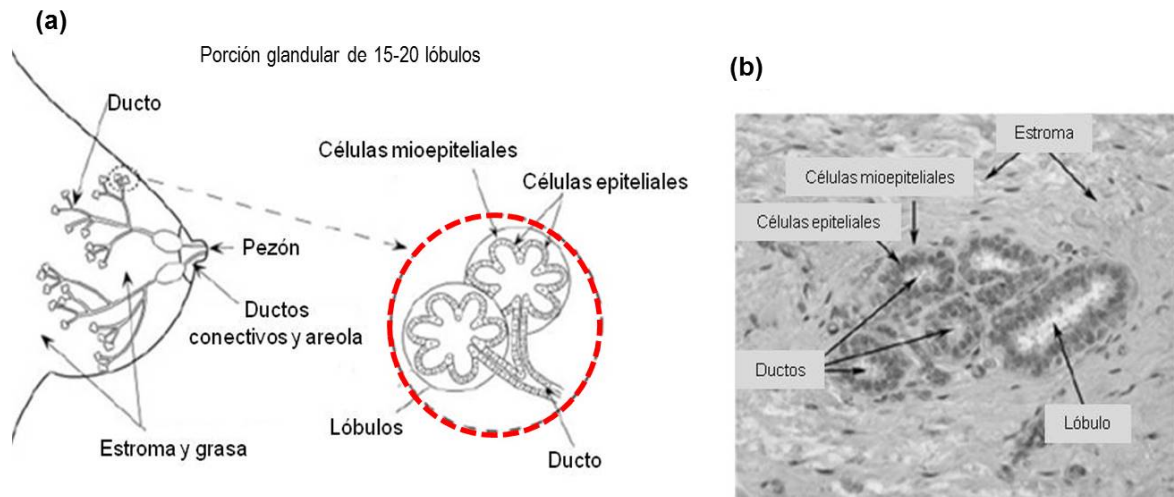


Figura 3. Anatomía (a) e histología (b) de la glándula mamaria. Compuesta de ductos y lóbulos, la parte glandular de la mama está formada por células epiteliales, rodeadas por células mioepiteliales y por otros tipos celulares presentes en el tejido conectivo (estroma): fibroblastos, células endoteliales, fibras nerviosas y adipocitos. Por lo tanto, la mama se presenta como un mosaico celular, en el que las interacciones célula-célula e interacciones matriz-célula juegan un papel fisiológicamente central durante el desarrollo y la carcinogénesis (Hondermarck, 2003).

A diferencia de muchos órganos, el desarrollo de la glándula mamaria ocurre predominantemente después del nacimiento y se completa en la edad adulta (Polyak y Kalluri, 2010). El desarrollo de la mama comienza durante la embriogénesis con la formación de un sistema ductal rudimentario que permanece así durante toda la infancia. Durante la pubertad los cambios hormonales inducen la formación de redes de conductos epiteliales que crecen hacia fuera, desde el pezón y se dividen en conductos primarios y secundarios que terminan en estructuras de brote. A partir de estos brotes y del sistema de ramificación ductal, las unidades terminales ductales lobulares alveolares (TDLUs) o lóbulos forman glándulas funcionales productoras de leche de una mama premenopáusicas. Cada lóbulo está recubierto por una capa de células epiteliales luminales rodeados por una capa basal de células mioepiteliales. El TDLU es el sitio desde el cual se cree se originan muchas hiperplasias epiteliales y carcinomas de mama. El desarrollo pleno de la glándula mamaria se

produce durante el embarazo con el crecimiento acelerado de los TDLUs para la preparación de la lactancia. Después del amamantamiento, se presenta una involución masiva y remodelamiento del tejido, regresando la glándula a un estado normal. Este ciclo de proliferación asociada con el embarazo, diferenciación, apoptosis y remodelamiento puede ocurrir varias veces durante la vida reproductiva (Figura 4). El proceso de desarrollo dentro de la mama puede explicarse por la presencia de una población de células madre tejido específicas de larga vida que tienen una propensión casi infinita para producir células funcionales. La longevidad de estas células hace susceptible la acumulación de cambios genéticos y la consecuente transformación de las células (Lamb *et al.*, 2007).

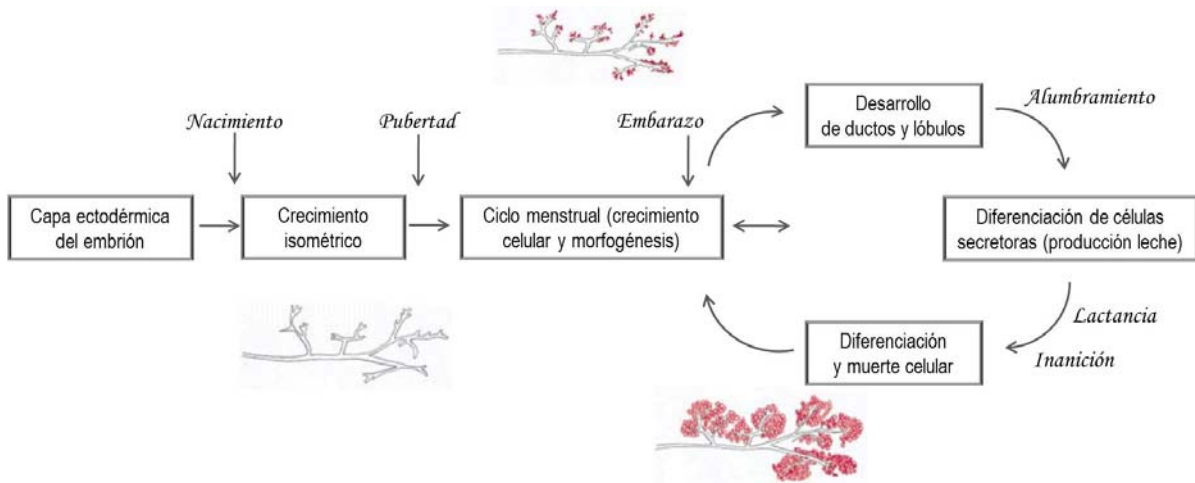


Figura 4. Desarrollo de la glándula mamaria. En la imagen se representan las diferentes etapas del desarrollo de la glándula mamaria, desde su comienzo en la embriogénesis hasta su pleno desarrollo en el embarazo (Modificado de Lamb *et al.*, 2007).

1.4 Cáncer de mama

Entre las patologías que pueden afectar a la glándula mamaria, la más recurrente y peligrosa es la carcinogénesis de mama (Kim *et al.*, 2009). El cáncer de mama es una enfermedad compleja durante la cual, la acumulación de numerosas alteraciones moleculares causa inestabilidad genética, proliferación celular y la

adquisición de un fenotipo resistente y altamente invasivo (Kim *et al.*, 2009). Existen evidencias que sugieren que el cáncer de mama es una colección de diferentes enfermedades con diferentes factores de riesgo, distintos cuadros clínicos y características patológicas particulares (Weigelt *et al.*, 2010).

La gran mayoría de los tumores malignos de mama son de origen epitelial y por lo tanto son clasificados como carcinomas (tumores epiteliales malignos). El término "cáncer de mama" comprende varios tipos de tumores que se clasifican tomando en cuenta su origen y características histológicas. Existen dos clases principales: carcinomas *in situ* y carcinomas invasivos; los primeros se caracterizan por tener células tumorales localizadas en los ductos [carcinoma ductal *in situ* (CDIS)] o en los lóbulos [carcinoma lobular *in situ* (CLIS)], sin la invasión a través de la membrana basal dentro del estroma circundante. En contraste, en los carcinomas invasivos la membrana basal es destruida parcial o totalmente. Los carcinomas ductales invasivos (65-80% de todos los cánceres de mama) y los carcinomas lobulares invasivos (5-15% de todos los casos) constituyen los dos principales tipos de cáncer de mama (Figura 5) (Hondermarck, 2003); sin embargo, existen otros tipos que cuentan con más del 25% de todos los cánceres de mama, en los cuales se reconoce la existencia de al menos 17 tipos con histología diferente (mucino, medular, tubular, carcinomas apocrinos, entre otros) (Weigelt *et al.*, 2010).

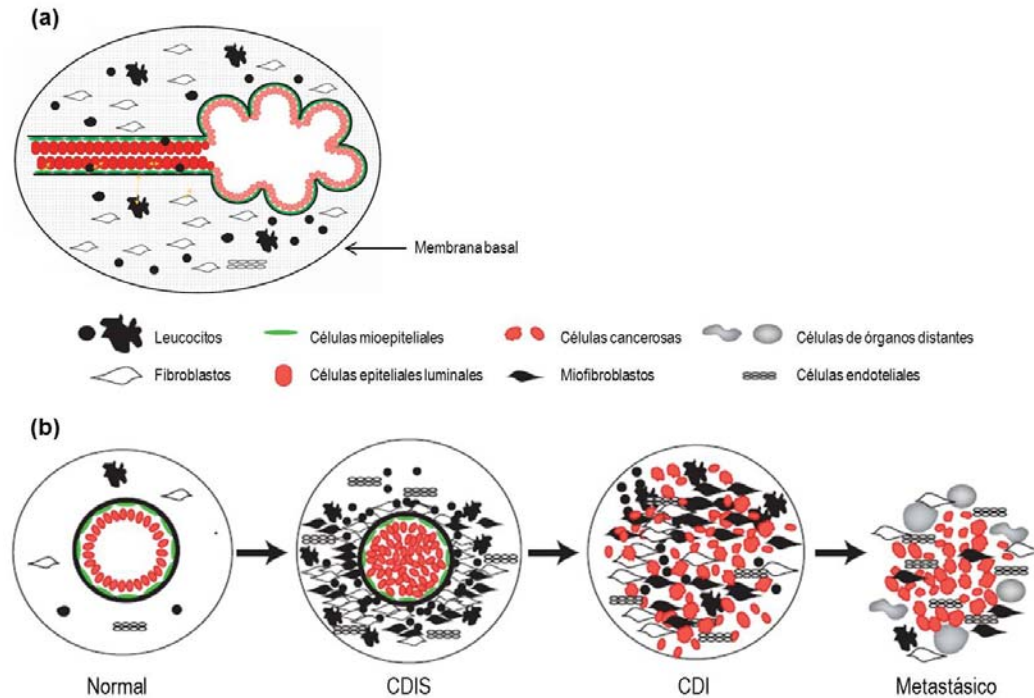


Figura 5. Principales clases de cáncer de mama. Representación esquemática de (a) Glándula mamaria normal, compuesta por varios tipos celulares delimitados por la membrana basal. (b) Desarrollo de los diferentes tipos de cáncer de mama. En el carcinoma ductal in situ (CDIS) se observa la presencia de células cancerosas localizadas. En el carcinoma ductal invasivo (CDI) la membrana basal se encuentra parcial o totalmente destruida y en el carcinoma metastásico, se puede apreciar la destrucción total de la membrana basal y por lo tanto, se presenta la migración de las células cancerosas a otros órganos (Modificada de Hondermarck, 2003).


Es importante destacar que la terminología “ductal” y “lobular” utilizada para definir distintas lesiones en mama, no implica el sitio de origen dentro de la glándula mamaria o los tipos celulares a partir de los cuales se formó la lesión; en realidad, la clasificación está basada en la arquitectura, citología e inmunohistoquímica de la lesión (Cichon *et al.*, 2010).



Debido a la variabilidad que los tumores presentan, estos se han clasificado mediante un sistema de estadificación propuesto por el Comité Americano Conjunto de Cáncer (American Joint Committee on Cancer o AJCC), proponiendo los estadios T, N y M o sistema TNM (T: Tumor, N: Nódulos o ganglios, M: Metastasis). Esta clasificación describe la gravedad de la enfermedad, basándose en la extensión del tumor original (primario) y si el cáncer se ha diseminado en el cuerpo. En la categoría

T se presentan distintas denominaciones como son: TX (no es posible encontrar un tumor), T0 (no hay indicio de algún tumor primario), Tis (tumor *in situ*), T1, T2, T3 y T4 (asignados dependiendo del tamaño del tumor, así como en qué medida han tomado tejidos circundantes). En la clasificación de N, también se encuentran diferentes subgrupos como son: NX (no es posible medir o encontrar los ganglios linfáticos (GL) cercanos), N0 (GL cercanos no presentan cáncer), N1, N2 y N3 (dependen de la cantidad de GL afectados y que nivel de cáncer presentan). Finalmente, la categoría de M indica si hay o no indicios de que el cáncer se haya propagado hacia otros órganos. En este caso también existen subclasificaciones como: MX (no es posible medir o encontrar metástasis), M0 (no existe metástasis distante) y M1 (se ha detectado metástasis distante).

La clasificación por estadios es importante debido a que permite un mejor manejo clínico, así como la elección de un tratamiento adecuado, además de que brinda información acerca del pronóstico de la enfermedad. Las diferentes combinaciones TNM permiten clasificar al CaMa en cinco estadios: 0, I, II, III y IV. En donde el estadio 0 corresponde a un cáncer no invasivo y el estadio IV pertenece a un cáncer invasivo (Greene *et al.*, 2002; <http://www.breastcancer.org/>)

Estadio 0. Es considerado un carcinoma *in situ*, este no presenta invasión de células cancerosas en otros tejidos. Este cáncer puede presentarse en el revestimiento de un conducto (CDIS) o en los lóbulos (CLIS) de la mama.

Estadio IA. El tumor mide ≤ 2 cm y no se disemina fuera de la mama (Figura 6a .

Estadio IB. En este estadio puede no encontrarse un tumor en la mama () o el tumor puede medir 2 cm o menos () y en ambos casos están presentes pequeños racimos de células cancerosas (mayores a 0.2 mm, pero menores a 2 mm) en los GL (Figura 6a).

Estadio IIA. En este caso puede no encontrarse un tumor en la mama, pero se detectan células cancerosas en los GL axilares o en los GL cercanos al esternón

(T0 N1 M0) (○). El tumor puede medir 2 cm o menos y puede propagarse hacia los GL axilares (T1 N1 M0) (⦿). Por otro lado, el tumor puede medir más de 2 cm, pero no más de 5 cm y este no se disemina en los GL axilares (T2 N0 M0) (⦿) (Figura 6b).

Estadio IIB. El tumor mide más de 2 cm, pero no más de 5 cm, y se observan grupos pequeños de células cancerosas (mayores a 0.2 mm, pero menores a 2 mm) en los GL axilares o en los GL cercanos al esternón (T2 N1 M0) (⦿). El tumor mide más de 5 cm, pero no se disemina hacia los GL axilares (T3 N0 M0) (⦿) (Figura 6b).

Estadio IIIA. Puede no encontrarse un tumor en la mama (T0 N2 M0) (○) o el tumor puede medir 2 cm o menos (T1 N2 M0) (⦿) o más de 2 cm, pero menos de 5 cm (T2 N2 M0) (⦿) o más de 5 cm (T3 N1 M0) (T3 N2 M0) (⦿) y el cáncer en cualquiera de las opciones mencionadas se disemina hasta los GL axilares que están unidos entre sí o a otras estructuras, o puede diseminarse hasta los GL cercanos al esternón (Figura 6c).

Estadio IIIB: En este estadio el tumor puede tener cualquier tamaño y el cáncer puede diseminarse hasta la pared torácica o a la piel de la mama (T4 N0 M0), o puede expandirse hasta los GL axilares o a los GL cercanos al esternón (T4 N1 M0) (T4 N2 M0) (Figura 6d ⦿).

Estadio IIIC: No se encuentra un tumor en la mama o el tumor puede tener cualquier tamaño y se puede haber diseminado hasta la pared torácica o a la piel de la mama. Asimismo, el cáncer puede diseminarse hasta los GL por arriba o debajo de la clavícula o puede expandirse hasta los GL axilares o hasta los GL cercanos al esternón (T: cualquiera N3 M0) (Figura 6d ⦿).

Estadio IV: Este estadio describe un cáncer de mama invasivo. El cáncer se diseminó hasta otros órganos del cuerpo, con mayor frecuencia hasta los huesos, pulmones, hígado o el cerebro (T: cualquiera, N: cualquiera y M1) (Figura 6e).

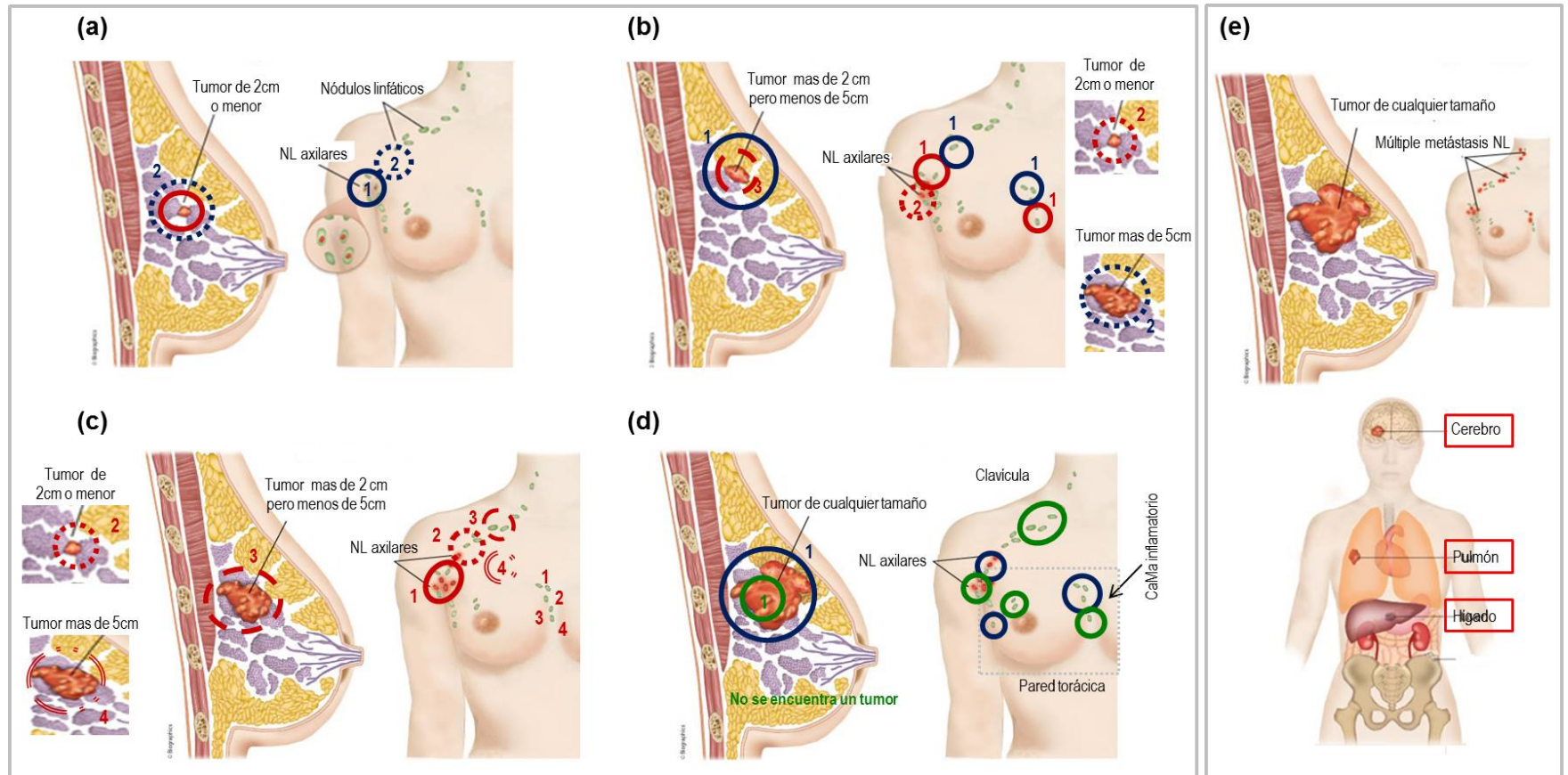


Figura 6. Clasificación de los estadios en cáncer de mama. En la imagen se representan las diferentes combinaciones que pueden presentarse en cada uno de los estadios de cáncer de mama. (a) Estadio IA (círculos rojos), IB (círculos azules), (b) Estadios IIA (círculos rojos) y IIB (círculos azules), (c) Estadio IIIA (círculos rojos), (d) Estadios IIIB (círculos azules) y estadios IIIC (círculos verdes), (e) Estadio IV, correspondiente a un cáncer de mama invasivo. (Modificado de [www. breastcancer.org](http://www.breastcancer.org)).

Los tumores también pueden ser clasificados molecularmente. Existen 5 categorías principales: Luminal A, Luminal B, Basal, Baja en Claudina y HER2 (Tabla 1) (Holliday y Speirs *et al.*, 2011).

Tabla 1. Clasificación molecular de tumores en cáncer de mama.

Clasificación	Inmunoperfil	Otras características	Respuesta a terapia	Ejemplos de líneas celulares
Luminal A 40-50% de casos	ER ⁺ , PR ^{+/-} , HER2 ⁻	Bajo Ki67	Sensible a terapia endocrina y frecuentemente sensible a quimioterapia	MCF7, T47D, SUM185
Luminal B 15-20% de casos	ER ⁺ , PR ^{+/-} , HER2 ⁺	Alto Ki67	Usualmente sensible a terapia endocrina, respuesta variable a quimioterapia. Los tumores HER2 ⁺ responden a terapia con Trastusumab	BT474, ZR-75
Basal 18% de casos	ER ⁻ , PR ⁻ , HER2 ⁻	Alto Ki67 EGFR ⁺ y/o Citoqueratina 5/6+	No sensibles a respuesta endocrina, frecuentemente responden a quimioterapia.	MDA-MB-468, SUM190
Baja en claudina 18% de casos	ER ⁻ , PR ⁻ , HER2 ⁻	Bajo Ki67 Baja E-cadherina, claudina 3,4, y 7	Respuesta intermedia a quimioterapia	BT549, MDA-MB-231, Hs578T, SUM1315
HER2 20-25% de casos	ER ⁻ , PR ⁻ , HER2 ⁺	Alto Ki67	Sensible a Trastusumab y quimioterapia.	SK-BR-3, MDA-MB-453

EGFR, Receptor del factor de crecimiento epidermal; ER, receptor de estrógeno; HER2, Receptor del factor de crecimiento epidermal humano 2; PR, receptor de progesterona.

Se cree que la gran mayoría de los tumores mamarios y sus precursores, independientemente de su tipo histológico, se originan en los TDLUs. Una de las hipótesis más aceptadas postula que el carcinoma invasivo inicia en los TDLUs y progresa de una enfermedad de mama benigna (BBD) hacia anormalidades celulares marcadas por una proliferación excesiva y atipia (Allred *et al.*, 2001). De acuerdo con esta hipótesis, la proliferación anormal en TDLU conduce inicialmente al desplegamiento de los lóbulos y/o estructuras quísticas, las cuales progresan a través de estadios donde el epitelio incrementa su proliferación sin adquirir características atípicas (enfermedad proliferativa sin atipia, PDWA). Eventualmente, la hiperplasia atípica

(incremento en el número de células con morfología anormal) puede manifestarse como forma lobular o ductal (hiperplasia ductal atípica (ADH) o hiperplasia lobular atípica (ALH) respectivamente), así como en estadios tempranos de carcinoma *in situ* ductal o lobular, respectivamente (Cichon *et al.*, 2010). Las lesiones ADH se derivan de excrecencias de células epiteliales lumbinales y están morfológicamente relacionadas con lesiones de grado bajo de DCIS, con la diferencia en el grado de espacio o volumen que ocupan (Allred *et al.*, 2001). DCIS es un grupo de lesiones cancerosas preinvasivas que surgen como resultado de una proliferación neoplásica de las células epiteliales lumbinales que no atraviesan la membrana basal. Tradicionalmente se considera que la transición de DCIS a carcinoma invasivo involucra la ruptura de la membrana basal y la capa que rodea a las células mioepiteliales; sin embargo, Sontag y Axerold en 2005 propusieron una vía diferente en la que el DCIS y el carcinoma invasivo se originan de una misma célula progenitora. Se considera que un DCIS histológicamente bien caracterizado progresa a un cáncer invasivo de grado bajo, mientras que un DCIS pobremente caracterizado evoluciona a través de distintas vías, progresando a un carcinoma invasivo de alto grado (Cichon *et al.*, 2010).

1.5 Factores de riesgo en cáncer de mama

Estudios epidemiológicos llevados a cabo en diferentes poblaciones han identificado diversos factores de riesgo para el desarrollo de cáncer de mama. La edad juega un papel importante, debido a que la incidencia de éste aumenta con ella, duplicándose cada diez años hasta la menopausia, etapa en la cual la tasa de incidencia disminuye drásticamente (McPherson *et al.*, 2000).

Además de la edad, existen otros factores de riesgo que se agrupan en dos categorías principales: exposición excesiva a estrógenos (hiperestrogenia) y la

deficiencia en el mantenimiento de la integridad genómica. La hiperestrogenia está relacionada con una sobrecarga de estrógenos (menarca temprana o menopausia tardía), y con un estilo de vida moderno, incluyendo baja paridad, edad avanzada en el primer alumbramiento, periodo corto de lactancia, uso prolongado de hormonas, comer en exceso y ejercicio limitado (Imyanitov y Hanson, 2004). Con lo que respecta a la deficiencia en el mantenimiento de la integridad genómica, algunos genes cuya mutación predispone al desarrollo de cáncer de mama contribuyen en la detección o reparación del ADN dañado (por ejemplo, *BRCA 1* y *BRCA 2*). Del mismo modo existen estudios de fenotipificación que demuestran la relación entre inestabilidad cromosomal y el riesgo a desarrollar cáncer de mama (Imyanitov y Hanson, 2004).

Se estima que al menos un 5% de todos los cánceres de mama presenta un componente hereditario. Las mutaciones en *BRCA1* y *BRCA2* se asocian con la mayor proporción de cáncer de mama hereditario (1-2% para cada gen) y con un 10% de los casos de cáncer de ovario (CO). Ambos genes son de alta penetrancia, es decir, un porcentaje elevado de los individuos que han heredado una mutación presenta la enfermedad a lo largo de su vida. Otra parte de los casos de cáncer de mama se asocia a síndromes infrecuentes debidos a mutaciones en otros genes de alta penetrancia: el gen *TP53* (mutado en pacientes con el síndrome de Li-Fraumeni), el gen *PTEN* (mutado en la mayoría de pacientes con el síndrome de Cowden), o el gen *LKB1-STK11* (causante del síndrome de Peutz-Jeghers), entre otros. Todos ellos presentan frecuencias poblacionales muy bajas y contribuyen escasamente al número de casos de cáncer de mama, que tan sólo es uno de los componentes de cada síndrome. La participación de otros genes de penetrancia menor, como *ATM* o *CHEK2*, puede ser mayor, pero es desconocida y controvertida, debido a resultados no coincidentes según el tipo y tamaño de población estudiada así como a la metodología utilizada (Diez *et al.*, 2006; Walsh *et al.*, 2006; Renwick *et al.*, 2006). Los polimorfismos comunes o las variantes en los genes *COMT*, *CYP1A1*, el gen del factor *TGF-beta-1* o las repeticiones en la secuencia del gen *AR* del receptor de

andrógenos confieren pequeños riesgos de cáncer de mama. Esto se ha observado en amplios estudios de casos y controles, aunque se requiere su confirmación en cohortes más amplias (Rebbeck *et al.*, 2002). Variantes en los genes de receptores hormonales y proteínas de unión a hormonas sexuales (*CYP19* o *SHBG*), reguladores del ciclo celular, del metabolismo de carcinógenos, de reparación de lesiones del ADN y de los genes *ATM*, *CHEK2* o *RAD51*, entre otros, son modificadores potenciales del riesgo de cáncer de mama (Narod, 2002; Hartge, 2003).

1.6 Métodos de detección en cáncer de mama

La autoexaminación o autoexploración de las mamas es ampliamente recomendada como primera detección, debido a que no es inusual para una mujer notar cambios anatómicos en los senos, además, el exámen clínico de las mamas es practicado comúnmente como parte de una revisión física de rutina por proveedores de la salud. Adicionalmente, varias técnicas de imagenología de mama son utilizadas ampliamente para la detección, diagnóstico y manejo clínico del cáncer. Las modalidades de imagen comúnmente utilizadas incluyen mamografía, ultrasonografía, imagen de resonancia magnética (MRI), cintimamografía, tomografía computarizada por emisión de fotones individuales (SPECT) y tomografía por emisión de positrones (PET) (Karellas y Vedantham, 2008).

1.6.1 Mamografía

La mamografía es la modalidad de imagen más ampliamente utilizada para detectar tumores u otras anormalidades. Esta técnica está recomendada para mujeres que no presentan síntomas o señales, o para aquéllas con anormalidades

observables en la mama. Consiste en que un haz de rayos X atraviesa la mama y crea una imagen proyectada en una placa; usualmente se requieren al menos dos mamogramas en ángulos diferentes de cada mama para el análisis (Prasad y Houserkova, 2007; Karellas y Vedantham, 2008). Esta técnica se asocia con una reducción de más del 40% en la tasa de mortalidad; sin embargo, la sensibilidad disminuye significativamente en mamas densas, con una sensibilidad reportada menor al 45% en mujeres con mamas extremadamente densas (Yang, 2011).

1.6.2 Mamografía digital

La mamografía digital emplea uno o varios detectores ensamblados para capturar una imagen electrónica de los rayos X transmitidos a través de la mama. La mamografía digital es superior a la mamografía de placa en mujeres jóvenes con mamas densas, debido a su capacidad para optimizar de forma selectiva el contraste en áreas de parénquima denso. Esta ventaja es especialmente importante en mujeres con una predisposición genética para el cáncer de mama, donde la intensificación de los programas de detección temprana puede comenzar desde los 25 o 30 años de edad (Bick y Diekmann, 2007).

1.6.3 Ultrasonografía

La ultrasonografía es la modalidad de imagen complementaria más importante para el diagnóstico de cáncer de mama. El ultrasonido es utilizado rutinariamente para diferenciar quistes (forma redonda u ovalada con márgenes circunscritos) de nódulos sólidos con una alta especificidad. En combinación con la mamografía, el ultrasonido se emplea para caracterizar masas sólidas, tanto benignas como malignas. Dentro de las características que ayudan a diferenciar los tumores se encuentran la forma, los márgenes y la ecogenicidad de la masa de la mama. Aunque el ultrasonido no está considerado como una prueba diagnóstica, es más

sensible que la mamografía en la detección de lesiones en mujeres con tejido denso en la mama. El ultrasonido puede ser empleado en mujeres jóvenes y mujeres con implantes de seno, que no siempre pueden detectar microcalcificaciones (Prasad y Houserkova, 2007).

1.6.4 Imagen de resonancia magnética (MRI)

La imagen de resonancia magnética es un método no invasivo que utiliza ondas de radio y campos magnéticos para observar órganos y estructuras que se encuentran dentro del cuerpo. La inyección de un agente de contraste es útil en este tipo de estudios, debido a que éste es preferencialmente absorbido por los tumores y la técnica puede utilizarse para mejorar el contraste de tejido normal circundante. Por otra parte, si la MRI es adquirida repetidamente después de la inyección de contraste, la naturaleza dinámica de este agente puede ser examinada, lo que aumenta la diferenciación entre las lesiones benignas y las malignas (MRI con contraste aumentado). La MRI es ampliamente empleada para el diagnóstico de mujeres con un alto riesgo de padecer la enfermedad por mutaciones genéticas congénitas y para la selección de tratamientos. Esta tecnología puede identificar focos de cáncer que no son detectables por la examinación física, mamografía o ultrasonografía (Morrow *et al.*, 2011).

1.6.5 Cintimamografía

La cintimamografía consiste en inyectar una cantidad pequeña de una sustancia radioactiva y, mediante el uso de una cámara especial, registra en qué regiones se acumuló la radiación en los senos. Este procedimiento no puede distinguir con precisión el cáncer, de lesiones no cancerosas, como en la mamografía de rutina y no se le utiliza como una prueba de detección inicial. Algunas ventajas de este método son que se puede utilizar en pacientes con mamas densas, para

obtener imágenes de grandes lesiones palpables que no aparecen con otras modalidades de imagen (Prasad y Houserkova, 2007).

1.6.6 Tomografía computarizada por emisión de fotones individuales (SPECT)

La ventaja principal de SPECT es su capacidad para proporcionar información en 3D, ya que en esta tecnología se hace rotar una cámara gama convencional alrededor del cuerpo, registrando una imagen en cada paso angular. Este sistema utiliza una trayectoria de exploración que sigue estrechamente el contorno de la mama (Karellas y Vedantham, 2008).

1.6.7 Tomografía por emisión de positrones

Es un estudio de medicina nuclear, en donde una cantidad pequeña de una sustancia radioactiva es inyectada y monitoreada por un escáner para formar una imagen. La sustancia utilizada en esta prueba es un azúcar, la cual se acumula en las células más activas, especialmente en el tejido canceroso. PET puede combinarse con una tomografía computarizada (CT) para proporcionar una vista anatómica y funcional de células sospechosas. El PET también se utiliza para detectar enfermedades metastásicas (Prasad y Houserkova, 2007).

1.7 Biomarcadores en cáncer de mama

Un biomarcador se define como un analito medible que correlaciona con un fenotipo específico, una condición biológica normal, un proceso patológico o a la respuesta farmacológica a una intervención terapéutica (Group biomarkers definition, 2001). En la investigación en cáncer, la detección de un biomarcador se refiere a una

sustancia o proceso que es indicativo de la presencia de cáncer en el cuerpo, esto puede ser una molécula secretada por el tumor o una respuesta específica del cuerpo a la presencia de cáncer (Kumar *et al.*, 2006). Las aplicaciones de los biomarcadores en oncología incluyen la evaluación de riesgos, tamizaje, diagnóstico diferencial, determinación del pronóstico, predicción de respuesta a tratamientos y monitoreo de la progresión de la enfermedad (Henry y Hayes, 2012).

Existen diversos tipos de biomarcadores, los cuales incluyen proteínas (enzima, receptores, entre otros), ácidos nucleicos (microRNAs u otros RNAs no codificantes), anticuerpos y péptidos, entre otros. Los biomarcadores pueden ser detectados en circulación (sangre, suero o plasma) o en excreciones o secreciones (orina, esputo o secreciones del pezón) y por lo tanto fácilmente evaluados y de forma no invasiva. Por otro lado, los marcadores pueden ser detectados en tejidos, donde se requiere de una biopsia o imágenes especiales para su evaluación (Henry y Hayes, 2012).

Los biomarcadores pueden ser divididos dentro de tres categorías: Diagnóstico, pronóstico y estratificación (predictivos). Los biomarcadores de diagnóstico son utilizados para detectar e identificar algún tipo de cáncer. Se espera que este tipo de biomarcadores tengan altos niveles de sensibilidad y especificidad en el diagnóstico. Los marcadores de pronóstico son comúnmente empleados una vez que el estado de la enfermedad ha sido establecido. Estos marcadores deben predecir el rumbo de la enfermedad y su recurrencia, por lo tanto influyen de manera importante en la agresividad de la terapia. Finalmente, los biomarcadores de estratificación o predictivos se basan frecuentemente en estudios en el ADN. Estos marcadores sirven para predecir la probable respuesta a una droga antes de comenzar un tratamiento, clasificando a los individuos en “respondedores” y “no respondedores” (Hamdan H, 2007).

Debido al papel crítico que los biomarcadores pueden llegar a desempeñar en las enfermedades es importante que estos sean evaluados rigurosamente. Pepe y colaboradores en 2001, propusieron una guía de cinco pasos para proporcionar un enfoque comprensivo y sistemático para evaluar y validar biomarcadores y guiar los resultados producidos en el descubrimiento de biomarcadores del laboratorio a la clínica. Las diversas fases incluyen: (1) Fase I (estudios preclínicos exploratorios) en la cual los perfiles de expresión de genes o proteínas son comparados entre tejidos cancerosos y tejidos normales. El objetivo de esta comparación es encontrar moléculas que estén sobre o subexpresados en los tejidos cancerosos en comparación con los normales. (2) Fase II, la cual abarca el desarrollo de ensayos clínicos, usualmente con muestras colectadas de manera no invasiva. (3) Fase III, la cual involucra un estudio retrospectivo, el cual determina la capacidad de un biomarcador para detectar enfermedades pre-clínicas utilizando un gran tamaño de muestra recolectada de individuos aparentemente sanos quienes son monitoreados para el desarrollo del cáncer. (4) Fase IV (estudio prospectivo), en la cual individuos usualmente asintomáticos con un alto riesgo para desarrollar cáncer son examinados para un biomarcador seleccionado, y los que dan positivo son generalmente objeto de seguimiento para determinar si tienen cáncer y, por lo tanto, ensayar el uso del biomarcador seleccionado en programas de detección temprana. (5) Fase V (estudios de control de cáncer), en la cual los estudios están diseñados para comprobar si el tamizaje de una gran población resulta en la disminución de la morbilidad y la mortalidad relacionada con la enfermedad (Pepe *et al.*, 2001). Los marcadores de cáncer aprobados por la FDA (US Food and Drug Administration) se enlistan en la tabla 2.

Tabla 2. Biomarcadores en cáncer aprobados por la FDA US.

Biomarcadores	Tipo	Fuente	Tipo de cáncer	Uso clínico
α -Fetoproteína	Glicoproteína	Suero	Testicular noseminomatoso	Estadificación
β -gonadotropina coriónica humana	Glicoproteína	Suero	Testicular	Estadificación
CA19-9	Carbohidrato	Suero	Pancreático	Monitoreo
CA125	Glicoproteína	Suero	Ovario	Monitoreo
Prueba de papanicolao	Frotis cervical	Cérvix	Cervical	Tamizaje
CEA	Proteína	Suero	Colon	Monitoreo
EGFR	Proteína	Colón	Colon	Selección de terapia
KIT	Proteína (IHC)	Tumor gastrointestinal	GIST	Diagnóstico y selección de terapia
Tiroglobulina	Proteína	Suero	Tiroides	Monitoreo
PSA (total)	Proteína	Suero	Próstata	Tamizaje y monitoreo
PSA (complejo)	Proteína	Suero	Próstata	Tamizaje y monitoreo
PSA (libre PSA%)	Proteína	Suero	Próstata	Diagnóstico de cáncer contra hiperplasia benigna prostática
CA15-3	Glicoproteína	Suero	Mama	Monitoreo
CA27-29	Glicoproteína	Suero	Mama	Monitoreo
Citoqueratinas	Proteína (IHC)	Tumor de mama	Mama	Pronóstico
RE y RP	Proteína (IHC)	Tumor de mama	Mama	Selección para terapia hormonal
HER2/NEU	Proteína (IHC)	Tumor de mama	Mama	Pronóstico y selección de terapia
HER2/NEU	Proteína	Suero	Mama	Monitoreo
HER2/NEU	ADN (FISH)	Tumor de mama	Mama	Pronóstico y selección de terapia
Cromosoma 3, 7, 9 y 17	ADN (FISH)	Orina	Vejiga	Tamizaje y monitoreo
NMP22	Proteína	Orina	Vejiga	Tamizaje y monitoreo
Fibrina/FDP	Proteína	Orina	Vejiga	Monitoreo
BTA	Proteína	Orina	Vejiga	Monitoreo
CEA de alto peso molecular y mucina	Proteína (IF)	Orina	Vejiga	Monitoreo

BTA, antígeno asociado a tumores de vejiga; CA, antígeno de cáncer; CEA, antígeno carcinoembrionario; FDP proteína de degradación de fibrina; FISH, hibridación fluorescente *in situ*; GIST, tumor estromal gastrointestinal; IHC, inmunohistoquímica; IF, inmunofluorescencia; NMP22, proteína de la matriz nuclear 22; RE, receptor de estrógeno; RP, receptor de progesterona; PSA, antígeno específico de próstata (Modificada de Ludwig y Weinstein, 2005).

1.7.1 Marcadores tisulares

Un número pequeño de biomarcadores han sido empleados por varios años en varios aspectos del manejo del cáncer de mama, incluidos el receptor de estrógenos (RE), receptor de progesterona (RP) y el receptor 2 del factor de crecimiento humano (HER2). La importancia biológica de estos marcadores establecidos ha sido reforzada en la última década por los resultados de clasificación genómica, en los cuales la presencia o ausencia de estos marcadores identifica tres grupos principales: luminal (RE positivo), similar a HER2 (principalmente RE negativo y HER2 positivo) y similar al estado basal o también conocido como triple negativo (principalmente RE-, RP-, HER2-) (Dowsett y Dunbier, 2008; Normanno *et al.*, 2009).

1.7.1.1 Receptor de estrógeno y progesterona

El receptor de estrógeno es el factor predictivo individual más poderoso, ya que dos tercios de mujeres menores de 50 años pueden presentar cáncer de mama RE+. Aproximadamente el 80% de los tumores en mujeres mayores de 50 años son RE+. Los tumores de mama RE+ y/o RP+ confieren un riesgo bajo de mortalidad comparado con tumores RE- y/o RP- (Payne *et al.*, 2008). La importancia del RE se deriva principalmente de los ligandos estrogénicos, fundamentalmente el estradiol, mitógeno más importante en el cáncer de mama RE+, de tal manera que este biomarcador es el blanco directo de las terapias endocrinas (Dowsett y Dunbier, 2007). La expresión del receptor de progesterona es muy dependiente de estrógeno y es raramente observado en tumores RE- (Payne *et al.*, 2008).

El propósito principal de la determinación de estos marcadores es seleccionar la probable respuesta a la terapia endocrina en pacientes con cáncer temprano o avanzado. Estos receptores también pueden ser empleados para fines pronósticos. En combinación con otros factores de pronóstico establecidos (estadio y grado del tumor, así como el número de nódulos linfáticos metastásicos), estos receptores

pueden utilizarse para determinar el pronóstico a corto plazo en pacientes con cáncer de mama recién diagnosticado, así como el riesgo de recurrencia (Sturgeon *et al.*, 2008).

1.7.1.2 HER2

El receptor HER2 se asocia con tumores de grado alto, con la presencia de nódulos linfáticos, así como alta recurrencia y mortalidad (Payne *et al.*, 2008). El gen *HER2* está amplificado y sobreexpresado en un 20 a 30% de los carcinomas ductales invasivos y se le reconoce ampliamente como un predictor de resultados clínicos pobres con una reducción en la tasa de supervivencia (Freudenberg *et al.*, 2009). HER2 es un marcador predictivo y de pronóstico, recomendado para la evaluación en cáncer de mama, así como para definir el pronóstico en pacientes con cáncer en estadios tempranos en ausencia de terapia sistémica y para predecir la respuesta a agentes quimioterapéuticos específicos (Harris *et al.*, 2007; Sturgeon *et al.*, 2008).

1.7.2 Marcadores en suero

Los usos potenciales de los marcadores de suero en cáncer de mama incluyen un diagnóstico temprano, determinación del pronóstico, eventual predicción de la respuesta o resistencia a terapias específicas, supervivencia posterior a una cirugía primaria y monitoreo de las terapias en pacientes con enfermedad avanzada (Duffy, 2006). Los marcadores tumorales séricos en cáncer de mama no están bien establecidos; sin embargo los marcadores utilizados más ampliamente son el CA 15-3 y el antígeno carcino-embriogénico (CEA). Otros marcadores menos empleados son BR 27.29 (también conocido como CA27.29), el antígeno polipeptídico tisular

(TPA), el antígeno polipeptídico específico tisular (TPS) y una forma de HER-2 (Duffy, 2006; Harris *et al.*, 2007).

1.7.2.1 CA15-3

El antígeno de carcinoma CA15-3 es un epítoto de una glicoproteína transmembranal denominada MUC1. En tejido normal de mama, MUC1 se expresa en los ductos y acinos, desde donde es liberada en forma soluble dentro de la leche o se une a los glóbulos de grasa de la leche. En la transformación neoplásica se asocia con la ruptura de la polarización normal de las células y de la arquitectura del tejido, conduciendo a que MUC1 entre al torrente sanguíneo, en donde puede ser detectada y cuantificada mediante inmunoensayos (Seregini *et al.*, 2004).

Como en casi todos los marcadores séricos, la falta de sensibilidad en estadios *in situ* o en estadios bajos de enfermedad invasiva y la falta de especificidad para el cáncer de mama, impide el uso de CA 15-3 para detectar la enfermedad de manera temprana. De hecho, en pacientes con cáncer de mama temprano o localizado, los niveles en suero de CA 15-3 correlacionan en gran medida con los niveles encontrados en mujeres sanas y en mujeres con enfermedades benignas de mama. Aunque niveles elevados de CA 15-3 se encuentran en la mayoría de pacientes con cáncer de mama metastásico, también se han encontrado en concentraciones altas en pacientes con diversos tipos de adenocarcinomas avanzados (Duffy *et al.*, 2010).

El uso principal de este biomarcador es para monitorear la terapia en pacientes con enfermedad metastásica; sin embargo, no se debe usar solo, sino que la medición debe ser acompañada por el diagnóstico con imágenes, historia clínica y examen físico. CA 15-3 es particularmente valioso para el seguimiento del tratamiento en pacientes que tienen enfermedades que no pueden ser evaluadas utilizando procedimientos radiológicos (Sturgeon *et al.*, 2008; Duffy *et al.*, 2010).

1.7.2.2 CEA

El antígeno carcino-embrionario es una glicoproteína que está presente en tejidos embrionarios y en algunos tumores epiteliales, particularmente en los malignos. En la vida embrionaria se sintetiza en el páncreas y en el tracto gastrointestinal, y sus valores disminuyen drásticamente en la vida adulta, pero pueden aumentar en diferentes carcinomas como el de mama, ovario, estómago y colon, entre otros. Este marcador también se ha encontrado aumentado en otras patologías como hepatitis, colitis ulcerosa, infecciones del tracto gastrointestinal, etc. Por lo tanto, este marcador no puede ser utilizado con fines diagnósticos en carcinoma mamario. Sin embargo, su utilidad clínica en el cáncer de mama reside en el monitoreo de pacientes con enfermedad metastásica durante una terapia activa, pero sólo en conjunto con otros métodos como son el diagnóstico a través de imagenología, historia clínica y examen físico (Harris *et al.*, 2007).

1.7.3 Marcadores de expresión génica

Las emergentes tecnologías genómicas han incrementado la tasa de descubrimiento de nuevos marcadores potenciales y han facilitado el desarrollo de perfiles de expresión génica, compuestos de varios genes que pretenden proveer información predictiva o de pronóstico acerca de los tumores. Dentro de las plataformas más reconocidas plataformas se encuentran Mammaprint, OnctypeDx y Rotterdam signature (Harris *et al.*, 2007).

1.7.3.1 Mamaprint

El Instituto del Cáncer de Holanda (*The Netherlands Cancer Institute*) mediante el estudio de microarreglos de ADN y empleando una clasificación

supervisada, definió 70 genes relacionados con la regulación de la proliferación, invasión, metástasis, integridad del estroma y angiogénesis, cuya determinación se lleva a cabo en una muestra de tejido tumoral fresco o congelado. Esta plataforma fue aprobada por la FDA en el 2007 y sólo ha sido validada como un análisis de pronóstico para mujeres menores de 61 años con cáncer de mama, independientemente del estado de los receptores hormonales (RE positivo o negativo) y con cáncer nódulo linfático negativo. Esta prueba sólo determina si el cáncer presenta un riesgo de recurrencia bajo o alto en el transcurso de los diez años posteriores al diagnóstico (Harris *et al.*, 2007; Dowsett y Dunbier, 2008).

1.7.3.2 Oncotype DX

Es un ensayo de RT-PCR de pronóstico y predictivo que mide la expresión de 21 genes (16 genes relacionados con cáncer y 5 genes de referencia) en muestras embebidas en parafina. Esta prueba determina la probabilidad de residencia a 10 años en mujeres diagnosticadas con cáncer de nódulo linfático negativo o receptor de estrógenos positivo. Oncotype ha sido aprobado por la Agencia de Laboratorios del Estado de California, pero no ha sido sometido a la FDA para su aprobación (Ross *et al.*, 2008).

1.7.3.3 Rotterdam signature

Es una prueba de expresión génica que consiste en un microarreglo de 76 genes, desarrollado por el Centro de Cáncer Universitario Erasmus en Rotterdam. Los genes en este arreglo no se encuentran en las pruebas de Oncotype o Mammaprint y básicamente están implicados en el proceso de proliferación. Este ensayo está validado para pacientes con nódulo linfático negativo, independientemente de sus receptores hormonales y requiere material biológico

fresco o congelado. Esta prueba ha sido aprobada por la FDA, pero aún no se ha lanzado a la venta (Dowsett y Dunbier, 2008; Ross *et al.*, 2008).

1.8 Tratamiento en cáncer de mama

El tratamiento primario para cáncer de mama localizado es cirugía conservadora de la mama y radiación o mastectomía. Seguido al tratamiento primario, muchas mujeres con carcinoma de mama invasivo reciben terapia adyuvante sistémica como quimioterapia, terapia hormonal o una combinación de quimioterapia y terapia hormonal. Tanto la quimioterapia adyuvante como la terapia hormonal han mostrado reducir la recurrencia sistémica y la mortalidad por este tipo de cáncer (Sturgeon *et al.*, 2008).

1.9 Proteómica y espectrometría de masas

La proteómica representa uno de los últimos avances tecnológicos para la búsqueda de biomarcadores en general y particularmente en cáncer, ya que posee el potencial para complementar la información generada por la genómica, debido a que los niveles de mRNA no necesariamente correlacionan con la abundancia de las proteínas correspondientes. Además, la complejidad del proteoma (conjunto de proteínas expresada por una célula en un momento dado) aumenta por las modificaciones postraduccionales (PTMs) como fosforilación, acetilación, glicosilación, entre otras más de cien existentes, las cuales no son detectadas a nivel de mRNA, pero desempeñan un papel importante en la función de las proteínas (Goncalves y Bertucci, 2011). Por lo tanto, la proteómica es el estudio global de las proteínas, que abarca la identificación, caracterización y cuantificación del conjunto

de proteínas expresadas por el genoma en un tiempo de la vida de una célula, tejido u organismo, incluyendo isoformas y modificaciones, interacciones proteína-proteína y la descripción estructural de proteínas y sus complejos (Lane, 2005). Una herramienta fundamental para el estudio de los proteomas es la espectrometría de masas (MS), la cual determina con alta sensibilidad la relación masa carga (m/z) de iones en fase gaseosa (Griffiths *et al.*, 2001).

Un experimento de proteómica consta de cinco etapas: (1) Extracción de proteínas a partir de tejidos, suero, líneas celulares o cualquier otra fuente biológica. Esta etapa incluye la separación de las proteínas a través de geles uni o bidimensionales o por cromatografía multidimensional. (2) Digestión de proteínas con alguna enzima (en general tripsina) para la generación de péptidos. (3) Enriquecimiento a través de cromatografía de fase reversa (C18) o separación de los péptidos por uno o más pasos de cromatografía líquida de alta resolución en capilares finos, de los cuales se eluye la muestra a la fuente de iones, donde son nebulizados en pequeñas gotas altamente cargadas. (4) Análisis de los péptidos a través del espectrómetro de masas, que genera un espectro de masas MS. (5) Selección de algunos iones precursores de los péptidos para su análisis por espectrometría de masas en tándem o MS/MS. En este paso la fragmentación de los iones precursores genera iones fragmento (Figura 7). Los análisis MS y MS/MS generan espectros de masas en donde se puede observar la intensidad relativa del ion como función de la relación masa/carga (m/z), la cual es definida como la relación del número de masa (m) de una partícula dada entre el número (z) de unidades cargadas electrostáticamente. Los datos obtenidos de los espectros MS y MS/MS se comparan en bases de datos, mediante algoritmos específicos para la identificación de las proteínas (Aebersold y Mann, 2003).

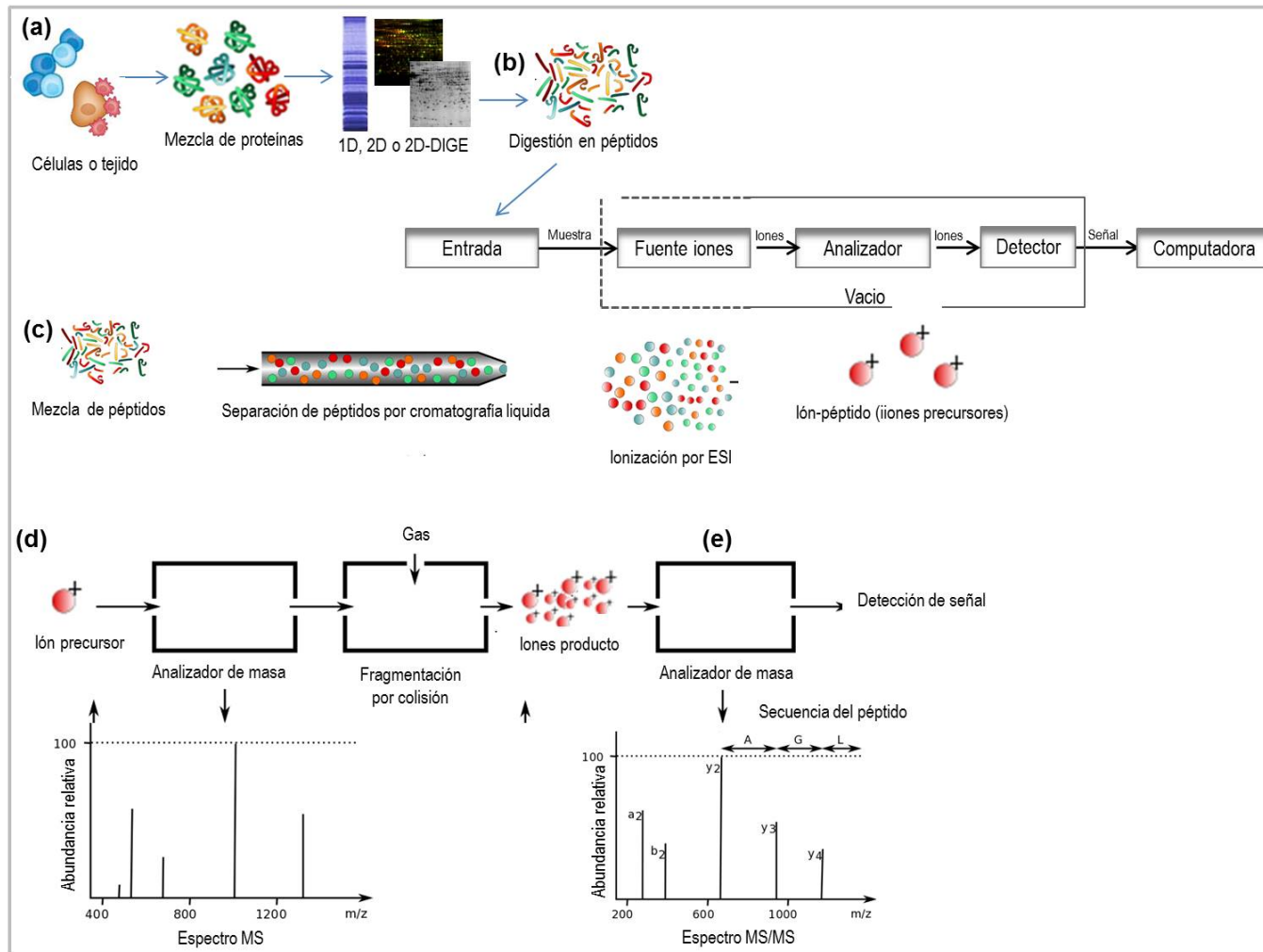


Figura 7. Análisis proteómico basado en espectrometría de masas. En la figura se representan las cinco etapas principales, de las que consta un experimento típico en proteómica. (a) Extracción y separación de proteínas, (b) digestión de proteínas en péptidos, (c) separación de péptidos por cromatografía acoplada al espectrómetro de masas, (d) generación de espectros MS y (e) MS/MS (Modificado de Aebersold y Mann, 2003).

1.9.1 Espectrómetro de masas

Un espectrómetro de masas es un instrumento que separa los iones que se desplazan rápidamente según su relación m/z . El espectrómetro de masas consiste de tres componentes básicos: una fuente de iones, un analizador de masas, que mide la relación m/z de los analitos ionizados y un detector que registra el número de iones de cada valor m/z (Figura 8) (Aebersold y Mann, 2003; Lane, 2005). Debido a que el análisis en el espectrómetro de masas requiere de campos electromagnéticos y de vacío, las moléculas son cargadas eléctricamente y transferidas a una fase gaseosa. Una vez en la fase gaseosa, la relación m/z es determinada por el tiempo que tardan en llegar al detector. El sistema de entrada, junto con la fuente de iones y el tipo de analizador de masas definen el tipo de espectrómetro y las capacidades del sistema (Skoog *et al.*, 2001) (Walther y Mann, 2010).

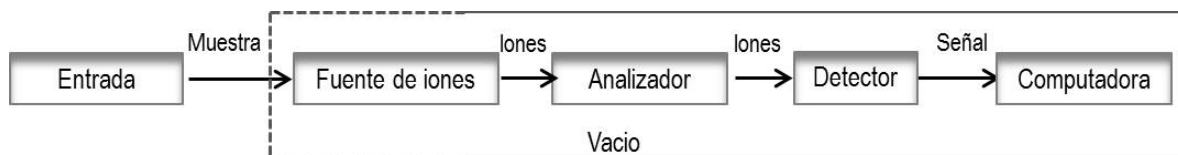


Figura 8. Componentes de un espectrómetro de masas. El espectrómetro de masas consta de un sistema de entrada de la muestra, una fuente de iones, un analizador y un detector (Modificado de Lane, 2005).

1.9.1.1 Fuentes de ionización

La función de las fuentes de ionización es convertir los componentes de las muestras en iones. Las técnicas de ionización son críticas para generar moléculas con carga o ionizadas que pueden ser manipuladas dentro de un campo eléctrico o magnético. Existen dos categorías principales de fuentes de ionización: *fuentes de fase gaseosa* y *fuentes de desorción*. En las primeras, la muestra es volatilizada y después es ionizada. Algunos ejemplos de esta son: Impacto de electrones (EI),

ionización química (CI) y la ionización por campo (FI). Por otro lado, en las fuentes de desorción, la muestra en estado sólido o líquido se transforma directamente en iones gaseosos. Dentro de estas se encuentran: Desorción por campo (FD), desorción por plasma (PD), bombardeo con átomos rápidos (FAB), espectrometría de masas de iones secundarios (SIMS) e ionización por termonebulización (TS). Sin embargo, las técnicas más comúnmente utilizadas en proteómica son las fuentes de desorción: Ionización y desorción de la matriz asistida por láser (MALDI) y Ionización por Electrospray (ESI) (Skoog *et al.*, 2001; Aebersold y Mann, 2003). Cabe mencionar, que las fuentes de iones también se clasifican en *fuentes duras o fuentes blandas*. Las primeras comunican energía suficiente a las moléculas, de manera que se encuentran en un estado de energía altamente excitado; la relajación posterior implica la rotura de las uniones, produciendo la fragmentación de los iones con una relación masa/carga menor que la del ion molecular. Las fuentes blandas dan lugar a poca fragmentación, generando así un espectro de masas con un pico del ion molecular y sólo alguno o incluso ningún otro pico (Skoog *et al.*, 2001; El-Aneed *et al.*, 2009).

1.9.1.1.1 MALDI

MALDI o "Matrix assisted laser desorption ionization" es una técnica de ionización suave que conduce a la ionización y vaporización de grandes biomoléculas no volátiles tales como proteínas intactas (Emonet *et al.*, 2010). En esta técnica de ionización se generan mayoritariamente iones de una sola carga ($z=1$) de manera que la relación de masa carga (m/z) del analito corresponde al valor de su masa (De Carolis *et al.*, 2014). En MALDI, las muestras se preparan mezclando el analito con una matriz orgánica que usualmente es una estructura conjugada con un anillo aromático, la cual absorbe energía en la longitud de onda del láser (Wysocki *et al.*, 2005). Las matrices DHB (ácido 2,5-dihidroxibenzoico) y CHCA (α -ciano-4-hidroxicinámico) son las matrices más utilizadas para la detección de iones de menor

masa y las más utilizadas en proteómica. Una vez que la muestra cocrystaliza con la matriz, esta última absorbe la energía del láser, dando lugar a la desorción e ionización de los analitos en la fase gaseosa. Posteriormente, los iones son acelerados a través de un campo electrostático dentro del tubo de vuelo al alto vacío hasta que estos alcanzan el detector. Los iones más pequeños viajan más rápidamente que los grandes y por lo tanto alcanzan primero el detector. Así que, el tiempo de vuelo (TOF) requerido para alcanzar el detector depende de la masa y la carga de los bioanalitos (Figura 9a) (Wysocki *et al.*, 2005; De Carolis *et al.*, 2014).

1.9.1.1.2 ESI

ESI o ionización por electroespray es una técnica de ionización suave ampliamente utilizada para la producción de iones en fase gaseosa (sin que se produzca fragmentación) de grandes supramoléculas térmicamente lábiles (Banerjee y Mazumdar, 2012). ESI utiliza energía eléctrica para ayudar a la transferencia de los iones de la solución en la fase gaseosa antes de que sean sometidos al análisis de espectrometría de masas (Ho *et al.*, 2003). La transferencia de especies iónicas de una solución a la fase gaseosa, implica la dispersión de un aerosol fino de gotas cargadas, seguido por la evaporación del solvente y la eyección de iones de las gotas altamente cargadas por una explosión colómbica (Ho *et al.*, 2003).

El proceso general de ESI, ocurre en la punta del emisor (punta de acero inoxidable o un tubo capilar de cuarzo), en donde, una corriente continua de la solución de analitos se dispersa en una fase gaseosa. La aplicación de un potencial de 3-4 kV genera que el líquido comience a salir del emisor, incrementando su carga y asumiendo una forma cónica (cono de Taylor). La aplicación de un gas de nebulización (por ejemplo, nitrógeno) y/o temperatura aumenta el grado de flujo de la muestra y a su vez provoca que las gotas cargadas reduzcan su tamaño continuamente, debido a la evaporación del disolvente. En la punta del cono, el líquido cambia de forma a una línea fina, que se vuelve inestable ya que es forzado a retener más y más cargas, y

finalmente llega a un punto crítico donde no puede soportar más carga eléctrica provocando la dispersión de la solución en forma de niebla de pequeñas gotas altamente cargadas que vuelan buscando una superficie de carga opuesta. Debido a que las gotas están altamente cargadas con la misma carga eléctrica se repelen fuertemente, las gotas vuelan y se dispersan cubriendo un área cada vez mayor y se van reduciendo de tamaño ya que las moléculas de solvente se evaporan en su superficie, y la distancia entre las moléculas cargadas disminuye dramáticamente. Si la gota no encuentra dónde disipar su carga, las cargas eléctricas llegan a un estado crítico y la gota explota violentamente. El proceso de evaporación y rompimiento de gotas se repite hasta que el tamaño y carga de las gotas desorbe los iones dentro de la fase gaseosa, donde pueden ser dirigidos al espectrómetro de masas por medio de campos eléctricos (Figura 9b). Finalmente, el sistema ESI está acoplado frecuentemente con sistemas cromatográficos o electroforesis capilar para el análisis de mezclas altamente complejas (Ho *et al.*, 2003), (Wysocki *et al.*, 2005; El-Aneed *et al.*, 2009; Yates *et al.*, 2009).

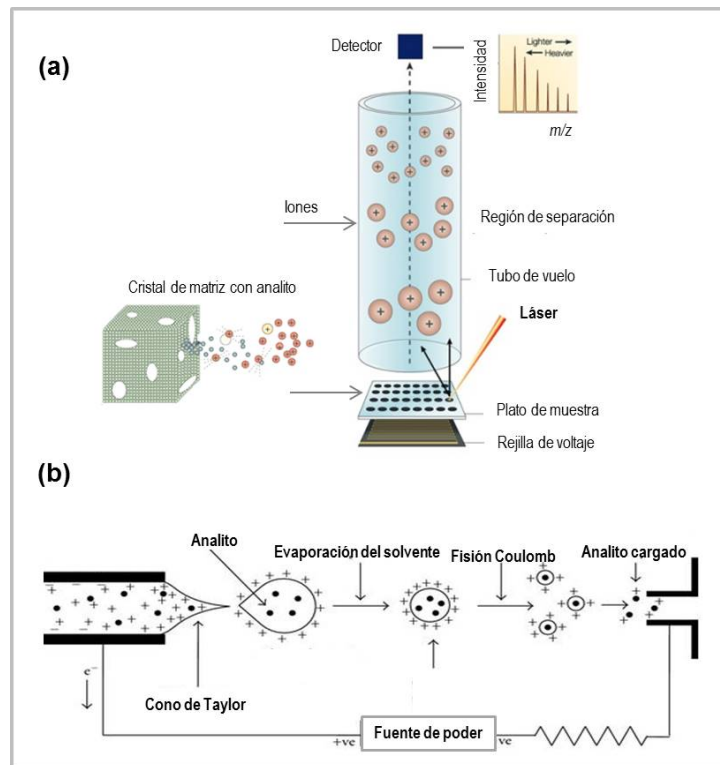


Figura 9. Técnicas de ionización utilizadas en proteómica. Las fuentes de ionización convierten los componentes de las muestras en iones. (a) Ionización a través de MALDI, en donde la muestra se co-cristaliza con una matriz, la cual absorbe la energía del láser, provocando la desorción e ionización de los analitos. Los iones generados son acelerados con un voltaje eléctrico constante en el tubo de vuelo hasta que estos alcanzan el detector (Modificada de Kim *et al.*, 2003). (b) Ionización por ESI, en donde una aguja de metal se mantiene a un alto potencial eléctrico, generando así una corriente continua de muestra y la formación de un cono de Taylor. La evaporación del solvente de la muestra provoca que las gotas altamente cargadas se aproximen, repelen y finalmente estallen debido a la repulsión de cargas. El proceso se repite hasta que el tamaño y carga de las gotas desorba los iones dentro de la fase gaseosa, donde son dirigidos al analizador de masas (Modificada de Banerjee y Mazumdar, 2011).

1.9.1.2 Analizadores de masas

El analizador de masas separa los iones de acuerdo a su relación masa/carga, basado en sus características de comportamiento en campos eléctricos y/o magnéticos (Smith *et al.*, 2002; El-Aneed *et al.*, 2009). Las fuentes de ionización ESI

y MALDI se pueden combinar con diferentes analizadores de masas; entre los más comúnmente utilizados en el análisis de biomoléculas se encuentran: Cuadrupolo (Q), trampa de iones (trampa de iones con cuadruplo QIT, trampa de iones lineal LIT o LTQ), tiempo de vuelo (TOF) y resonancia de ión en ciclotrón con transformada de Fourier (FTICR), los cuales varían en sus principios físicos y rendimientos analíticos. También existen los instrumentos “híbridos”, que han sido diseñados para combinar las capacidades de diferentes analizadores de masas Q-q-Q, Q-q-LIT, Q-TOF, TOF-TOF y LTQ-FTICR (Figura 10) (Yates *et al.*, 2004; Domon y Aebersold, 2006).

1.9.1.2.1 Analizador de trampa de iones

En los analizadores ion-trap o trampa de iones, los iones son capturados o “atrapados” y pueden ser acumulados en un cierto intervalo de tiempo, para luego ser sometidos a un análisis MS o MS/MS. Las trampas de iones también son llamadas “trampas de iones cuadrupolo” (Aebersold y Mann, 2003). En este tipo de analizador, los iones se llevan al interior de la trampa y se someten a campos eléctricos oscilantes para atrapar los iones en forma controlada. Los iones son atrapados en un espacio tridimensional compuesto por: un electrodo circular y dos electrodos hiperbólicos en los extremos. La trampa de iones es operada al aplicar un potencial sinusoidal (rf) con una frecuencia determinada al electrodo circular, mientras que los de los extremos se mantienen constantes (frecuentemente en cero), o mantenidos a un potencial oscilante. Las oscilaciones de los iones atrapados son reducidas por colisiones con helio a alta presión. Dado que las frecuencias de oscilación de los iones son en función de su masa, los iones de valores diferentes de m/z pueden salir de la trampa a diferentes voltajes y, por lo tanto, a diferentes tiempos. Para un análisis MS/MS los iones son seleccionados aplicando voltajes a los electrodos de los extremos para expulsar selectivamente todos los iones en la trampa a excepción del ion precursor seleccionado. Posteriormente, el ion precursor comienza a oscilar y colisiona con el helio, que eventualmente induce la fragmentación del ion precursor (CAD). Los iones producto de la fragmentación

iónica tienen frecuencias de resonancia diferentes que el ion primario, por lo tanto, estos iones se amortiguan en el centro de la trampa debido al helio. Después de un período de tiempo, tanto el voltaje como la radiofrecuencia (Rf) en el electrodo circular es aumentado, causando la expulsión de los iones (Aebersold y Mann, 2003; Lane, 2005).

1.9.1.2.2 Analizador Cuadrupolo

Este tipo de analizador está compuesto por un conjunto de cuatro barras cilíndricas paralelas que actúan como electrodos. Las barras opuestas se conectan eléctricamente, un par está unido al polo positivo de una fuente variable de corriente continua y el otro par se une con la terminal negativa. Además, se aplican a cada par de barras potenciales variables de corriente alterna de radiofrecuencia.

En este tipo de analizadores de masas, los iones se extraen de la fuente de iones y son acelerados dentro del espacio central formado por el cuadrupolo a lo largo del eje longitudinal hacia el detector. Entre tanto, las tensiones de corriente continua y de corriente alterna se incrementan simultáneamente, mientras se mantiene constante su relación. En cualquier momento, todos los iones excepto aquéllos que tengan un determinado valor m/z inciden en las barras y se convierten en moléculas neutras. Por lo tanto, sólo los iones cuyo valor de m/z esté dentro de un intervalo limitado alcanzarán el detector (Figura 10c) (Skoog *et al.*, 2001; Aebersold y Mann, 2003).

Para un análisis MS/MS, tres cuadrupolos pueden configurarse juntos (triple cuadrupolo). El primero (Q1) y el tercer cuadrupolo (Q3) son utilizados para el escaneo de iones, mientras que el segundo cuadrupolo (q2) se utiliza como una celda de colisión. Los iones en q2 se fragmentan por la disociación activada por colisión (CAD), la cual es una colisión de baja energía con un gas de fondo como el nitrógeno (Lane, 2005).

1.9.1.2.3 Analizador TOF

El principio de operación del analizador de tiempo de vuelo (TOF) involucra la medición del tiempo requerido por un ion para viajar desde la fuente de iones hasta el detector. Teóricamente, todos los iones son formados a la vez y todos reciben la misma energía cinética durante la aceleración, pero debido a que tienen valores de m/z diferentes, se separan en grupos de acuerdo con su velocidad (por lo tanto, m/z). Los iones de m/z baja llegan al detector antes que aquéllos con m/z alta, esto, debido a que entre mayor sea el valor de m/z que tengan los iones tendrán una velocidad menor (Aebersold y Mann, 2003; Lane, 2005). La resolución de masa es afectada por ligeras variaciones en el tiempo de vuelo, así como por factores que crean una distribución en los tiempos de vuelo entre iones con la misma relación m/z , dando como resultado una resolución de masas pobre. Se utilizan dos técnicas para compensar la distribución temporal (tiempo deformación de iones), espacial (localización de la formación de iones) y cinética (energía de formación). La primera de ellas consta de la introducción de un tiempo de retardo entre la formación de iones y la extracción de estos de la fuente, proceso denominado “retraso de extracción”. En este proceso, los iones son expandidos dentro de una región libre de campo en la fuente de iones y después de un cierto retardo (nanosegundos a microsegundos) se aplica un pulso de voltaje para extraer los iones fuera de la fuente. La segunda estrategia es el uso de espejos de iones o reflectrones, los cuales compensan las variaciones en la distribución de energía. El reflectrón crea un campo de retardo que desvía los iones, enviándolos de vuelta a través del tubo de vuelo. Aquellos iones con más energía que otros, penetran más profundamente dentro del reflectrón, antes de ser reflectados. Por lo tanto, un ion con más energía tendrá una trayectoria de vuelo más larga y llegará al detector al mismo tiempo que aquellos iones menos energéticos de la misma masa (Figura 10a) (Lane, 2005).

1.9.1.2.4 Analizador TOF-TOF

Los analizadores TOF-TOF incorporan una celda de colisión entre las dos secciones TOF que poseen. Los iones de una relación m/z dada (ion precursor), son seleccionados en la primera sección TOF, fragmentados por la celda de colisión y finalmente los iones fragmento son separados en la segunda sección TOF (Figura 10b). Esta configuración permite sensibilidad y resolución altas, tanto en MS como en MS/MS (Guerrera y Kleiner, 2005).

1.9.1.2.5 Analizador Q-TOF

Los instrumentos híbridos Q-TOF poseen una celda de colisión que se encuentra entre un filtro de masas cuadrupolo y el analizador TOF (Aebersold y Mann, 2003). Los iones de una relación m/z particular (ion precursor) son seleccionados en el filtro de masas cuadrupolo (Q1), fragmentados en la celda de colisión (q2) y los iones fragmento generados son finalmente analizados en la sección TOF (Figura 10d). Este tipo de instrumentos puede intercambiar la fuente de ionización entre MALDI y ESI. El instrumento Q-TOF tiene una sensibilidad alta, gran poder de resolución, y una gran precisión en masa (Loboda *et al.*, 2000; Guerrera y Kleiner, 2005).

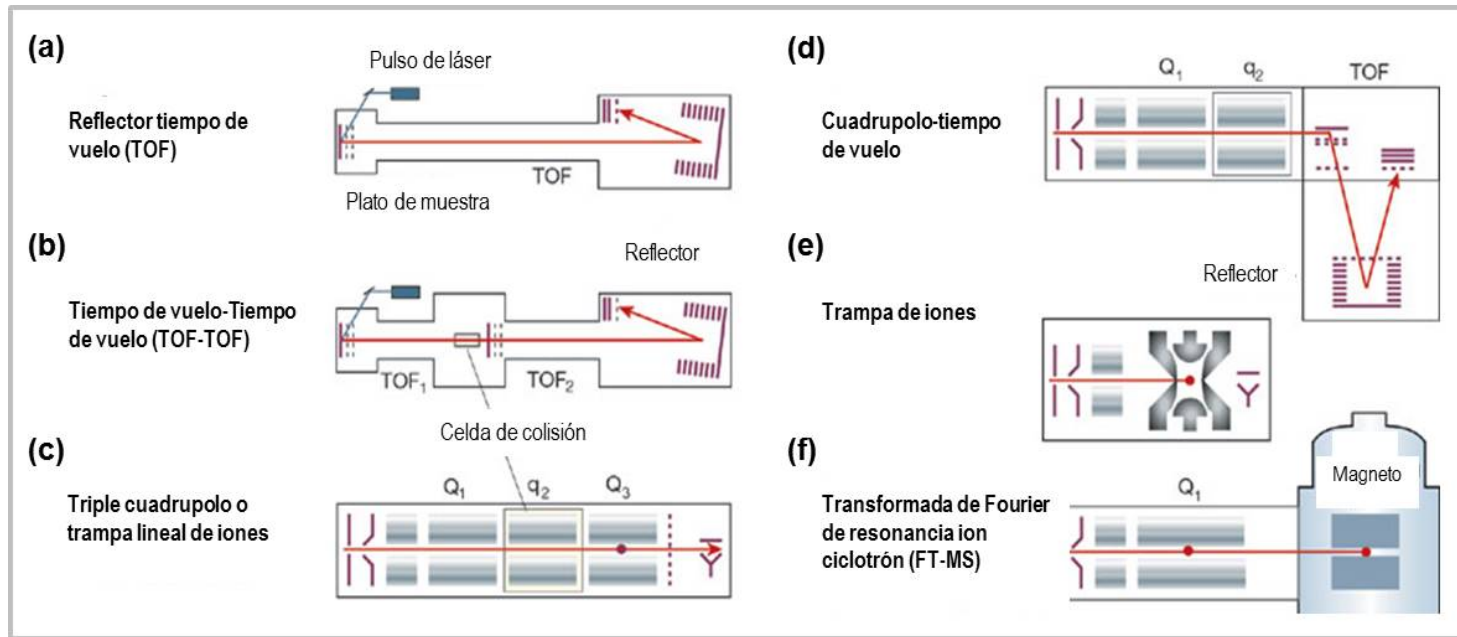


Figura 10. Analizadores de masas utilizados en proteómica. En la figura se representan diversos analizadores de masas. (a) TOF: Los iones son acelerados a una energía cinética alta y son separados a lo largo de un tubo de vuelo como resultado de sus diferentes velocidades. Los iones son reflejados en un reflector para compensar las ligeras diferencias en la energía cinética. (b) TOF-TOF: Incorporan una celda de colisión entre dos secciones TOF. Los iones de una m/z dada son seleccionados en la primera sección TOF, fragmentados en la celda de colisión y las masas de los iones fragmentos generados son separadas en la segunda sección TOF. (c) Triple cuadrupolo: Selecciona por tiempo los iones entre cuatro varillas. Los iones de una dada m/z son seleccionados (Q₁), fragmentos en una celda de colisión (Q₂) y los iones fragmento son separados (Q₃). En la trampa de iones lineal, los iones son capturados en Q₃, excitados por medio de campos eléctricos y los iones fragmento son escaneados. (d) Q-Tof: Combina la parte frontal de un instrumento triple cuadrupolo con una sección TOF-reflectrón para medir la masa de los iones. (e) Trampa de iones: Captura los iones como en el caso de la trampa de iones lineal. (f) FT-MS: Atrapa los iones con ayuda de campos magnéticos fuertes. La figura muestra la combinación de FT-MS con una trampa de iones lineal para el aislamiento eficiente, la fragmentación y la detección de iones fragmento en la sección FT-MS (Modificado de Aebersold y Mann, 2003).

1.9.1.3 Detectores de masas

El detector de masas convierte el haz de iones en una señal eléctrica que puede ser procesada y almacenada. La sensibilidad, precisión y tiempo de respuesta son criterios importantes que distinguen los diferentes sistemas de detección de iones. Una precisión alta y tiempo de respuesta rápido son características mutuamente excluyentes (Oda *et al.*, 2001). Existen diversos tipos de detectores como son: Copa de Faraday, canal multiplicador de electrones, Daly, detector de microcanales, entre otros (Watson y Sparkman 2007).

1.10 Estrategias de análisis en Proteómica

En años recientes, la MS ha emergido como una poderosa herramienta para identificar proteínas de forma rápida y eficiente en muestras biológicas (Aebersold y Mann, 2003). Debido a esto, las tecnologías proteómicas pueden clasificarse de acuerdo con la implicación o no de la MS. Típicamente las aproximaciones no basadas en MS corresponden a técnicas de microarreglos (tejido y proteínas), y requieren de una hipótesis preliminar o una selección de proteínas para ser analizadas, mientras que las tecnologías basadas en MS no requieren ningún conocimiento biológico previo y además permiten la examinación y cuantificación de un gran número de parámetros proteicos inicialmente desconocidos (Goncalves y Bertucci, 2011). Por otro lado, la proteómica basada en MS es una disciplina que ha sido posible por la creación y disposición de bases de datos de secuencias de genes y genomas y se ha establecido por sí misma en una herramienta indispensable para interpretar la información codificada por los genomas (Aebersold y Mann, 2003). Esta tecnología puede utilizarse para determinar la composición de proteínas de una célula, para caracterizar a los miembros de complejos proteicos, su arquitectura, la composición de proteínas de organelos y la dinámica de estos procesos biológicos

(Walther y Mann, 2010). En la identificación de proteínas vía MS existen dos aproximaciones principales: Top-down (TD) y Bottom-up (BU) “proteomics” (Han *et al.*, 2008; Yates *et al.*, 2009).

1.10.1 Aproximación Top-down (TD)

Top-Down (de arriba hacia abajo), se centra en la caracterización e identificación de proteínas intactas a partir del análisis por MS de las mismas (Figura 11a) (Kelleher, 2004; Bogdanov y Smith, 2005). TD identifica las proteínas determinando su masa molecular relativa y/o junto con la fragmentación de los iones proteicos obtenidos, generando así información sobre su secuencia (Bagdanov y Smith, 2005). La fragmentación de las proteínas se da a través de disociación por captura de electrones (ECD) y disociación por transferencia de electrones (ETD). Tanto ECD y ETD producen una secuenciación más completa y conservan las modificaciones postraduccionales (PTMs) lábiles (Kelleher *et al.*, 1999; Yates *et al.*, 2009). Los espectrómetros de masas utilizados en esta aproximación son: Transformada de Fourier (FTICR) (Bogdanov y Smith, 2005) o trampas iónicas tipo Orbitrap (Hu *et al.*, 2005). La identificación de proteínas en TD se realiza a través de búsqueda en bases de datos de espectros de iones fragmento sin interpretar (Meng *et al.*, 2001), o a través de la determinación “de novo” de la secuencia completa de amino ácidos (Horn *et al.*, 2000).

1.10.2 Aproximación Bottom-up

En BU (de abajo hacia arriba) las proteínas son digeridas con proteasas para generar péptidos, posteriormente las muestras se analizan por MS y las subsiguientes relaciones de m/z y las secuencias de péptidos son utilizadas para identificar las proteínas correspondientes. La mayoría de las aplicaciones BU requieren adquisiciones de masas en tándem (MS/MS), en las cuales los péptidos

son sujetos a una disociación activada por colisión (CAD o CID) (Figura 11b). El método utilizado más ampliamente en la identificación de datos en tándem es la búsqueda en bases de datos, en las cuales los espectros de masas experimentales son comparados con espectros de masas teóricos o con otros experimentales (Han *et al.*, 2008; Yates *et al.*, 2009).

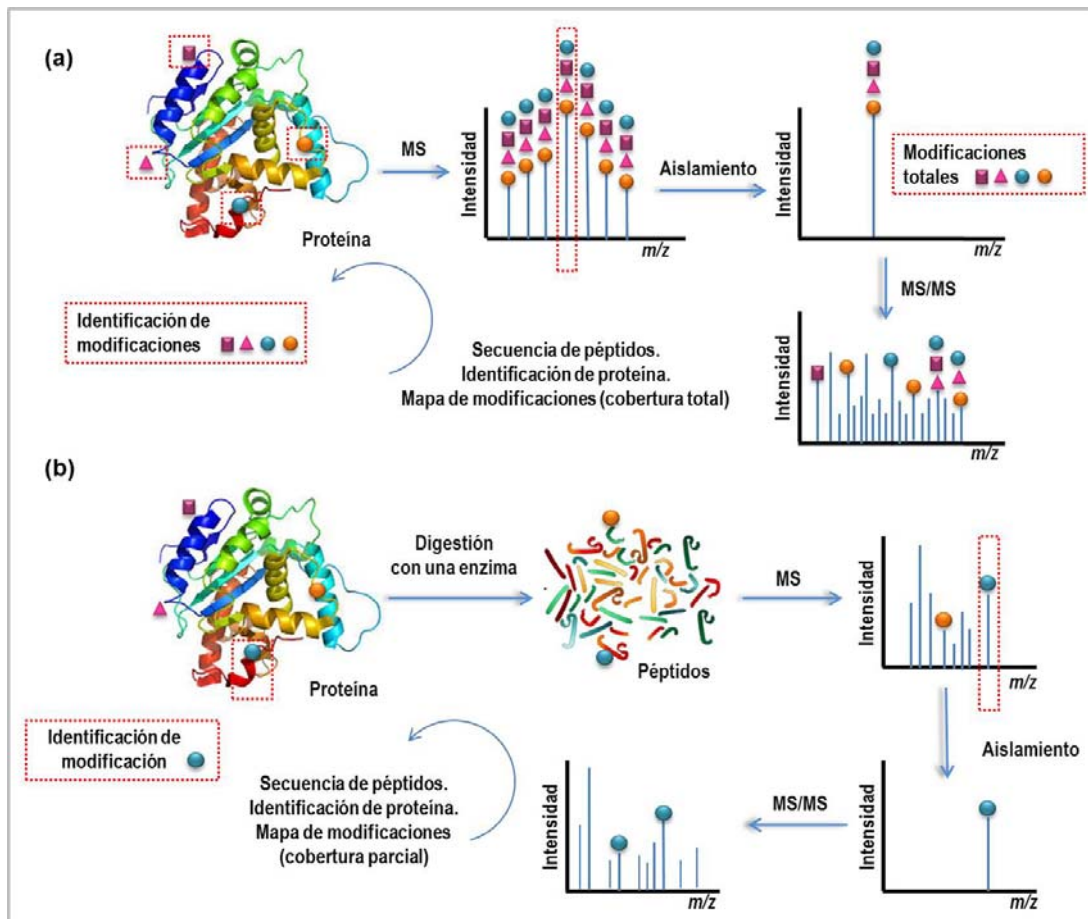


Figura 11. Estrategias de análisis en proteómica. Principales aproximaciones para la identificación de proteínas. (a) Top-Down: La proteína completa es analizada directamente por MS, no se requiere de digestión. La proteína se aísla y fragmenta por MS/MS para localizar sitios de modificaciones. Todas las modificaciones pueden ser identificadas con una cobertura total en la secuencia. (b) Bottom-up. Las proteínas provenientes de gel o solución son digeridas con tripsina u otra enzima para generar péptidos. Los péptidos son detectados por MS y los péptidos de una m/z específica pueden aislarse y fragmentarse por MS/MS para la identificación de las proteínas. No todas las modificaciones pueden ser detectadas, debido a que muchos péptidos no son recuperados ni detectados por MS, resultado en una cobertura parcial de la secuencia (Modificada de Zhang y Ge, 2011).

1.11 Proteómica cuantitativa

La capacidad de determinar cambios cuantitativos en la expresión de proteínas en respuesta a una variedad de estímulos es una de las metas más importantes en proteómica. El desarrollo de métodos para una cuantificación precisa de proteínas es actualmente una de las áreas en proceso de cambio. La proteómica cuantitativa se encuentra dividida en dos tipos: absoluta y relativa (Elliott *et al.*, 2009). La cuantificación absoluta determina los cambios en la expresión de proteínas en términos de cantidad o concentración exacta (Elliott *et al.*, 2009). El uso de estándares internos ha sido durante mucho tiempo una herramienta para la cuantificación absoluta de moléculas pequeñas en experimentos de dilución de isótopos. La cuantificación se realiza adicionando cantidades conocidas de alguna marca isotópica a un analito conocido, posteriormente, las muestras se analizan a través de espectrometría de masas y finalmente los niveles relativos de marcaje son calculados. Debe tenerse en cuenta que en este caso la identidad del analito es conocida antes del análisis. Variaciones en el método de estándares internos han sido desarrollados para su aplicación en proteómica, siendo la cuantificación absoluta de proteínas (AQUA) de los más representativos (Simpson *et al.*, 2009). La MS en tándem ha mejorado considerablemente la precisión y especificidad de la cuantificación de analitos a través del uso de un monitoreo selectivo de una reacción (SRM), principalmente de los pares de iones precursores (Elliott *et al.*, 2009).

La cuantificación relativa puede comparar dos o más muestras relacionadas con una muestra control y los resultados se expresan como “veces de expresión” en incremento o disminución (Elliott *et al.*, 2009). Este tipo de cuantificación emplea métodos estables de marcaje isotópico o métodos libres de marcaje. Los marcajes isotópicos pueden ser introducidos metabólicamente (*in vivo*), químicamente (*in vitro*) o enzimáticamente (Yates *et al.*, 2009). En el marcaje metabólico, elementos isotópicamente enriquecidos (^{13}C , ^{15}N , ^{18}O , introducidos como sales o aminoácidos) se añaden al medio de cultivo de forma que estén disponibles para su incorporación en todo el organismo, tejido o célula. Este tipo de marcaje incluye: marcaje con isótopos,

marcaje con ^{15}N , ^{13}C , y marcaje isotópico por medio de aminoácidos en cultivos celulares (SILAC). En contraste, el marcaje químico ocurre después de la extracción de proteínas, por lo que es completamente independiente de la procedencia y preparación de la muestra, lo cual representa una ventaja, ya que prácticamente cualquier tipo de muestra biológica puede ser etiquetada, incluyendo tejidos y fluidos corporales; además, el tiempo requerido para este tipo de marcaje es menor. Dentro de este tipo de marcajes se encuentran: la electroforesis diferencial en gel (DIGE), el marcaje isotópico diferencial (ICAT), el marcaje dimetilado y el marcaje con etiquetas isobáricas para cuantificación relativa y absoluta (iTRAQ) (Gouw *et al.*, 2010).

1.11.1 SILAC

SILAC implica el crecimiento de dos poblaciones de células, una en un medio que contiene aminoácidos "ligeros" (normal) y otra en un medio que contiene aminoácidos "pesados". El crecimiento de estas poblaciones celulares, tiene que llevarse a cabo durante varias generaciones (5 o 6 pases), para asegurar la incorporación de los aminoácidos marcados (Mann, 2006). La introducción de isótopos pesados estables resulta en un cambio en la masa de los péptidos. Por lo tanto, las muestras marcadas diferencialmente se pueden mezclar y analizar juntas (Vermeulen y Selbach, 2009). Una vez combinadas las muestras, se lisan y las proteínas obtenidas se digieren con tripsina. Los péptidos resultantes de la digestión se analizan por espectrometría de masas. El primer análisis MS (escáner de iones precursor) se utiliza para cuantificar la abundancia relativa de los péptidos marcados como ligeros en comparación con los pesados (Figura 12a) (Gingras, 2005). En el espectro de masas resultante, cada péptido aparece como un par de picos, en donde uno de ellos corresponde al péptido marcado con aminoácidos ligeros y el otro a los péptidos marcados con aminoácidos pesados. Si el par de picos presentan una proporción 1:1, no existe diferencia significativa en la abundancia de esta proteína, en cambio, si uno de los picos presenta una intensidad diferente, la

proteína se encuentra sobreexpresada o subexpresada. Por lo tanto, la intensidad relativa de los picos obtenidos es directamente proporcional a la abundancia de los péptidos en las dos muestras (Mann, 2006). La identificación de las proteínas, se realiza a través de experimentos MS/MS de diversos iones precursores (Gingras *et al.*, 2005).

1.11.2 DIGE

La electroforesis diferencial en gel es una estrategia de marcaje químico, la cual consiste en el marcaje de proteínas con diferentes fluorocromos y la separación de las muestras mediante SDS-PAGE. La innovación de la tecnología DIGE es su capacidad para etiquetar dos o más muestras con fluorocromos distintos y separarlos en el mismo gel minimizando la variabilidad de gel a gel, disminuyendo el tiempo de análisis y permitiendo una cuantificación del perfil de expresión precisa (Marouga *et al.*, 2005). De manera general, en DIGE las proteínas son marcadas con los colorantes fluorescentes Cy2 (generalmente utilizado como control interno), Cy3 (muestra 1) y Cy5 (muestra 2). Cy2 es utilizado como un control de normalización entre las muestras (muestra 1 + muestra 2). Después del marcaje, las muestras se mezclan, se prefraccionan a través de isoelectroenfoque y se separan simultáneamente en un gel SDS-PAGE. Los diferentes extractos de proteína marcados con los distintos fluoróforos pueden ser visualizados por separado, al excitar los colorantes en las longitudes de onda de excitación específica para cada uno. Cada fluoróforo genera imágenes digitales individuales, además se puede generar una imagen de solapamiento, en donde se puede observar la expresión diferencial de las proteínas (Figura 12b). Finalmente, las proteínas con expresión diferencial son sometidos a un análisis de espectrometría de masas para su identificación (Marouga *et al.*, 2005; Paul, D *et al.*, 2013).

1.11.3 ICAT

ICAT se une de manera específica a las cisteínas y está compuesto de un grupo reactivo (utilizado para unir covalentemente la etiqueta a los péptidos que poseen una característica química específica), un grupo enlazador (que contiene isótopos ligeros o pesados) y un “brazo” de afinidad (usualmente biotina, utilizado para la purificación de péptidos) (Gringas *et al.*, 2005). De manera general, en ICAT se pueden comparar dos muestras, una de ellas es marcada con ICAT ligero y la otra con ICAT pesado. Posteriormente, las dos muestras mezcladas, se digieren con tripsina, se separan en una columna de afinidad y se analizan por espectrometría de masas. En el espectro de masas, la intensidad relativa de los picos muestra la abundancia de las proteínas. La identificación de estas proteínas se da mediante experimentos MS/MS (Figura 12c) (Gringas *et al.*, 2005; Gringas *et al.*, 2007).

1.11.4 iTRAQ

iTRAQ consiste de un grupo reportero cargado que es único para cada uno de los cuatro u ocho reactivos, un grupo reactivo a péptidos y una porción neutra para mantener una masa global de 145 (Figura 13a) (Zieske, 2006). El grupo reactivo a péptidos fue diseñado para reaccionar con todas las aminas primarias, incluyendo el N-terminal y el grupo amino ϵ de la cadena lateral de lisina (Figura 13b). La selección de la región reportera en la zona de masas baja, se debe a dos razones principales: (1) Mantener la masa aditiva de los fragmentos tan insignificante como sea posible, con el fin de minimizar cualquier efecto en el modo de análisis MS o MS/MS y (2) eliminar cualquier interferencia con iones imonio u otros iones fragmento, permitiendo así a un alto grado de confianza en la interpretación. iTRAQ consta de dos presentaciones: 4plex (114, 115, 116 y 117) y 8plex ((113, 114, 115, 116, 117, 118, 119, 121) (Ross *et al.*, 2004; Zieske, 2006).

De manera general, la estrategia de marcaje con iTRAQ consiste en la reducción, alquilación y digestión de las muestras. Los péptidos resultantes se marcan con las etiquetas de iTRAQ, por la formación de un enlace amido entre la etiqueta y cualquier amina. Posteriormente, las muestras se mezclan en una sola para finalmente ser analizada por espectrometría de masas en tándem (Figura 12d) (Ross *et al.*, 2004; Zieske, 2006). Debido a la naturaleza isobárica de los reactivos de iTRAQ, el mismo péptido de cada muestra se presenta como un solo pico en el espectro de masas, lo que reduce la complejidad del análisis. Después de la CID, los fragmentos de péptidos marcados con iTRAQ liberan los iones reporteros (114, 115, 116, y 117 m/z para el caso del 4plex y 113, 114, 115, 116, 117, 118, 119 y 121 m/z para el 8plex) en iones de series *b*, *y*, entre otros fragmentos. El área del pico del ión reportero se utiliza para evaluar la abundancia relativa de péptidos y consecuentemente, la de las proteínas de los cuales se derivan (Aggarwal *et al.*, 2006).

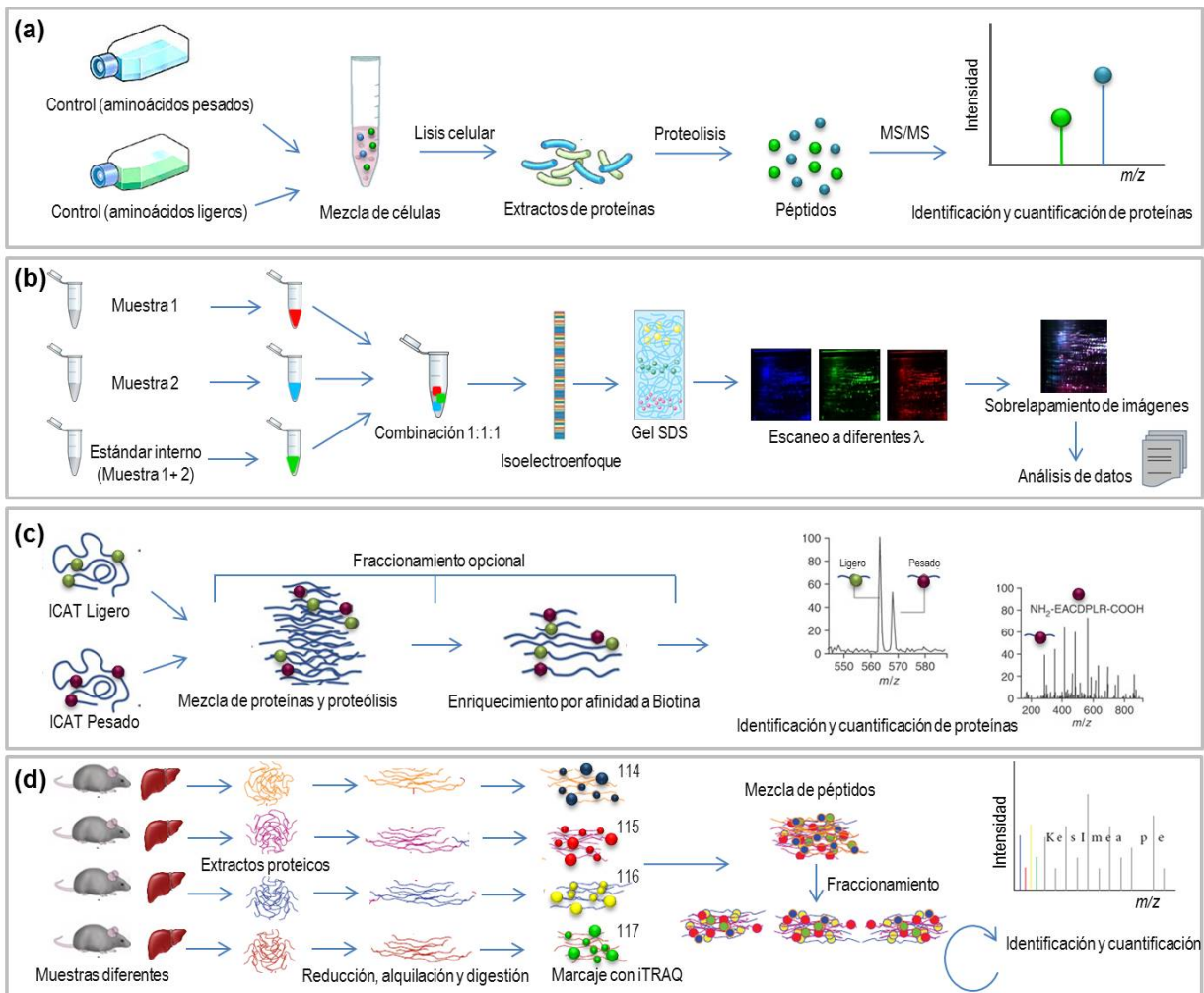


Figura 12. Tecnologías en proteómica cuantitativa. (a) SILAC, emplea el marcaje isotópico de aminoácidos (ligeros y pesados) para su incorporación en cultivos celulares. Las células marcadas se mezclan, lisan y los péptidos obtenidos se analizan por MS y MS/MS (Modificada de Mourelatos, 2008). (b) DIGE, consiste en el marcaje de proteínas con fluoróforos diferentes (Cy2, Cy3 y Cy5), posteriormente las muestras se combinan y se separan por punto isoeléctrico y SDS-PAGE. Las muestras pueden visualizarse de manera independiente al excitar los colorantes en las diferentes λ (Modificada de Paul, D *et al.*, 2013). (c) ICAT implica el marcaje de cisteínas con un reactivo ICAT ligero e ICAT pesado. Las muestras se combinan, se digieren con tripsina y se enriquecen por afinidad y se analizan por espectrometría de masas en tándem (Modificado de Shio y Aebersold, 2006). (d) iTRAQ consiste en el marcaje de lisinas con diferentes etiquetas isobáricas. Para ello, los extractos proteicos se reducen, alquilan y digieren con tripsina. Los péptidos obtenidos se marcan con las etiquetas de iTRAQ, prefractionan y analizan por MS y MS/MS (Modificada de Chandramouli y Qian, 2009).

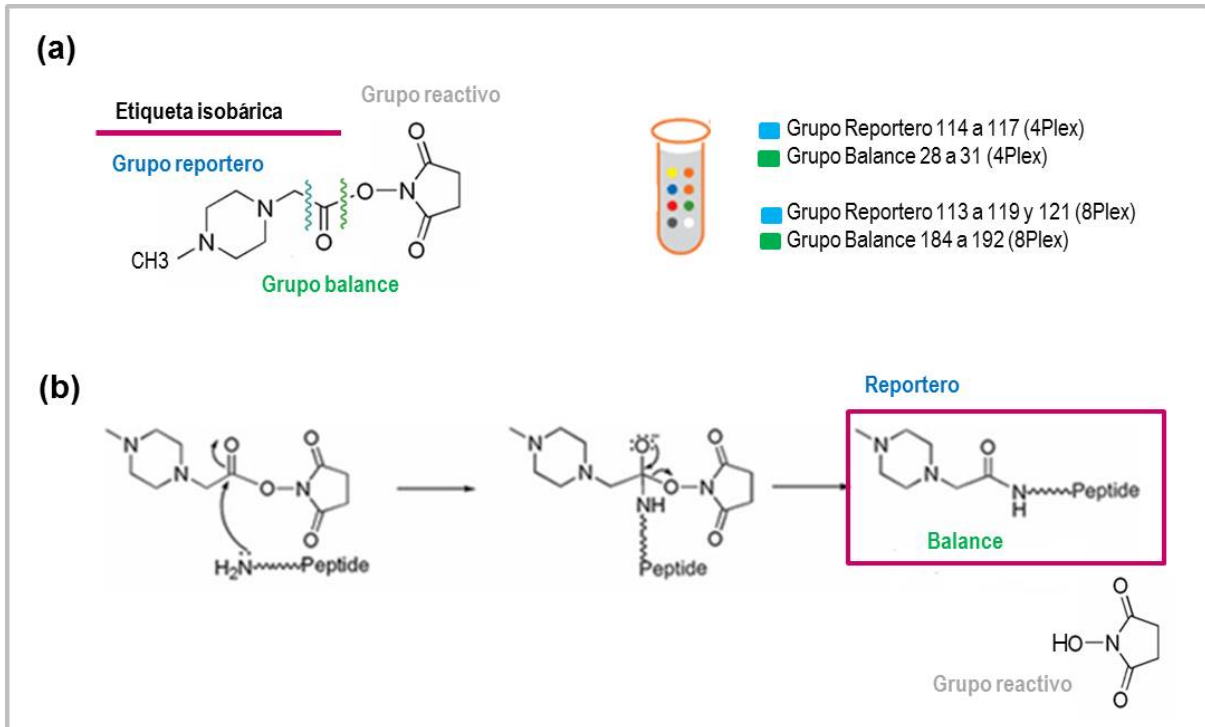


Figura 13. Composición y mecanismo de acción de iTRAQ. En el esquema se observa (a) la estructura de iTRAQ, el cual consiste de grupo reportero, un grupo reactivo a péptidos y una porción neutral para mantener una masa global de 145. (b) iTRAQ se une al extremo N-terminal o al grupo amino ϵ de lisina y forma un enlace amido, desplazando así al grupo reactivo (Modificada de http://proteomics.sysu.edu.cn/html/list_821.html).

CAPÍTULO II

ANTECEDENTES

Antecedentes

Se han realizado diversos estudios con diferentes aproximaciones proteómicas en cáncer de mama, con el objetivo de identificar moléculas que puedan servir como biomarcadores de diagnóstico o pronóstico, o para la comprensión de vías de señalización u otros procesos implicados en la carcinogénesis. A continuación, se mencionan algunos de estos trabajos.

Kim, DH y colaboradores en 2009 realizaron un estudio en tejidos de cáncer de mama (carentes de albúmina e inmunoglobulinas) con DIGE. Identificaron 53 proteínas por MS. Entre las proteínas con niveles de expresión incrementados en tumores encontraron: coronina 1A, subunidad 2 del complejo Arp2/3 de 34 kDa (p34-Arc) y la proteína 1 con repetidos WD (Wdr1). Estas proteínas poseen motivos repetidos de WD en su extremo N-terminal y un cierre de leucinas en el C-terminal, por lo que están relacionadas en la interacción con actina; por lo tanto, fueron clasificadas en el grupo de mantenimiento y crecimiento celular. El aumento significativo de p34-Arc y coronina 1 en cáncer de mama correlacionó con un análisis de Western blot y fue consistente con los niveles de mRNA encontrados. Experimentos posteriores con siRNA para *p34-Arc* mostraron una atenuación de la invasión en SK-BR3 (línea celular de cáncer de mama); en contraste, la sobreexpresión de coronina A1 incrementó la actividad invasiva. Por lo tanto, en conjunto, los niveles celulares de estas proteínas están involucrados en el desarrollo y migración del cáncer.

En otro estudio, se determinaron los perfiles de expresión diferencial de proteínas totales y proteínas secretadas extracelularmente en las líneas celulares MCF 10A (línea celular control), MCF7 (células de cáncer de mama no invasivo) y MDA-MB-231 (células de cáncer de mama invasivo) a través de 2D-DIGE y MALDI-TOF. Se identificaron 50 polipéptidos únicos expresados diferencialmente en los tres medios utilizados de las líneas celulares y 133 proteínas únicas en los extractos de

proteínas totales. De manera interesante, se identificaron 14 proteínas secretadas y 51 polipéptidos totales que no habían sido reportados previamente en cáncer de mama (Lai *et al.*, 2010).

Larsen y colaboradores (2009), realizaron un estudio proteómico cuantitativo utilizando SILAC y LC-MSMS en líneas celulares metastásicas (M-4A4) contra no metastásicas (NM-2C5), para identificar proteínas de membrana plasmática con patrones de expresión diferencial. Lograron identificar 13 proteínas sobreexpresadas y 3 subexpresadas en M-4A4 con respecto a NM-2C5. Entre las proteínas identificadas se encuentran: ecto-5-nucleotidasa (CD73), NDRG1, Integrina β 1, CD44, CD74 y proteínas del complejo mayor de histocompatibilidad clase II. Los niveles de expresión se validaron por citometría de flujo, Western blot, inmunocitoquímica e inmunohistoquímica. Estos resultados llevaron a un análisis más profundo en muestras de tumores de mama y esto confirmó que la expresión de la integrina β 1 y ecto-5-NT correlaciona significativamente con metástasis de enfermedades malignas humanas.

Por otro lado, se llevó a cabo un análisis para la identificación y cuantificación de proteínas secretadas diferencialmente en líneas celulares de cáncer de mama. Las células tanto normales como malignas fueron marcadas en el medio de cultivo con SILAC, posteriormente, el sobrenadante celular se analizó con SDS-PAGE y LC-MS en un sistema Q-TOF. Alrededor de 380 proteínas no redundantes fueron cuantificadas en medio libre de suero, de las cuales, 62% se clasificaron como secretadas según las bases de datos. Se encontraron algunos factores de crecimiento regulados diferencialmente, tales como el factor de crecimiento tumoral (TNF), el factor de diferenciación epitelial y el precursor del factor de crecimiento de células troncales cuya expresión se encontró disminuida en líneas de cáncer de mama, en tanto que la inhibina beta y el factor inhibidor de migración de macrófagos mostraron una expresión incrementada. Interesantemente, los inhibidores de proteasas incluyendo al inhibidor de proteasa de plasma (C1), el precursor PZP, y la serpina E2 se encontraron significativamente reprimidos en líneas celulares de

cáncer, así como los factores angiostáticos de la matriz extracelular (ECM) tales como la endorepilina. Además, el fragmento C-terminal de la colágena tipo XVIII, la endostatina (potente marcador angiostático), se encontraron disminuidas, mientras que colágenas extracelulares y el factor específico de osteoblastos (OSF-2) se encontraron sobreexpresados (Liang *et al.*, 2009).

Para identificar proteínas relacionadas con el proceso de iniciación y progresión de la carcinogénesis, Xu y colaboradores en 2010 realizaron un estudio con líneas celulares de mama normal, tumor primario (NT:16N) y tumor metastásico (MT2:NT), todas obtenidas de un paciente. En este estudio se utilizó la estrategia de SILAC y LC-MS/MS, identificándose 1228 proteínas con expresión diferencial. Para la clasificación de proteínas de acuerdo con su ontología y la generación de redes de interacción proteína-proteína, se aplicó un valor de corte del 90% de confianza, obteniendo así 33 proteínas sobreexpresadas y 35 subexpresadas en NT:16N comparada con la línea de células normales y 21 proteínas sobre y 23 subexpresadas en NT:16N comparada con MT2:NT. Con estos datos determinaron la subexpresión de diversas proteínas del citoesqueleto en la tumorigénesis, así como la sobreexpresión de ERBB2 y su papel central en este proceso. Por otro lado, identificaron que los procesos metabólicos se encuentran altamente desregulados en el proceso de tumorigénesis y metástasis. Finalmente encontraron que el nivel de expresión de la proteína TIMM17A correlaciona directamente con la progresión del tumor.

Sutton y colaboradores (2010), realizaron un estudio de proteómica cuantitativa, empleando un método de extracción de proteínas dual (solución RIPA y urea) y un marcaje con iTRAQ para ocho reacciones en biopsias de tumores y tejido normal de mama. Este enfoque experimental consistió de digestión, marcaje, separación de péptidos por punto isoeléctrico (OffGel), nanoHPLC y MS MALDI TOF-TOF. Se lograron identificar 481 proteínas únicas (con al menos dos MS/MS y dos valores de iTRAQ para cada ion reportero); sin embargo, se observó redundancia a nivel de péptidos, así como en las proteínas identificadas, por lo que sólo 30 de éstas

fueron consideradas (identificadas con un único péptido). Las proteínas identificadas fueron agrupadas en proteínas en sangre, celulares y extracelulares. Entre las proteínas celulares incrementadas en los tejidos cancerosos se encuentran: Tropomiosina β (estabilización de los filamentos de actina en el citoesqueleto), citoqueratinas K1, 6E, 7, 8, 19 y 22, proteínas ribosomales y factores de iniciación de la traducción, factor de elongación 1-beta (síntesis de proteínas); las proteínas con expresión disminuida fueron, la proteína de unión a selenio (detección de xenobióticos reactivos), proteína Rcl (metabolismo de nucleótidos), enzimas dependientes de NAD^+ y NADP^+ (procesos metabólicos), entre otras. En el grupo de proteínas extracelulares, las proteínas incrementadas incluyen, periostina (induce anclaje y difusión celular), región 5' del factor neurotrópico derivado de astrocitos mesencefálicos (ARMET) (promueve la supervivencia de neuronas dopaminérgicas); en las disminuidas encontraron a la proteína 1 relacionada con frizzled (regulador de la vía Wnt), con la dermatopontina (inhibe proliferación) y con la superóxido dismutasa extracelular (protege espacio extracelular de especies reactivas de oxígeno), entre otras.

En otro estudio se determinaron los perfiles de expresión de fibroadenomas, CDIS y carcinoma ductal invasivo. Se identificaron 826 proteínas, de las cuales 402 fueron comunes en los carcinomas. Se excluyeron las proteínas de plasma, obteniendo así 59 proteínas con cambios significativos en los tumores en comparación con los tejidos normales, la mayoría de los cambios estuvieron asociados con los carcinomas invasivos. Las proteínas con expresión diferencial estuvieron involucradas en procesos como: transporte y metabolismo de lípidos, plegamiento de proteínas, ensamble, morfología y organización celular, así como en localización mitocondrial. Los polipéptidos Cofilina-1 (subexpresadas en carcinoma invasivo), p23 (sobreexpresada en carcinoma invasivo) y amina oxidasa 3 (subexpresada en carcinoma invasivo) fueron validadas. A través de microarreglos de tejidos, se determinó que la tinción positiva de Cofilina-1 correlaciona con el grado

del tumor y la tinción positiva de p23 correlaciona con un estado RE positivo (Shaheed *et al.*, 2013).

Finalmente, el antecedente más directo de este proyecto es un trabajo realizado en el laboratorio del Dr. Luna Arias, en el cual se realizó un marcaje con iTRAQ en extractos totales de líneas celulares de cáncer de mama (MCF 10A, MCF7, MDA MB-231, SKBR3). En este trabajo se utilizó la estrategia de MudPit (separación por intercambio catiónico fuerte seguido de fase reversa) en un espectrómetro de masas Q-Trap 3200 (AB Sciex). En el estudio se identificaron 59 proteínas, de las cuales 28 presentan un umbral de confianza mayor al 95%. Algunas de las proteínas sobreexpresadas en las líneas celulares de cáncer de mama son gliceraldehído-3-fosfato deshidrogenasa, proteína de envejecimiento asociada al gen 9, piruvato cinasa 3 isoforma 1, entre otras (Tesis de maestría de Cesar Bazán Méndez).

CAPÍTULO III

JUSTIFICACIÓN, PREGUNTA DE INVESTIGACIÓN, HIPÓTESIS, OBJETIVOS

1. Justificación

1) El cáncer representa un serio problema de salud en el mundo. En México ocupa el segundo lugar como causa de muerte de la población. Específicamente el CaMa representa la primera causa de muerte por cáncer en mujeres, por lo que resulta importante entender sus bases moleculares para poder elaborar estrategias de diagnóstico, monitoreo y terapia.

2) El CaMa es una enfermedad que puede ser tratada con buenos resultados si se diagnostica en estadios tempranos de la enfermedad, por lo que es sumamente importante implementar estrategias de diagnóstico que ofrezcan una detección oportuna. Los biomarcadores ofrecen una forma sencilla y sensible de detectar la enfermedad, además también pueden ser utilizados en el pronóstico y terapia de la misma, por lo que resulta indispensable la investigación en este campo.

3) Existen muy pocos trabajos sobre proteómica cuantitativa que se hayan realizado en varias líneas celulares de cáncer de mama, usando iTRAQ y espectrometría de masas en tándem.

2. Pregunta de investigación

¿Existen proteínas que se expresen diferencialmente en las líneas celulares de cáncer de mama en comparación con una línea celular control, que pudieran utilizarse para el diagnóstico y/o pronóstico de esta enfermedad?

3. Hipótesis

Se observarán patrones de expresión diferenciales en las líneas celulares de cáncer de mama con diferente clasificación molecular en comparación con la línea celular control no transformada.

4. Objetivos

4.1 Objetivo general

Identificar proteínas que se expresan diferencialmente en las líneas celulares de cáncer de mama con diferente clasificación molecular que puedan ser útiles en el diagnóstico y/o pronóstico de esta enfermedad.

4.2 Objetivos específicos

1. Estandarizar las metodologías para la obtención de proteínas totales de las líneas celulares.
2. Determinar los perfiles de expresión diferenciales de las líneas celulares de cáncer de mama.
3. Clasificar las proteínas expresadas diferencialmente en todas las líneas celulares de cáncer de mama de acuerdo con los procesos y enfermedades en los que estén implicadas a través de diferentes plataformas bioinformáticas.
4. Identificar proteínas candidatas a biomarcadores en todas las líneas celulares de cáncer de mama.
5. Validar las proteínas candidatas a biomarcadores a través de Western blot.
6. Clasificar las proteínas expresadas diferencialmente en las líneas celulares de cáncer de mama con clasificación molecular diferente de acuerdo con los procesos y enfermedades en los que estén implicadas a través de varias plataformas bioinformáticas.

CAPÍTULO IV

METODOLOGÍA

1. Cultivo celular

Se utilizaron las siguientes líneas celulares: MCF 10A de estirpe epitelial no tumorigénica de una enfermedad fibroquística de mama (línea control); las líneas MCF7, SK-BR-3 y MDA-MB-231 provenientes de células epiteliales de adenocarcinoma de mama derivadas de un sitio metastásico de efusión pleural y la línea T47D proveniente de un carcinoma ductal, obtenida de igual manera de una efusión pleural. Estas líneas celulares fueron elegidas, debido a que las características moleculares de éstas se asemejan a tumores tipo: Luminal A (MCF7, T47D), Baja en claudina (MDA-MB-231) y HER2⁺ (SK-BR-3). Todas las líneas celulares fueron obtenidas de American Type Culture Collection (ATCC).

Se cultivaron 4 botellas (Corning) de 75 cm³ de cada una de las líneas celulares en medios de cultivo suplementados. La línea celular MCF 10A se creció en medio DMEM F12 suplementado con suero fetal bovino (SFB) al 10%, factor de crecimiento epidermal humano 10 ng/mL, insulina 4 µg/mL, hidrocortisona 0.4 µg/mL, glutamina 1X y penicilina-estreptomicina 1X. Las líneas MCF7 y MDA MB-231 se crecieron en medio DMEM (suplementado con SFB al 10%, piruvato de sodio 1X, glutamina 1X, penicilina-estreptomicina 1X y una mezcla de aminoácidos no esenciales 1X). La línea T47D en medio RPMI suplementado de la misma manera que el medio DMEM, con la adición de NaHCO₃ 7.5% y la línea SK-BR-3 en medio McCoy (SFB al 10%, glutamina 1X y penicilina-estreptomicina 1X). Todas las líneas celulares se incubaron a 37°C en una atmósfera de aire humidificada conteniendo CO₂ al 5% (v/v), hasta que alcanzaron una confluencia aproximada del 90%.

Para cosechar las células, se retiró el medio de cultivo con una pipeta estéril y se lavaron dos veces con 10 mL de PBS 1X. Las células se desprendieron por adición de 10 mL de EDTA 5 mM en PBS 1X incubando a 37°C por 10 min. Las células se transfirieron a un tubo Falcon de 50 mL frío y se centrifugaron a 1,500 rpm (453 x g) a 4°C por 10 min. Posteriormente, se retiró el EDTA y el paquete celular se

colocó en hielo. Este procedimiento se realizó dos veces más, para obtener la mayor cantidad de células posibles.

2. Extracción de proteínas

Para la estandarización de la extracción de proteínas, se probaron los siguientes protocolos: Tritón-NP40 (Tesis de maestría de Bazán-Méndez, C), RIPA-Urea (extracción dual) (Sutton y colaboradores, 2010) e hidrocloreuro de guanidina, utilizando un paquete celular homogéneo de la línea MDA MB-231 para cada uno de ellos.

En el método de lisis con Tritón-NP40 (NP-40 0.2%, Tritón X-100 0.2%, CHAPS 0.2%, NaCl 75 mM, EDTA 1 mM, NaF 50 mM, Na₂VO₄ 1 mM, Complete 1X, TCEP 10 mM, Tris·HCl 50 mM, pH 8.0) se agregó solución de extracción en una relación 1:3 (v/v) con respecto al volumen del paquete celular. La pastilla se sonicó 3 veces con un sonicador de punta CPX130PB (Cole-Parmer, Vernon Hills, IL, USA), utilizando una amplitud del 60% durante 10 s, con un reposo de 1 min en hielo entre cada ronda de sonicación. Finalmente, la muestra se dejó en agitación por 2 h a 4°C.

Para la extracción de proteínas con el método dual, RIPA-Urea (PBS pH 7.4, SDS 0.1%, desoxicolato de sodio 0.25%, Urea 7 M, Tiourea 2 M, CHAPS 4%, DTT 50 mM) el paquete celular se resuspendió en 100 µL de RIPA, se incubó 30 min en un agitador de vórtice a temperatura ambiente y se sonicó dos veces como se describió antes. Posteriormente, las muestras se centrifugaron a 12,000 x g por 20 min a 4°C, se recuperó el sobrenadante (SN) y el paquete celular remanente se resuspendió en 100 µL de la solución de urea y se incubó, sonicó y centrifugó como se mencionó anteriormente. El sobrenadante obtenido se combinó con el SN de la extracción con RIPA. El paquete celular remanente se incubó nuevamente con la solución de lisis de urea y se repitió el mismo procedimiento. Al final se obtuvo un volumen total de 300 µL.

En la lisis celular con hidrocloreuro de guanidina (GuHCl 6 M), el paquete celular se resuspendió en 300 μ L y se sonicó 3 veces siguiendo el procedimiento anteriormente descrito.

Después de la extracción de proteínas, los ácidos nucleicos se eliminaron en cada uno de los métodos utilizados, adicionando 10 U de benzonasa (Sigma) e incubando a 37°C por 1 h. Al término de la incubación las muestras se centrifugaron a 12,000 x g durante 15 min y se conservó el sobrenadante.

El método de extracción de proteínas finalmente elegido para la realización de este proyecto fue el de hidrocloreuro de guanidina, sin embargo este se modificó ligeramente. La extracción de proteínas se realizó con GuHCl 6 M-RapiGest 0.1% (Waters, México) en bicarbonato de amonio 100 mM. La sonicación de las muestras se llevó a cabo como se describió anteriormente.

3. Deslipidación

La deslipidación de las muestras se realizó en metanol-cloroformo. Por cada 50 μ L de muestra se añadieron 200 μ L de metanol y se mezcló en un agitador de vórtice por 1 min a máxima velocidad. Después se adicionaron 50 μ L de cloroformo, se agitó de igual manera y se adicionaron 150 μ L de agua desionizada (dH₂O) agitándose de nuevo por 1 min. Posteriormente, las muestras se centrifugaron a 12,000 x g durante 10 min y al término de la centrifugación el SN se retiró. El procedimiento se repitió dos veces más y al término se dejó secar parcialmente la pastilla a 37°C. Finalmente las muestras se resuspendieron en 400 μ L de Urea 8 M.

4. Precipitación con TCA en acetona

Las muestras de proteína se llevaron a un volumen final de 900 μ L con dH₂O, se agregaron 100 μ L de ácido tricloroacético (TCA) al 77% en acetona fría, se

mezcló y se incubó a -20°C durante 30 min. Al término de la incubación, las muestras se centrifugaron a 12,000 x g por 15 min a 4°C y se descartó el SN. Finalmente, se realizaron tres lavados del precipitado con 1 mL de acetona fría y el paquete de proteínas se resuspendió en Urea 8 M.

5. Electroforesis en geles de acrilamida

Los extractos totales de proteína se analizaron en geles de poliacrilamida en condiciones desnaturalizantes (SDS-PAGE) de acuerdo con el método descrito por Laemmli (1970). Para la comparación de los perfiles proteicos obtenidos con diferentes métodos de extracción (Tritón-NP40, RIPA-Urea, Hidrocloruro de guanidina) el gel se preparó al 10%. Los extractos obtenidos con Hidrocloruro de guanidina e Hidrocloruro de guanidina –Rapigest se analizaron en un gel al 12%. Finalmente, los extractos totales obtenidos con Hidrocloruro de guanidina –Rapigest de cada una de las líneas celulares se analizaron en un gel al 10%. Los geles se corrieron a un voltaje de 100 durante 2 horas. Al término de la electroforesis, los geles de comparación de métodos de extracción fueron teñidos con nitrato de plata y el gel de comparación de extractos totales de todas las líneas celulares se tiño con Azul de Coomassie.

6. Tinción de geles de poliacrilamida con nitrato de plata

Los geles de poliacrilamida se tiñeron con nitrato de plata de acuerdo con una modificación realizada en el protocolo establecido por Schevencko *et al.*, 1996. Al término de la electroforesis los geles se enjuagaron brevemente con dH₂O y se colocaron en 100 mL de solución fijadora (etanol 30%, ácido acético 5%, H₂O 65%) durante toda la noche en agitación suave. Al término de la fijación, los geles se lavaron con 100 mL de dH₂O durante 1 h, haciendo cambio de solución cada 15 min. Los geles se sensibilizaron con 100 mL de tiosulfato de sodio (Na₂S₂O₃ 0.02%)

durante 1 min y se lavaron dos veces con dH₂O por 1min. Los geles se tiñeron con 100 mL de una solución de nitrato de plata [AgNO₃ 0.1%, formaldehído 0.028% (v/v)] durante 45 min en la oscuridad. Al término de la tinción, se descartó la solución, los geles se lavaron con dH₂O por 1 min y se incubaron en 100 mL de solución reveladora (NaHCO₃ 2.4%, formaldehído 0.028% (v/v), Na₂S₂O₃ 0.0125%) hasta que las bandas presentaron la intensidad requerida. Finalmente la reacción se detuvo con 100 mL de solución de paro (Tris base 4%, ácido acético 2%) durante 15 min y al término de la incubación la solución se reemplazó con dH₂O. El gel fue digitalizado con el equipo GelDoc EZ Imager (BioRad).

7. Digestión de proteínas

Antes de la digestión con tripsina, las muestras se precipitaron con TCA al 10% por 30 min en hielo, se centrifugaron a 12,000 x g durante 5 min a 4 °C, se lavaron dos veces las pastillas con acetona fría (-20 °C) y se secaron a temperatura ambiente. Los polipéptidos se solubilizaron por sonicación en 20 µL de una solución de urea 8 M, detergente potenciador de la tripsina al 0.05% (p/v) (ProteaseMAX Surfactant Trypsin Enhancer, Promega) y bicarbonato de trietilamonio (TEAB) 0.5 M pH 8.0. La cuantificación de las proteínas se realizó siguiendo el método de Bradford (Bradford, 1976).

De cada una de las muestras se tomaron 110 µg de proteínas, los cuales se llevaron a un volumen final de 40 µL con TEAB 0.5 M. Posteriormente, las muestras se redujeron con 2 µL de Tris 2-carboxietil fosfina (TCEP) 50 mM a 37°C por 1 h y se alquilaron con 1 µL de S-metilmetanotiosulfonato (MMTS) 200 mM durante 10 min a temperatura ambiente. Posteriormente, se diluyó la concentración de urea hasta 2 M por adición de TEAB 0.5 M. Las muestras se digirieron con 10 µg de tripsina grado espectrometría de masas (Promega) a 37°C durante 16 h y posteriormente, los péptidos se desecaron en un concentrador (Savant SpeedVac)

8. Marcaje de péptidos con iTRAQ

Para realizar el marcaje con iTRAQ, se utilizó el kit iTRAQ Reagents- 8plex (AB SCIEX), realizando pequeñas modificaciones al protocolo. Los viales de las etiquetas se reconstituyeron con 50 μ L de isopropanol y se añadieron a cada una de las muestras de péptidos digeridos. El marcaje se realizó de la siguiente forma: MCF 10A (etiqueta 113), MCF7 (114 y 118), MDA-MB-231 (115 y 119), SK-BR-3 (116 y 121) y T47D (117) durante 3 h a temperatura ambiente. Para detener la reacción, se adicionaron 100 μ L de dH_2O . Posterior al marcaje, todas las muestras se mezclaron y combinaron en un solo tubo.

9. Prefraccionamiento de péptidos

El prefraccionamiento de los péptidos se realizó mediante isoelectroenfoque (IEF) preparativo. Para ello, los 440 μ g de péptidos digeridos y marcados se resuspendieron en 250 μ L de solución de rehidratación (urea 7M, tiourea 2M, CHAPS 2%, anfolitos 1.6% pH 3-10, azul de bromofenol 0.002%), se depositaron en una tira de pH 3-10 NL de 13 cm (GE Healthcare) y se incubaron toda la noche a temperatura ambiente. El IEF se llevó a cabo a través de un protocolo de variación de voltaje con el sistema IPGphor a 20°C y 75 μ A por tira en dos pasos, el primero de ellos de 5000 V hasta 30000 Vh y el segundo a 500 V por 30 h. Al término del IEF, la tira se lavó con dH_2O y se dividió en 12 secciones de 1 cm de ancho cada una. De cada uno de estos fragmentos, se extrajeron secuencialmente los péptidos con 50 μ L de cada una de las siguientes soluciones: 1) ácido trifluoroacético (TFA) al 0.1%; 2) TFA al 0.1% en acetonitrilo (ACN) al 50% en agua y 3) TFA al 0.1% en ACN, incubando en cada caso 10 min a temperatura ambiente. Al término de cada incubación, los péptidos se recuperaron en un tubo por cada sección y se concentraron al vacío (SpeedVac). Posteriormente, se realizó una cromatografía en fase reversa (C18) con puntas Poros R2 (hechas en casa). Los péptidos se eluyeron de las puntas en 40 μ L de ACN al 75% y TFA al 0.1% y se evaporaron en SpeedVac hasta obtener un

volumen de 20 μL . Finalmente, las muestras se llevaron a un volumen de 125 μL con ACN al 5% y TFA al 0.1% y se analizaron mediante espectrometría de masas.

10. Análisis por espectrometría de masas

El análisis de espectrometría de masas se realizó en el Laboratorio de Proteómica del Centro de Investigación del Príncipe Felipe, miembro de ProteoRed en el espectrómetro de masas nano ESI-QTOF (QSTAR ESI XL, AB Sciex). Previo al análisis de las muestras, el equipo se calibró con 200 femtomoles de una mezcla de seis proteínas previamente digeridas (β -galactosidasa (*E. coli*), serotransferrina (*Bos Taurus*), albumina, citocromo c; alcohol deshidrogenasa (*Saccharomyces cerevisiae*), lisosima c (*Gallus gallus*)). La identificación de los péptidos de estas proteínas se realizó a través del programa Mascot.

Para el análisis de las muestras, estas se introdujeron en una precolumna (LC Packings PepMap C18, 300 μm x 5 mm) a través de un flujo isocrático de TFA 0.1% a 30 $\mu\text{L}/\text{min}$ durante 10 min. Posteriormente los péptidos se inyectaron en una columna analítica (Silica nanoBaume 75 μm x 50 cm, 5 μm de tamaño de partícula, hecha en casa) a un flujo de 250 nL/min. La elución de los péptidos de la columna se realizó con tres gradientes lineales de ACN en ácido fórmico 0.1% a un flujo de 250 nL/min: 1) 1.9% a 28.5% en 105 min, 2) 28.5% a 38% en 10 min y 3) 38% a 85% en 5 min. Los péptidos eluidos se analizaron a través de espectrometría de masas en tándem. El equipo se operó en el modo IDA (Information Dependent Acquisition) con el programa Analyst QS 1.1 (AB Sciex), realizando escaneos MS cada segundo, seguido de tres escaneos MS/MS cada tres segundos.

11. Identificación y cuantificación de proteínas

La identificación y cuantificación de proteínas se realizó con el programa ProteinPilot v 4.0 (AB Sciex) a través de los algoritmos Paragon y ProGroup, respectivamente. La base de datos utilizada para el análisis fue UniProtSwiss-Prot2012 seleccionando la base de datos de humano. Los parámetros considerados fueron: marcaje de péptidos (iTRAQ 8 Plex), alquilación de cisteínas (MMTS), digestión enzimática (tripsina), factores especiales (desnaturalización con urea), tipo de búsqueda (exhaustiva), análisis de falsos descubrimientos o asignaciones (FDR) y tipo de instrumento (ESI Q-STAR). También se asignaron modificaciones fijas (carbamidometilación de cisteínas (C[CAM])) y modificaciones variables (oxidación de metioninas, deamidación de asparagina y glutamina). La tolerancia en medición de masa en modo MS fue de 100 ppm y ± 0.5 Da para los iones obtenidos en la fragmentación MS/MS. La identificación de proteínas se realizó con un porcentaje de confiabilidad del 95% (score de 1.30). En la cuantificación de proteínas se consideró la etiqueta 113 (MCF-10A) como denominador y se aplicó automáticamente la corrección de "bias".

Para determinar los cambios significativos en el nivel de expresión de las proteínas identificadas se realizó un análisis con la plantilla PDST (ProteinPilot Descriptive Statistics Template). Los datos se exportaron a Excel para su análisis manual, en el cual se fijaron los valores de corte >1.2 y <0.8 como sobreexpresión y subexpresión, respectivamente (Afkarian *et al.*, 2010).

12. Análisis bioinformático

Las proteínas sobreexpresadas y subexpresadas en las líneas celulares de cáncer de mama se agruparon en diagramas de Venn-Euler utilizando el programa Venny (<http://bioinfogp.cnb.csic.es/tools/venny>). Por otro lado, se utilizó el programa PANTHER (Análisis de proteínas a través de relaciones evolutivas) versión 9.0

(<http://www.pantherdb.org/>) para clasificar a las proteínas de acuerdo con su papel biológico. Adicionalmente, se obtuvieron los interactomas de las proteínas sobreexpresadas y subexpresadas con el programa STRING (Interacciones proteína-proteína predichas y conocidas) versión 9.1 (<http://string-db.org/>).

Para identificar enfermedades y funciones biológicas, vías canónicas, redes moleculares y posibles proteínas candidatas a biomarcadores de las proteínas que son expresadas comúnmente en todas las líneas celulares de cáncer de mama, se utilizó el programa IPA (Análisis de vías de señalización de Ingenuity) (QIAGENs Redwood City, www.qiagen.com/ingenuity). Se realizaron tres análisis diferentes, denominados Core I, II y III. En todos los casos se utilizó un filtro astringente y una prueba de Fisher para determinar la significancia estadística ($p \leq 0.05$) de las asignaciones. Core I se realizó para clasificar a las proteínas identificadas en todos los tejidos, células primarias y todas las líneas celulares de cáncer: Cervical, del sistema nervioso central, colon, hepatoma, inmune, riñón, leucemia, pulmón, linfoma, macrófagos, melanoma, mieloma, neuroblastoma, osteosarcoma, ovario, páncreas, próstata, teratocarcinoma y cáncer de mama, entre otras. El análisis Core II se enfocó en glándula mamaria y líneas celulares de cáncer de mama. Finalmente, el análisis Core III se realizó en todas las líneas celulares de cáncer a excepción de las líneas de cáncer de mama. Los parámetros utilizados en el análisis fueron: (i) Base de datos de ingenuity (genes), considerando relaciones directas e indirectas. (ii) Redes de interacción, que incluyeron endógenos químicos y 25 interactomas por análisis con 35 moléculas por red. (iii) Todas las bases de datos. (iv) Confidencia experimentalmente observada, alta y moderadamente predicha. (v) Especie humano con filtro astringente. (vi) Todas las mutaciones.

Para la búsqueda de biomarcadores se utilizaron dos estrategias: (I) Análisis con el módulo Biomarcadores de IPA y (II) búsqueda manual en la base de datos de PubMed. Para la primera estrategia, se realizaron tres análisis con filtros diferentes,

denominados Biomarcadores I, II y III. En el primer filtro se consideraron todos los tejidos, células primarias y todas las líneas celulares de cáncer. Biomarcadores II se acotó a glándula mamaria y líneas celulares de cáncer de mama, y Biomarcadores III se centró en todas las líneas celulares de cáncer excluyendo cáncer de mama. Los parámetros utilizados en todos los filtros fueron: (i) Especie, humano, (ii) todos los tipos de moléculas, (iii) todos los biofluidos, (iv) todas las enfermedades para Biomarcadores I y cáncer para Biomarcadores II y III, (v) todas las aplicaciones de biomarcadores y enfermedades para Biomarcadores I. En el caso de Biomarcadores II, se estableció el filtro de todas las aplicaciones de biomarcadores y enfermedades de cáncer de mama (incluyendo cáncer de mama, carcinoma, carcinoma ductal, carcinoma ductal *in situ*, carcinoma ductal infiltrante, cáncer ductal invasivo, carcinoma lobular y cáncer metastásico). Finalmente, en Biomarcadores III, se consideraron todas las aplicaciones de biomarcadores y enfermedades de cáncer, excluyendo cáncer de mama.

La segunda estrategia para priorizar proteínas candidatas a biomarcadores fue a través de la búsqueda manual en la base de datos PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). Se realizaron búsquedas de las 206 (78 sobreexpresadas y 128 subexpresadas) proteínas expresadas en común en todas las líneas celulares de cáncer de mama. En el análisis de búsqueda se consideraron todos los nombres alternativos de las proteínas. Los parámetros de búsqueda fueron palabras clave, como biomarcadores, cáncer y cáncer de mama. Las proteínas seleccionadas fueron aquéllas que no han sido reportadas previamente en cáncer de mama. Finalmente, se realizó un análisis de comparación de los resultados obtenidos con el módulo IPA-Biomarcadores y PubMed con el programa Venny.

Por otro lado, se identificaron las vías metabólicas en las que están involucradas algunas de las 206 proteínas con expresión diferencial a través de la plataforma KEGG (Enciclopedia Kyoto de genes y genomas) (<http://www.genome.jp/kegg/>).

13. Inmunodetección sobre membrana de nitrocelulosa (Western blot)

Los extractos totales de proteína se analizaron en geles de poliacrilamida al 12% en un sistema Mini-PROTEAN Tetra Cell (BioRad) y se transfirieron a membranas de nitrocelulosa (GE Healthcare Life Sciences) con un sistema Mini Trans-Blot Cell (Bio-Rad). Los sitios libres en las membranas se bloquearon con 5 mL de leche descremada al 5% en PBS 1X-Tween 20 al 0.05% (PBS-T) por 4 h a temperatura ambiente y se incubaron con las diluciones apropiadas de los anticuerpos primarios en 1 mL leche al 5% en PBS-T toda la noche a 4 °C con agitación suave. La dilución y anticuerpos utilizados fueron los siguientes: 1:1 000, anti-Proteína Ribosomal S7 (E-1), anti-BAG-6 (D-1), anti-COX4 (F-8) (Santa Cruz) anti-DDX39, anti-Anexina A8 (abcam). Después de la incubación, las membranas se lavaron 10 veces con 10 mL de PBS-T cada 10 min a temperatura ambiente en agitación constante. Subsecuentemente las membranas se incubaron con los anticuerpos secundarios correspondientes a una dilución 1:20,000, anti-ratón (Ribosomal S7, BAG6 y COX4) y anti-conejo (DDX39 y ANXA8) (Zymed) en 5 mL de leche al 5% en PBS-T con 2 µL de estreptactina (BioRad) para la visualización de marcadores de peso, durante 2 h a temperatura ambiente. Posteriormente, las membranas se lavaron 10 veces con 10 mL de PBS-T cada 5 min a temperatura ambiente en agitación constante. La inmunoreactividad se detectó por quimioluminiscencia con una dilución 1:8 de SuperSignal West Femto Maximum Sensitivity Substrate kit (Pierce) utilizando placas High Performance Chemiluminescence (GE Healthcare). Finalmente, las placas fueron digitalizadas (GelDoc EZ Imager, BioRad) y el análisis densitométrico se realizó con el programa Quantity One versión 4.6.6 (BioRad).

CAPÍTULO V

RESULTADOS

1. Estandarización de protocolos de extracción de proteínas

Los extractos proteicos obtenidos con diferentes métodos de extracción (Tritón-NP40, RIPA-Urea e hidrocloreuro de guanidina) se precipitaron con TCA 77% en acetona y se resuspendieron en Urea 8 M. Los extractos de proteínas se analizaron en un gel de poliacrilamida al 10% en condiciones desnaturalizantes para determinar cuál de ellos es el que proporcionó más bandas de proteínas, así como en cuál se observa una mayor definición de las mismas. Se observó que en el método de extracción con hidrocloreuro de guanidina 6 M se obtuvo una mayor definición e intensidad en las bandas (Figura 14, carril 3). En el análisis se incluyeron extractos totales, extractos totales incubados con bensonaza (para la eliminación de ácidos nucleicos) y extractos totales deslipidados, para verificar que en estos pasos no se presentaba una pérdida significativa de muestra (Figura 14).

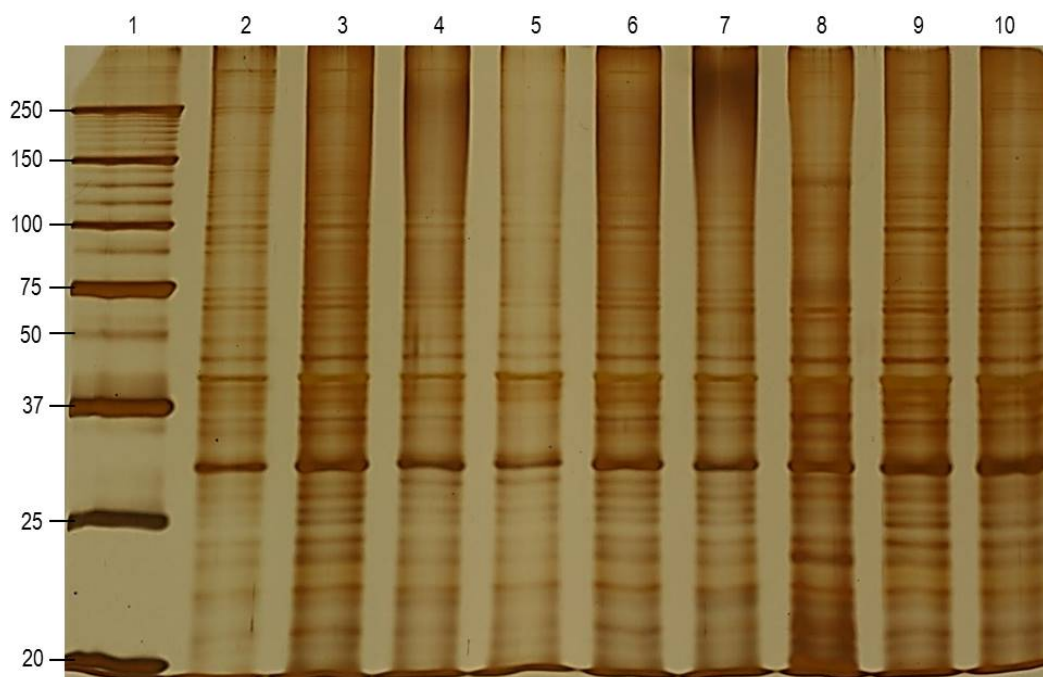


Figura 14. Extractos totales de proteína obtenidos con Tritón-NP40, RIPA-Urea e hidrocloreuro de guanidina. Las proteínas se separaron en un gel SDS-PAGE al 10% y se tñieron con nitrato de plata. Marcadores de masa molecular (Carril 1); 2 µg de proteínas obtenidas con Triton-NP40 (Carril 2), Guanidina (Carril 3) y Ripa-Urea (Carril 4). En el mismo orden se cargaron los extractos incubados con bensonaza (Carriles 5-7) y deslipidados (Carriles 8-10).

Los resultados obtenidos en el patrón electroforético de los extractos deslipidados, se verificaron por densitometría con el programa Image J (Rasband, W. 1997). El análisis densitométrico indicó que los tres métodos representan una buena opción para la extracción de proteínas (Figura 15a). Por otro lado, se realizó una detección del perfil de bandas de las proteínas más abundantes con el programa ImageLab. De acuerdo con los resultados obtenidos, en el método de Tritón-NP40 se lograron detectar 25 bandas, con GuHCl 30, mientras que con Ripa-Urea 27 (Figura 15b). Con base en estos resultados, se determinó que aun cuando los tres métodos son óptimos para la extracción de proteínas, el de hidrocloreuro de guanidina representa la mejor opción para la extracción de proteínas totales.

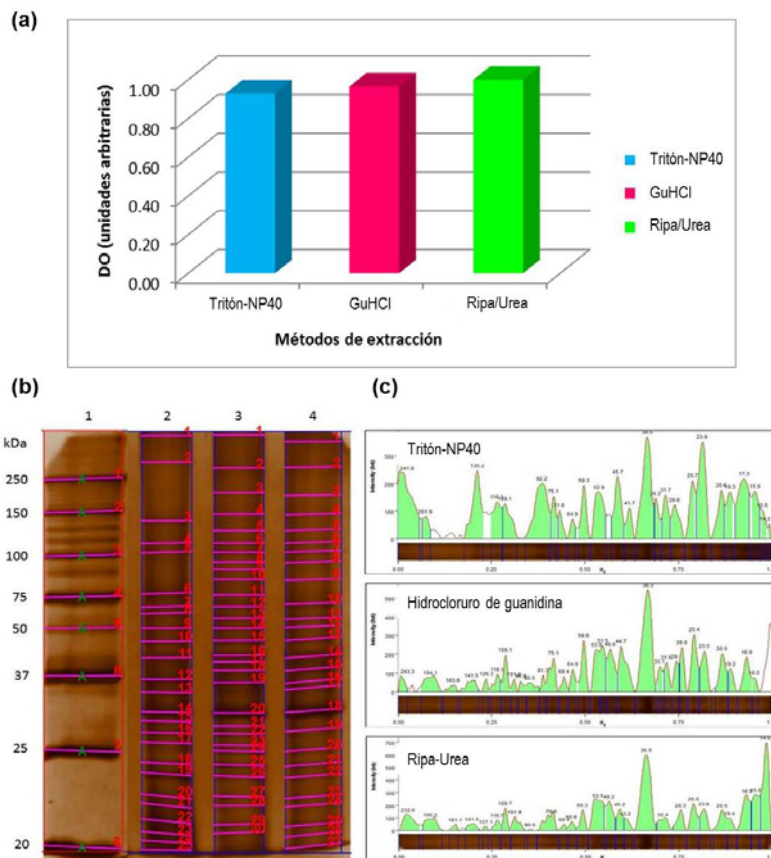


Figura 15. Análisis del patrón electroforético de extractos totales obtenidos con Tritón-NP40, RIPA-Urea e hidrocloreuro de guanidina. (a) Densitometría de bandas de proteína. (b) Perfil de bandas proteicas. Carril 1, Marcadores de masa molecular; Carril 2, Triton-NP40; Carril 3, Guanidina; Carril 4, Ripa-Urea. (c) Electroferograma de bandas proteicas obtenidas con los diferentes métodos de extracción.

En cuanto a la eliminación de ácidos nucleicos de las muestras, ésta se verificó a través de electroforesis en un gel agarosa al 1%. En este análisis se observó la presencia de ADN en cada uno de los extractos obtenidos con los tres métodos de extracción de proteína, por lo que se concluyó que la benzonaza no estaba digiriendo el ADN (Figura 16), probablemente debido a que ésta sólo tolera una concentración máxima de urea 6 M y sólo es compatible con detergentes no iónicos. Para verificar la eficiencia de la enzima se realizaron incubaciones con benzonaza al ADN λ y el resultado obtenido indicó que la enzima trabaja adecuadamente (Figura 16). Debido a lo anterior, se concluyó que las soluciones de extracción de proteína no proporcionan un medio adecuado para el óptimo funcionamiento de la enzima. Adicionalmente, la migración electroforética de las proteínas no se vio afectada por la presencia de los ácidos nucleicos, además de que estos no interfieren en el análisis por espectrometría de masas, debido a que estos poseen carga negativa y el espectrómetro de masas trabaja en modo positivo. Debido a esto se determinó que la eliminación de los ácidos nucleicos no es necesaria en el procesamiento de la muestra.

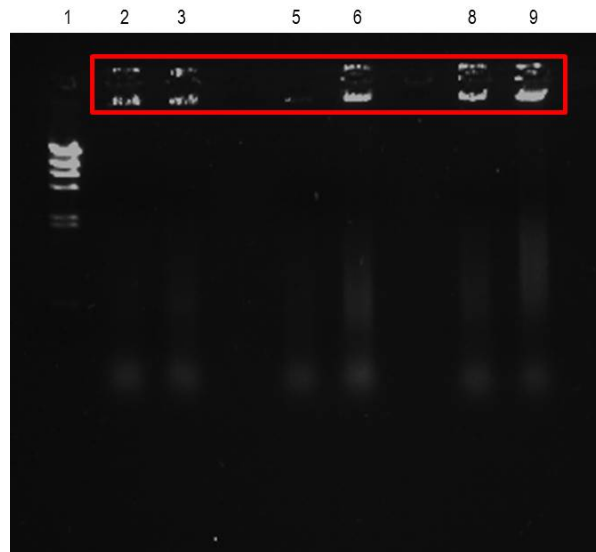


Figura 16. La benzonaza no elimina el ADN en extractos totales de proteína obtenidos con diferentes soluciones de lisis. Marcadores de alta masa molecular (Carril 1), extracto total de proteína obtenido con Tritón-NP40 (Carriles 2-3), GuHCl (Carriles 5-6) y Ripa-Urea (Carriles 8-9) sin y con benzonaza, respectivamente. Las muestras se analizaron en un gel de agarosa al 1% y se tiñeron con bromuro de etidio.

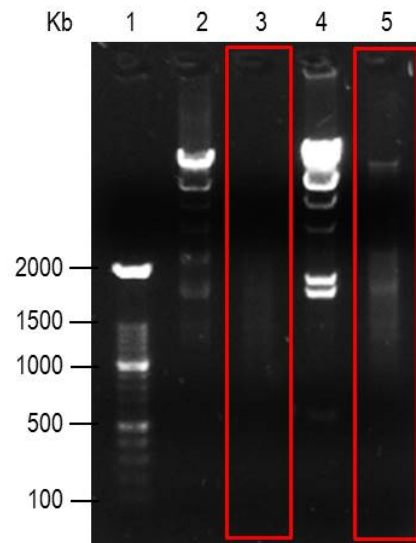


Figura 17. Eliminación de ADN del fago λ . Marcadores de 100 pb (Carril 1), 250 ng (Carriles 2-3) y 500 ng de ADN de λ sin y con benzonaza respectivamente (Carriles 4-5). Las muestras se analizaron en un gel de agarosa al 1%, teñido con bromuro de etidio.

Con base en otros trabajos de proteómica realizados, se decidió enriquecer el método de extracción de proteínas elegido (hidrocloruro de guanidina) con un detergente, con la finalidad de obtener el mayor número de proteínas posibles, es decir, extraer proteínas membranales. El detergente elegido fue RapiGest, debido a que es ampliamente reportado en protocolos de extracción de proteínas para estudios proteómicos. Este detergente ayuda a solubilizar proteínas membranales y además es compatible con espectrometría de masas. Para esto, se obtuvieron extractos proteicos de la línea celular HeLa con hidrocloruro de guanidina e hidrocloruro de guanidina-RapiGest (0.1%), de acuerdo con el protocolo ya establecido. Los extractos proteicos obtenidos en cada caso se analizaron en un gel de poliacrilamida, para determinar cuál de ellos es el que proporcionaba mayor número de bandas, así como cuál de ellos ofrecerá mayor definición de éstas. Se observó que en el método de extracción GuHCl-RapiGest se enriquecieron algunas bandas de alto peso molecular, sin embargo los patrones obtenidos en ambos métodos son muy parecidos (Figura 18).

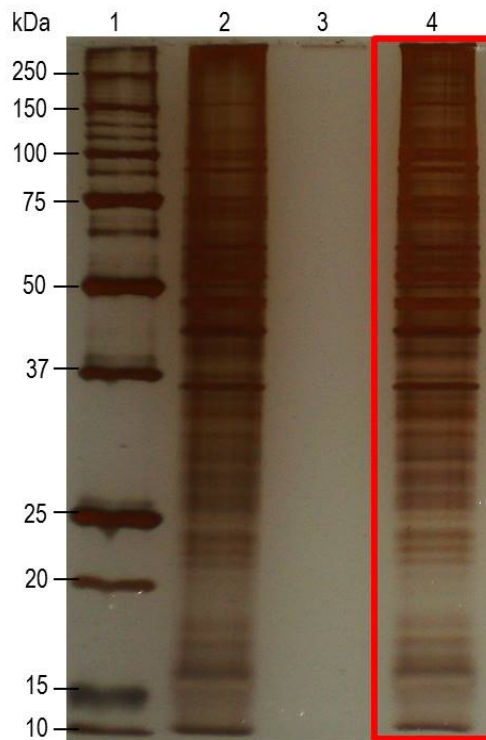


Figura 18. Extractos totales de proteína obtenidos con GuHCl y GuHCl-RapiGest. Las proteínas se separaron en un gel SDS-PAGE al 12% y se tiñeron con nitrato de plata. Marcadores de masa molecular (Carril 1); 2 μ g de proteína de extractos totales obtenidos con GuHCl (Carril 2), GuHCl-RapiGest (Carril 4)

Los resultados obtenidos en el patrón electroforético se analizaron con densitometría (Figura 19a). Se observó que el patrón del número de bandas obtenidas con el método de hidrocloreuro de guanidina fue ligeramente mayor con respecto al otro método de extracción; sin embargo, al realizar la detección del perfil proteico se identificaron 26 bandas con GuHCl y 35 con GuHCl-RapiGest (Figura 19b). Por esta razón decidimos que la mezcla de GuHCl-RapiGest es la mejor opción como método de extracción al enriquecer proteínas de alto peso molecular.

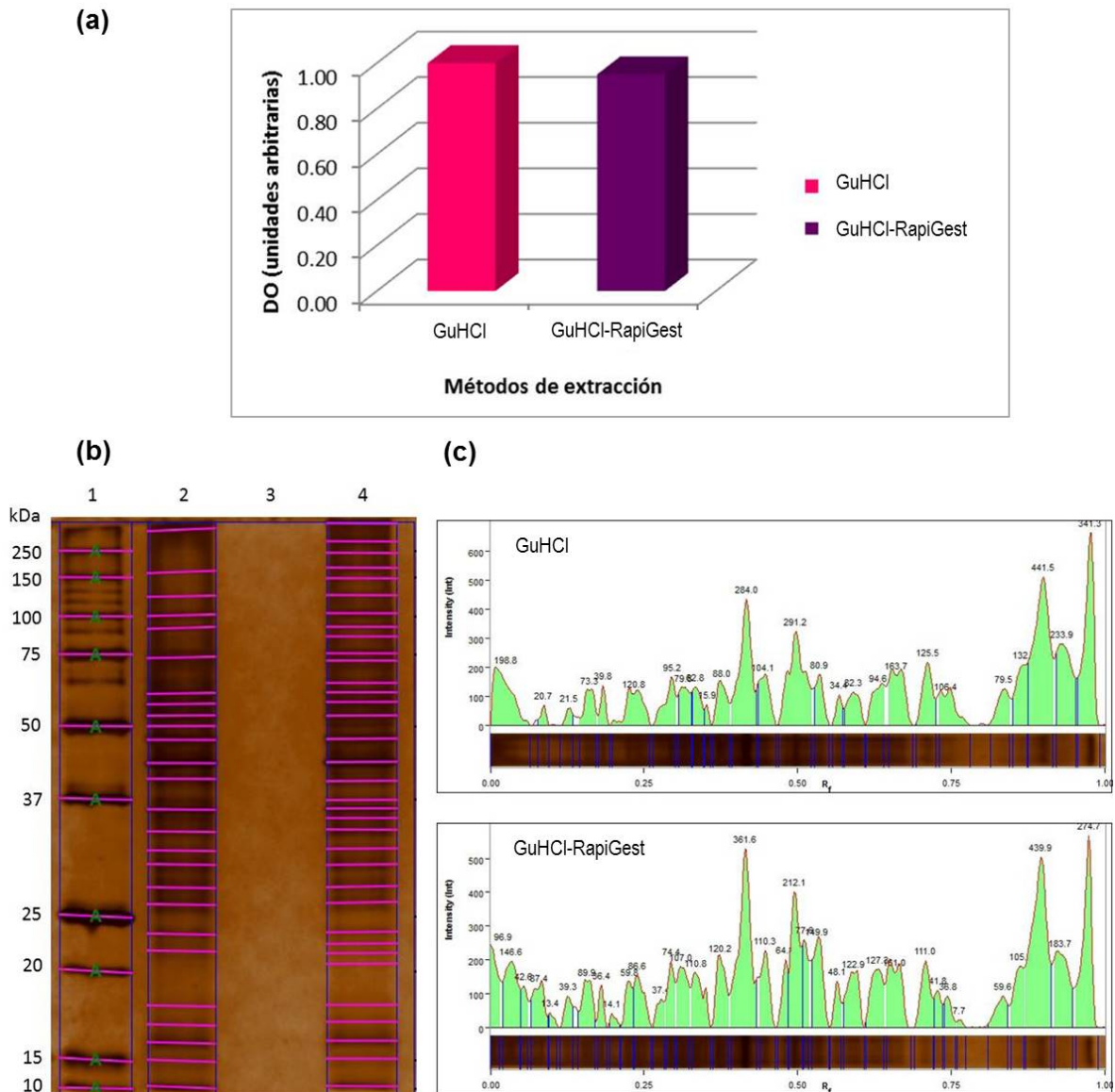


Figura 19. Análisis del patrón electroforético de extractos totales obtenidos con GuHCl e GuHCl-RapiGest. (a) Densitometría de bandas de proteína. (b) Perfil de bandas proteicas. Carril 1, Marcadores de masa molecular; Carril 2, GuHCl; Carril 4, GuHCl-RapiGest. (c) Electroferograma de bandas proteicas obtenías con cada uno de los métodos de extracción.

Una vez elegido el método de extracción de proteínas (GuHCl-RapiGest), se obtuvieron extractos totales de cada una de las líneas celulares empleadas en este trabajo. Los extractos obtenidos se cuantificaron por el método de Lowry y analizados en un gel de poliacrilamida al 10% (Figura 20).

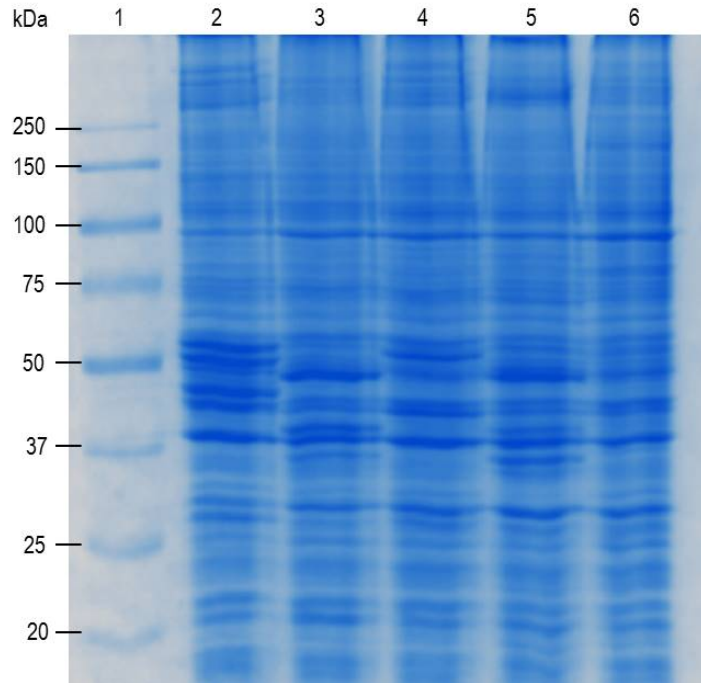


Figura 20. Perfil electroforético de extractos totales de proteína de líneas celulares de cáncer de mama. Los extractos proteicos obtenidos con GuHCl-RapiGest se separaron en un gel de poliacrilamida al 10% y se tiñeron con Azul de Coomassie. Marcadores de masa molecular (Carril 1); 20 µg de proteína de extractos totales de la línea celular MCF 10A (Carril 2), MCF7 (Carril 3), MDA-MB-231 (Carril 4), SK-BR-3 (Carril 5) y T47D (Carril 6).

2. Perfiles de expresión diferencial de las líneas celulares de CaMa

Para determinar el patrón de expresión diferencial de las líneas celulares de cáncer de mama (MCF7, MDA-MB-231, SK-BR3 y T47D), y compararlo con el de la línea celular control (MCF 10A) se realizó una aproximación de proteómica cuantitativa a través de marcaje con iTRAQ. Brevemente, los extractos de proteína de cada una de las líneas celulares se digirieron, marcaron, prefraccionaron y analizaron por espectrometría de masas en tándem. Los patrones de elución de los iones se obtuvieron a partir del minuto 25 hasta el minuto 150 (Figura 21a). La asignación y cuantificación de los péptidos se llevó a cabo en los espectros MS/MS obtenidos. Los espectros MS/MS de los péptidos LADHFGGK Y DHPLPEVAHVK correspondientes a las proteínas LUC7 y COX4, respectivamente, se muestran en la figura 21 (b-c).

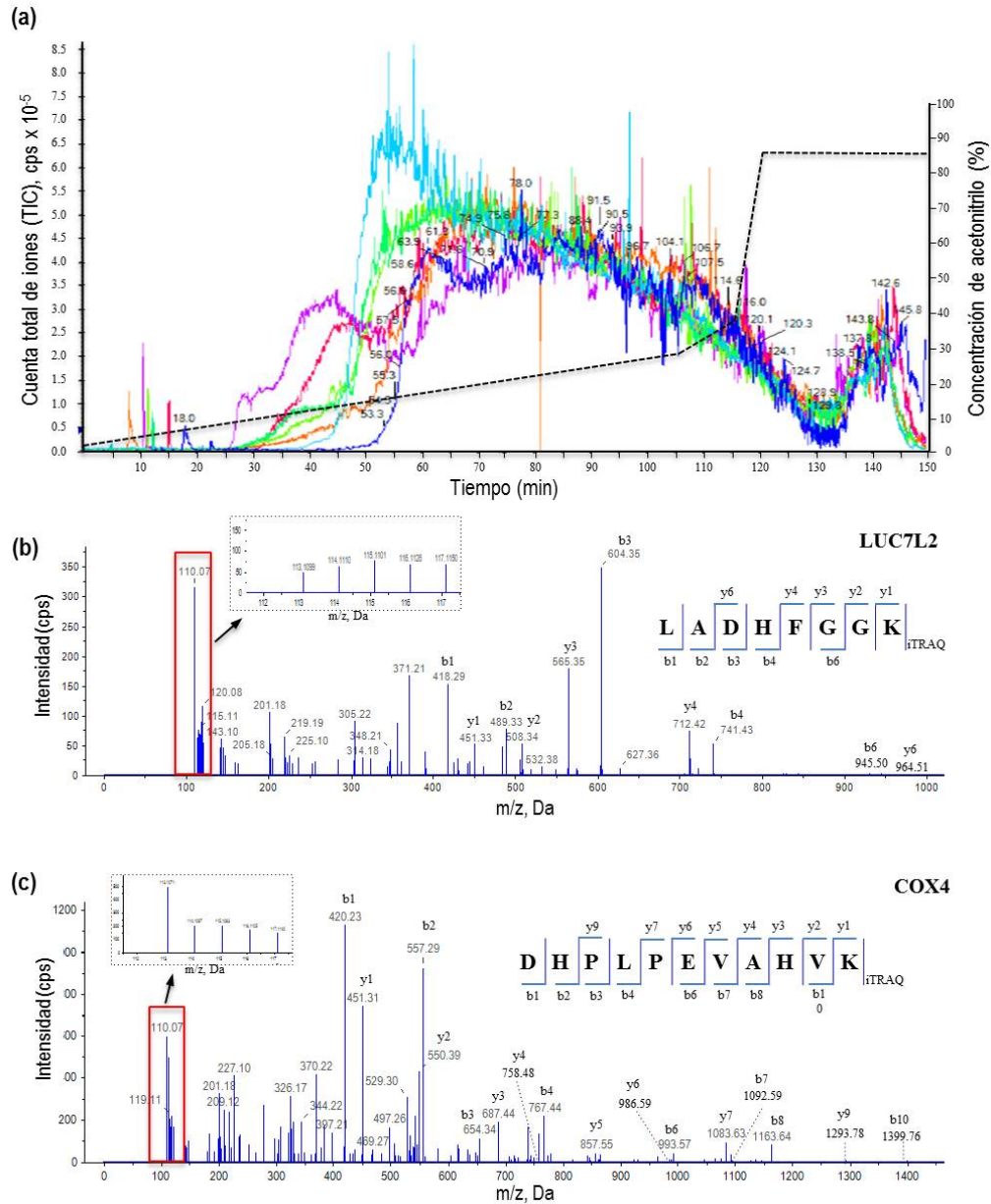


Figura 21. Identificación y cuantificación de péptidos de las líneas celulares de cáncer de mama a través de marcaje con iTRAQ y análisis por espectrometría de masas en tándem. (a) Cuenta total de iones como una función del tiempo de elución de la columna de fase reversa de las secciones 4-10 de la tira de IEF. Las diferentes secciones se indican con colores: sección 4 (azul), sección 5 (morado), sección 6 (rojo), sección 7 (naranja), sección 8 (verde claro), sección 9 (verde oscuro) y sección 10 (cian). (b) Espectro MS/MS del péptido LADHFGGK, correspondiente a la proteína LUC7L2. (c) Espectro MS/MS del péptido DHPLPEVAHVK, correspondiente a la proteína COX4. En las cajas insertadas en los espectros de masa, se observan los niveles de expresión relativa obtenida para cada péptido de cada una de las líneas celulares. Relación masa-carga (m/z). Las etiquetas de iTRAQ se asignaron de la siguiente manera: 113, MCF 10A; 114, MCF7; 115, MDA-MB-231; 116, SK-BR-3; 117, T47D. La intensidad para B y C se encuentra en cuentas por segundo (cps).

Se identificaron 1,020 proteínas marcadas con al menos un péptido con 95% de confianza (Calderón-González *et al.*, Data in brief. 2015). La lista de péptidos y proteínas se analizaron con la plantilla PDST para la generación de los FDR (False Discovery Rate), así como para graficar las proteínas de acuerdo con su cambio de expresión y el valor de p obtenido con la prueba de corrección de Bonferroni (Figura 22).

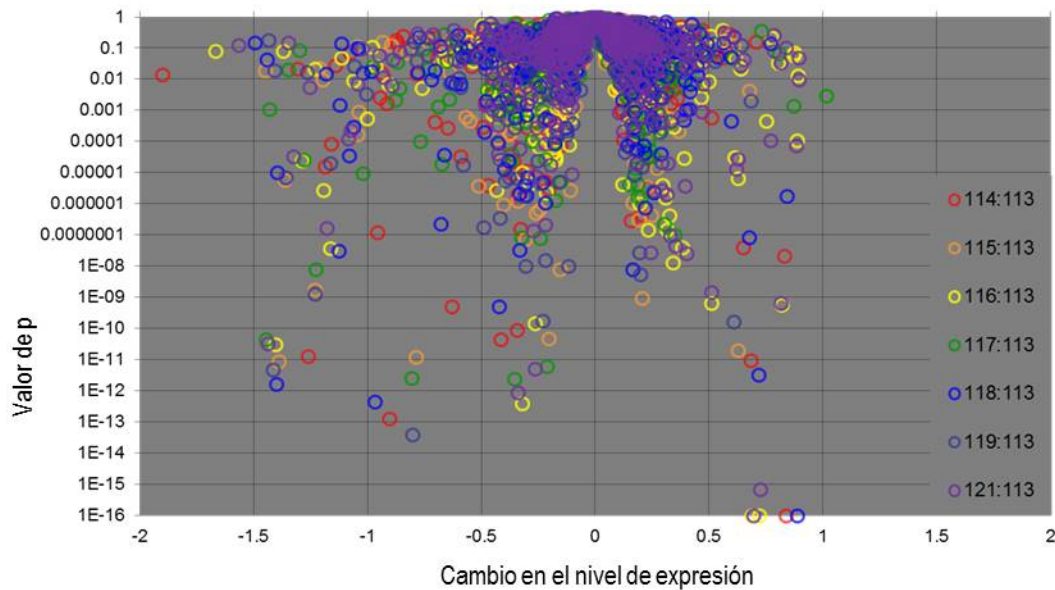


Figura 22. Gráfica de volcán de las proteínas expresadas en las líneas celulares de cáncer de mama. Todas las proteínas marcadas con las etiquetas de iTRAQ fueron graficadas. El eje X representa el cambio en el nivel de expresión de las proteínas identificadas. A la derecha del eje X se representan las proteínas sobreexpresadas (valores positivos) y a la izquierda se encuentran los polipéptidos subexpresados (valores negativos) en las líneas celulares de cáncer de mama en comparación con la línea celular control. El eje Y corresponde a la significación estadística (valor de p) obtenida para el cambio de expresión de las proteínas. Las muestras fueron etiquetadas de la siguiente manera: 113, MCF 10A; 114, MCF7; 115, MDA-MB-231; 116, SK-BR-3; 117, T47D. Las etiquetas 118, 119 y 121 fueron duplicados de las etiquetas 114, 115 y 116.

Posteriormente, los datos se exportaron a Excel y se analizaron manualmente, se establecieron los valores de corte >1.2 y <0.8 para proteínas sobreexpresadas y subexpresadas, respectivamente. Con estos valores de corte, se obtuvieron 336

proteínas sobreexpresadas en la línea celular MCF7, 326 en T47D, 319 en la MDA-MB-231 y 313 en la línea SK-BR-3 (Figura 23a). También se encontraron grupos de proteínas específicas para cada una de las líneas celulares: 41, 59, 55 y 74 polipéptidos para MCF7, T47D, MDA-MB-231 y SK-BR-3, respectivamente. Cabe mencionar que también se encontraron grupos de proteínas compartidas entre las líneas celulares: 34 proteínas para MCF7 y T47D, 30 para T47D y MDA-MB-231, 24 para MDA-MB-231 y SK-BR-3, mientras que MCF7 and SK-BR-3 compartieron 35, MCF7 y MDA-MB-231 compartieron 35 y T47D y SK-BR-3 tuvieron 19 polipéptidos en común. Otro conjunto de proteínas compartidas fueron 50, 36, 20, y 27 para MCF7/T47D/MDA-MB-231, MCF7/T47D/SK-BR-3, T47D/MDA-MB-231/SK-BR-3, MCF7/MDA-MB-231/SK-BR-3, respectivamente. Finalmente, todas las líneas celulares compartieron 78 proteínas (Tabla 3).

En relación con las proteínas subexpresadas, se encontraron 337 proteínas en MCF7, 321 en T47D, 310 en MDA-MB-231 y 301 en SK-BR-3 (Figura 23b). De igual manera, se encontraron conjuntos de proteínas para cada línea celular: 39, 32, 44, y 49 polipéptidos para MCF7, T47D, MDA-MB-231, y SK-BR-3, respectivamente. Las líneas MCF7 y T47D compartieron 22 polipéptidos, T47D y MDA-MB-231 tuvieron 28 proteínas en común, MDA-MB-231 y SK-BR-3 compartieron 9, MCF7 y SK-BR-3 tuvieron 33, MCF7 y MDA-MB-231 solo 19, T47D y SK-BR-3 tuvieron 22 proteínas en común. Se encontraron 49, 27, 13, y 20 polipéptidos compartidos en las líneas MCF7/T47D/MDA-MB-231, MCF7/T47D/SK-BR-3, T47D/MDA-MB-231/SK-BR-3, MCF7/MDA-MB-231/SK-BR-3, respectivamente. Finalmente, se identificaron 128 proteínas en común en todas las líneas celulares de cáncer de mama (Tabla 4). Con estos resultados, se decidió analizar primero las proteínas sobre y subexpresadas en común en todas las líneas celulares de cáncer de mama. Posteriormente se analizaron los conjuntos de proteínas exclusivas para cada una de las líneas celulares.

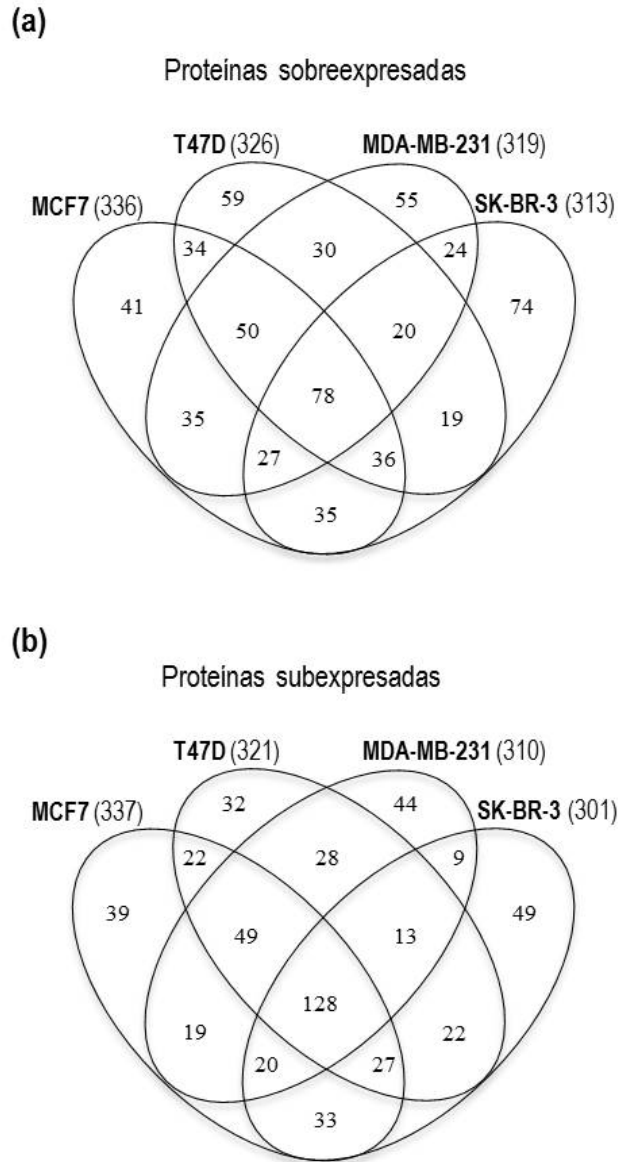


Figura 23. Diagramas de Venn-Euler de las proteínas diferencialmente expresadas en las líneas celulares de cáncer de mama. (a) polipéptidos sobreexpresados. (b) proteínas subexpresadas. Los valores de corte utilizados fueron >1.2 para las proteínas sobreexpresadas y <0.8 para los polipéptidos subexpresados.

Tabla 3. Proteínas sobreexpresadas encontradas en común en todas las líneas celulares de cáncer de mama por iTRAQ y espectrometría de masas en tándem.

N	Nombres de proteínas	Genes	Localización	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (≥95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
1	Glyceraldehyde-3-phosphate dehydrogenase	<i>GAPDH</i>	C	P04406	35.92	8.58	43.39	58.2	50	1.36	1.46	2.07	1.64
2	Heat shock protein HSP 90-alpha	<i>HSP90AA1</i>	C	P07900	84.53	4.94	32.98	28.8	37	1.71	1.69	1.56	2.02
3	Triosephosphate isomerase	<i>TPI1</i>	C	P60174	30.79	5.65	27.59	45.8	27	1.48	1.49	1.70	1.70
4	Cofilin-1	<i>CFL1</i>	N	P23528	18.37	8.26	13.75	53.6	8	2.20	1.75	1.56	1.45
5	Annexin A6	<i>ANXA6</i>	PM	P08133	75.74	5.41	13.08	12.5	8	1.57	4.75	3.14	2.28
6	Rab GDP dissociation inhibitor beta	<i>GDI2</i>	C	P50395	50.66	6.10	11.37	20.5	8	1.51	1.50	1.85	1.55
7	T-complex protein 1 subunit delta	<i>CCT4</i>	C	P50991	57.79	8.13	11.25	14.5	6	1.38	1.25	1.22	1.54
8	ATP-dependent RNA helicase DDX39A	<i>DDX39A</i>	N	O00148	49.00	5.46	11.19	12.9	8	2.29	3.25	1.27	1.41
9	Proteasome activator complex subunit 1	<i>PSME1</i>	C	Q06323	28.72	5.78	10.12	28.9	6	2.14	1.27	2.91	2.17
10	Eukaryotic translation initiation factor 5	<i>EIF5</i>	C	P55010	49.22	5.41	9.58	11.4	6	1.43	1.28	1.50	1.94
11	Phosphatidylethanolamine-binding protein 1	<i>PEBP1</i>	C	P30086	20.93	7.43	8.78	25.1	4	1.66	1.11	1.24	1.66
12	High mobility group protein B1	<i>HMGB1</i>	N	P09429	24.76	5.60	8.57	17.2	5	3.25	1.89	1.98	1.93
13	DNA topoisomerase 2-alpha	<i>TOP2A</i>	N	P11388	174.39	8.82	8.26	2.9	4	1.25	1.41	1.68	1.46
14	Putative RNA-binding protein Luc7-like 2	<i>LUC7L2</i>	Other	Q9Y383	46.51	10.02	8	8.7	5	1.54	1.63	1.35	1.54
15	6-phosphogluconate dehydrogenase, decarboxylating	<i>PGD</i>	C	P52209	53.01	6.88	7.62	7.0	3	1.92	1.36	1.91	1.89
16	Lactoylglutathione lyase	<i>GLO1</i>	C	Q04760	20.65	5.12	7.58	18.5	4	1.48	1.84	1.47	2.27
17	ATP-dependent RNA helicase DDX1	<i>DDX1</i>	N	Q92499	82.43	6.80	7.55	8.2	6	1.60	1.96	1.37	1.31
18	Exportin-1	<i>XPO1</i>	N	O14980	123.39	5.71	6.86	4.9	4	1.75	1.46	1.67	2.01
19	Proteasome activator complex subunit 2	<i>PSME2</i>	C	Q9UL46	27.27	5.54	6.46	16.3	4	1.63	1.41	1.84	1.43
20	Coatomer subunit gamma	<i>COPG1</i>	C	Q9Y678	97.72	5.32	5.7	3.5	3	1.56	1.38	1.64	1.27
21	Solute carrier family 2, facilitated glucose transporter member 1	<i>SLC2A1</i>	PM	P11166	54.08	8.93	4.57	1.8	2	1.83	1.75	1.25	1.24
22	Histone H1.4	<i>HIST1H1E</i>	N	P10412	21.73	11.03	4.42	10.1	2	1.60	1.68	1.49	2.72
23	Serine/arginine-rich splicing factor 2	<i>SRSF2</i>	N	Q01130	25.35	11.86	4.23	11.3	2	1.61	1.58	1.33	1.73
24	Acidic leucine-rich nuclear phosphoprotein 32 family member B	<i>ANP32B</i>	N	Q92688	28.79	3.93	4.22	8.4	2	1.91	1.74	1.26	1.63
25	DNA topoisomerase 2-beta	<i>TOP2B</i>	N	Q02880	183.14	8.16	4.05	2.6	4	1.67	1.97	1.96	1.69
26	Acetyl-CoA acetyltransferase	<i>ACAT2</i>	C	Q9BWD1	41.35	6.46	4.05	4.8	3	2.17	2.76	1.26	2.88
27	Thioredoxin reductase 1	<i>TXNRD1</i>	C	Q16881	70.91	7.16	4.05	5.2	2	2.37	1.49	5.94	5.36
28	Eukaryotic translation initiation factor 4 gamma 2	<i>EIF4G2</i>	C	P78344	102.36	6.69	4.03	2.1	2	1.64	1.26	1.53	2.73
29	Vesicle-fusing ATPase	<i>NSF</i>	C	P46459	82.59	6.52	4	4.0	2	1.26	2.33	3.56	2.32
30	WD40 repeat-containing protein SMU1	<i>SMU1</i>	N	Q2TAY7	57.54	6.74	4	3.9	2	1.72	1.31	1.59	1.99
31	Ribose-phosphate pyrophosphokinase 1	<i>PRPS1</i>	C	P60891	34.70	6.56	4	8.5	3	1.32	1.84	2.18	3.06
32	ATP-dependent RNA helicase DDX42	<i>DDX42</i>	C	Q86XP3	102.98	6.54	4	3.0	2	1.30	1.36	1.94	1.37
33	6-phosphogluconolactonase	<i>PGLS</i>	C	O95336	27.42	5.71	4	12.0	2	2.03	5.76	1.68	1.26
34	Guanine nucleotide-binding protein subunit beta-2-like 1	<i>GNB2L1</i>	C	P63244	35.08	7.60	3.94	12.6	3	1.43	1.32	1.41	1.62
35	UPF0160 protein MYG1, mitochondrial	<i>C12orf10</i>	N, M	Q9HB07	37.11	5.50	3.89	5.9	2	2.39	1.56	1.39	1.85
36	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	<i>PPP2R5D</i>	N	Q14738	69.99	8.24	3.85	4.2	2	3.20	2.10	1.34	3.93
37	UDP-glucose 4-epimerase	<i>GALE</i>	C	Q14376	38.28	6.26	3.6	13.2	2	1.53	1.53	4.29	1.53
38	DnaJ homolog subfamily A member 1	<i>DNAJA1</i>	N	P31689	44.55	6.65	3.56	2.0	1	2.16	1.95	1.47	1.48

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

39	Vacuolar protein sorting-associated protein 26A	<i>VPS26A</i>	C	O75436	38.17	6.13	3.22	6.1	2	1.56	1.96	2.33	2.28
40	HLA class I histocompatibility antigen, A-69 alpha chain	<i>HLA-A</i>	PM	P10316	40.98	6.42	3.21	5.5	2	1.71	4.04	1.96	2.96
41	Histidyl-tRNA synthetase	<i>HARS</i>	C	P12081	57.28	5.73	3.18	3.3	2	1.20	1.37	1.72	1.65
42	Large proline-rich protein BAG6	<i>BAG6</i>	N	P46379	119.41	5.40	3.17	2.1	2	2.40	2.42	1.38	1.91
43	Endoplasmic reticulum resident protein 29	<i>ERP29</i>	C	P30040	25.85	6.08	3.1	8.4	2	1.43	1.34	1.26	1.40
44	Eukaryotic translation initiation factor 4B	<i>EIF4B</i>	C	P23588	69.15	5.54	2.64	2.5	2	2.46	1.97	1.29	1.83
45	Adenylyl cyclase-associated protein 2	<i>CAP2</i>	PM	P40123	52.69	5.96	2.41	2.1	1	1.99	2.61	2.36	1.28
46	Activated RNA polymerase II transcriptional coactivator p15	<i>SUB1</i>	N	P53999	14.26	9.60	2.17	10.2	1	1.30	1.83	1.79	1.68
47	Proteasome subunit beta type-2	<i>PSMB2</i>	C	P49721	22.84	6.52	2.15	9.5	1	2.41	2.06	1.99	2.20
48	ATP-dependent RNA helicase DDX19A	<i>DDX19A</i>	N	Q9NUU7	53.84	6.21	2.14	4.0	1	1.91	1.21	1.27	1.95
49	Peptidyl-prolyl cis-trans isomerase FKBP3	<i>FKBP3</i>	N	Q00688	25.05	9.29	2.14	3.6	1	2.39	2.00	1.56	1.54
50	Eukaryotic translation initiation factor 4A-III	<i>EIF4A3</i>	N	P38919	46.87	6.30	2.11	6.1	2	1.60	1.39	1.46	1.35
51	Constitutive coactivator of PPAR-gamma-like protein 1	<i>FAM120A</i>	C	Q9NZB2	121.89	9.07	2.1	1.3	1	1.33	1.43	1.57	1.82
52	40S ribosomal protein S27	<i>RPS27</i>	C	P42677	9.33	9.58	2.1	14.3	2	1.71	1.28	1.21	1.27
53	Cysteine and glycine-rich protein 1	<i>CSRP1</i>	N	P21291	20.57	8.90	2.09	7.8	2	1.62	1.24	3.62	2.84
54	Cleavage and polyadenylation specificity factor subunit 5	<i>NUDT21</i>	N	O43809	26.10	8.85	2.08	10.1	3	1.76	1.79	1.64	1.69
55	Nuclear transport factor 2	<i>NUTF2</i>	N	P61970	14.48	5.10	2.08	6.3	1	2.55	1.68	1.22	1.21
56	Poly(rC)-binding protein 2	<i>PCBP2</i>	N	Q15366	38.58	6.33	2.05	14.2	4	1.51	1.42	1.43	1.42
57	Serine/arginine-rich splicing factor 3	<i>SRSF3</i>	N	P84103	19.33	11.64	2.03	5.5	1	2.68	1.75	1.88	2.18
58	Tropomyosin beta chain	<i>TPM2</i>	C	P07951	32.85	4.66	2	8.5	3	1.38	1.66	1.43	1.41
59	C-Jun-amino-terminal kinase-interacting protein 4	<i>SPAG9</i>	C	O60271	146.21	5.04	2	0.8	1	1.76	1.43	1.26	1.34
60	Ubiquitin-associated protein 2-like	<i>UBAP2L</i>	Other	Q14157	114.53	6.61	2	0.8	1	1.23	1.45	1.68	2.38
61	Selenide, water dikinase 1	<i>SEPHS1</i>	Other	P49903	42.78	5.64	2	3.8	1	1.65	1.26	1.41	2.16
62	Acyl-CoA-binding protein	<i>DBI</i>	C	P07108	9.91	6.11	2	18.4	1	3.47	1.32	2.33	1.40
63	Carnitine O-palmitoyltransferase 1, liver isoform	<i>CPT1A</i>	C	P50416	88.24	8.85	2	2.2	1	14.85	3.46	26.29	8.85
64	Myelin expression factor 2	<i>MYEF2</i>	N	Q9P2K5	64.12	8.86	2	1.7	1	6.78	3.09	3.48	2.79
65	DNA replication licensing factor MCM2	<i>MCM2</i>	N	P49736	101.76	5.34	2	1.9	1	7.20	5.77	4.52	4.98
66	Calcium-regulated heat stable protein 1	<i>CARHSP1</i>	C	Q9Y2V2	15.76	8.40	2	10.9	2	2.75	3.06	3.23	4.68
67	Prostaglandin E synthase 3	<i>PTGES3</i>	C	Q15185	18.70	4.32	2	6.3	1	1.69	1.44	1.57	1.44
68	Glycogen debranching enzyme	<i>AGL</i>	C	P35573	174.76	6.31	1.92	1.1	1	1.35	1.83	5.32	3.44
69	Heterogeneous nuclear ribonucleoprotein H3	<i>HNRNPH3</i>	N	P31942	36.93	6.37	1.89	2.9	1	1.33	1.46	1.30	1.44
70	ATP-binding cassette sub-family F member 3	<i>ABCF3</i>	Other	Q9NUQ8	79.61	5.96	1.89	1.8	1	1.89	1.73	1.58	2.04
71	N-terminal kinase-like protein	<i>SCYL1</i>	C	Q96KG9	89.63	5.93	1.85	1.9	1	4.28	1.74	2.30	1.78
72	Beta-catenin-like protein 1	<i>CTNBL1</i>	N	Q8WYA6	65.17	4.96	1.83	1.4	1	1.58	2.01	1.65	1.73
73	26S proteasome non-ATPase regulatory subunit 5	<i>PSMD5</i>	Other	Q16401	56.06	5.34	1.82	2.0	1	1.75	1.31	1.56	1.67
74	Nuclear pore membrane glycoprotein 210	<i>NUP210</i>	N	Q8TEM1	202.66	6.27	1.77	0.4	1	2.67	1.54	1.40	3.04
75	Basic leucine zipper and W2 domain-containing protein 2	<i>BZW2</i>	C	Q9Y6E2	48.16	6.26	1.67	2.1	1	1.65	1.36	1.44	2.05
76	High mobility group protein B2	<i>HMGB2</i>	N	P26583	23.90	7.77	1.47	6.2	1	4.16	3.87	2.14	2.49
77	Protein phosphatase 1G	<i>PPM1G</i>	N	O15355	59.27	4.27	1.44	1.5	1	2.34	1.92	1.65	2.23
78	Myosin-XVIIIa	<i>MYO18A</i>	C	Q92614	233.12	5.95	1.41	1.0	2	1.24	1.63	2.94	1.46

Tabla 4. Proteínas subexpresadas encontradas en común en todas las líneas celulares de cáncer de mama por iTRAQ y espectrometría de masas en tándem.

Nombres de proteínas		Genes	Localización	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (≥95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
1	Plectin OS	<i>PLEC</i>	C	Q15149	531.79	5.74	83.39	10.4	47	0.38	0.78	0.48	0.61
2	Myosin-9	<i>MYH9</i>	C	P35579	226.40	5.50	52.32	13.2	37	0.23	0.70	0.77	0.57
3	Filamin-B	<i>FLNB</i>	C	O75369	278.16	5.47	48.96	11.8	28	0.45	0.62	0.54	0.44
4	Prelamin-A/C	<i>LMNA</i>	N	P02545	73.81	6.57	31.76	32.4	23	0.62	0.77	0.79	0.78
5	Stress-70 protein, mitochondrial	<i>HSPA9</i>	M,N	P38646	68.76	5.44	29.02	27.1	23	0.47	0.46	0.67	0.79
6	ATP synthase subunit beta, mitochondrial	<i>ATP5B</i>	M	P06576	51.77	5.00	28.95	37.4	28	0.43	0.31	0.53	0.52
7	Annexin A2	<i>ANXA2</i>	PM	P07355	38.47	7.56	28.44	42.8	21	0.32	0.70	0.35	0.38
8	ATP synthase subunit alpha, mitochondrial	<i>ATP5A1</i>	M	P25705	55.21	8.28	22.76	21.3	11	0.50	0.42	0.54	0.58
9	Heterogeneous nuclear ribonucleoprotein U	<i>HNRNPU</i>	N	Q00839	90.45	5.76	21.44	11.9	11	0.63	0.73	0.57	0.72
10	Trifunctional enzyme subunit alpha, mitochondrial	<i>HADHA</i>	M	P40939	79.01	8.98	19.15	18.4	10	0.34	0.38	0.61	0.42
11	4F2 cell-surface antigen heavy chain	<i>SLC3A2</i>	PM	P08195	67.99	4.89	16.6	15.7	10	0.50	0.27	0.37	0.78
12	10 kDa heat shock protein, mitochondrial	<i>HSPA1</i>	M	P61604	10.80	8.91	14.04	69.6	8	0.46	0.39	0.51	0.60
13	Voltage-dependent anion-selective channel protein 1	<i>VDAC1</i>	C	P21796	30.64	8.63	13.22	27.6	7	0.23	0.38	0.34	0.40
14	Trifunctional enzyme subunit beta, mitochondrial	<i>HADHB</i>	M	P55084	47.48	9.24	12.59	11.2	7	0.44	0.52	0.67	0.50
15	Heterogeneous nuclear ribonucleoprotein Q	<i>SYNCRIP</i>	N	O60506	69.47	8.68	12.47	10.0	6	0.64	0.67	0.32	0.63
16	Cytochrome b-c1 complex subunit 1, mitochondrial	<i>UQCRC1</i>	M	P31930	49.13	5.43	12.26	19.4	8	0.49	0.47	0.63	0.43
17	EH domain-containing protein 2	<i>EHD2</i>	N	Q9NZN4	61.16	6.02	12.06	10.3	6	0.21	0.30	0.11	0.10
18	Core histone macro-H2A.1	<i>H2AFY</i>	N	O75367	39.62	9.80	11.54	18.5	7	0.43	0.57	0.50	0.46
19	Prohibitin	<i>PHB</i>	N	P35232	29.80	5.57	10.75	32.7	9	0.40	0.52	0.40	0.63
20	Exportin-2	<i>CSE1L</i>	N	P55060	110.42	5.51	10.59	8.8	8	0.68	0.49	0.62	0.45
21	Glutamate dehydrogenase 1, mitochondrial	<i>GLUD1</i>	M	P00367	56.01	6.71	10.57	10.9	6	0.35	0.39	0.29	0.30
22	Eukaryotic translation initiation factor 5B	<i>EIF5B</i>	C	O60841	138.83	5.39	10.02	6.1	6	0.50	0.51	0.41	0.41
23	Desmoplakin	<i>DSP</i>	PM	P15924	331.77	6.44	9.97	1.9	4	0.64	0.36	0.40	0.44
24	Junction plakoglobin	<i>JUP</i>	PM	P14923	81.74	5.75	9.95	7.9	9	0.33	0.37	0.40	0.52
25	Isoleucyl-tRNA synthetase, mitochondrial	<i>IARS2</i>	M	Q9NSE4	108.74	6.14	8.59	7.1	5	0.44	0.40	0.54	0.49
26	Fumarate hydratase, mitochondrial	<i>FH</i>	M	P07954	50.08	6.99	8.38	12.2	5	0.20	0.28	0.38	0.23
27	Sulfide:quinone oxidoreductase, mitochondrial	<i>SQRDL</i>	M	Q9Y6N5	49.96	9.18	8.14	11.3	4	0.26	0.42	0.37	0.21
28	Glutathione S-transferase P	<i>GSTP1</i>	C	P09211	23.22	5.44	8.13	36.7	5	0.11	0.23	0.09	0.20
29	Superoxide dismutase [Mn], mitochondrial	<i>SOD2</i>	M	P04179	22.20	6.86	7.84	28.4	6	0.13	0.13	0.23	0.13
30	Mitochondrial inner membrane protein	<i>IMMT</i>	M	Q16891	83.68	6.08	7.71	8.2	6	0.48	0.60	0.55	0.62
31	Phosphate carrier protein, mitochondrial	<i>SLC25A3</i>	M	Q00325	34.90	9.29	7.37	10.8	7	0.63	0.72	0.76	0.54
32	NADH-cytochrome b5 reductase 3	<i>CYB5R3</i>	C	P00387	34.10	7.31	7.06	17.9	6	0.45	0.78	0.59	0.48
33	Myosin-1c	<i>MYO1C</i>	C	O00159	121.68	9.46	7.03	6.5	5	0.57	0.79	0.51	0.34
34	Electron transfer flavoprotein subunit alpha, mitochondrial	<i>ETFA</i>	M	P13804	35.07	8.62	6.59	15.9	7	0.36	0.23	0.30	0.29
35	Adenylate kinase 2, mitochondrial	<i>AK2</i>	M	P54819	26.48	7.67	6.58	23.0	6	0.36	0.44	0.71	0.60
36	Polymerase I and transcript release factor	<i>PTRF</i>	N	Q6NZI2	43.48	5.50	6.1	7.9	3	0.12	0.68	0.17	0.13
37	Cytochrome c1, heme protein, mitochondrial	<i>CYC1</i>	M	P08574	27.35	6.49	6.06	8.3	3	0.42	0.38	0.41	0.40
38	RuvB-like 1	<i>RUVBL1</i>	N	Q9Y265	50.23	6.02	6.04	8.3	3	0.73	0.73	0.60	0.71
39	Nuclear pore complex protein Nup133	<i>NUP133</i>	N	Q8WUM0	128.98	4.98	6	2.9	3	0.32	0.29	0.66	0.34
40	Dihydrolipoyl dehydrogenase, mitochondrial	<i>DLD</i>	M	P09622	50.17	6.50	6	7.7	3	0.35	0.52	0.78	0.44

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

41	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	<i>SDHA</i>	M	P31040	68.01	6.25	5.71	5.6	3	0.54	0.49	0.74	0.48
42	Nucleolar protein 58	<i>NOP58</i>	N	Q9Y2X3	59.58	9.03	5.51	8.3	4	0.71	0.61	0.60	0.68
43	Presequence protease, mitochondrial	<i>PITRM1</i>	M	Q5JRX3	115.64	6.26	5.38	3.2	3	0.57	0.39	0.46	0.42
44	Myosin light polypeptide 6	<i>MYL6</i>	C	P60660	16.80	4.56	5.35	21.9	7	0.33	0.76	0.79	0.79
45	Aconitate hydratase, mitochondrial	<i>ACO2</i>	M	Q99798	82.43	6.85	5.26	5.0	4	0.42	0.41	0.79	0.50
46	Creatine kinase U-type, mitochondrial	<i>CKMT1A</i>	M	P12532	43.08	7.31	5.1	4.3	4	0.28	0.15	0.51	0.16
47	ATP synthase subunit gamma, mitochondrial	<i>ATP5C1</i>	M	P36542	30.17	9.02	4.66	11.1	3	0.54	0.37	0.64	0.61
48	Cytochrome b-c1 complex subunit 2, mitochondrial	<i>UQCRC2</i>	M	P22695	46.78	7.74	4.5	6.2	2	0.59	0.62	0.66	0.63
49	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	<i>DLST</i>	M	P36957	41.39	5.90	4.32	4.4	2	0.55	0.52	0.69	0.35
50	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	<i>COX4II</i>	M	P13073	17.20	9.16	4.23	7.1	2	0.43	0.47	0.44	0.35
51	Mitochondrial 2-oxoglutarate/malate carrier protein	<i>SLC25A11</i>	M	Q02978	33.93	9.92	4.22	9.2	3	0.36	0.71	0.73	0.59
52	Catenin alpha-1	<i>CTNNA1</i>	PM	P35221	99.94	5.95	4.15	3.3	2	0.50	0.34	0.42	0.51
53	DNA (cytosine-5)-methyltransferase 1	<i>DNMT1</i>	N	P26358	183.17	7.99	4.04	1.7	2	0.44	0.29	0.11	0.25
54	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	<i>NDUFS8</i>	M	O00217	20.29	5.10	4.04	9.5	2	0.31	0.50	0.50	0.44
55	Tumor-associated calcium signal transducer 2	<i>TACSTD2</i>	PM	P09758	33.12	8.98	4.03	8.0	3	0.42	0.23	0.47	0.34
56	Calcium-binding mitochondrial carrier protein SCaMC-1	<i>SLC25A24</i>	M	Q6NUK1	53.35	6.00	4.02	6.1	2	0.48	0.58	0.58	0.46
57	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	<i>SDHB</i>	M	P21912	28.81	8.78	4.02	8.2	2	0.44	0.53	0.77	0.44
58	Alpha-aminoadipic semialdehyde dehydrogenase	<i>ALDH7A1</i>	C	P49419	55.58	6.47	4.02	5.8	3	0.27	0.74	0.61	0.40
59	Inorganic pyrophosphatase 2, mitochondrial	<i>PPA2</i>	M	Q9H2U2	34.71	5.97	4	11.7	3	0.45	0.17	0.59	0.39
60	3-ketoacyl-CoA thiolase, mitochondrial	<i>ACAA2</i>	M	P42765	41.92	8.32	4	5.5	2	0.20	0.32	0.47	0.27
61	5'-3' exoribonuclease 2	<i>XRN2</i>	N	Q9H0D6	108.58	7.26	4	2.6	2	0.43	0.46	0.35	0.34
62	Cytochrome c oxidase subunit 2	<i>MT-CO2</i>	C	P00403	25.57	4.67	4	13.2	2	0.28	0.43	0.33	0.33
63	Ragulator complex protein LAMTOR1	<i>LAMTOR1</i>	PM	Q61AA8	17.61	5.01	3.89	18.0	2	0.58	0.18	0.45	0.35
64	Polyribonucleotide nucleotidyltransferase 1, mitochondrial	<i>PNPT1</i>	M	Q8TCS8	80.87	6.26	3.65	2.3	2	0.77	0.47	0.71	0.80
65	SWI/SNF complex subunit SMARCC2	<i>SMARCC2</i>	N	Q8TAQ2	132.88	5.49	3.48	1.9	2	0.45	0.34	0.32	0.44
66	Protein S100-A2	<i>S100A2</i>	N	P29034	11.12	4.68	3.45	33.7	4	0.05	0.04	0.04	0.04
67	Glycerol-3-phosphate dehydrogenase, mitochondrial	<i>GPD2</i>	M	P43304	76.36	6.30	3.43	4.1	2	0.79	0.46	0.78	0.52
68	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	<i>NDUFA10</i>	M	O95299	37.15	6.87	3.42	5.6	2	0.49	0.57	0.76	0.40
69	39S ribosomal protein L38, mitochondrial	<i>MRPL38</i>	M	Q96DV4	41.63	6.50	3.4	4.7	2	0.47	0.43	0.74	0.50
70	ATP synthase subunit g, mitochondrial	<i>ATP5L</i>	M	O75964	11.30	9.65	3.33	23.3	2	0.55	0.38	0.72	0.64
71	ATP synthase subunit e, mitochondrial	<i>ATP5I</i>	M	P56385	7.80	9.35	3.08	30.4	2	0.32	0.22	0.36	0.39
72	ATP synthase subunit d, mitochondrial	<i>ATP5H</i>	M	O75947	18.36	5.22	3.01	14.3	2	0.43	0.38	0.49	0.49
73	Cytochrome c oxidase subunit 5A, mitochondrial	<i>COX5A</i>	M	P20674	12.50	4.88	2.99	10.0	2	0.64	0.50	0.58	0.51
74	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	<i>NDUFA9</i>	M	Q16795	38.90	9.68	2.92	8.2	2	0.36	0.50	0.68	0.48
75	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	<i>NDUFS2</i>	M	O75306	49.17	6.26	2.92	6.5	3	0.42	0.55	0.54	0.38
76	Myosin regulatory light chain 12B	<i>MYL12B</i>	C	O14950	19.78	4.69	2.91	9.9	2	0.68	0.80	0.71	0.61
77	Alpha-2-macroglobulin receptor-associated protein	<i>LRPAP1</i>	PM	P30533	37.77	6.88	2.84	4.8	2	0.69	0.65	0.49	0.72
78	Mitochondrial import inner membrane translocase subunit TIM50	<i>TIMM50</i>	M	Q3ZCQ8	35.08	6.01	2.49	5.1	2	0.58	0.35	0.49	0.56
79	Annexin A8-like protein 2	<i>ANXA8L2</i>	PM	Q5VT79	36.86	5.45	2.39	2.1	1	0.08	0.06	0.09	0.25

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

80	Dipeptidyl peptidase 1	<i>CTSC</i>	C	P53634	18.47	5.94	2.26	2.4	1	0.09	0.11	0.24	0.22
81	Neurofilament light polypeptide	<i>NEFL</i>	C	P07196	61.39	4.63	2.17	3.3	2	0.09	0.06	0.09	0.06
82	Glutamine synthetase	<i>GLUL</i>	C	P15104	41.93	6.42	2.14	2.1	1	0.19	0.04	0.54	0.03
83	ATP synthase subunit b, mitochondrial	<i>ATP5F1</i>	M	P24539	24.63	9.10	2.13	5.9	1	0.46	0.41	0.51	0.51
84	Serum deprivation-response protein	<i>SDPR</i>	PM	O95810	47.04	5.14	2.12	3.3	1	0.08	0.34	0.26	0.10
85	28S ribosomal protein S7, mitochondrial	<i>MRPS7</i>	M	Q9Y2R9	24.12	9.72	2.12	5.4	1	0.51	0.48	0.68	0.50
86	Mitochondrial-processing peptidase subunit alpha	<i>PMPCA</i>	M	Q10713	54.65	5.88	2.09	3.2	1	0.67	0.15	0.40	0.37
87	OCLIA domain-containing protein 1	<i>OCLAD1</i>	C	Q9NX40	27.63	7.02	2.08	7.3	1	0.57	0.52	0.32	0.43
88	Prenylcysteine oxidase 1	<i>PCYOX1</i>	C	Q9UHG3	53.85	5.89	2.08	3.6	1	0.66	0.78	0.60	0.74
89	Acetyl-CoA acetyltransferase, mitochondrial	<i>ACAT1</i>	M	P24752	41.39	8.16	2.07	4.0	1	0.51	0.50	0.54	0.70
90	Sideroflexin-1	<i>SFXN1</i>	C	Q9H9B4	35.49	9.22	2.07	2.8	1	0.74	0.77	0.78	0.72
91	Serpin B5	<i>SERPINB5</i>	ES	P36952	42.10	5.72	2.06	3.2	1	0.13	0.18	0.17	0.22
92	Condensin complex subunit 1	<i>NCAPD2</i>	N	Q15021	157.18	6.19	2.04	0.7	1	0.60	0.54	0.50	0.36
93	Protein transport protein Sec31A	<i>SEC31A</i>	C	O94979	133.01	6.43	2.04	1.5	1	0.38	0.27	0.35	0.33
94	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	<i>GNB1</i>	PM	P62873	37.25	5.60	2.04	2.9	1	0.33	0.32	0.33	0.24
95	Mitochondrial import inner membrane translocase subunit TIM44	<i>TIMM44</i>	M	O43615	51.35	8.51	2.02	3.3	1	0.49	0.65	0.25	0.55
96	Glutaminase kidney isoform, mitochondrial	<i>GLS</i>	C or M	O94925	71.56	7.01	2.02	2.2	2	0.62	0.57	0.51	0.55
97	Protein QIL1	<i>C19orf70</i>	C	Q5XKP0	13.09	9.44	2.02	16.9	1	0.28	0.49	0.38	0.37
98	Tubulin beta-2A chain	<i>TUBB2A</i>	C	Q13885	49.91	4.78	2	33.9	27	0.79	0.72	0.60	0.45
99	Ras-related protein Rab-10	<i>RAB10</i>	C	P61026	22.54	8.58	2	11.0	2	0.62	0.74	0.63	0.52
100	Protein AHNAK2	<i>AHNAK2</i>	Other	Q81VF2	616.63	5.20	2	1.1	1	0.12	0.28	0.14	0.16
101	Microsomal glutathione S-transferase 1	<i>MGST1</i>	C	P10620	17.47	9.41	2	7.7	1	0.15	0.18	0.44	0.12
102	Quinone oxidoreductase PIG3	<i>TP53I3</i>	Other	Q53FA7	35.54	6.66	2	3.6	1	0.27	0.46	0.33	0.21
103	DNA-directed RNA polymerases I, II, and III subunit RPABC5	<i>POLR2L</i>	N	P62875	7.65	7.65	2	16.4	1	0.80	0.79	0.57	0.75
104	Electron transfer flavoprotein subunit beta	<i>ETFB</i>	C	P38117	27.71	8.29	2	4.7	1	0.52	0.62	0.37	0.42
105	Insulin-like growth factor 2 mRNA-binding protein 2	<i>IGF2BP2</i>	C	Q9Y6M1	66.12	8.48	2	2.3	1	0.27	0.76	0.22	0.53
106	Epidermal growth factor receptor	<i>EGFR</i>	PM	P00533	132.01	6.17	2	0.9	1	0.10	0.38	0.28	0.40
107	ATP synthase subunit f, mitochondrial	<i>ATP5J2</i>	M	P56134	10.79	9.70	2	11.7	1	0.54	0.33	0.73	0.55
108	Thioredoxin-related transmembrane protein 4	<i>TMX4</i>	C	Q9H1E5	36.76	4.29	2	3.7	1	0.72	0.35	0.54	0.41
109	ATP synthase subunit epsilon-like protein, mitochondrial	<i>ATP5EP2</i>	M	Q5VTU8	5.81	10.15	2	15.7	1	0.47	0.40	0.49	0.45
110	Acyl-coenzyme A thioesterase 13	<i>ACOT13</i>	C	Q9NPJ3	14.96	9.23	2	9.3	1	0.32	0.50	0.69	0.69
111	Ubiquitin fusion degradation protein 1 homolog	<i>UFDIL</i>	C	Q92890	34.50	6.27	2	3.3	1	0.65	0.77	0.63	0.63
112	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	<i>NDUFB9</i>	M	Q9Y6M9	21.70	8.59	2	8.9	1	0.37	0.19	0.61	0.35
113	Diablo homolog, mitochondrial	<i>DIABLO</i>	M	Q9NR28	20.77	4.78	2	4.6	1	0.49	0.56	0.56	0.44
114	Serum paraoxonase/arylesterase 2	<i>PON2</i>	PM	Q15165	39.40	5.33	2	2.5	1	0.24	0.21	0.63	0.34
115	Small nuclear ribonucleoprotein Sm D1	<i>SNRPD1</i>	N	P62314	13.28	11.56	2	10.9	1	0.42	0.39	0.34	0.38
116	Translocator protein	<i>TSPO</i>	C	P30536	18.83	9.43	2	4.7	1	0.27	0.29	0.48	0.35
117	Monocarboxylate transporter 4	<i>SLC16A3</i>	PM	O15427	49.34	8.25	2	1.5	1	0.09	0.47	0.44	0.79
118	Cytochrome c oxidase subunit 7C, mitochondrial	<i>COX7C</i>	M	P15954	5.36	8.98	2	14.3	1	0.40	0.66	0.68	0.51
119	Nucleoside diphosphate-linked moiety X motif 19, mitochondrial	<i>NUDT19</i>	M	A8MXV4	40.58	6.48	2	3.7	1	0.02	0.62	0.56	0.13
120	Metaxin-1	<i>MTX1</i>	C	Q13505	51.48	9.80	1.96	3.0	1	0.24	0.23	0.50	0.49
121	39S ribosomal protein L15, mitochondrial	<i>MRPL15</i>	M	Q9P015	31.32	9.93	1.92	3.0	1	0.59	0.65	0.61	0.67

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

122	Nicotinamide N-methyltransferase	<i>NNMT</i>	C	P40261	29.57	5.56	1.8	7.2	1	0.11	0.36	0.09	0.77
123	Interleukin-18	<i>IL18</i>	ES	Q14116	18.22	5.01	1.72	4.7	1	0.10	0.69	0.23	0.30
124	Dimethyladenosine transferase 2, mitochondrial	<i>TFB2M</i>	M	Q9H5Q4	43.26	9.11	1.69	2.3	1	0.36	0.32	0.47	0.36
125	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	<i>NDUFV1</i>	M	P49821	48.51	7.53	1.64	1.9	1	0.55	0.44	0.68	0.41
126	Tripartite motif-containing protein 29	<i>TRIM29</i>	C	Q14134	65.83	6.73	1.57	2.2	1	0.18	0.11	0.09	0.10
127	Protein ETHE1, mitochondrial	<i>ETHE1</i>	M	O95571	27.10	6.37	1.54	3.9	1	0.45	0.23	0.55	0.32
128	Junctional adhesion molecule A	<i>F11R</i>	PM	Q9Y624	29.68	6.72	1.48	3.7	1	0.56	0.24	0.77	0.48

^a UniProtKB ID, numero de acceso en la base de datos de UniProt; ^b MW, peso molecular; ^c pI, punto isoeléctrico; ^d Unused, Puntaje de proteínas, el cual es calculado utilizando solo los péptidos de espectros que no han sido usados en otras proteínas; ^e % Cov. (≥ 95), Porcentaje de cobertura; ^f Pep. Ident. (>95), Numero de péptidos con al menos 95% de confianza en su identificación; ^g Expresión diferencial entre las diferentes líneas celulares de cáncer de mama y la línea control (113): 114:113, MCF7/MCF 10A; ^h 115:113, MDA-MB-231/MCF 10A; ⁱ 116:113, SK-BR-3/MCF 10A; ^j 117:113, T47D/MCF 10A. C, Citoplasma. N, Núcleo. M, Mitocondria. PM, Membrana plasmática. Localización de proteínas obtenida con la base de datos de IPA.

3. Clasificación funcional de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.

Para obtener una visión global de las proteínas identificadas, se clasificaron con base en su ontología génica con el programa PANTHER. Las 78 proteínas sobreexpresadas se clasificaron de acuerdo con la clase de proteína a la que pertenecen, obteniendo 21 categorías con 89 hits (Figura 24a). La clase con mayor número de genes involucrados fue la de unión a ácidos nucleicos (XII, con 25 polipéptidos). Las proteínas también se clasificaron de acuerdo con los procesos biológicos en los que están involucradas, obteniendo 11 categorías con 132 hits, siendo los procesos metabólicos (VII con 55 polipéptidos) la categoría más relevante (Figura 24b). Las 128 proteínas subexpresadas, estas se clasificaron de igual manera en clases de proteínas, obteniendo 24 categorías con 148 hits. Las clases de proteínas más relevantes fueron: Oxidorreductasa (XV con 21 polipéptidos), Hidrolasa (VIII con 16) y unión a ácidos nucleicos (XIV con 16 proteínas) (Figura 24c). En cuanto a la clasificación de procesos biológicos, se obtuvieron 12 clases con 178 hits, siendo los procesos metabólicos (IX con 78 polipéptidos) la clase más representativa (Figura 24d).

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

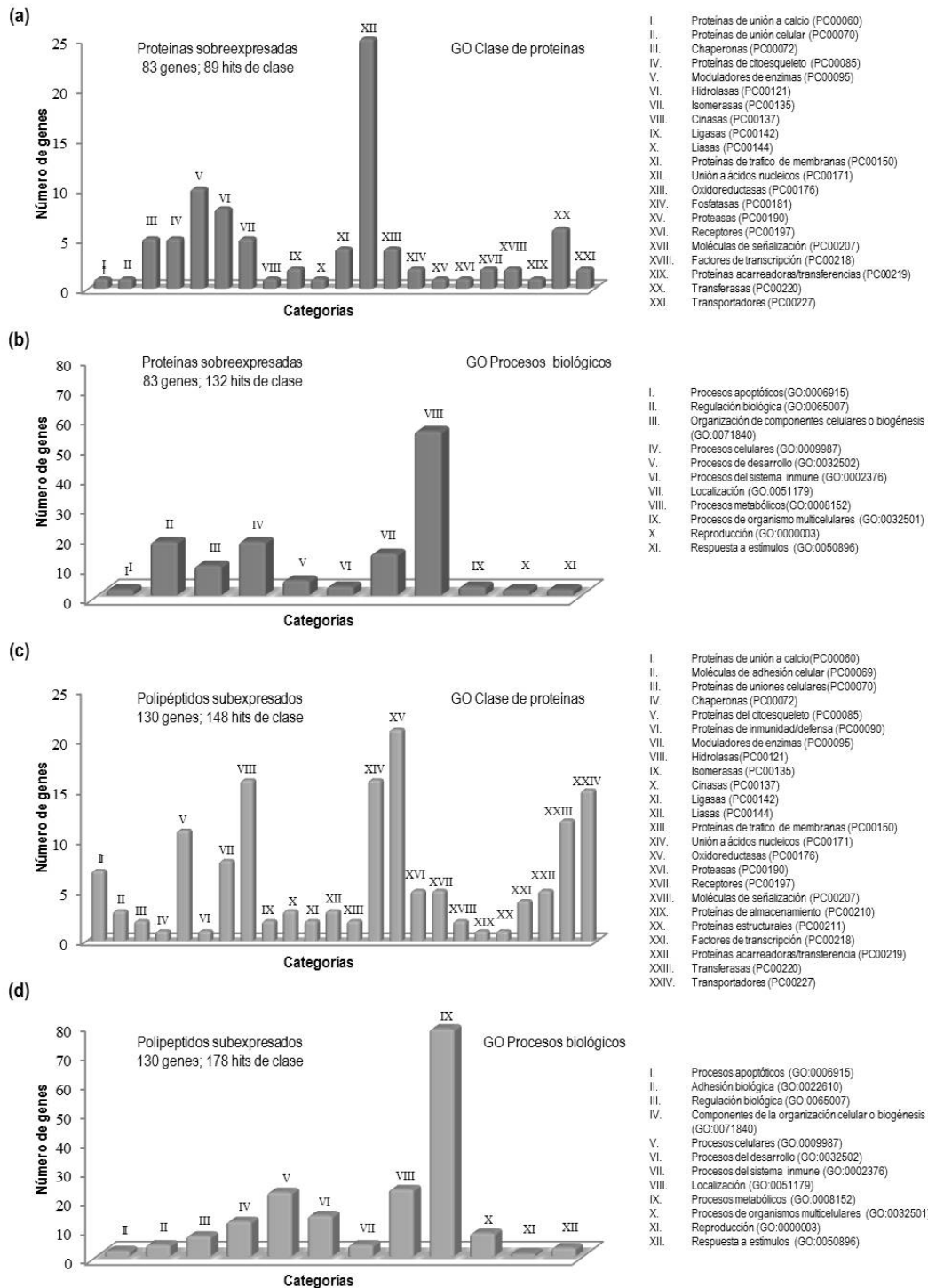


Figura 24. Clasificación funcional con PANTHER de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama. Las proteínas se clasificaron de acuerdo a la clase de proteína a la que pertenecen (a,c), así como en los procesos biológicos en los que participan (b,d). Las barras representan las diversas categorías en las que se clasificaron las proteínas (eje x) y el número de genes involucrados en cada una de estas categorías (eje Y). (a,b) Proteínas sobreexpresadas. (c,d) Polipéptidos subexpresados.

4. Redes de interacción proteína-proteína de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama

Para determinar las interacciones proteína-proteína entre las proteínas diferencialmente expresadas en las líneas celulares de cáncer de mama, se realizó un análisis con la plataforma STRING. En las redes de interacción de las proteínas sobreexpresadas, se identificaron nodos de interacción relacionados con: 1) Topología del ADN, 2) Vía de las pentosas, 3) Regulación del inicio de la traducción, 4) Proteínas involucradas en degradación por el proteasoma, 5) Corte y empalme del RNA, y 6) Glicólisis (Figura 25). La red de interacción de las proteínas subexpresadas está caracterizada por nodos de proteínas relacionados con procesos mitocondriales: 1) Respiración celular, 2) Transporte mitocondrial, 3) Fosforilación oxidativa (Figura 26). También se muestran aquellas proteínas que no formaron nodos de interacción, pero que están involucradas en diferentes procesos biológicos, los cuales tuvieron una significancia estadística ($p\text{-value} \leq 0.05$).

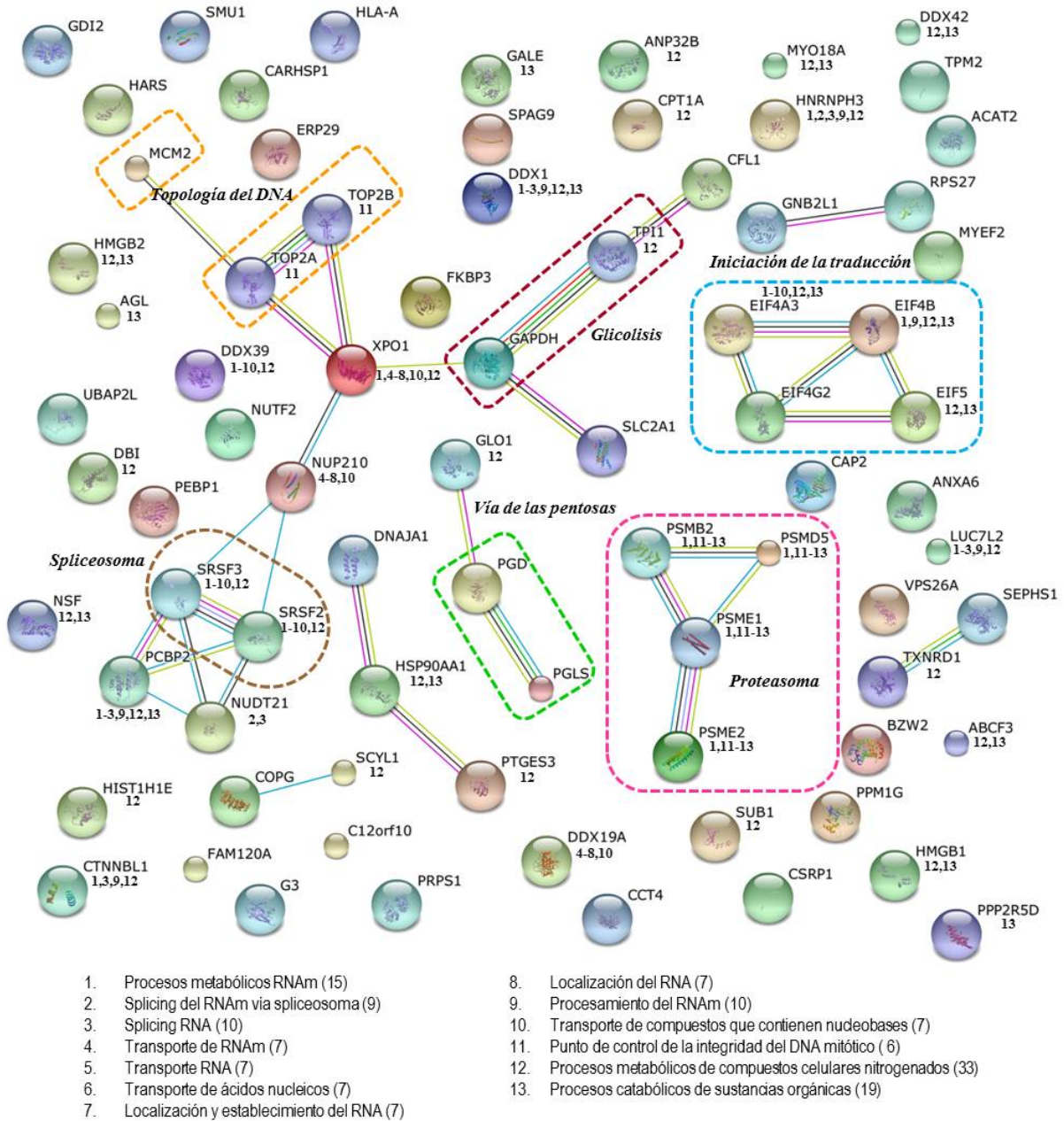


Figura 25. Redes de interacción proteína-proteína de los polipéptidos sobreexpresados en común en todas las líneas celulares de cáncer de mama. Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los polipéptidos implicados en el mismo proceso biológico se agruparon como se muestra. Topología del ADN: TOP2A, TOP2B, y MCM2. Glucólisis: GAPDH y TPI1. Regulación del inicio de la traducción: EIF5, EIF4B, EIF4G2 y EIF4A3. Espliceosoma: SRSF2 y SRSF3. Vía de las pentosas: PGD y PGLS. Degradación por el proteasoma: PSMD5, PSMB2, PSME1 y PSME2. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$).

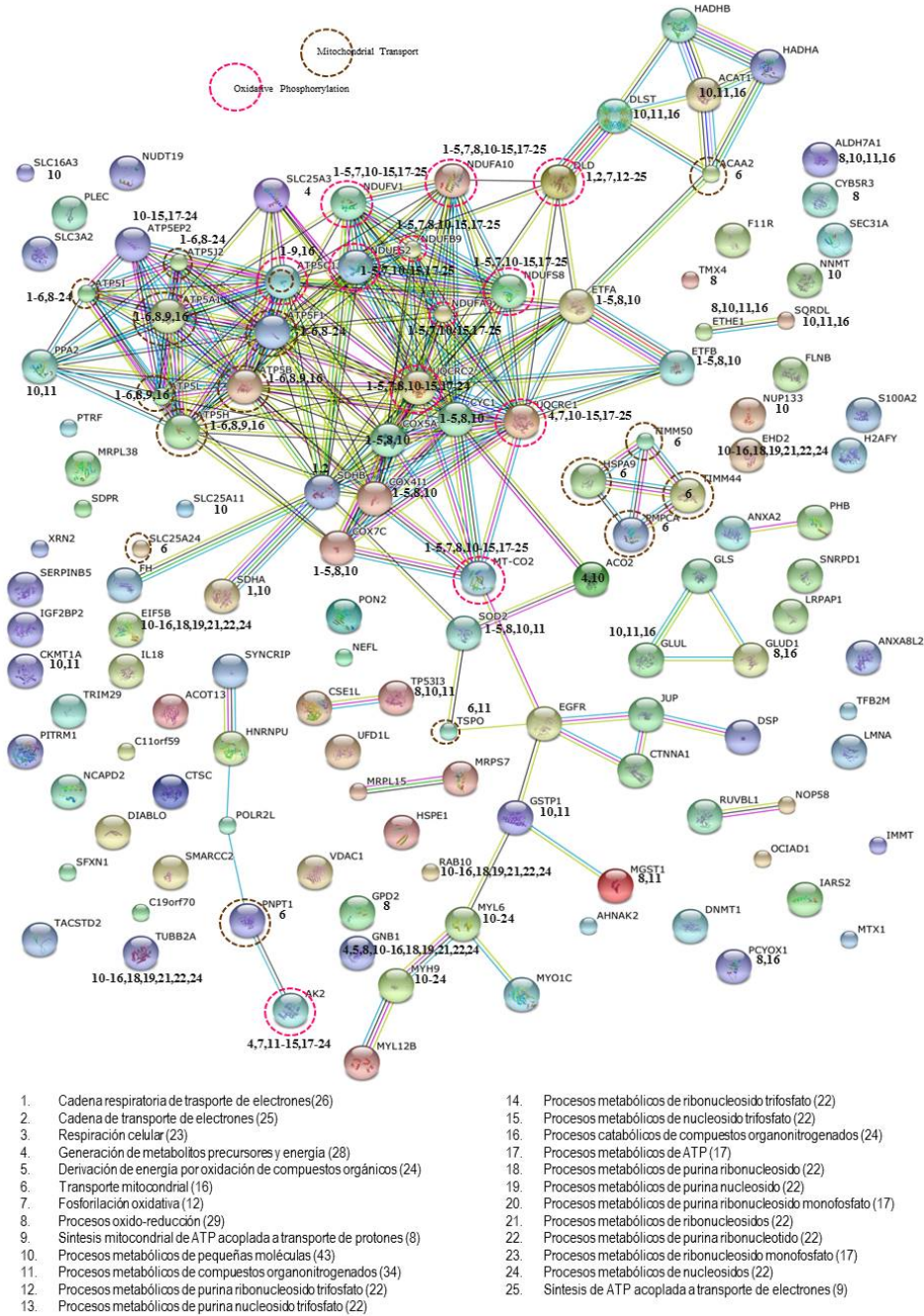


Figura 26. Redes de interacción proteína-proteína de los polipéptidos subexpresados en común en todas las líneas celulares de cáncer de mama. Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$). Las proteínas involucradas en el proceso de fosforilación oxidativa se representan con círculos rosas, mientras que los polipéptidos implicados en el transporte mitocondrial se representan con círculos de color café.

5. Localización de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama en vías metabólicas.

Las proteínas sobre y subexpresadas se localizaron en las vías bioquímicas en las que están involucradas, empleando la base de datos de KEGG. En este análisis, los procesos más afectados fueron las vías metabólicas (Figura 27). Algunas de las vías representadas son: Metabolismo de aminoácidos (naranja), metabolismo energético (morado), metabolismo de hidratos de carbono (azul oscuro), metabolismo de cofactores y vitaminas (rosa), metabolismo de nucleótidos (rojo claro), metabolismo de lípidos (verde menta), biosíntesis y metabolismo de glicanos (azul claro), entre otros. Las proteínas sobreexpresadas se muestran en color rojo, mientras que la subexpresadas en verde. Se encontraron 51 proteínas sobreexpresadas alteradas en las vías metabólicas: AGL (metabolismo de almidón y sacarosa), GALE (metabolismo de galactosa y metabolismo de aminoácidos y nucleótidos de azúcar), PGLS (metabolismo del carbono y vía de la pentosa fosfato), DGP (metabolismo del carbono, vía de la pentosa fosfato y metabolismo del glutatión), PRPS1 (metabolismo del carbono, biosíntesis de aminoácidos, vía de la pentosa fosfato y metabolismo de purina), TPI1 (metabolismo del carbono, biosíntesis de aminoácidos, glucólisis/gluconeogénesis, metabolismo de fructosa y manosa y metabolismo de inositol fosfato), GAPDH (metabolismo del carbono, biosíntesis de aminoácidos, glucólisis/gluconeogénesis), PTGES3 (metabolismo del ácido araquidónico) y SEPHS1 (metabolismo de selenocompuestos).

Por otro lado, las proteínas subexpresadas alteradas en vías metabólicas fueron: ALDH7A1 (biosíntesis de aminoácidos, glucólisis/gluconeogénesis, metabolismo de ascorbato y aldarato, metabolismo de piruvato, degradación de ácidos grasos, metabolismo de glicerolípidos, metabolismo de glicina, serina, treonina, arginina, prolina, histidina, triptófano y beta-alanina, degradación de valina, leucina, isoleucina y lisina, así como biosíntesis de lisina), Ak2 (metabolismo de purinas), POLR2L (metabolismo de purina y pirimidina), CKMT1A (metabolismo de arginina y prolina), DNMT1 (metabolismo de cisteína y metionina), GLS (metabolismo

de alanina, aspartato, glutamato, metabolismo de arginina y prolina, metabolismo de D-glutamina y D- glutamato), NNMT (metabolismo de nicotinato y nicotinamida), HADHA (metabolismo del carbono, metabolismo de ácidos grasos, metabolismo de propanoato y butanoato, elongación y la degradación de ácidos grasos, biosíntesis de ácidos grasos insaturados, degradación de valina, leucina, isoleucina y lisina, metabolismo de triptófano y beta-alanina), ACAA2 y HADHB (metabolismo de ácidos grasos, la elongación y la degradación de ácidos grasos, degradación de valina, leucina e isoleucina), GLUL (biosíntesis de aminoácidos, metabolismo de glioxilato y dicarboxilato, metabolismo del nitrógeno, metabolismo de alanina, aspartato, glutamato, arginina y prolina), GLUD1 (metabolismo del nitrógeno, metabolismo de alanina, aspartato, glutamato, arginina, prolina, D-glutamina y D-glutamato), ACO2 (metabolismo del carbono, metabolismo del ácido 2-oxocarboxílico, biosíntesis de aminoácidos, ciclo del citrato, metabolismo de glioxilato y dicarboxilato), DLD (metabolismo del carbono, glucólisis / gluconeogénesis, ciclo del citrato, metabolismo de piruvato, metabolismo de glicina, serina y treonina, degradación de valina, leucina e isoleucina), DLST (metabolismo del carbono, ciclo del citrato y degradación de lisina), FH (metabolismo del carbono, ciclo del citrato y metabolismo de piruvato), SDHA y SDHB (metabolismo del carbono, ciclo del citrato, fosforilación oxidativa), ACAT1 y ACAT2 (metabolismo del carbono, metabolismo de ácidos grasos, metabolismo de piruvato, metabolismo de glioxilato y dicarboxilato, metabolismo de propanoato y butanoato, degradación de ácidos grasos, síntesis y degradación de cuerpos cetónicos, degradación de valina, leucina, isoleucina y lisina, metabolismo de triptófano, biosíntesis de columna vertebral terpenoide) y PON2 (respuesta a estrés oxidativo, respuesta a sustancias tóxicas, procesos catabólicos de compuestos aromáticos).

Sorprendentemente, se encontraron 24 polipéptidos subexpresados implicados en el proceso de fosforilación oxidativa (Figura 27). La localización específica de estas proteínas en cada complejo afectado se representa en la figura 28. Los polipéptidos NDUFA9, NDUFA10, NDUFB9, NDUFS2, NDUFV1 y NDUFS8

se identificaron como elementos del complejo I (NADH deshidrogenasa). En el caso del complejo II (succinato deshidrogenasa) se encontró a las proteínas SDHA y SDHB. UQCRC1 (COR1), UQCRC2 (QCR2) y CYC1 (Cyt1) se localizaron en el complejo III (ubiquinol-citocromo c reductasa). Como parte del complejo IV (citocromo c oxidasa) se ubicaron las proteínas COX2, COX4I1 (COX4), COX5A y COX7C. Finalmente, se localizaron 9 proteínas pertenecientes al complejo V (ATP sintasa): ATP5H (subunidad d), ATP5L (subunidad g), ATP5C1 (subunidad gamma), ATP5I (subunidad e), ATP5J2 (subunidad f2), ATP5A1 (subunidad alfa), ATP5F1 (subunidad b1), ATP5B (subunidad beta) y PPA2 (pirofosfatasa).

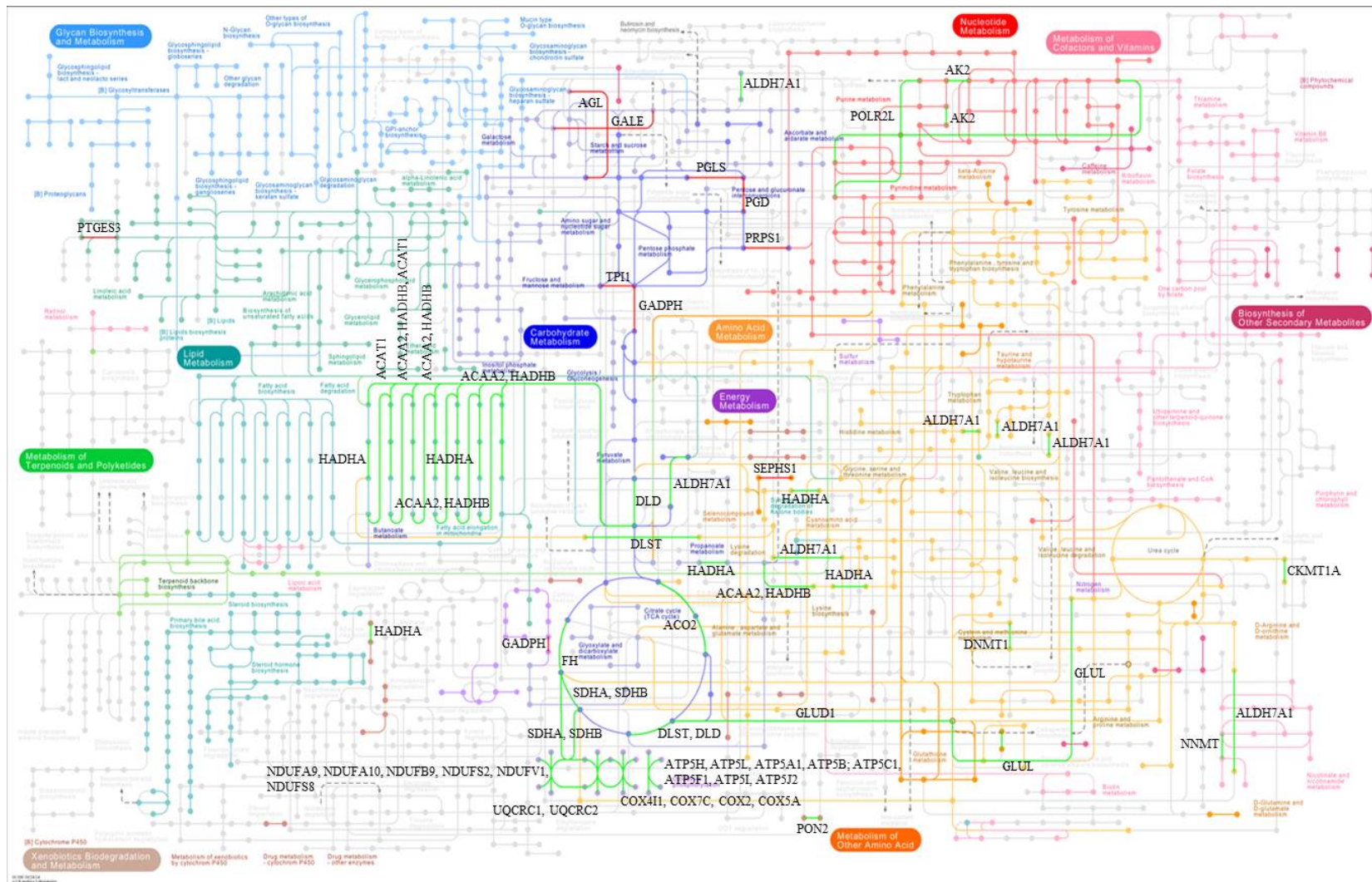


Figura 27. Visión global de las vías metabólicas afectadas en todas las líneas celulares de cáncer de mama. Los 206 polipéptidos que mostraron un cambio en su nivel de expresión en todas las líneas celulares de cáncer de mama fueron representados en las vías de la plataforma KEGG. Las proteínas sobreexpresadas se muestran en color rojo y los polipéptidos subexpresados en color verde.

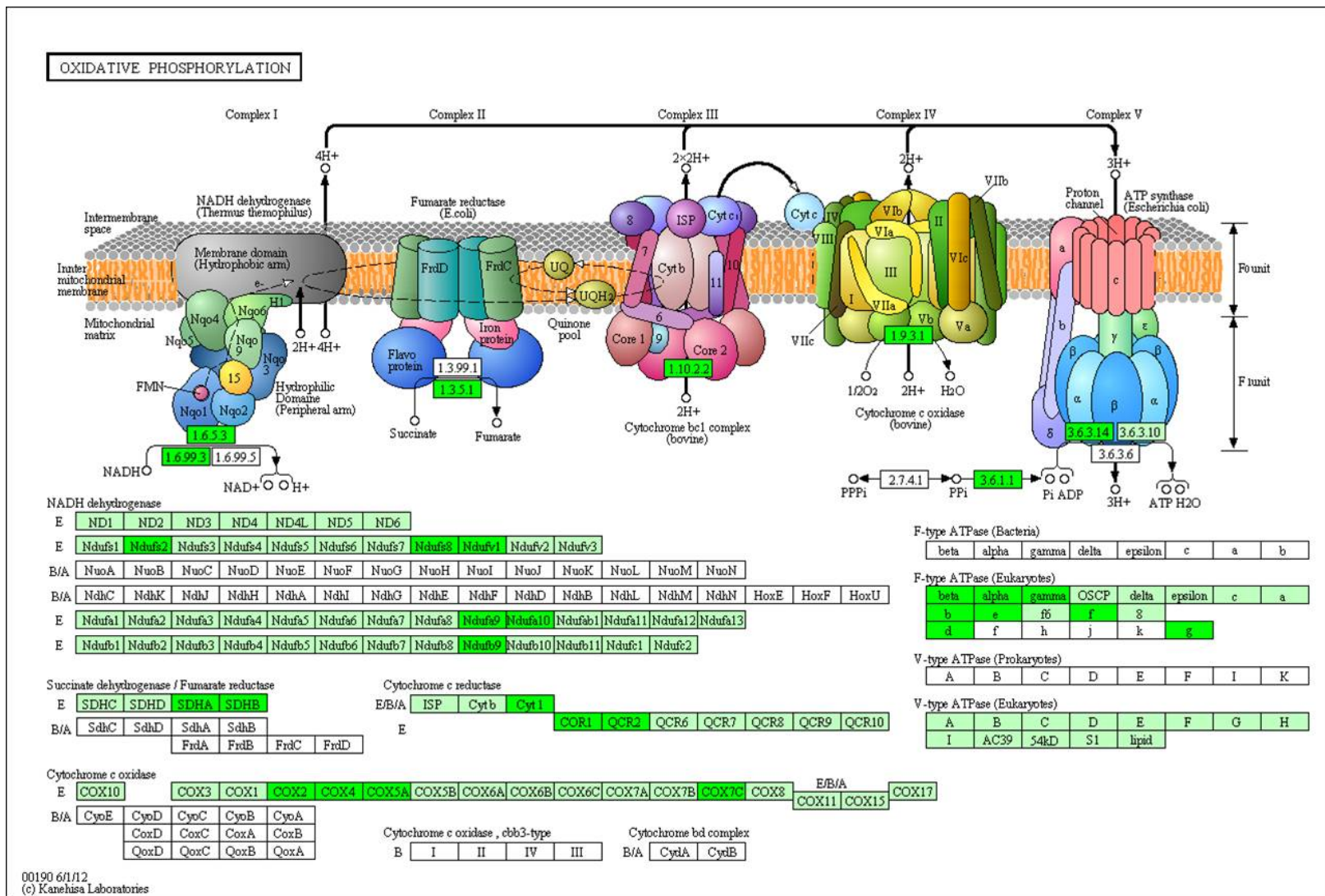


Figura 28. Localización de los polipéptidos subexpresados en la cadena respiratoria obtenida con la base de datos KEGG. Las proteínas subexpresadas (verde brillante) en todas las líneas celulares de cáncer de mama involucradas en el proceso de fosforilación oxidativa son representadas en cada complejo de la cadena respiratoria.

6. IPA análisis de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama.

Para determinar si las proteínas sobre o subexpresadas están implicadas en enfermedades y biofunciones, o localizadas en redes, se utilizó el programa IPA. Se realizaron tres análisis: a) Core I, en el cual se consideraron todos los tejidos, células primarias y todas las líneas celulares. b) Core II, para glándula mamaria y líneas celulares de cáncer de mama. c) Core III para todas las líneas celulares de cáncer, con excepción de las líneas celulares de cáncer de mama. Los resultados obtenidos se muestran en la Tabla 5.

Los análisis Core I y III compartieron similitudes en los resultados obtenidos. Core I incluyó 78 categorías de enfermedades y biofunciones, mientras que Core II 11 categorías y Core III 72. Las categorías más representativas en Core I fueron: metabolismo de ácidos nucleicos (16 proteínas y un valor de p entre 4.46×10^{-10} y 2.53×10^{-2}), bioquímica de pequeñas moléculas (36 polipéptidos y un valor de p que oscila entre 4.46×10^{-10} a 3.77×10^{-2}), replicación, recombinación y reparación del ADN (19 proteínas y un valor de p entre 7.24×10^{-10} y 3.77×10^{-2}) y producción de energía (10 polipéptidos y un valor de p de 7.24×10^{-10}). En Core II, las categorías fueron: respuesta celular a terapéuticos (2 proteínas con un valor de p de 2×10^{-3}), movimiento celular (9 polipéptidos y un valor de p de 1.72×10^{-2} a 3.67×10^{-2}), morfología celular, ensamble y organización celular (2 polipéptidos y un valor de p de 1.85×10^{-2}). Finalmente, en Core III las categorías más representativas fueron las mismas que en Core I: metabolismo de ácidos nucleicos con 16 proteínas y un valor de p de 4.6×10^{-10} a 4.02×10^{-2} ; bioquímica de pequeñas moléculas (34 proteínas, con un valor de p de 4.6×10^{-10} y 4.02×10^{-2}); replicación, recombinación y reparación del ADN (21 proteínas con un valor de p de 7.46×10^{-10} a 3.78×10^{-2}), producción de energía (12 proteínas con un valor de p de 7.46×10^{-10} a 4.02×10^{-2}). De estos análisis Core, se generaron varias redes entre las proteínas sobre y subexpresadas. En Core I, las principales redes estuvieron relacionadas con: 1)

Cáncer, enfermedades del sistema reproductivo, enfermedades y afecciones dermatológicas y enfermedades hematológicas con un score de 38 y con 25 moléculas involucradas en este proceso. 2) Trastornos del tejido conectivo, enfermedades neurológicas, trastornos óseos y musculares, replicación, recombinación y reparación del ADN, cáncer, enfermedades gastrointestinales con 27 de marca y 20 moléculas involucradas (Figura 29). En el análisis Core II las redes más representativas participaron en: 1) el desarrollo celular, crecimiento y proliferación celular, ciclo celular con una puntuación de 15 y 15 moléculas involucradas en estos procesos. 2) Muerte celular y supervivencia, morfología del tumor y desarrollo celular con un score de 9 y 11 moléculas implicadas (Figura 30). Por último, en Core III, las proteínas identificadas participaron en: 1) Ciclo celular, cáncer, enfermedades hereditarias, enfermedades hematológicas, con 41 de puntaje y 26 moléculas involucradas. 2) Trastornos del desarrollo, trastornos hereditarios, enfermedades metabólicas, muerte celular y supervivencia con marcación de 24 y 18 moléculas de implicadas (Figura 31).

Tabla 5. Clasificación de las proteínas sobre y subexpresadas con el programa IPA.

Core I				Core II				Core III			
Categoría	p-value	N	Moléculas	Categoría	p-value	N	Moléculas	Categoría	p-value	N	Moléculas
Metabolismo de ácidos nucleicos	4.46 x 10 ⁻¹⁰ to 2.53 x 10 ⁻²	16	ATP5A1, PGLS ,ATP5L,TP53I3,ATP5H,ATP5C1, PGD , PRPS1 ,ATP5B, GALE ,ATP5J2,MYH9, HSP90AA1 ,VDAC1,ATP5F1,ATP5I	Respuesta celular a terapéuticos	2 x 10 ⁻³	2	SOD2,PNPT1	Metabolismo de ácidos nucleicos	4.6 x 10 ⁻¹⁰ to 4.02 x 10 ⁻²	16	ATP5A1, PGLS ,ATP5L,TP53I3,ATP5H,ATP5C1, PGD , PRPS1 ,ATP5B, GALE ,ATP5J2,MYH9, HSP90AA1 ,VDAC1,ATP5F1,ATP5I
Bioquímica de pequeñas moléculas	4.46 x 10 ⁻¹⁰ to 3.77 x 10 ⁻²	36	PEBP1 ,SDHB,ACAT2,ACO2,TIMM50,ATP5L, DNAJA1 , PTGES3 ,ATP5H, PGD ,SOD2, PRPS1 ,ATP5J2,ATP5F1,ATP5I,EGFR,SDHA, SLC2A1 ,GLS,ATP5A1, PGLS ,GLUD1,ANXA2, TXNRD1 ,TP53I3,SLC3A2,ATP5C1,IL18,ATP5B, GALE ,ACAT1,MYH9, HSP90AA1 ,FH,VDAC1,GSTP1	Movimiento celular	1.72 x 10 ⁻² to 3.67 x 10 ⁻²	9	CSE1L, GNB2L1 ,MYH9,RUVBL1,ANXA2,SERPINB5,JUP,LRPAP1,EGFR	Bioquímica de pequeñas moléculas	4.6 x 10 ⁻¹⁰ to 4.02 x 10 ⁻²	34	PEBP1 ,SDHB,ACAT2,ACO2,TIMM50,ATP5L, DNAJA1 , PTGES3 ,ATP5H, PGD , PRPS1 ,SOD2,ATP5J2,ATP5F1,ATP5I,SDHA, SLC2A1 ,GLS,ATP5A1, PGLS ,GLUD1,ANXA2, TXNRD1 ,TP53I3,ATP5C1,IL18,ATP5B,IMMT, GALE ,ACAT1,MYH9, HSP90AA1 ,VDAC1,GSTP1
Replicación, recombinación y reparación del DNA	7.24 x 10 ⁻¹⁰ to 3.77 x 10 ⁻²	19	HNRNPU,ATP5A1,XRN2,ATP5L,NCAPD2,ATP5H, TOP2B ,ATP5C1, HMGBI ,ATP5B, TOP2A ,ATP5J2, HSP90AA1 ,MYH9,ATP5F1,ATP5I,DNMT1,EGFR, HMG2	Morfología celular	1.85 x 10 ⁻²	2	CSE1L,JUP	Replicación, recombinación y reparación del DNA	7.46 x 10 ⁻¹⁰ to 3.78 x 10 ⁻²	21	HNRNPU,ATP5A1,XRN2,ATP5L,NCAPD2,ATP5H, TOP2B ,ATP5C1,SD2, HMGBI ,ATP5B, TOP2A ,ATP5J2, HSP90AA1 ,MYH9, DDXI ,ATP5F1,ATP5I,DNMT1,EGFR, HMG2
Producción de energía	7.24 x 10 ⁻¹⁰	10	ATP5C1,ATP5B,ATP5A1,ATP5L,ATP5J2, HSP90AA1 ,MYH9,ATP5F1,ATP5L,ATP5H	Ensamble y organización celular	1.85 x 10 ⁻²	2	CSE1L,JUP	Producción de energía	7.46 x 10 ⁻¹⁰ to 4.02 x 10 ⁻²	12	ATP5C1,ATP5B,IMMT,ATP5A1,ATP5L,ATP5J2, HSP90AA1 ,MYH9,VDAC1,ATP5F1,ATP5L,ATP5H
Enfermedades neurológicas	5.21 x 10 ⁻⁷ to 2.72 x 10 ⁻²	55	SRSF2 ,NDUFA9,ACO2,ATP5L,ETHE1, HARS ,DNMT, PRPS1 ,SOD2,CSE1L, SUBI ,NDUFA10,DNMT1,HADHA,ALDH7A1, SLC2A1 ,TUBB2A,LMNA,ANXA2,COX7C,ATP5C1,UQCRC2, HSP90AA1 ,DIABLO,AK2,GSTP1, HMG2 , PEBP1 ,SDHB,NEFL, BAG6 , GAPDH , PTGES3 , TOP2B ,HADHB, TOP2A ,NDUFS2,EGFR,SDHA,ETFB,HNRNPU, DBI , DDX42 ,GLUD1,F11R,NDUFS8, TPH1 , SEPHS1 ,CYC1,PNPT1,ACAT1,COX5A,MYH9,TSP0,UQCRC1	Mantenimiento y función celular	1.85 x 10 ⁻²	2	CSE1L,JUP	Enfermedades neurológicas	2.15 x 10 ⁻⁷ to 2.54 x 10 ⁻²	45	SRSF2 ,SDHB,NDUFA9,NEFL, BAG6 ,ACO2,ETHE1,ATP5L, GAPDH , HARS , TOP2B ,HADHB, PRPS1 , SUBI , TOP2A ,NDUFA10,NDUFS2,HADHA,DNMT1,TACSTD2,ALDH7A1,EGFR,SDHA,ETFB,HNRNPU, SLC2A1 ,TUBB2A,LMNA,GLUD1,COX7C,F11R,ATP5C1,NDUFS8, TPH1 ,CYC1, SEPHS1 ,PNPT1,ACAT1, TSP0 , HSP90AA1 ,MYH9,DIABLO,UQCRC1,GSTP1, HMG2
Desordenes psicológicos	5.21 x 10 ⁻⁷ to 3.66 x 10 ⁻²	39	PEBP1 ,SDHB,NEFL, GAPDH ,ATP5L, PTGES3 ,HADHB,SOD2,CSE1L, SUBI ,NDUFS2,DNMT1,SDHA,ETFB,HNRNPU, SLC2A1 , DBI , DDX42 ,TUBB2A,ANXA2,COX7C,ATP5C1,F11R,ANXA6,NDUFS8, TPH1 ,UQCRC2,CYC1, SEPHS1 ,COX5A,ACAT1,MYH9, HSP90AA1 ,TSP0,DIABLO,AK2,UQCRC1,GSTP1, HMG2	Expresión de genes	3.67 x 10 ⁻²	1	SOD2	Desordenes hereditarios	2.55 x 10 ⁻⁷ to 3.78 x 10 ⁻²	61	NDUFA9,ACO2,ATP5L,ETHE1,MT-CO2,SOD2, PRPS1 , SUBI ,NDUFA10,JUP,HADHA,DSP,DNMT1,ALDH7A1, CPH1A , SLC2A1 ,TUBB2A,PLEC,LMNA,COX7C,ATP5C1,DLG,GLUL, HSP90AA1 ,AK2,FH,DIABLO, HMG2 ,FLNB,SDHB,NEFL, GAPDH , TOP2B ,HADHB,CYB5R3, TOP2A ,NDUFS2,TACSTD2,EGFR,SDHA,NDUFV1,ETFB,HNRNPU,ETFA, TPM2 ,GLUD1,SLC25A3,AGL,F11R,NDUFS8, TPH1 , GALE ,CYC1, SEPHS1 ,PNPT1,ACAT1,MYH9,TSP0,UQCRC1,CTSC,PTRF

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Condiciones y enfermedades dermatológicas	8.33 x 10 ⁻¹ to 3.77 x 10 ⁻²	37	DDX39A,SYNCRIP,HLA-A,GAPDH,ANXA8L2 (incluye otros), MTX1, TOP2B,PGD,SOD2,H MGB1,PHB,CSE1L,SUB1,E1F5,HSP E1,CYB5R3, TOP2A,JUP,SERPINB5 ,PCBP2,DSP,PON2,CFL1,TUBB2A, PLEC,PSME2,LMNA,ANXA2,S100 A2,IL18,H2AFY,E1F4A3, FH,VDAC 1,CTSC,PTRF,GSTP1	Transporte molecular	3.67 x 10 ⁻²	1	SOD2	Desordenes psicológicos	2.55 x 10 ⁻¹ to 2.54 x 10 ⁻²	26	SDHB,NEFL, GAPDH,ATP5L,HADHB, SUB1,NDUFS2,DNMT1,SDHA,ETFB,H NRNP, DBI,SLC2A1,TUBB2A,COX7C ,ATP5C1,F11R,TPH1,CYC1,SEPHS1,ACAT1,HSP90AA1,TSPO,UQCRC1,GST P1,HMGB2
Desordenes musculares y esqueléticos	9.16 x 10 ⁻⁷ to 3.77 x 10 ⁻²	42	PEBP1,SRSF2,FLNB,SDHB,NDUF A9,NEFL,BAG6,GAPDH,ATP5L,PTGES3, TOP2B,HADHB,PRPS1,H MGB1,CSE1L,SUB1, TOP2A,NDUF A10,NDUFS2,SDHA,ETFB,HNRNP U,DDX42,PLEC,TPM2,LMNA,ANX A2,COX7C,F11R,ATP5C1,IL18,ND UFS8,TPH1,CYC1,SEPHS1,ACAT1, HSP90AA1,TSPO,CTSC,UQCRC1, GSTP1,HMGB2	Trafico de proteínas	3.67 x 10 ⁻²	1	SOD2	Desordenes musculares y esqueléticos	2.55 x 10 ⁻¹ to 3.78 x 10 ⁻²	37	SRSF2,FLNB,SDHB,NDUFA9,NEFL, B AG6,GAPDH,ATP5L, TOP2B,HADHB, PRPS1,HMGB1,SUB1, TOP2A,NDUFA 10,NDUFS2,SDHA,ETFB,HNRNP,PL EC,TPM2,LMNA,COX7C,F11R,ATP5C 1,IL18,NDUFS8,TPH1,CYC1,SEPHS1,AC AT1,HSP90AA1,TSPO,CTSC,UQCRC 1,GSTP1,HMGB2
Desordenes hereditarios	8.37 x 10 ⁻⁶ to 3.77 x 10 ⁻²	61	NDUFA9,ACO2,ATP5L,ETHE1,MT- CO2,SOD2,PRPS1,SUB1,NDUFA10 ,JUP,HADHA,DSP,DNMT1,ALDH7 A1,CPT1A,SLC2A1,TUBB2A,PLEC ,LMNA,COX7C,ATP5C1,DLG,GLU L,HSP90AA1,AK2, FH,DIABLO,HM GB2,FLNB,SDHB,NEFL, GAPDH, T OP2B,HADHB,CYB5R3, TOP2A,N DUFS2,EGFR,SDHA,NDUFV1,ETF B,HNRNP,DDX42,ETFA,TPM2,G LUD1,SLC25A3,AGL,F11R,NDUFS 8,TPH1,GALE,CYC1,SEPHS1,PNPT 1,ACAT1,MYH9,TSPO,UQCRC1,CT SC,PTRF	Ciclo celular	4.3 x 10 ⁻²	2	TOP2A,SERPINB5	Cáncer	9.02 x 10 ⁻⁶ to 4.56 x 10 ⁻²	79	SRSF2,MYL6,GNB2L1,MT- CO2,ANXA8L2 (incluye otros), GLO1,NNMT,SOD2,HMGB1,CSE1L,S DPR,E1F5,XPO1,SERPINB5,JUP,DNM T1,ALDH7A1,MGST1,CPT1A,MYEF2, TIMM44,SLC2A1,TUBB2A,FAM120A, LMNA,ANXA2,TXNRD1,ATP5C1,PS MB2,LUC7L2,GLUL,HSP90AA1,AK2, FH,DIABLO,GSTP1,HMGB2,PEBP1,S DHB,SEC31A,DDX39A,HLA- A,CAP2,MRPL15,GAPDH,CTNNA1,R UVBL1,DNAJA1,CCT4, TOP2B,SLC16 A3,PHB,HSPE1, TOP2A,ATP5J2,TRIM2 9,AHNAK2,COX4I1,EGFR,TACSTD2,S DHA,DBI,GLUD1,TPM2,S100A2,COP G1,UBAP2L,F11R,PSME1,ANXA6,MC M2,TPH1,H2AFY,COX5A,ACAT1,MYH 9,DDX1,CTSC,NUTF2
Cáncer	8.89 x 10 ⁻⁶ to 3.77 x 10 ⁻²	82	SRSF2,MYL6,NUP133,GNB2L1,M T-CO2,ANXA8L2 (incluye otros), GLO1,NNMT,GNB1,SOD2,H MGB1,CSE1L,SDPR,E1F5,XPO1,S MARCC2,JUP,SERPINB5,DSP,DN MT1,E1F4B,MGST1,CPT1A,SLC2 A1,TIMM44,MYEF2,TUBB2A,FA M120A,PLEC,LMNA,ANXA2,TXN RD1,ATP5C1,IL18,PSMB2,LUC7L 2,NUP210,GLUL,HSP90AA1, FH,A K2,DIABLO,GSTP1,PEBP1,SDHB,S EC31A,MRPL15,CAP2,HLA- A,BAG6,GAPDH,CTNNA1, TOP2B, SLC16A3,PHB,ATP5J2, TOP2A,AH NAK2,COX4I1,TACSTD2,EGFR,SD HA,ETFA,GLUD1,TPM2,S100A2,C OPG1,UBAP2L,F11R,PSME1,ANX A6,CTNBNB1,MCM2,H2AFY,TPH1 ,GALE,COX5A,ACAT1,MYH9,DD X1,CTSC,NUTF2	Interacción y señalización célula-célula	4.3 x 10 ⁻²	2	TOP2A,SERPINB5	Desordenes del sistema endocrino	9.02 x 10 ⁻⁶ to 3.78 x 10 ⁻²	28	SDHB,ACAT2,GAPDH,ANXA8L2 (incluye otros), LRPAP1, TOP2B, NNMT,SLC16A3,SOD2,GPD2, TOP2A, DNMT1,HADHA,EGFR,TACSTD2,CPT 1A,PON2,TUBB2A,LMNA,GLUD1,S10 0A2,IMMT,CTNBNB1,H2AFY,IGF2BP 2,COX5A, HSP90AA1,GSTP1

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Desordenes del sistema endocrino	8.89 x 10 ⁻⁶ to 3.77 x 10 ⁻²	30	SDHB,ACAT2,BAG6,GAPDH,ANXA8L2 (incluye otros),LRPAP1,TOP2B,NNMT,SLC16A3,SOD2,GPD2,TOPT2A,DNMT1,HADHA,EGFR,TACSTD2,PON2,CPT1A,TUBB2A,LMNA,GLUD1,S100A2,IMMT,CTNBL1,H2AFY,IGF2BP2,COX5A,HSP90AA1,DIABLO,GSTP1	Crecimiento y proliferación celular	4.3 x 10 ⁻²	2	TOP2A,SERPINB5	Enfermedades del sistema reproductivo	9.02 x 10 ⁻⁶ to 4.32 x 10 ⁻²	43	SRSF2,BAG6,NUP133,ETHE1,GAPDH,ANXA8L2 (incluye otros),TOP2B,SLC16A3,CSE1L,GPD2,DLST,TOPT2A,NDUFA10,XPO1,SERPINB5,JUP,DNMT1,EGFR,TACSTD2,ALDH7A1,MGST1,HNRNPU,SLC2A1,HSPA9,TUBB2A,PSME2,LMNA,TPM2,ANXA2,S100A2,COPG1,POLR2L,ATP5B,MCM2,H2AFY,COX5A,HSP90AA1,DDX1,UQCRC1,FH,VDAC1,CTSC,GSTP1
Enfermedades del sistema reproductivo	8.89 x 10 ⁻⁶ to 3.77 x 10 ⁻²	43	SRSF2,BAG6,NUP133,ETHE1,GAPDH,ANXA8L2 (incluye otros),TOP2B,SLC16A3,CSE1L,GPD2,DLST,TOPT2A,NDUFA10,XPO1,JUP,DNMT1,EGFR,TACSTD2,MGST1,HNRNPU,HSPA9,TUBB2A,PSME2,LMNA,TPM2,ANXA2,S100A2,COPG1,POLR2L,IL18,ATP5B,MCM2,H2AFY,GALE,COX5A,HSP90AA1,DDX1,UQCRC1,FH,VDAC1,DIABLO,CTSC,GSTP1					Ciclo celular	2.01 x 10 ⁻³ to 4.48 x 10 ⁻²	11	TOP2B,SRSF2,HMGB1,NEFL,CSE1L,PNPT1,TOPT2A,LMNA,ANXA2,NCAPD2,EGFR
Enfermedades hematológicas	1.55 x 10 ⁻⁵ to 3.77 x 10 ⁻²	33	SRSF2,HLA-A,GNB2L1,MT-CO2,NNMT,GNB1,TOPT2B,SOD2,HMGB1,CSE1L,TOPT2A,CYBSR3,XPO1,SERPINB5,AHNAK2,DNMT1,EGFR,CPT1A,SLC2A1,FAM120A,TUBB2A,ANXA2,TXNRD1,F11R,PSME1,ANXA6,PSMB2,LUC7L2,MYH9,DIABLO,AK2,NUTF2,GSTP1					Enfermedades hematológicas	4.07 x 10 ⁻³ to 4.94 x 10 ⁻²	30	SRSF2,HLA-A,GNB2L1,MT-CO2,TOPT2B,NNMT,HMGB1,CSE1L,CYBSR3,TOPT2A,SERPINB5,AHNAK2,DNMT1,EGFR,CPT1A,SLC2A1,FAM120A,TUBB2A,ANXA2,TXNRD1,F11R,PSME1,ANXA6,PSMB2,LUC7L2,MYH9,DIABLO,AK2,NUTF2,GSTP1
Ciclo celular	2 x 10 ⁻⁵ to 3.77 x 10 ⁻²	12	TOPT2B,SRSF2,HMGB1,CSE1L,NEFL,PNPT1,TOPT2A,LMNA,ANXA2,SERPINB5,NCAPD2,EGFR					Enfermedades gastrointestinales	6.93 x 10 ⁻³ to 4.32 x 10 ⁻²	41	SDHB,SEC31A,DDX39A,CAP2,GNB2L1,CTNNA1,RUVBL1,MT-CO2,DNAJA1,CCT4,TOPT2B,NNMT,SOD2,HMGB1,PHB,CSE1L,HSPE1,TOPT2A,TRIM29,HADHA,TACSTD2,EGFR,CPT1A,MYEF2,DBI,SLC2A1,TUBB2A,TPM2,GLUD1,LMNA,ANXA2,AGL,TXNRD1,COPG1,IL18,H2AFY,TPI1,HSP90AA1,CTSC,GSTP1,HMGB2
Expresión de genes	2.07 x 10 ⁻³ to 2.53 x 10 ⁻²	23	CFL1,SYNCRIP,EIF5B,GNB2L1,GAPDH,LMNA,TXNRD1,IL18,SOD2,HMGB1,PHB,EIF4G2,EIF4A3,EIF5,PNPT1,IGF2BP2,JUP,PTRF,GSTP1,DNMT1,HMGB2,EGFR,EIF4B					Enfermedades del sistema hepático	6.93 x 10 ⁻³ to 3.96 x 10 ⁻²	21	CPT1A,SLC2A1,CAP2,GNB2L1,GLUD1,ANXA2,AGL,TXNRD1,COPG1,NNMT,TOPT2B,IL18,SOD2,HMGB1,TPI1,CSE1L,TOPT2A,HSP90AA1,HADHA,TACSTD2,EGFR
Síntesis de proteínas	2.07 x 10 ⁻³ to 2.16 x 10 ⁻²	19	ACOT13,SYNCRIP,EIF5B,GAPDH,PITRM1,PCYOX1,TXNRD1,COPG1,SOD2,EIF4G2,EIF4A3,EIF5,IGF2BP2,MYH9,CTSC,PMPCA,UFD1L,EGFR,EIF4B					Enfermedades renales y urológicas	1.3 x 10 ⁻⁴ to 2.65 x 10 ⁻²	22	ETFB,DBI,ETFA,TUBB2A,ANXA2,COX7C,TOPT2B,IL18,SOD2,HMGB1,MCM2,H2AFY,ACAT1,TOPT2A,DLSD,MYH9,HSP90AA1,FH,ATP5L,GSTP1,TACSTD2,EGFR
Enfermedades gastrointestinales	7.55 x 10 ⁻³ to 3.77 x 10 ⁻²	29	SEC31A,CAP2,GNB2L1,CTNNA1,TOPT2B,NNMT,SOD2,HMGB1,PHB,CSE1L,TOPT2A,HADHA,TACSTD2,EGFR,CPT1A,SLC2A1,TUBB2A,LMNA,GLUD1,ANXA2,AGL,TXNRD1,COPG1,IL18,H2AFY,TPI1,HSP90AA1,CTSC,GSTP1					Supervivencia y muerte celular	1.57 x 10 ⁻⁴ to 4.43 x 10 ⁻²	34	PEBP1,SRSF2,FLNB,BAG6,GNB2L1,ETHE1,GAPDH,GLO1,TOPT2B,SOD2,EIF4G2,CSE1L,DLST,TOPT2A,XPO1,SERPINB5,JUP,TRIM29,PCBP2,EGFR,MGST1,ATP5A1,HSPA9,LMNA,ANXA2,TXNRD1,SLC3A2,IL18,IMMT,PNPT1,COX5A,VDAC1,DIABLO,GSTP1

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Enfermedades del sistema hepático	7.55 x 10 ⁻³ to 3.77 x 10 ⁻²	23	CPT1A, SLC2A1, CAP2, GNB2L1, GLUD1, ANXA2, AGL, TXNRD1, COPG1, NNMT, TOP2B, IL18, SOD2, HMGB1, H2AFY, CSE1L, TPI1, TOP2A, HSP90AA1, GSTP1, HADHA, TACSTD2, EGFR	Ensamble y organización celular	1.62 x 10 ⁻⁴ to 3.78 x 10 ⁻²	24	FLNB, CFL1, NEFL, NUP133, TIMM50, GABDH, CTNNA1, LMNA, ANXA2, NCAPD2, PTGES3, SLC3A2, MYO18A, TOP2B, NSF, SOD2, IMMT, PHB, MYL12B, TOP2A, MYH9, DSP, DNMT1, EGFR
Enfermedades renales y urológicas	1.56 x 10 ⁻⁴ to 2.77 x 10 ⁻²	22	ETFB, DBI, ETFA, TUBB2A, ANXA2, COX7C, TOP2B, IL18, SOD2, HMGB1, MCM2, H2AFY, ACAT1, TOP2A, LDL, MYH9, HSP90AA1, FH, ATP5I, GSTP1, TACSTD2, EGFR	Compromiso celular	1.62 x 10 ⁻⁴ to 2.54 x 10 ⁻²	4	TOP2B, SOD2, TOP2A, VPS26A
Supervivencia y muerte celular	1.56 x 10 ⁻⁴ to 3.77 x 10 ⁻²	50	SRSF2, GNB2L1, TIMM50, ETHE1, LRPAP1, GLO1, MYO18A, NSF, SOD2, EIF4G2, HMGB1, CSE1L, DLST, SUB1, XPO1, SERPINB5, JUP, MGST1, CFL1, ATP5A1, HSPA9, LMNA, ANXA2, TXNRD1, SLC3A2, IL18, HSP90AA1, VDAC1, DIABLO, GSTP1, PEBP1, FLNB, SDHB, BAG6, GAPDH, ACA2, TOP2B, PHB, TOP2A, PCBP2, TRIM29, EGFR, SDHA, IMMT, MCM2, CTNBL1, COX5A, PNPT1, MYH9, TSPO	Desordenes del desarrollo	1.62 x 10 ⁻⁴ to 3.78 x 10 ⁻²	29	FLNB, SDHB, MT-CO2, HADHB, PRPS1, JUP, HADHA, DSP, TACSTD2, SDHA, ETFB, HNRNPU, CPT1A, SLC2A1, ETFA, PLEC, TPM2, GLUD1, LMNA, AGL, IL18, GALE, PNPT1, DLG, ACAT1, GLUL, FH, AK2, PTRF
Ensamble y organización celular	1.61 x 10 ⁻⁴ to 3.77 x 10 ⁻²	20	FLNB, NEFL, NUP133, CTNNA1, TIMM50, LMNA, ANXA2, MYO18A, UBAP2L, TOP2B, NSF, SOD2, IMMT, PHB, CSE1L, TOP2A, MYH9, DNMT1, DSP, EGFR	Función y desarrollo del sistema endocrino	1.62 x 10 ⁻⁴ to 2.54 x 10 ⁻²	6	ACAT2, ACAT1, TIMM50, DNAJA1, PTGES3, GSTP1
Compromiso celular	1.61 x 10 ⁻⁴ to 2.53 x 10 ⁻²	6	TOP2B, SOD2, TOP2A, LMNA, VPS26A, EGFR	Enfermedades infecciosas	1.62 x 10 ⁻⁴ to 3.78 x 10 ⁻²	33	SRSF2, BAG6, NUP133, ETHE1, PTGES3, TOP2B, PRPS1, HMGB1, CSE1L, GPD2, DLST, SUB1, TOP2A, NDUFA10, XPO1, PCBP2, TRIM29, EGFR, HNRNPU, MYE2, HSPA9, TUBB2A, PSME2, LMNA, ANXA2, COPG1, POLR2L, ANXA6, ATP5B, EIF4A3, HSP90AA1, SMU1, UQCRC1
Desordenes del desarrollo	1.61 x 10 ⁻⁴ to 2.53 x 10 ⁻²	27	SDHA, FLNB, ETFB, SDHB, CPT1A, HNRNPU, SLC2A1, ETFA, PLEC, LMNA, GLUD1, TPM2, MT-CO2, AGL, HADHB, PRPS1, GALE, ACAT1, DLG, PNPT1, GLUL, JUP, AK2, FH, PTRF, HADHA, DSP	Enfermedades inflamatorias	1.62 x 10 ⁻⁴ to 4.32 x 10 ⁻²	18	CFL1, HLA-A, TUBB2A, PSME2, RUVBL1, S100A2, TOP2B, IL18, SOD2, PRPS1, HMGB1, TPI1, TOP2A, MYH9, SERPINB5, JUP, CTSC, GSTP1
Función y desarrollo del sistema endocrino	1.61 x 10 ⁻⁴ to 2.53 x 10 ⁻²	6	ACAT2, ACAT1, TIMM50, DNAJA1, PTGES3, GSTP1	Metabolismo de lípidos	1.62 x 10 ⁻⁴ to 2.54 x 10 ⁻²	10	SDHA, PEBP1, SDHB, ACAT2, IMMT, ACAT1, ANXA2, DNAJA1, PTGES3, GSTP1
Enfermedades infecciosas	1.61 x 10 ⁻⁴ to 3.77 x 10 ⁻²	39	SRSF2, BAG6, NUP133, GAPDH, ETHE1, MTX1, LRPAP1, PTGES3, TOP2B, PRPS1, HMGB1, CSE1L, GPD2, SUB1, DLST, TOP2A, NDUFA10, XPO1, PCBP2, TRIM29, EGFR, HNRNPU, SLC2A1, MYEF2, HSPA9, TUBB2A, PSME2, LMNA, ANXA2, COPG1, POLR2L, F1R, ANXA6, ATP5B, EIF4A3, HSP90AA1, GLUL, SMU1, UQCRC1	Enfermedades metabólicas	1.62 x 10 ⁻⁴ to 2.54 x 10 ⁻²	40	SDHB, ACAT2, NDUFA9, HLA-A, BAG6, MT-CO2, LRPAP1, NNMT, HADHB, PRPS1, SOD2, GPD2, NDUFS2, NDUFA10, HADHA, TACSTD2, SDHA, ETFB, NDUFV1, PON2, CPT1A, SLC2A1, ETFA, SLC25A3, LMNA, GLUD1, AGL, IMMT, CTNBL1, NDUFS8, TPI1, GALE, IGF2BP2, DLG, ACAT1, PNPT1, TSPO, GLUL, FH, PTRF
Enfermedades inflamatorias	1.61 x 10 ⁻⁴ to 3.77 x 10 ⁻²	12	TOP2B, IL18, PRPS1, CFL1, HMGB1, HLA-A, PSME2, TOP2A, MYH9, JUP, CTSC, S100A2	Enfermedades oftalmológicas	1.62 x 10 ⁻⁴ to 2.54 x 10 ⁻²	4	TOP2B, ACO2, TOP2A, TACSTD2
Metabolismo de lípidos	1.61 x 10 ⁻⁴ to 2.53 x 10 ⁻²	12	SDHA, PEBP1, SDHB, IL18, ACAT2, ACAT1, ANXA2, FH, DNAJA1, PTGES3, GSTP1, EGFR	Metabolismo de aminoácidos	4.83 x 10 ⁻⁴ to 3.78 x 10 ⁻²	3	GLS, GLUD1, TXNRD1

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Enfermedades metabólicas	1.61 x 10 ⁻⁴ to 2.16 x 10 ⁻²	39	SDHB,ACAT2,NDUFA9,HLA-A,BAG6,MT-CO2,LRPAP1,NNMT,HADHB,PRPS1,SOD2,GPD2,NDUF S2,NDUFA10,HADHA,SDHA,ETFB,NDUFV1,PON2,CPT1A,SLC2A1,E TFA,SLC25A3,LMNA,GLUD1,AGL,IMMT,CTNBL1,NDUFS8,TPI1,GALE,IGF2BP2,DL2,ACAT1,PNPT1,TSPO,GLUL,PH,PTRF		Metabolismo de carbohidratos	4.83 x 10 ⁻⁴ to 3.78 x 10 ⁻²	6	IL18,PGD,SLC2A1,GALE,PGLS,VDA C1
Transporte molecular	1.61 x 10 ⁻⁴ to 3.76 x 10 ⁻²	23	PEBP1,SDHB,PON2,DDX39A,SLC 2A1,NUP133,DNAJA1,SLC3A2,IL1 8,SLC16A3,SOD2,ERP29,SLC25A1 1,ACAT1,XPO1,MYH9,SERPINB5,J UP,PH,VDAC1,GSTP1,NUTF2,EGF R		Transporte molecular	7.9 x 10 ⁻⁴ to 3.78 x 10 ⁻²	21	PEBP1,SDHB,PON2,SLC2A1,DDX39A, NUP133,DNAJA1,IL18,SLC16A3,SOD 2,ERP29,SLC25A11,ACAT1,XPO1,MY H9,SERPINB5,JUP,VDAC1,GSTP1,NU TF2,EGFR
Enfermedades oftalmológicas	1.61 x 10 ⁻⁴ to 1.27 x 10 ⁻²	3	TOP2B,ACO2, TOP2A		Trafico de proteínas	7.9 x 10 ⁻⁴ to 2.88 x 10 ⁻²	6	ERP29,MYH9,JUP,DNAJA1,NUTF2,E GFR
Interacción y señalización célula-célula	1.87 x 10 ⁻⁴ to 3.77 x 10 ⁻²	14	FLNB,CTNNA1,ANXA2,PTGES3,U BAP1L,NSF,IL18,SOD2,HMGB1,TO P2A,MYH9,SERPINB5,DSP,EGFR		Interacción y señalización célula-célula	1.02 x 10 ⁻³ to 4.32 x 10 ⁻²	10	NSF,SOD2,HMGB1,HLA-A,GNB2L1, CTNNA1,MYH9,JUP,DSP,EGFR
Desarrollo de tejidos	1.87 x 10 ⁻⁴ to 3.77 x 10 ⁻²	11	FLNB,SOD2,CTNNA1,MYH9,SERP INB5,TXNRD1,PPM1G,PTGES3,DS P,HMGB2,EGFR		Desarrollo de tejidos	1.02 x 10 ⁻³ to 4.32 x 10 ⁻²	9	SOD2,GNB2L1,CTNNA1,MYH9,JUP,T XNRD1,PPM1G,DSP,EGFR
Mantenimiento y función celular	2.28 x 10 ⁻⁴ to 3.77 x 10 ⁻²	15	SDHB,NEFL,NUP133,TIMM50,SOD 2,HMGB1,IMMT,PHB,UQCRC2,CO X5A,DIABLO,UQCRC1,VDAC1,DS P,EGFR		Enfermedades inmunológicas	1.27 x 10 ⁻³ to 2.54 x 10 ⁻²	22	CFL1,HLA-A,FAM120A,PSME2, RUVBL1,ANXA2,S100A2,F11R,IL18,S OD2,ANXA6,TPI1,PSMB2, TOP2A,MY H9,JUP,SERPINB5,AK2,DIABLO,AHN AK2,NUTF2,EGFR
Supervivencia del organismo	2.33 x 10 ⁻⁴ to 1.99 x 10 ⁻²	15	SRSF2,CFL1,MYL6,GNB1,IL18,SO D2,MCM2,HMGB1,ATP5J2,GLUL, HSP90AA1,JUP,DDX1,GSTP1,EGF R		Condiciones y enfermedades dermatológicas	1.29 x 10 ⁻³ to 3.78 x 10 ⁻²	31	HLA-A,SYNCRIP,GAPDH,RUVBL1, ANXA8L2 (incluye otros),TOP2B, SOD2,HMGB1,CSE1L,SUB1, TOP2A,CYBSR3,SERPINB5,JUP,PCBP 2,DSP,CFL1,TUBB2A,PSME2,PLEC,L MNA,ANXA2,S100A2,IL18,TPI1,H2AF Y,PH,CTSC,VDAC1,PTRF,GSTP1
Metabolismo de aminoácidos	4.8 x 10 ⁻⁴ to 3.77 x 10 ⁻²	4	GLS,GLUD1,TXNRD1,SLC3A2		Expresión de genes	1.36 x 10 ⁻³ to 4.02 x 10 ⁻²	17	CFL1,EIF5B,GNB2L1,LMNA,TXNRD1, EIF4G2,PHB,HMGB1,EIF5,PNPT1,TF B2M,JUP,PTRF,DNMT1,EGFR,HMGB 2,EIF4B
Metabolismo de carbohidratos	4.8 x 10 ⁻⁴ to 2.53 x 10 ⁻²	6	IL18,PGD,SLC2A1,GALE,PGLS,VDAC1		Síntesis de proteínas	1.36 x 10 ⁻³ to 3.6 x 10 ⁻²	16	ACOT13,EIF5B,PITRM1,PCYOX1,TXN RD1,COPG1,SOD2,EIF4G2,EIF4A3,EIF 5,MYH9,CTSC,PMPCA,UFD1L,EGFR, EIF4B
Enfermedades inmunológicas	5.02 x 10 ⁻⁴ to 2.53 x 10 ⁻²	20	CFL1,HLA-A,FAM120A,PSME2, ANXA2,S100A2,F11R,IL18,SOD2,ANXA6,PSMB2, TOP2A,MYH9,JUP, SERPINB5,AK2,DIABLO,AHNAK2, NUTF2,EGFR		Mantenimiento y función celular	2.35 x 10 ⁻³ to 4.66 x 10 ⁻²	17	FLNB,SDHB,CFL1,NEFL,GNB2L1,NU P133,TIMM50,GAPDH,PTGES3,SLC3 A2,SRSF3,SOD2,HMGB1,COX5A,MY H9,DSP,EGFR
Respuesta celular a terapéuticos	9.52 x 10 ⁻⁴	2	SOD2,PNPT1		Metabolismo de drogas	2.35 x 10 ⁻³ to 2.54 x 10 ⁻²	3	DNAJA1,PTGES3,GSTP1
Trafico de proteínas	9.8 x 10 ⁻⁴ to 3.17 x 10 ⁻²		ERP29,MYH9,JUP,DNAJA1,NUTF 2,EGFR		Desarrollo celular	3.27 x 10 ⁻³ to 4.69 x 10 ⁻²	14	SDHB,TUBB2A,ANXA2,SLC3A2,SOD2, HMGB1,MCM2,H2AFY,DLST,TSPO,J UP,VDAC1,EGFR,TACSTD2
Metabolismo de drogas	2.34 x 10 ⁻³ to 2.53 x 10 ⁻²	6	IL18,ANXA2,DNAJA1,PTGES3,GS TP1,EGFR		Proliferación y crecimiento celular	3.27 x 10 ⁻³ to 4.69 x 10 ⁻²	14	SDHB,TUBB2A,ANXA2,SLC3A2,SOD2, HMGB1,MCM2,H2AFY,DLST,TSPO,J UP,VDAC1,TACSTD2,EGFR

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Modificación postranscripcional de RNA	3.21 x 10 ⁻³ to 1.38 x 10 ⁻²	9	SRSF2,HNRNPU,DDX39A,SYNCRIP,NUDT21,XRN2,NOP58,SNRPD1,HNRNP3		Modificación postranscripcional de RNA	3.27 x 10 ⁻³ to 1.39 x 10 ⁻²	9	SRSF2,HNRNPU,DDX39A,SYNCRIP,NUDT21,XRN2,NOP58,SNRPD1,HNRNP3
Desarrollo celular	3.25 x 10 ⁻³ to 3.71 x 10 ⁻²	4	SOD2,HMGB1,EGFR,HMGB2		Enfermedades respiratorias	3.84 x 10 ⁻³ to 3.78 x 10 ⁻²	13	HLA-A, TUBB2A, GNB2L1, ANXA2, TOP2B,SOD2,H2AFY,TP11,TOP2A,HS P90AA1,GSTP1,TACSTD2,EGFR
Crecimiento y proliferación celular	3.25 x 10 ⁻³ to 3.77 x 10 ⁻²	51	SRSF2,GNB2L1,SOD2,HMGB1,CS E1L,DLST,JUP,SERPINB5,ATP5F1,DSP,HADHA,E1F4B,SLC2A1,ATP5 A1,TUBB2A,PSME2,PLEC,ANXA2, SLC3A2,IL18,ATP5B,PSMB2,GLU L,AK2,VDAC1,DIABLO,GSTP1,H MGB2,PEBP1,SDHB,RUVBL1,DN AJA1,HADHB,PHB,NOP58,TOP2A, EGFR,TACSTD2,ETFB,NDUVF1,H NRNPU,PPM1G,F11R,MCM2,IMM T,H2AFY,ACAT1,PNPT1,MYH9,TS PO,CTSC		Eliminación de radicales libres	3.97 x 10 ⁻³ to 3.78 x 10 ⁻²	9	IL18,PON2,SOD2,HMGB1,IMMT,DLS T,GNB2L1,PNPT1,SERPINB5
Movimiento celular	3.79 x 10 ⁻³ to 3.77 x 10 ⁻²	22	PEBP1,FLNB,SLC2A1,GNB2L1,L MNA,ANXA2,RUVBL1,LRPAP1,P TGES3,F11R,IL18,SOD2,PHB,CSE1 L,ACAT1,TOP2A,HSP90AA1,MYH 9,JUP,SERPINB5,FE,EGFR		Enfermedades auditivas	4.59 x 10 ⁻³ to 1.28 x 10 ⁻²	6	PRPS1,PNPT1,HARS,MYH9,DIABLO, DNMT1
Enfermedades respiratorias	3.8 x 10 ⁻³ to 3.77 x 10 ⁻²	19	MYEF2,HLA-A, TUBB2A, NUP133,GNB2L1,ANXA2,TOP2B, SOD2,CTNNB1,MCM2,H2AFY,T P11,TOP2A,MYH9,HSP90AA1,SM ARCC2,GSTP1,TACSTD2,EGFR		Modificaciones postraduccionales	4.64 x 10 ⁻³ to 2.17 x 10 ⁻²	6	SOD2,ERP29,ACOT13,HSP90AA1,DN AJA1,LRPAP1
Eliminación de radicales libres	4.11 x 10 ⁻³ to 3.76 x 10 ⁻²	10	IL18,PON2,SOD2,HMGB1,IMMT,D LST,GNB2L1,PNPT1,SERPINB5,V DAC1		Plegamiento de proteínas	4.64 x 10 ⁻³ to 5.51 x 10 ⁻³	4	ERP29,HSP90AA1,DNAJA1,LRPAP1
Modificaciones postraduccionales	4.59 x 10 ⁻³ to 3.77 x 10 ⁻²	6	ERP29,SOD2,ACOT13,HSP90AA1, DNAJA1,LRPAP1		Desarrollo y función del Sistema hematológico	5.51 x 10 ⁻³ to 2.54 x 10 ⁻²	2	HMGB1,HLA-A
Plegamiento de proteínas	4.59 x 10 ⁻³ to 5.48 x 10 ⁻³	4	ERP29,HSP90AA1,DNAJA1,LRPA P1		Trafico inmune celular	5.51 x 10 ⁻³ to 2.54 x 10 ⁻²	2	HMGB1,HLA-A
Enfermedades auditivas	4.71 x 10 ⁻³ to 1.27 x 10 ⁻²	6	PRPS1,PNPT1,HARS,MYH9,DIAB LO,DNMT1		Respuesta inflamatoria	5.51 x 10 ⁻³ to 2.81 x 10 ⁻²	15	CFL1,HLA-A, TUBB2A, RUVBL1, S100A2,TOP2B,IL18,SOD2,HMGB1,T P11,TOP2A,MYH9,JUP,SERPINB5,GS TP1
Morfología celular	5.58 x 10 ⁻³ to 3.77 x 10 ⁻²	22	PEBP1,SDHB,NEFL,BAG6,LMNA, ANXA2,VPS26A,GNB1,F11R,IL18, SOD2,PHB,IMMT,CSE1L,COX5A, MYH9,HSP90AA1,SERPINB5,DIA BLO,VDAC1,DNMT1,EGFR		Enfermedades cardiovasculares	6.83 x 10 ⁻³ to 3.78 x 10 ⁻²	16	SDHA,ACAT2,PON2,CPT1A,PPP2R5 D,TUBB2A,LMNA,LRPAP1,NNMT,SO D2,CTNNB1,TSPO,JUP,DSP,GSTP1,E GFR
Enfermedades cardiovasculares	6.79 x 10 ⁻³ to 3.77 x 10 ⁻²	6	SDHA,SOD2,LMNA,JUP,DSP,GSTP 1		Supervivencia del organismo	8.19 x 10 ⁻³ to 1.77 x 10 ⁻²	11	SRSF2,IL18,SOD2,HMGB1,MCM2,GL UL,HSP90AA1,JUP,DDX1,GSTP1,EGF R
Función y desarrollo del tejido conectivo	8.23 x 10 ⁻³	2	SLC3A2,EGFR		Función y desarrollo del tejido conectivo	8.27 x 10 ⁻³	2	SLC3A2,EGFR
Morfología del tumor	8.23 x 10 ⁻³ to 3.77 x 10 ⁻²	4	SERPINB5,JUP,GSTP1,EGFR		Morfología del tumor	8.27 x 10 ⁻³ to 2.54 x 10 ⁻²	2	GSTP1,EGFR
Lesiones y anomalías del organismo	8.38 x 10 ⁻³ to 3.77 x 10 ⁻²	10	TOP2B,SRSF2,PRPS1,NEFL,TOP2 A,PLEC,LMNA,CTSC,GSTP1,DSP		Lesiones y anomalías del organismo	8.45 x 10 ⁻³ to 4.94 x 10 ⁻²	10	TOP2B,SRSF2,PRPS1,NEFL,TOP2A,P LEC,LMNA,CTSC,GSTP1,DSP

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Trafico de RNA	9.1 x 10 ⁻² to 1.27 x 10 ⁻²	3	DDX39A,NUP133,XPO1	Trafico de RNA	9.18 x 10 ⁻² to 1.28 x 10 ⁻²	3	DDX39A,NUP133,XPO1
Función y desarrollo del Sistema cardiovascular	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	2	GNB1,IL18	Movimiento celular	9.85 x 10 ⁻² to 3.79 x 10 ⁻²	13	PEBP1,FLNB,SLC2A1,GNB2L1,LMNA,ANXA2,SOD2,PHB,ACAT1, TOP2A, HSP90AA1,SERPINB5,EGFR
Señalización celular	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	2	PHB,VDAC1	Morfología celular	1.08 x 10 ⁻² to 3.78 x 10 ⁻²	13	PEBP1,F11R,SDHB,SOD2,CSE1L,NEFL,COX5A,HSP90AA1,LMNA,ANXA2, VPS26A,DNMT1,EGFR
Respuesta inmune mediada por células	1.27 x 10 ⁻² to 2.53 x 10 ⁻²	3	F11R,IL18, HMGB1	Señalización celular	1.28 x 10 ⁻² to 3.78 x 10 ⁻²	2	PHB,VDAC1
Desordenes del tejido conectivo	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	12	FLNB,IL18, PRPS1,HNRNPU,HMGB1,PLEC,TPM2,MYH9,LMNA,CTSC,PTRF,DSP	Desordenes del tejido conectivo	1.28 x 10 ⁻² to 3.78 x 10 ⁻²	12	FLNB,IL18, PRPS1,HNRNPU,HMGB1, PLEC,TPM2,MYH9,LMNA,CTSC,PTRF,DSP
Desarrollo embrionario	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	3	SOD2, HMGB2,EGFR	Desarrollo embrionario	1.28 x 10 ⁻² to 3.92 x 10 ⁻²	5	SDHB,SOD2,DLST,VDAC1,EGFR
Desarrollo y función de la piel y el cabello	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	1	EGFR	Desarrollo y función del Sistema nervioso	1.28 x 10 ⁻²	1	IL18
Desarrollo y función del Sistema hematológico	1.27 x 10 ⁻² to 2.53 x 10 ⁻²	3	F11R,IL18, HMGB1	Desarrollo de órganos	1.28 x 10 ⁻²	1	EGFR
Trafico celular inmune	1.27 x 10 ⁻² to 2.53 x 10 ⁻²	2	F11R,IL18	Metabolismo de vitaminas y minerales	1.28 x 10 ⁻²	1	VDAC1
Respuesta inflamatoria	1.27 x 10 ⁻² to 2.53 x 10 ⁻²	3	HMGB1, TOP2A, MYH9	Respuesta inmune mediada por células	2.54 x 10 ⁻²	1	HMGB1
Función y desarrollo del sistema nervioso	1.27 x 10 ⁻²	1	IL18	Hematopoyesis	2.54 x 10 ⁻²	1	HMGB1
Desarrollo de órganos	1.27 x 10 ⁻² to 3.71 x 10 ⁻²	2	HMGB2,EGFR	Respuesta inmune humoral	2.54 x 10 ⁻²	1	HMGB1
Morfología de órganos	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	2	GNB1,IL18	Estructura y desarrollo del tejido linfoide	2.54 x 10 ⁻²	1	HMGB1
Desarrollo del organismo	1.27 x 10 ⁻² to 3.71 x 10 ⁻²	5	CSE1L,PNPT1, PCBP2, HMGB2,EGFR	Reparación y daño del RNA	2.65 x 10 ⁻²	2	XRN2,PNPT1
Función y desarrollo del sistema reproductivo	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	3	UBAP2L,NSF,EGFR	Desarrollo del organismo	2.87 x 10 ⁻²	3	CSE1L,PNPT1, PCBP2
Función y desarrollo del Sistema respiratorio	1.27 x 10 ⁻² to 2.53 x 10 ⁻²	2	HMGB1,EGFR	Degradación de proteínas	3.6 x 10 ⁻²	8	MYH9,PCYOX1,PITRM1,CTSC,UFD1L,PMPCA, COPG1,EGFR
Función y desarrollo del Sistema muscular y esquelético	1.27 x 10 ⁻² to 3.77 x 10 ⁻²	2	GNB1,IL18	Función y desarrollo del cabello y piel	3.78 x 10 ⁻² to 4.24 x 10 ⁻²	5	SDHB,DLST,VDAC1,EGFR,TACSTD2
Metabolismo de vitaminas y minerales	1.27 x 10 ⁻²	1	VDAC1	Función y desarrollo del sistema urológico y renal	3.78 x 10 ⁻²	1	SOD2
Comportamiento	2.53 x 10 ⁻²	1	SOD2				
Hematopoyesis	2.53 x 10 ⁻²	1	HMGB1				
Respuesta inmune humoral	2.53 x 10 ⁻²	1	HMGB1				

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

Estructura y desarrollo del tejido linfoide	2.53 x 10 ⁻² to 3.71 x 10 ⁻²	3	HMGB1,EGFR,HMGB2		
Reparación y daño del RNA	2.64 x 10 ⁻²	2	XRN2,PNPT1		
Función y desarrollo del sistema urológico y renal	3.77 x 10 ⁻²	1	SOD2		

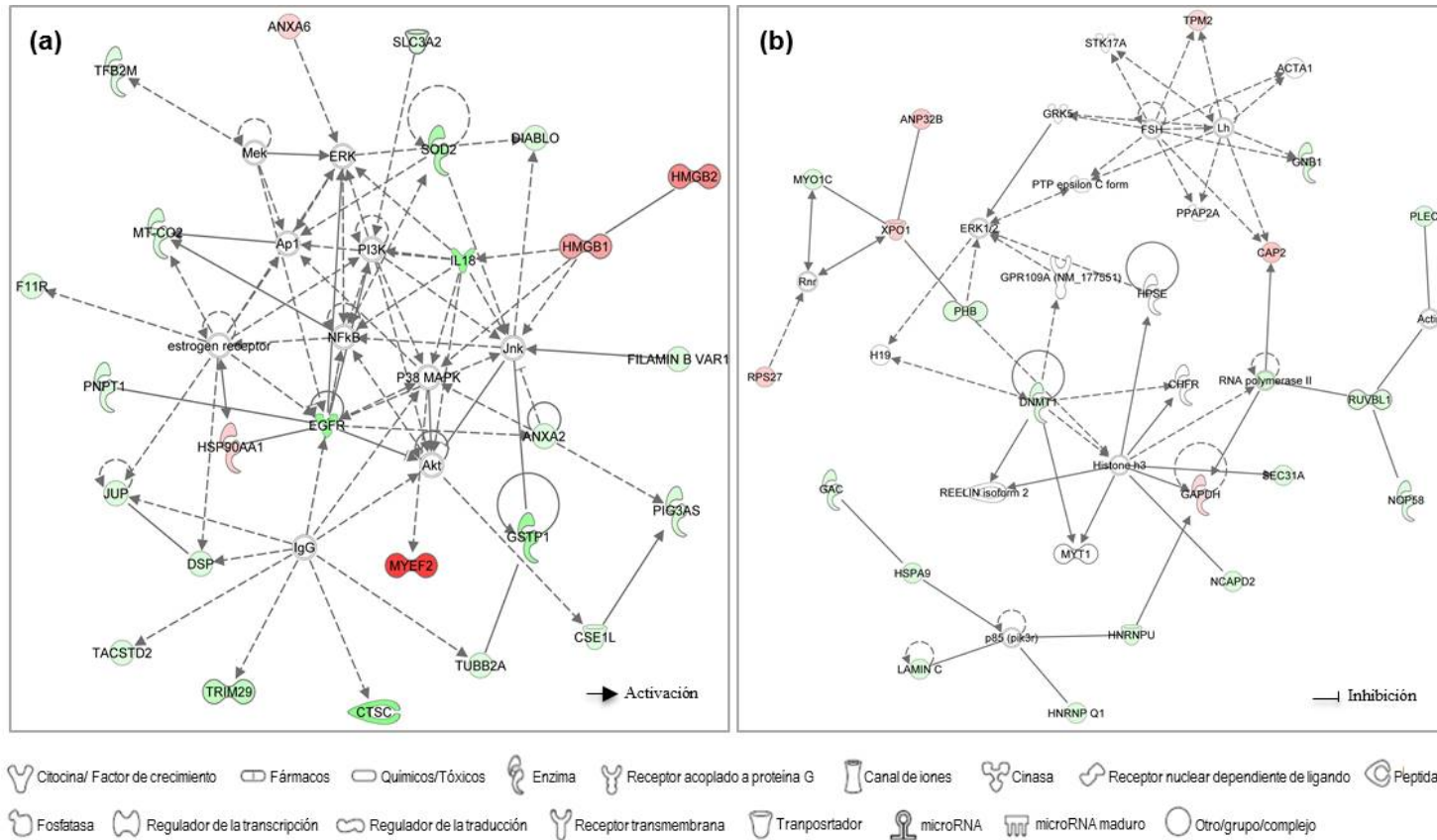


Figura 29. Redes de interacción funcional del análisis Core I de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama. El interactoma de proteínas estuvo relacionado con (a) Cáncer; enfermedades del sistema reproductivo; enfermedades y afecciones dermatológicas y enfermedades hematológicas, así como con (b) Enfermedades del tejido conectivo; enfermedades neurológicas; trastornos esqueléticos y musculares; replicación recombinación y reparación del ADN; cáncer y enfermedades gastrointestinales. Las diferentes formas de las moléculas, indican la clase funcional a la que pertenecen las proteínas. Las moléculas en rojo representan a las proteínas sobreexpresadas y las verdes representan a las subexpresadas. Las moléculas de color gris, las cuales no fueron especificadas, se incorporaron a las redes a través de IPA, debido a su relación con las proteínas de interés. Las relaciones moleculares entre los polipéptidos son indicadas con líneas. Una línea continua representa una interacción directa y una línea punteada se utiliza para las interacciones indirectas.

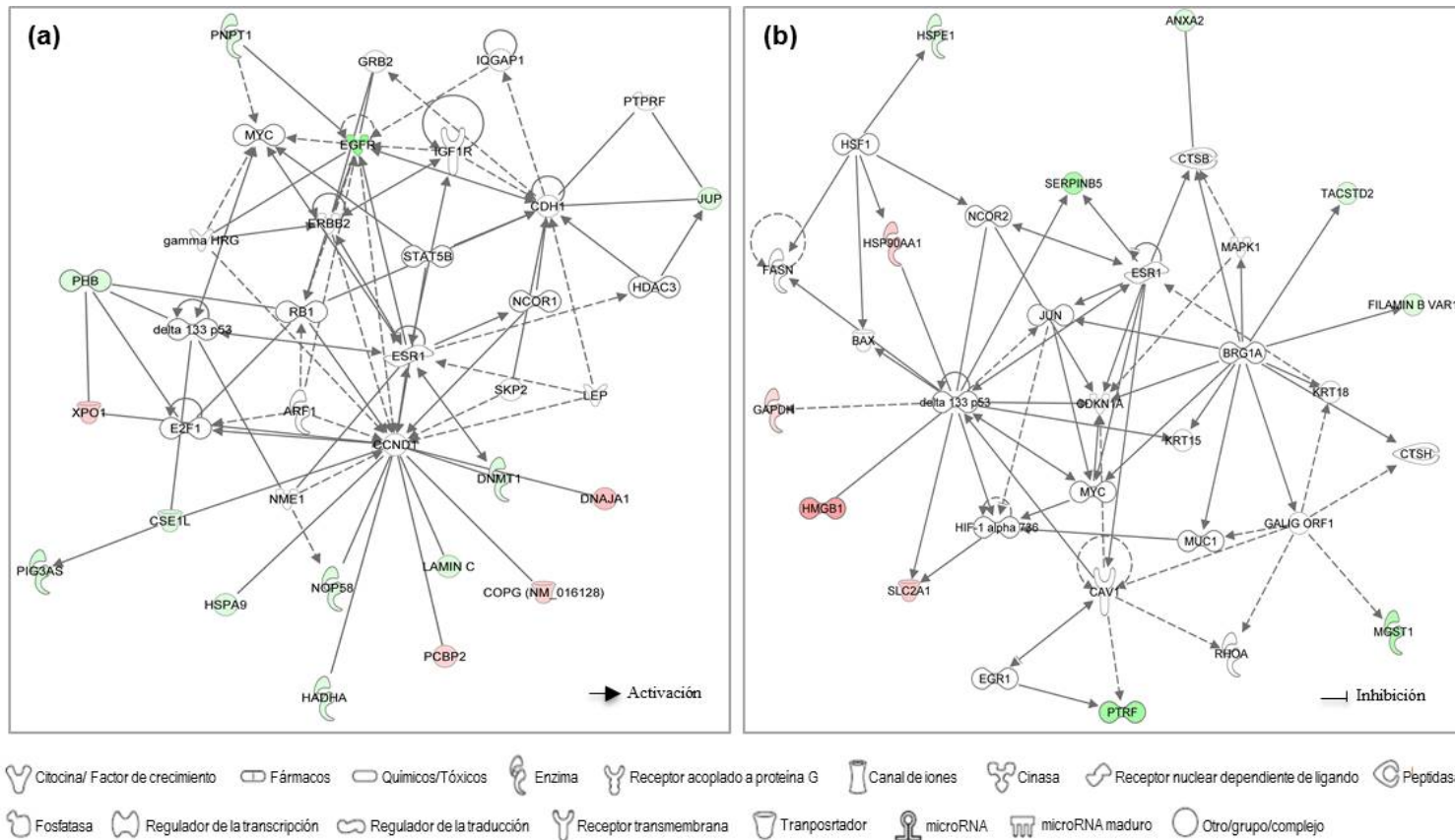


Figura 30. Redes de interacción funcional del análisis Core II de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama. El interactoma de proteínas estuvo involucrado en (a) Desarrollo celular; crecimiento y proliferación celular; ciclo celular. (b) Muerte celular y supervivencia; morfología del tumor y desarrollo celular. Las diferentes formas de las moléculas, indican la clase funcional a la que pertenecen las proteínas. Las moléculas en rojo representan a las proteínas sobreexpresadas y las verdes representan a las subexpresadas. Las moléculas de color gris, las cuales no fueron especificadas, se incorporaron a las redes a través de IPA, debido a su relación con las proteínas de interés. Las relaciones moleculares entre los polipéptidos son indicadas con líneas. Una línea continua representa una interacción directa y una línea punteada se utiliza para las interacciones indirectas.

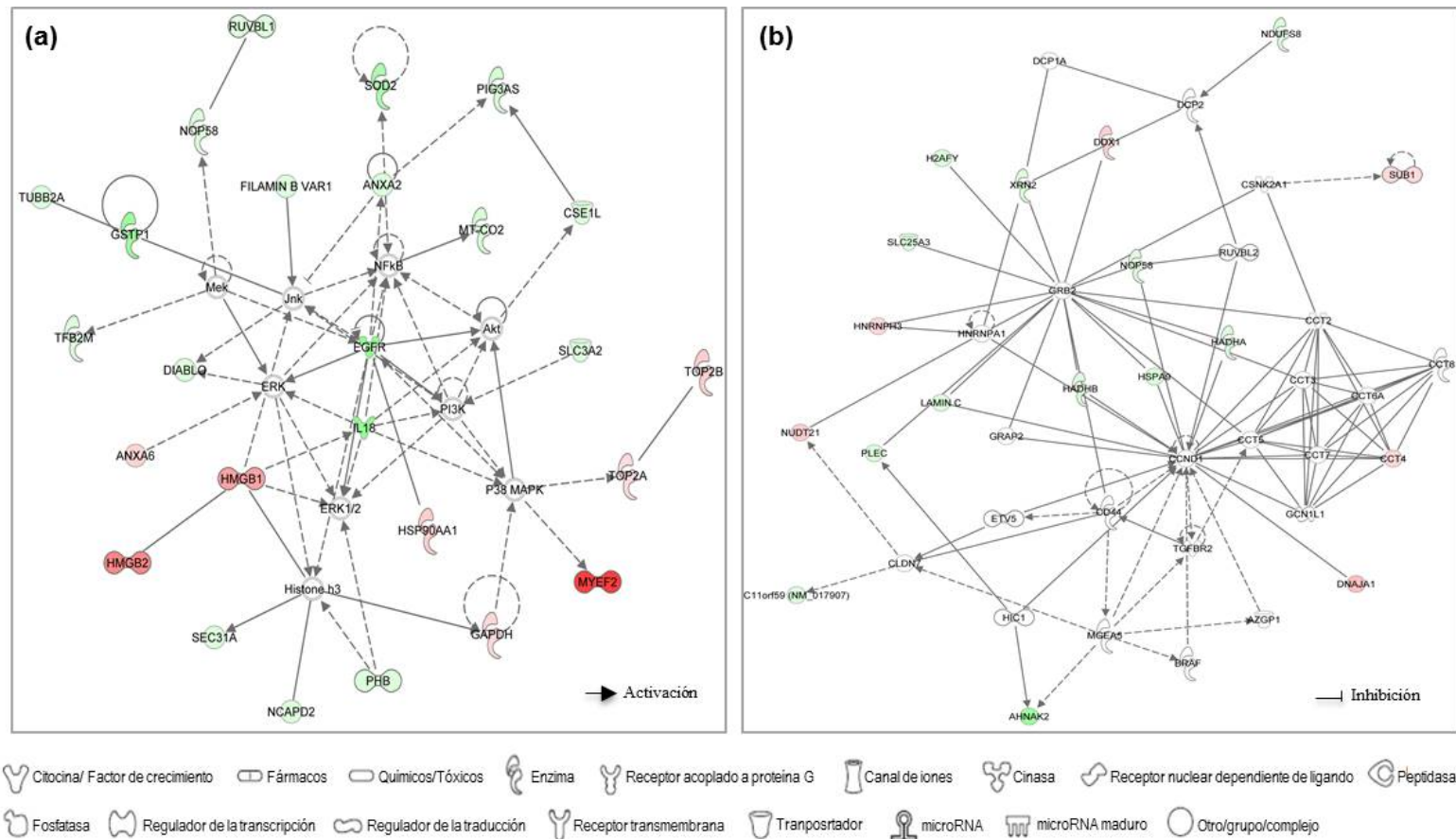


Figura 31. Redes de interacción funcional del análisis Core III de las proteínas expresadas diferencialmente en común en todas las líneas celulares de cáncer de mama. El interactoma de proteínas estuvo implicado en (a) Ciclo celular; cáncer; trastornos hereditarios; enfermedades hematológicas. (b) Trastornos del desarrollo, trastornos hereditarios; enfermedades metabólicas; muerte celular y supervivencia. Las diferentes formas de las moléculas, indican la clase funcional a la que pertenecen las proteínas. Las moléculas en rojo representan a las proteínas sobreexpresadas y las verdes representan a las subexpresadas. Las moléculas de color gris, las cuales no fueron especificadas, se incorporaron a las redes a través de IPA, debido a su relación con las proteínas de interés. Las relaciones moleculares entre los polipéptidos son indicadas con líneas. Una línea continua representa una interacción directa y una línea punteada se utiliza para las interacciones indirectas.

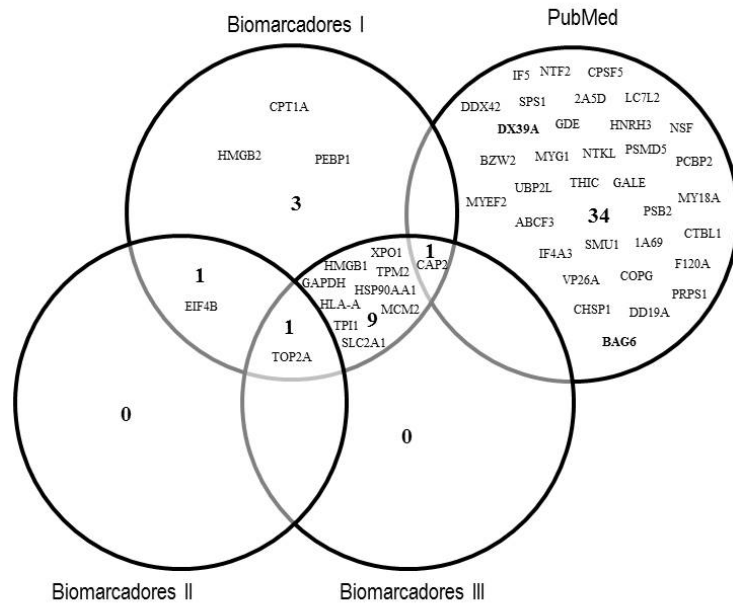
7. Selección de probables biomarcadores en cáncer de mama.

Se realizaron dos aproximaciones en la búsqueda de biomarcadores: 1) Filtro de posibles proteínas candidatas con el programa IPA. 2) Búsqueda de proteínas en la base de datos PubMed. Para la primera estrategia, se realizaron 3 análisis: Biomarcadores I, en el cual se consideraron todos los tejidos, células primarias y todas las líneas celulares. Biomarcadores II, para líneas celulares de cáncer de mama y glándula mamaria. Biomarcadores III, para todas las líneas celulares de cáncer, con excepción de las líneas celulares de cáncer de mama. Los resultados obtenidos en Biomarcadores I revelaron 39 posibles candidatos: ACAA2, ACAT1, ANXA2, CAP2, CPT1A, CTNNA1, CTSC, CYB5R3, DNMT1, EGFR, EIF4B, ETFB, F11R, FLNB, GAPDH, GLUD1, GSTP1, HADHA, HADHB, HLA-A, HMGB1, HMGB2, HNRNPU, HSP90AA1, IL18, JUP, MCM2, MYH9, NNMT, PEBP1, PHB, SERPINB5, SLC2A1, SOD2, TOP2A, TPI1, TPM2, TRIM29, XPO1. En el caso de Biomarcadores II, se obtuvieron 9 proteínas candidatas: CTSC, CYB5R3, EGFR, EIF4B, FLNB, GSTP1, IL18, MYH9, TOP2A. Finalmente, en Biomarcadores III, se encontraron 24 polipéptidos: ANXA2, CAP2, CTNNA1, DNMT1, EGFR, GAPDH, GLUD1, GSTP1, HLA-A, HMGB1, HSP90AA1, IL18, JUP, MCM2, NNMT, PHB, SERPINB5, SLC2A1, SOD2, TOP2A, TPI1, TPM2, TRIM29 y XPO1. Estos paneles de biomarcadores se proponen para todas las líneas celulares de cáncer de mama.

En cuanto a la búsqueda de biomarcadores en la base de datos de PubMed, se encontraron 35 proteínas sobreexpresadas y 53 subexpresadas que no han sido previamente reportadas en cáncer de mama, con excepción de la proteína ANXA8, de la cual sólo existe un reporte (Stein *et al.*, 2005). Finalmente, se realizó un análisis comparativo entre los posibles biomarcadores obtenidos con el filtro de biomarcadores de IPA y la búsqueda de las proteínas en PubMed (Figura 32). En este análisis, se encontraron 3 proteínas sobreexpresadas específicas para el filtro de Biomarcadores I (CPT1A, HMGB2, PEBP1) y 34 para el análisis con PubMed (IF5, NTF2, CPSF5, DDX42, SPS1, 2A5D, LC7L2, DX39A, GDE, HNRH3, NSF, BZW2, MYG1, NTKL, PSMD5, PCBP2, UBP2L, THIC, GALE, MYEF2, ABCF3,

PSB2, MY18A, IF4A3, SMU1, IA69, CTBL1, VP26A, COPG, F120A, CHSP1, DD19A, PRPS1, BAG6). En los filtros de Biomarcadores II y III, no se encontraron marcadores exclusivos (Figura 32a). En el caso de las proteínas subexpresadas, se encontraron sólo 6 biomarcadores exclusivos para el análisis de Biomarcadores I (ACAA2, HADHB, ACAT1, HADHA, F11R, HNRNPU) y 51 polipéptidos únicos para la búsqueda con PubMed (VDAC1, NU133, AT5EL, IMMT, ATPG, EHD2, H2AY, DHE3, UFD1, MYL6, IF2P, MYOC1, SYIM, SQRD, ACO13, M2OM, XRN2, AL7A1, NDUS8, ATP51, COX41, NUD19, NDUA9, RM38, SFXN1, LTOR1, AHNK2, NDUA9, NDUS2, AT5F1, CND1, OCAD1, MPPA, PCYOX, AXA82, TIM44, ETHE1, QIL1, TMX4, MTX1, SC31A, RM15, COX7C, RAB10, TFB2M, NDUB9, RPAB5, NOP58, ETFA, DLDH, ACON). De igual forma que en el análisis de las proteínas sobreexpresadas, en el filtro de Biomarcadores II y III no se obtuvieron marcadores exclusivos (Figura 32b). En este análisis, se redujo el número de proteínas encontradas exclusivamente en la base de datos de PubMed, obteniendo así 34 proteínas sobreexpresadas (Figura 32a) y 51 subexpresadas (Figura 32b) que no han sido previamente reportadas en cáncer de mama. Debido a que nuestra meta en este trabajo fue la identificación de nuevos biomarcadores en cáncer de mama, decidimos trabajar con estas proteínas que no han sido reportadas. Proponemos a BAG6, DDX39A (sobreexpresadas), ANXA8 y COX4 (subexpresadas) como biomarcadores. Hasta donde sabemos, no existen reportes en donde se relacione a BAG6 y DDX39 con cáncer de mama. En el caso de la proteína ANXA8, solo existe un reporte de ésta en cáncer de mama (Stein *et al.*, 2005). COX4 ha sido analizada en algunas líneas celulares de cáncer de mama (Radde *et al.*, 2014), sin embargo ésta no ha sido propuesta como biomarcador. Otro de los criterios para la selección de estas proteínas como biomarcadores en cáncer de mama fue su función. BAG6 se ha relacionado con procesos apoptóticos, la regulación de la estabilidad y la degradación de proteínas por el proteasoma. DDX39 interactúa con componentes de la maquinaria de transcripción y el espliceosoma. ANXA8 puede funcionar como organizador de dominios de membrana, así como andamio para el reclutamiento de proteínas. COX4 es un componente importante de la cadena respiratoria.

(a) Proteínas sobreexpresadas



(b) Proteínas subexpresadas

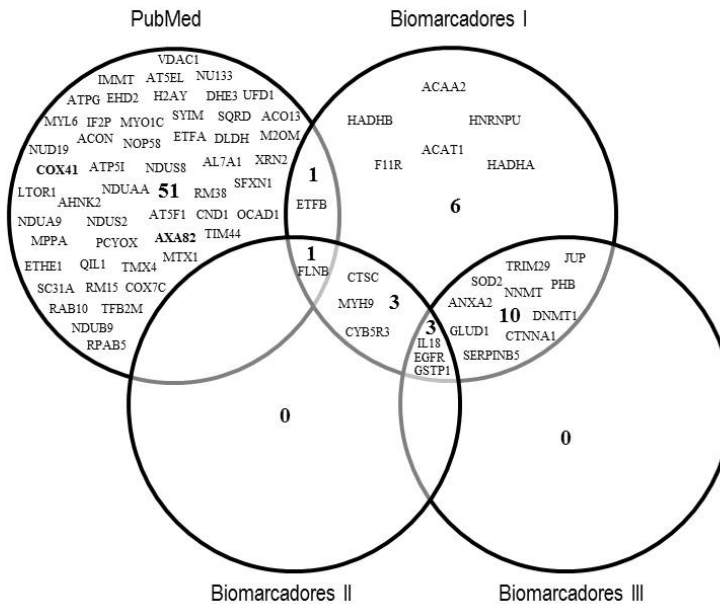


Figura 32. Diagramas de Venn-Euler de probables paneles de biomarcadores obtenidos con el programa IPA y la base de datos de PubMed. Los diagramas muestran lo posibles marcadores encontrados con el filtro de Biomarcadores I (todos los tejidos, células primarias y todas las líneas celulares), con Biomarcadores II (glándula mamaria y cáncer de mama) y con el filtro de Biomarcadores III (todas las líneas celulares de cáncer excluyendo las de cáncer de mama), así como los posibles biomarcadores obtenidos con la búsqueda en PubMed. (a) Panel de marcadores de las proteínas sobreexpresadas y (b) subexpresadas obtenidos en común en todas las líneas celulares de cáncer de mama.

8. Validación de los probables biomarcadores para cáncer de mama.

De acuerdo con la información y funciones de las proteínas identificadas, se decidió validar a las proteínas BAG6 y DDX39A como sobreexpresadas y ANXA8 y COX4 como subexpresadas a través de Western blot (Figura 33). A pesar de que los valores obtenidos con el análisis densitométrico no corresponden con los obtenidos en la cuantificación con iTRAQ, se observó una tendencia en la sobreexpresión de las proteínas BAG6 y DDX39 (Figura 33 a y b), así como una subexpresión de los polipéptidos ANXA8 y COX4 (Figura 33 c y d).

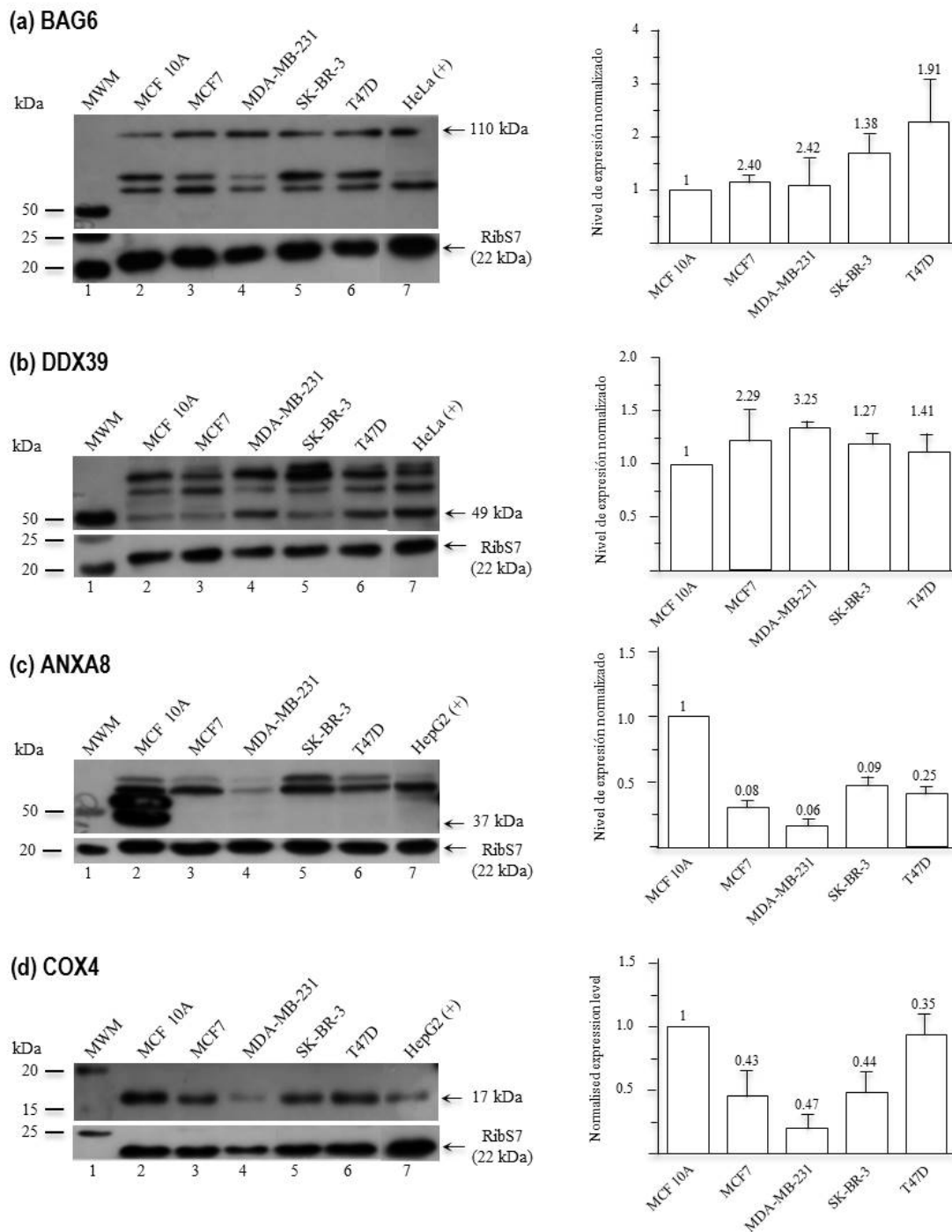


Figura 33. Validación de proteínas candidatas a biomarcadores en cáncer de mama. Western blot y análisis densitométrico de las proteínas sobreexpresadas BAG6 (A) y DDX39 (B) así como de los polipéptidos subexpresados ANXA8 (C) y COX4 (D) en las líneas celulares de cáncer de mama comparadas con la línea celular control. Los valores densitométricos obtenidos fueron normalizados respecto a la línea celular control. Las barras corresponden a la desviación estándar. Los números arriba de las barras son los valores obtenidos con la cuantificación de iTRAQ.

9. Perfil de expresión diferencial de las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+.

Independientemente de las proteínas compartidas entre las diferentes líneas celulares estudiadas en este trabajo, se encontraron proteínas específicas para cada una de ellas. En el caso de las proteínas sobreexpresadas se encontraron 41 proteínas exclusivas para MCF7 y 59 para T47D (líneas luminal A). Estas líneas luminales comparten 34 proteínas en común (Figura 34a). La línea tipo baja en claudina (MDA-MB-231) tuvo 55 proteínas específicas y en la línea que sobreexpresa HER2+ (SK-BR-3) se encontraron 74 polipéptidos únicos (Figura 34b). Por otro lado, en las proteínas subexpresadas se identificaron 39 y 32 proteínas únicas en las líneas luminales MCF7 y T47D, respectivamente. Estas líneas comparten 22 proteínas (Figura 35a). También se encontraron 44 y 49 polipéptidos exclusivos para las líneas MDA-MB-231 (baja-claudina) y SK-BR-3 (HER2+) (Figura 35b).

10. Clasificación funcional de las proteínas expresadas diferencialmente en las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+.

Las proteínas sobre y subexpresadas específicas de cada una de las líneas celulares de cáncer de mama (Luminal A, Baja en claudina y HER2+) se clasificaron con PANTHER de acuerdo con la clase de proteína a la que pertenecen, así como en los procesos biológicos en los que participan. Las proteínas sobreexpresadas fueron agrupadas dentro de 25 categorías, siendo la unión a ácidos nucleicos (XIV) la categoría más representativa para las líneas celulares tipo Luminal A y Baja en claudina, mientras que la clase de Hidrolasa (VIII) y Oxidorreductasa (XV) fue la clase con mayor número de proteínas para la línea celular HER2+ (Figura 36a). En cuanto a los procesos biológicos en los que están involucradas, se obtuvieron 12 categorías, siendo los procesos metabólicos (IX) la categoría más representativa en

cada una de las líneas celulares de cáncer de mama (Figura 36b). Por otra parte, en las proteínas subexpresadas se obtuvieron 25 categorías en la clasificación por clase de proteína y 11 en los procesos biológicos. En el caso de la clasificación por tipo de proteína, la clase con mayor número de genes involucrados fue la de unión a ácidos nucleicos (XIV) para las líneas celulares Baja en claudina y HER2+, mientras que las clases de Hidrolasa (VIII) y proteínas del citoesqueleto (V) fueron las más representativas para la líneas celulares de tipo Luminal A (Figure 36c). Finalmente, el proceso biológico más afectado en todas las líneas celulares fue el de procesos metabólicos (Figura 36d).

11. Redes de interacción proteína-proteína de las proteínas encontradas expresadas diferencialmente en las líneas celulares de cáncer de mama Luminal A, Baja en claudina y HER2+

Las redes de interacción proteína-proteína de cada una de las líneas celulares se generaron con el programa STRING. En el caso de las proteínas de la línea celular MCF7, el nodo de interacción principal estuvo relacionado con procesos metabólicos del RNAm (Figura 37a), mientras que el interactoma de los polipéptidos de la línea T47D presentó dos nodos; el primero de ellos estuvo involucrado en la biogénesis del ribosoma y el segundo estuvo relacionado con funciones, como el transporte de proteínas intracelular, y el direccionamiento de proteínas al retículo endoplásmico, entre otros (Figura 37b). En cuanto a la red de interacción de las proteínas encontradas en común entre las líneas celulares Luminales A (MCF7 y T47D), el nodo de interacción principal estuvo relacionado con procesamientos del RNAm (Figura 38a). En el caso de las proteínas sobreexpresadas únicas encontradas en la línea celular Baja claudina (MDA-MB-231), el nodo de interacción principal estuvo implicado en procesos que incluyen el ensamble de complejos de ribonucleoproteínas, formación del complejo de preinicio de la traducción, así como

en la regulación del inicio de la traducción (Figura 38b). Finalmente, el interactoma de los polipéptidos sobreexpresados específicos para la línea celular HER2+ (SK-BR-3) estuvo implicado en procesos metabólicos de pequeñas moléculas (Figura 37c).

Por otro lado, el análisis de interacción de los polipéptidos subexpresados específicos para MCF7 no reveló ningún nodo significativo (Figura 39a). Por el contrario, el nodo de la línea celular T47D, estuvo implicado en procesos como la regulación de la polimerización y despolimerización de actina, regulación de la longitud de los filamentos de actina, organización del citoesqueleto de actina, así como en la activación de procesos del sistema inmunológico (Figura 39b). Para las proteínas subexpresadas compartidas entre las líneas celulares Luminales A, no se encontró ningún nodo significativo (Figura 40a). Sin embargo, la red de interacciones proteína-proteína de la línea celular Baja en claudina reveló un pequeño nodo relacionado con 10 procesos biológicos, entre los cuales se encuentran: transporte intracelular de proteínas y organización de membrana (Figura 40b). Finalmente, el interactoma de la línea HER2+ mostró dos nodos principales, uno de ellos implicado en casi todos los procesos biológicos especificados y el otro nodo estuvo relacionado con la biogénesis del ribosoma (Figura 40c).

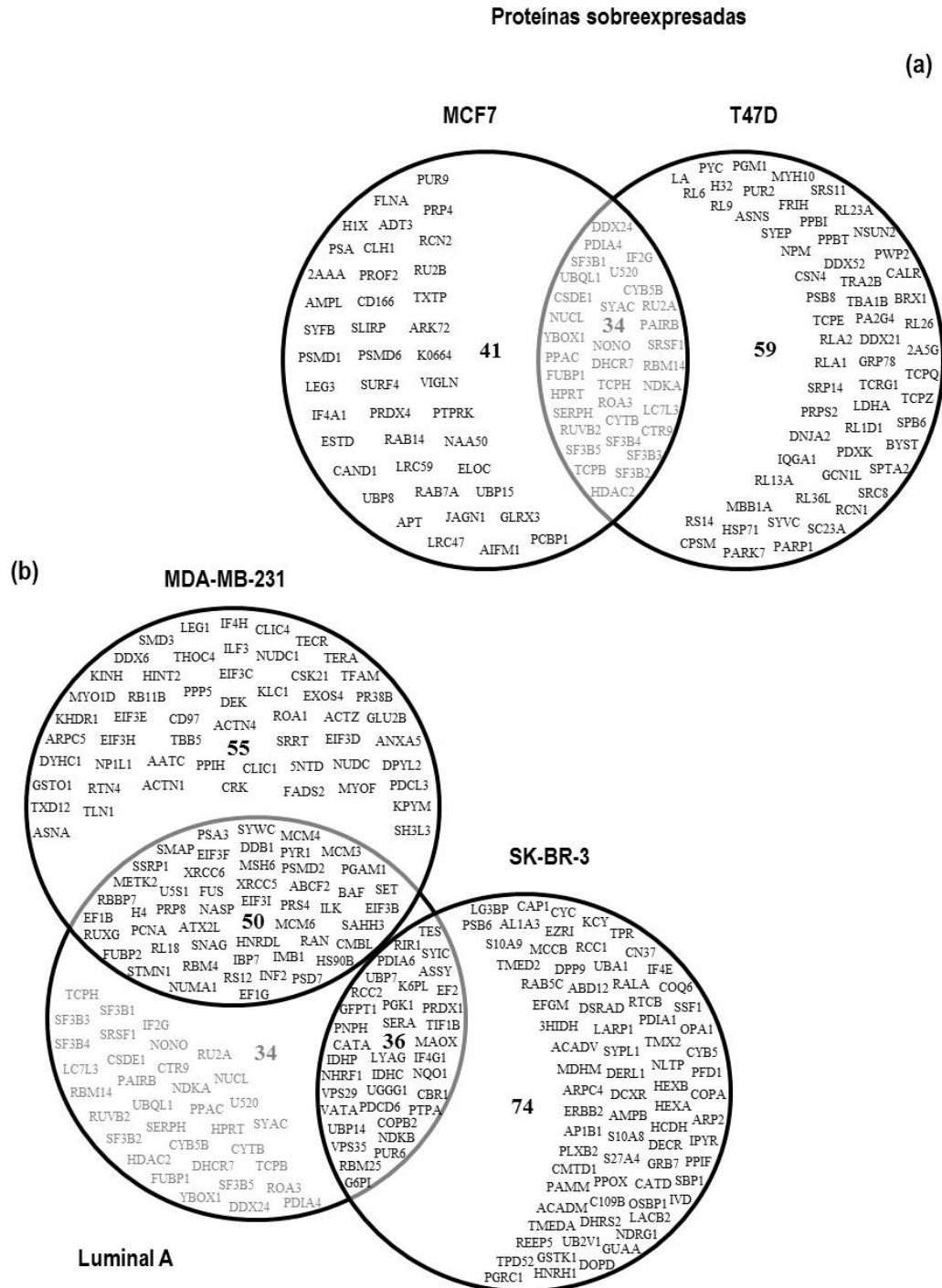


Figura 34. Diagramas de Venn-Euler de las proteínas sobreexpresadas en los diferentes tipos moléculas de líneas celulares de cáncer de mama. (a) El diagrama muestra las proteínas específicas para las líneas celulares luminales A (MCF7 y T47D), así como las proteínas que comparten estas líneas. (b) El diagrama muestra las proteínas únicas para la línea Baja-claudina (MDA-MB-231) y la línea HER2+ (SK-BR-3). También se observa la comparación de las proteínas comunes entre las líneas luminales A con los otros dos tipos de líneas celulares.

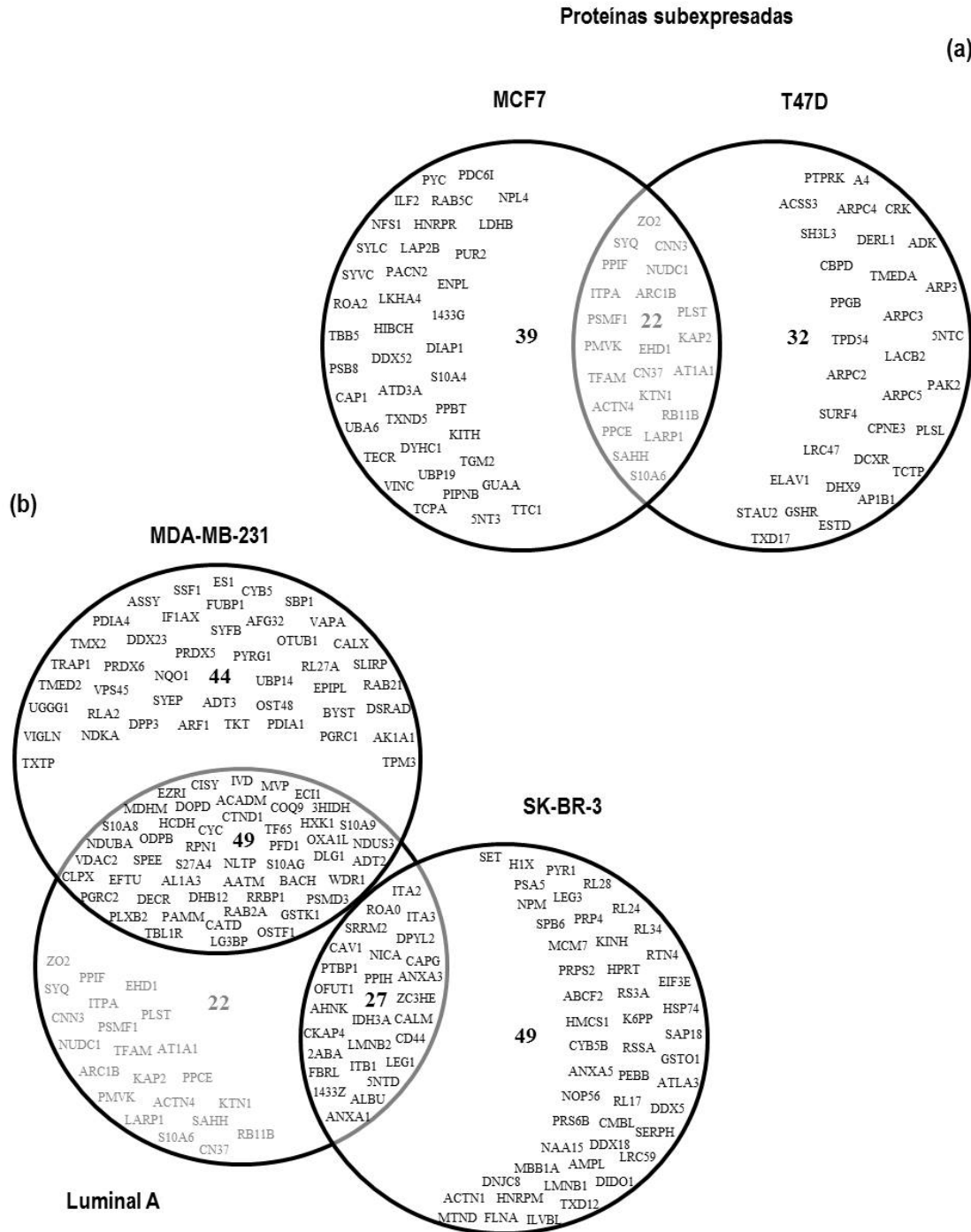


Figura 35. Diagramas de Venn-Euler de las proteínas subexpresadas en los diferentes tipos moléculas de líneas celulares de cáncer de mama. (a) El diagrama muestra las proteínas específicas para las líneas celulares luminales A (MCF7 y T47D), así como las proteínas que comparten estas líneas. (b) El diagrama muestra las proteínas únicas para la línea Bajacaudina (MDA-MB-231) y la línea HER2+ (SK-BR-3). También se observa la comparación de las proteínas comunes entre las líneas luminales A con los otros dos tipos de líneas celulares.

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

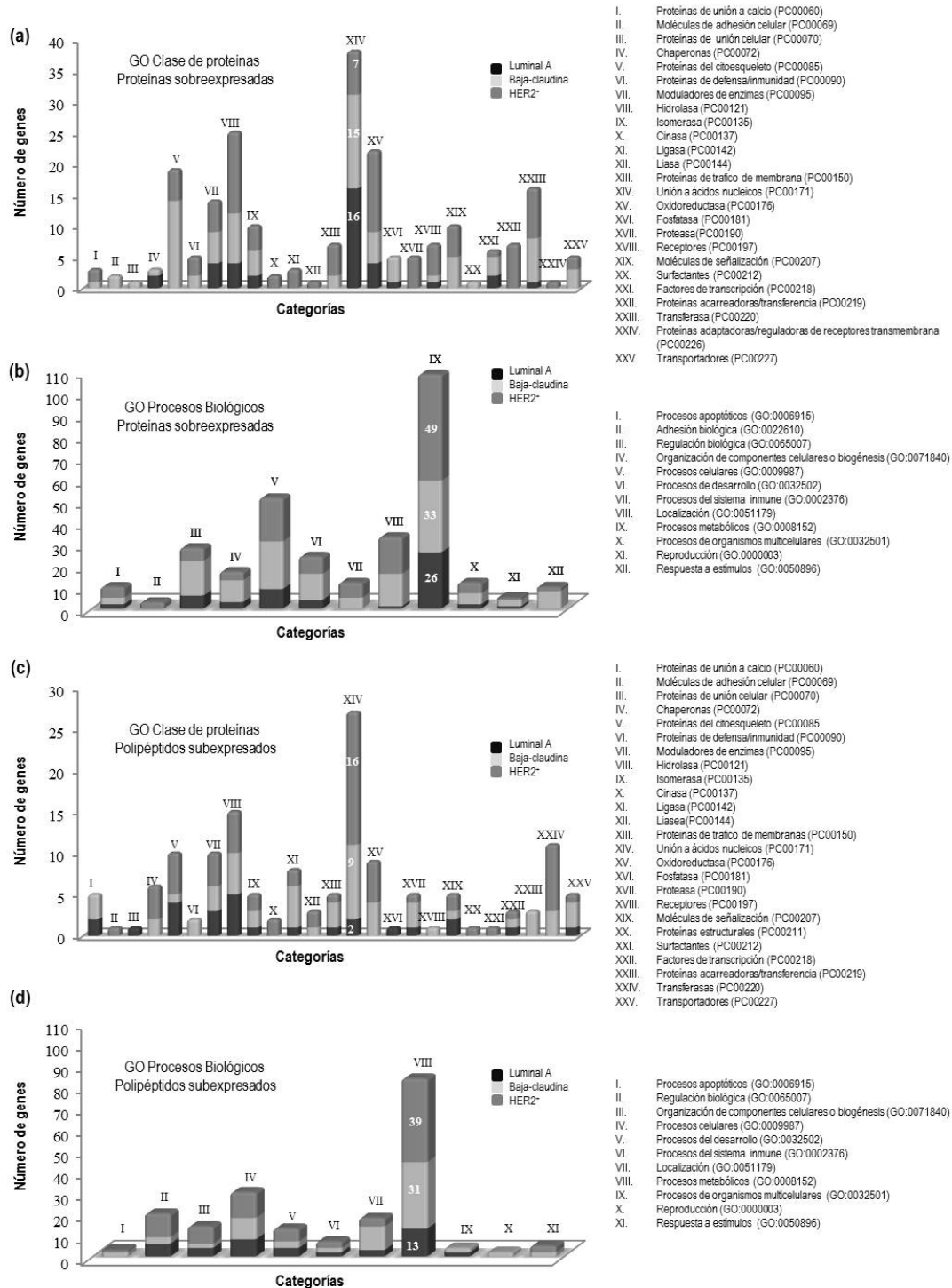
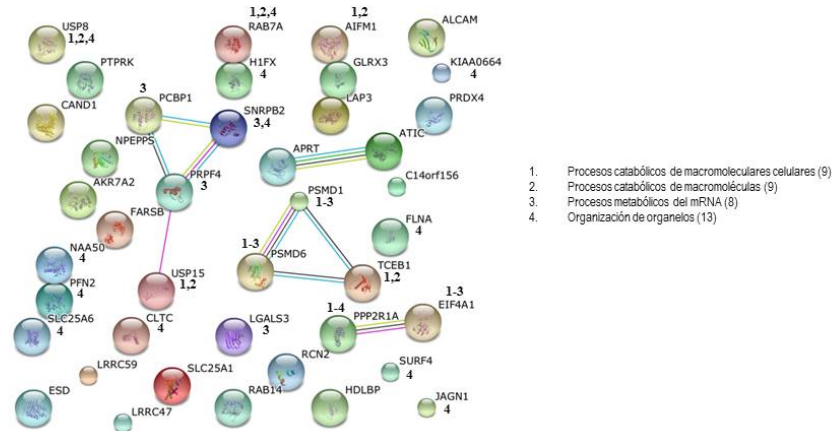


Figura 36. Clasificación funcional con PANTHER de las proteínas sobreexpresadas y subexpresadas exclusivamente encontradas en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama. Las proteínas se clasificaron de acuerdo a la clase de proteína a la que pertenecen (a,c), así como en los procesos biológicos en los que participan (b,d). Las barras representan las diversas categorías en las que se clasificaron las proteínas (eje x) y el número de genes involucrados en cada una de estas categorías (eje Y). (a,b) Proteínas sobreexpresadas. (c,d) Polipéptidos subexpresados.

(a) MCF7



(b) T47D

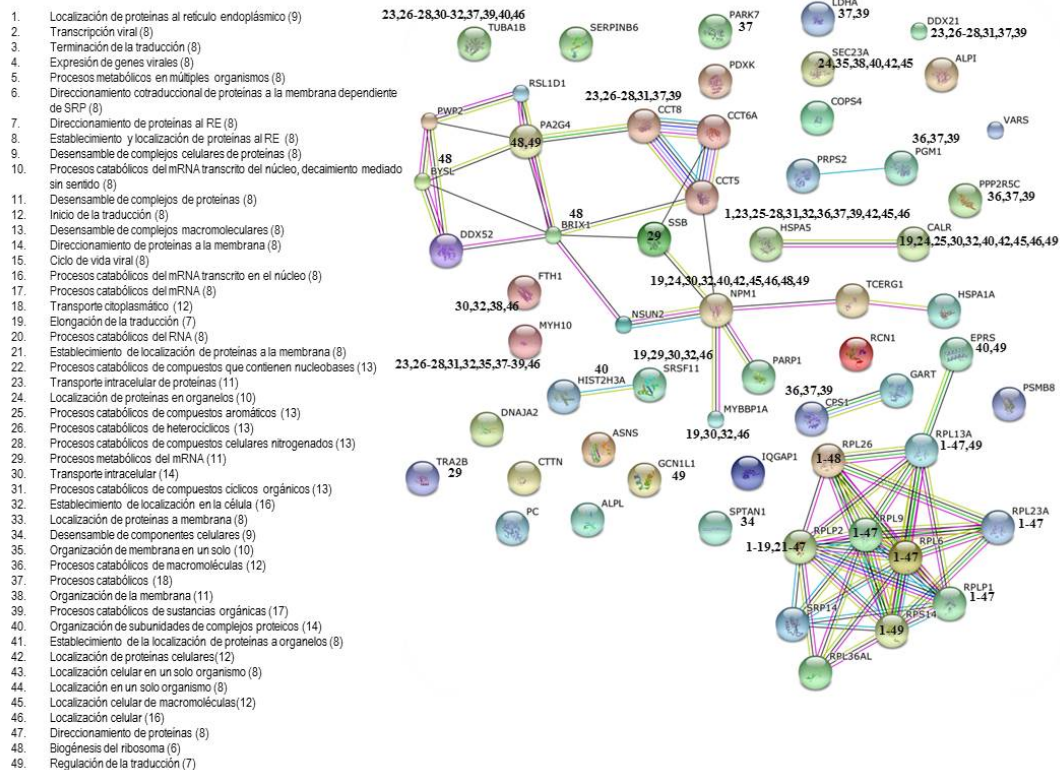


Figura 37. Interactoma de los polipéptidos sobreexpresados exclusivos en cada una de las líneas celulares Luminal A de cáncer de mama. Las redes muestran las interacciones funcionales predichas para las líneas celulares de cáncer de mama MCF7 (A) y T47D (B). Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$).

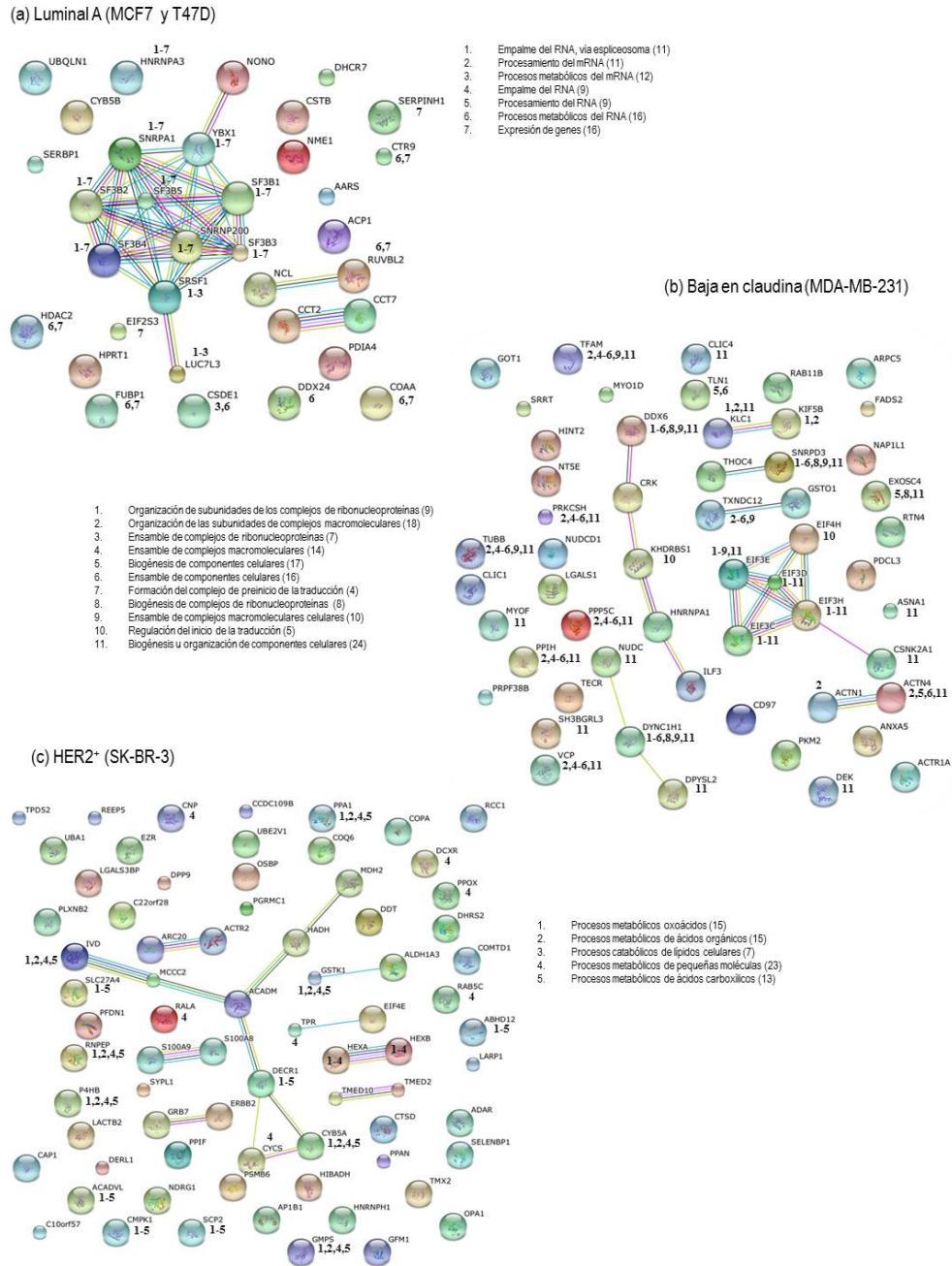


Figura 38. Interactoma de las proteínas sobreexpresadas exclusivas encontradas en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama. (A) Mapa de interacción de los polipéptidos sobreexpresados compartidos en las líneas celulares Luminales A (MCF7 y T47D). (B) Interactoma de las proteínas sobreexpresadas exclusivas en la línea celular Baja en claudina (MDA-MB-231). (C) Redes de interacción de las proteínas sobreexpresadas específicas para la línea HER2+ (SK-BR-3). Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$).

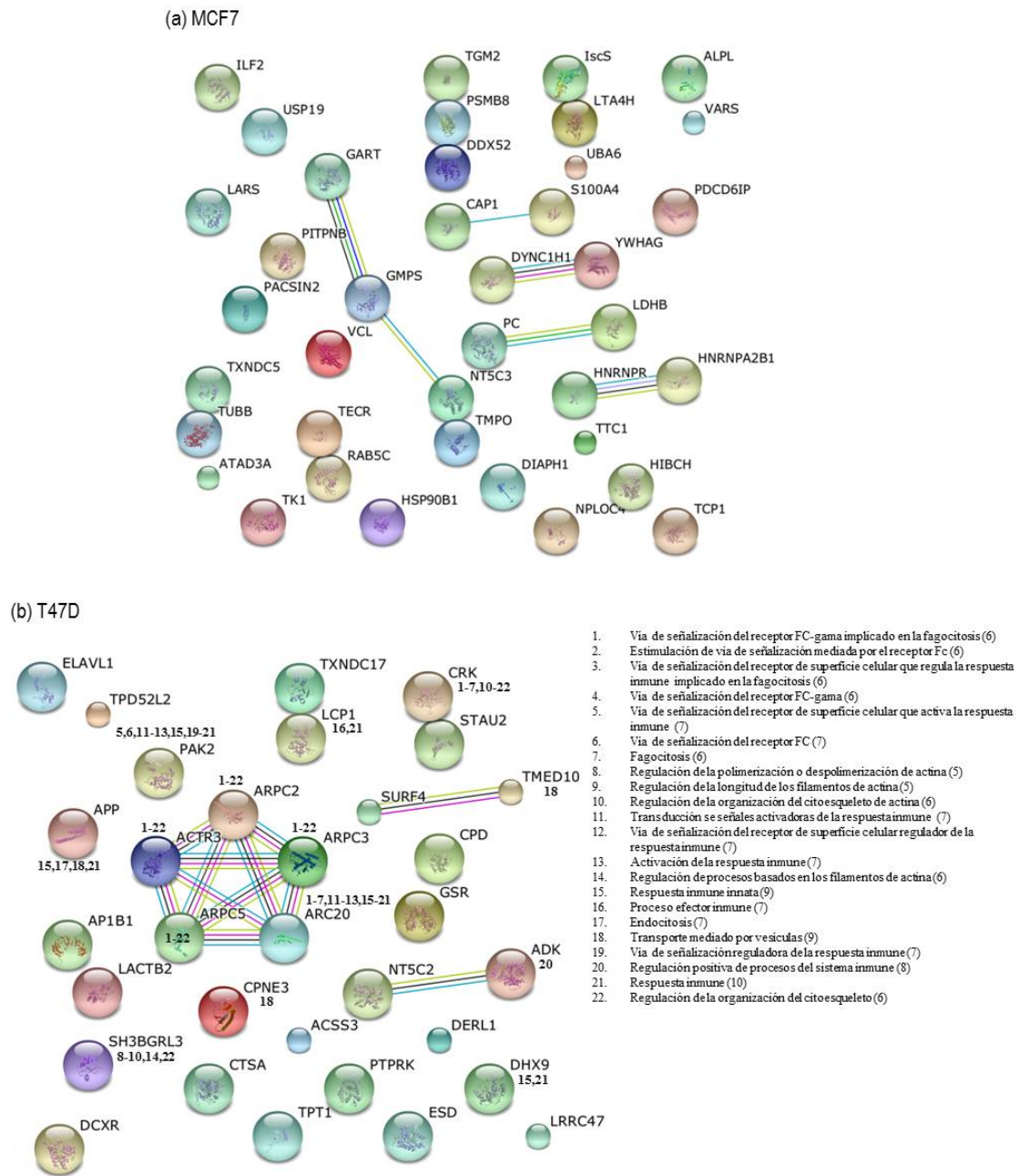
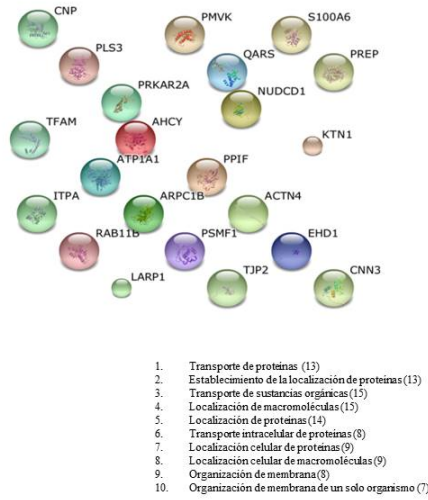


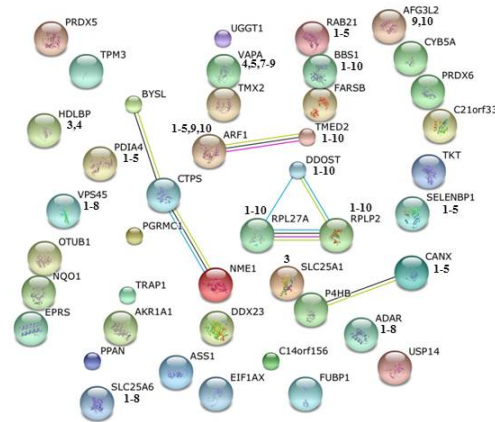
Figura 39. Interactoma de los polipéptidos subexpresados exclusivos en cada una de las líneas celulares Luminal A de cáncer de mama. Las redes muestran las interacciones funcionales predichas para las líneas celulares de cáncer de mama MCF7 (A) y T47D (B). Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$).

Análisis de la expresión diferencial de proteínas totales de células de cáncer de mama para la identificación de probables biomarcadores útiles para el diagnóstico y/o pronóstico de esta enfermedad.

(a) Luminal A (MCF7 y T47D)



(b) Baja en claudina (MDA-MB-231)



(c) HER2+ (SK-BR-3)

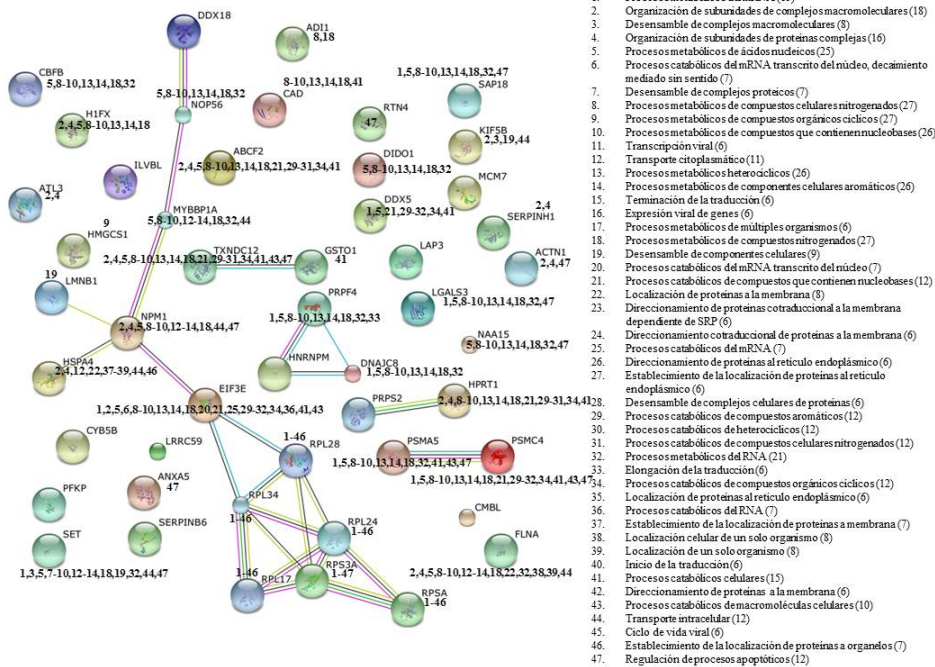


Figura 40. Interactoma de los polipéptidos subexpresados exclusivos encontrados en las líneas celulares Luminal A, Baja en claudina y HER2+ de cáncer de mama. (A) Mapa de interacción de los polipéptidos subexpresados compartidos en las líneas celulares Luminales A (MCF7 y T47D). (B) Interactoma de las proteínas subexpresadas exclusivas en la línea celular Baja en claudina (MDA-MB-231). (C) Redes de interacción de las proteínas subexpresadas específicas para la línea HER2+ (SK-BR-3). Las líneas en colores, representan las diferentes evidencias de interacción: línea roja, fusión; línea verde, vecindad; línea azul, co-ocurrencia; línea morada, evidencia experimental; línea amarilla, información en literatura; línea de color azul claro, base de datos; línea negra, co-expresión. Los números en negrita asociados a las proteínas corresponden a los diferentes procesos biológicos que tuvieron una significancia estadística (p -valor $\leq 0,05$).

CAPÍTULO VI

DISCUSIÓN, CONCLUSIONES, PERSPECTIVAS

Discusión

En este estudio, se determinaron los perfiles de expresión diferencial de cuatro líneas celulares de cáncer mama con características moleculares diferentes: MCF7 y T47D (Luminal A), MDA-MB-231 (Baja Claudina) y SK-BR-3 (HER2 +) (Holliday, 2011). Estas líneas celulares se compararon con la línea celular no tumorigénica MCF 10A. Para esto, se estandarizó el método de extracción de proteínas, comparando tres protocolos: Tritón-NP40, Ripa-Urea e hidrocloreuro de guanidina. Los extractos de proteínas obtenidos en cada uno de estos métodos de extracción se incubaron con benzonaza para la eliminación de ácidos nucleicos y se deslipidaron con una precipitación de Metanol-Cloroformo. Los patrones electroforéticos de los tres métodos de extracción fueron comparados (Figura 14). Los resultados mostraron un mayor número de bandas y mejor definición de las mismas en el método de extracción con hidrocloreuro de guanidina (Figura 14 y 15). Durante la verificación de la eliminación del ADN en geles de agarosa, se observó que la enzima no digería el ADN (Figura 16) y que esto no era efecto de un inadecuado funcionamiento de la enzima (Figura 17). Al parecer la enzima no tolera concentraciones de urea altas. Aunado a esto, se observó que la presencia de ácidos nucleicos en las muestras no afecta el corrimiento electroforético (Figura 14 carriles 2-4); además estos tampoco interfieren en el análisis de espectrometría de masas, por lo que se determinó que la eliminación de ácidos nucleicos no es un paso indispensable en la extracción de proteínas. Sin embargo, se decidió complementar el método de extracción de hidrocloreuro de guanidina con la adición del detergente RapiGest (0.1%), debido a que este ayuda a solubilizar proteínas de membrana y es compatible con espectrometría de masas. El resultado de la comparación de ambos métodos arrojó patrones electroforéticos muy parecidos; sin embargo, se observa un ligero enriquecimiento en bandas de peso molecular alto en la extracción con GuHCl-RapiGest (Figura 18). En cuanto al análisis densitométrico de algunas de las bandas de proteína, no mostraron gran diferencia (Figura 19). Por lo que se determinó que ambos métodos son eficaces; sin embargo, se optó por GuHCl-RapiGest como

método de extracción. Una vez elegido el método de extracción, las muestras se digirieron, marcaron, prefraccionaron a través de IEF y fase reversa y se analizaron por espectrometría de masas en tándem. Se identificaron 1,020 polipéptidos marcados con iTRAQ con al menos un péptido identificado con más del 95% de confianza. De este universo, se encontraron 78 proteínas sobreexpresadas y 128 polipéptidos subexpresados en común para todas las líneas celulares de cáncer de mama. Estas proteínas se analizaron con el programa IPA en el módulo filtro de biomarcadores para la selección de posibles marcadores en cáncer de mama. A través de IPA, se establecieron 3 filtros de biomarcadores: Biomarcadores I (todos los tejidos, células primarias y todas las líneas celulares), Biomarcadores II (glándula mamaria y líneas celulares de cáncer de mama) y Biomarcadores III (todas las líneas de cáncer excluyendo las de cáncer de mama). El análisis de Biomarcadores I fue el único en donde se identificaron proteínas exclusivas (Figura 32). Algunas de las proteínas seleccionadas en este filtro han sido detectadas en tejidos como corazón, pulmón y riñón, así como en líneas celulares derivadas de sistema nervioso central, colon, mama, riñón, pulmón, mieloma, ovario y próstata, entre otras. Además, ya se han desarrollado algunas drogas inhibitoras específicas para algunas de estas proteínas como: CPT1A (perhexilina), DNMT1 (5-azacytidina, decitabina), HSP90AA1 (17-dimethylaminoethylamino-17-demethoxygeldanamycina, IPI-504, cisplatino), NNMT (ácido nicotínico). Cabe destacar que no se encontraron proteínas exclusivas pertenecientes al filtro de Biomarcadores II y que algunos de los probables biomarcadores propuestos por IPA se han estudiado ampliamente. Por esta razón, se propuso como objetivo encontrar proteínas que no han sido previamente reportados en cáncer de mama a través de una búsqueda en la base de datos PubMed. En este análisis, se encontraron 35 proteínas sobreexpresadas y 53 subexpresadas (Figura 32). De estas proteínas, se seleccionó a BAG6 y DDX39 (proteínas sobreexpresadas) y a ANXA8 y COX4 (proteínas subexpresadas) para su validación por Western blot (Figura 33).

La proteína BCL2-asociada al Athanogeno 6 (BAG6) es una proteína rica en prolina de 1132 residuos en longitud, la cual contiene en el N-terminal un dominio parecido a ubiquitina, dos dominios ricos en prolina, un dominio de función desconocida (DUF3538), un sitio de escisión a caspasa 3, una señal de localización nuclear y un dominio BAG conservado en el C-terminal (Kumar *et al.*, 2004; Desmots *et al.*, 2005). Diversos reportes, han sugerido un papel de BAG6 (también llamada Scythe/BAT3) en la apoptosis (Kumar *et al.*, 2004; Desmots *et al.*, 2005; Tsukahara *et al.*, 2009). Se ha reportado, que la interacción del gen IEX-1 (Immediate Early gene X-1) con BAG6 y otras proteínas modulan la apoptosis y el crecimiento celular (Kumar *et al.*, 2004). La inactivación de BAG6 en ratones mostró fuertes defectos en el desarrollo del pulmón, riñón y cerebro, conduciendo a una letalidad embrionaria. Estos defectos se relacionaron con un incremento de la apoptosis y la proliferación celular (Desmots *et al.*, 2005). Un incremento en el nivel de PBF (Papiloma Binding Factor) ha sido asociado con un pobre pronóstico en osteosarcoma. El complejo BAG6-PBF inhibe la apoptosis a través de una vía independiente de la caspasa-9. Interesantemente, BAG6 fue localizada en el citoplasma de tejido normal en contraste con su colocalización nuclear con PBF en células de osteosarcoma (Tsukahara *et al.*, 2009). Otro papel asociado a la proteína BAG6 es la progresión del ciclo celular, en donde BAG6 es capaz de encender y apagar los papeles pro y antiproliferativos de p21 durante el ciclo celular, regulando así su actividad, abundancia y localización intracelular para facilitar la transición de G1/S y la progresión de G2/M (Yong *et al.*, 2012). Por otra parte, en ausencia de BAG6, el complejo p300-p53 no se forma y la acetilación de p53 a través de p300 es abatida, afectando así las funciones de los genes blanco de p53, incluyendo p21 y PUMA (Sasaki *et al.*, 2007). BAG6 también ha sido relacionado con la inmunovigilancia de las células tumorales a través de su liberación en respuesta al estrés celular. BAG6 interactúa con el receptor NKp30, presente en las células NK a través de un sitio de unión localizado en los residuos 686-936 de BAG6. Este nuevo dominio de BAG6 es suficiente para el acoplamiento del receptor NKp30 y la inhibición de la citotoxicidad de las células NK (Pogge *et al.*, 2007; Imhadri *et al.*, 2008; Binici *et al.*, 2013).

DDX39 pertenece a una familia de helicasas de RNA y es altamente regulada en células malignas pancreáticas y mesoteliales pleurales (Kuramitsu, Tominaga *et al.*, 2013; Kuramitsu, Suenaga *et al.*, 2013), así como en células escamosas de cáncer de pulmón y en cáncer de vejiga urinaria (Sugiura T, Nagano *et al.*, 2007; Kato *et al.*, 2012). Por otro lado, se ha reportado, que DDX39 interactúa con componentes de la maquinaria de transcripción y splicing como son ALY, CIP29 y FUS/TLS (Sugiura T, Sakurai *et al.*, 2007). Se ha encontrado que la sobreexpresión de DDX39 acelera la proliferación celular en células HeLa, lo que sugiere un papel importante de esta proteína en la patogénesis del cáncer (Sugiura T, Nagano *et al.*, 2007). Sin embargo, DDX39 también puede tener un papel como gen supresor de tumores. Cuando, DDX39 fue eliminada en células de cáncer de vejiga, se observó un aumento en la capacidad de invasión en estas células en comparación con las células control (Kato *et al.*, 2012). Por otro lado, estudios en pacientes han demostrado que la expresión de DDX39 es inversamente proporcional a la tasa de supervivencia libre de enfermedad en tumores gastrointestinales estromales, indicando un mal pronóstico en pacientes con un nivel elevado de DDX39 (Kikuta *et al.*, 2012; Kubota *et al.*, 2012). Finalmente, cabe mencionar, que DDX39 interactúa con hTERT y la subunidad TRF2 de shelterina, donde TRF2 ha sido implicada en la protección de los extremos de los cromosomas (Yoo *et al.*, 2011).

ANXA8 pertenece a la familia de proteínas de unión a fosfolípidos dependientes de Ca^{2+} . Las proteínas ANXA pueden unirse a algunos fosfolípidos de membrana y formar redes de interacción, a través de las cuales, las anexinas pueden funcionar como organizadores de dominios de membrana, así como andamios para el reclutamiento de proteínas (Gerke *et al.*, 2005). En células pancreáticas humanas, se han encontrado altos niveles de RNAm de la proteína ANXA8 (Karanjawala *et al.*, 2008) y se ha observado que la inhibición de esta proteína a través de siRNA, reduce el número de células migrantes y también disminuye los niveles de RNAm de HIF-1 α (Hata *et al.*, 2014). Sin embargo, estos hallazgos contrastan con nuestros resultados,

los cuales demostraron que la proteína ANXA8 se encuentra subexpresada en líneas celulares de cáncer de mama (Figura 30c). De acuerdo con nuestros resultados, un análisis inmunohistoquímico reveló que ANXA8 es expresada en bajos niveles en el epitelio de lóbulos y ductos, en algunas células basales, así como en algunos tejidos mamarios humanos normales. Sin embargo, se encontró una sobreexpresión de ANXA8 en una baja frecuencia en las células tumorales (Stein *et al.*, 2005).

La citocromo c oxidasa (COX4 o COX4i1) es la enzima terminal de la cadena respiratoria mitocondrial. Es un complejo conformado por 13 subunidades y su función principal es acoplar la transferencia de electrones del citocromo c al oxígeno molecular. La regulación de la citocromo c oxidasa subunidad IV isoforma 1 (COX4i1) ha sido asociada a p73, un miembro de la familia de p53. El control de la expresión del gen COX4i1 se basa en dos promotores, lo cual permite la expresión de TAp73 y Δ Np73, conteniendo o no, el dominio de transactivación, respectivamente. Por lo tanto, solo TAp73 tiene la capacidad de transactivar COX4i1. Se ha reportado que ratones deficientes de TAp73, los cuales tuvieron una reducción en los niveles de expresión COX4i1, mostraron una disfunción mitocondrial y un incremento en los niveles de ROS, conduciendo a un envejecimiento prematuro (Rufini *et al.*, 2012). El papel de COX4i1 en enfermedades, ha sido principalmente asociado a trastornos metabólicos como la diabetes tipo 2 y el cáncer. En ratones transgénicos MKR que desarrollan diabetes tipo 2, se ha observado una baja regulación de múltiples proteínas implicadas en el metabolismo oxidativo en islotes pancreáticos en comparación con los ratones control no diabéticos (Lu *et al.*, 2008). Algunas de las proteínas afectadas son: COX2, COX4i1 y COX5A, las cuales forman parte del complejo IV. En cáncer, se han encontrado que los mRNAs de COX4i1, COX5B y COX7C son altamente expresados, principalmente en el carcinoma cutáneo escamoso celular (Dang *et al.*, 2006). Por otro lado, la subunidad COX4i1 ha sido encontrada disminuida en respuesta a los tratamientos anti-tumorales de carcinoma adrenocortical y otros tumores sólidos mediante el uso de mitotano y gemcitabina

(van de Wiel *et al.*, 2005). Otro estudio demostró que las células T47D exhiben una notablemente mayor capacidad de reserva mitocondrial en comparación con las células MCF7 en ausencia de hormonas en suero, las cuales aumentaban después del tratamiento con 4-hidroxitamoxifeno. Bajo estas condiciones, se observó un bajo nivel de COX4i1 (Radde *et al.*, 2014).

Hasta donde sabemos, no hay estudios que muestren una relación de BAG6 y DDX39 con cáncer de mama. En el caso de ANXA8, esta fue mencionada sólo en un artículo hasta ahora (Stein *et al.*, 2005). En este reporte, así como en nuestro trabajo, ANXA8 es propuesta como parte de un panel de marcadores en cáncer de mama. En cuanto a COX4, recientemente se observó que su expresión es sensible a estrógenos en células T47D (Radde *et al.*, 2014), pero no ha sido reportada como un posible biomarcador en cáncer de mama. En resumen, proponemos a BAG6, DDX39, ANXA8 y COX4 como posibles biomarcadores en cáncer de mama. Sin embargo, es necesaria la validación de estos polipéptidos en una amplia cohorte de tumores y suero de pacientes con cáncer de mama. No descartamos las otras moléculas identificadas en nuestro estudio, pero es necesaria su validación.

Nuestro análisis mostró de manera global que algunas de las proteínas identificadas están involucradas en procesos metabólicos. Las proteínas sobreexpresadas estuvieron implicadas en diversos procesos, como son: metabolismo de azúcar y almidón, biosíntesis de aminoácidos, vía de las pentosas, metabolismo de purinas, glicólisis, entre otros (Figura 27). Por otro lado, la mayoría de los polipéptidos subexpresados estuvieron implicados en diferentes procesos mitocondriales. El ciclo del ácido tricarboxílico (TCA) genera productos vitales como NADH y FADH para la producción de energía a través de la fosforilación oxidativa. En nuestro análisis, encontramos diversos polipéptidos subexpresados que son críticos para la generación de NADH y FADH (Figura 27). Se determinó que 24

polipéptidos subexpresados forman parte de los complejos I, II, III, IV y V de la cadena respiratoria (Figura 28). En consecuencia, suponemos que la fosforilación oxidativa está gravemente afectada. En resumen, los procesos alterados están relacionados con el efecto Warburg, que se caracteriza por la generación de ATP a través de la glucólisis aeróbica en lugar de la fosforilación oxidativa (Warburg, 1956; Vander *et al.*, 2009).

Todos los análisis en este estudio estuvieron basados en las proteínas sobreexpresadas y subexpresadas comunes en todas las líneas celulares de cáncer de mama. Sin embargo, se identificaron conjuntos de polipéptidos que son únicos para cada una de las líneas celulares de cáncer de mama (Figura 34, 35). Estos resultados mostraron diferencias importantes en la expresión de proteínas entre las líneas celulares. La similitud entre algunos tejidos tumorales y sus correspondientes líneas celulares han sido establecidas a través de los perfiles de expresión de genes. El índice de similitud de tejido (TSI) mide la similitud en la expresión de genes entre las líneas celulares y tejidos tumorales. Una puntuación alta en el TSI, indica una estrecha similitud en los niveles de expresión génica. Las líneas celulares MCF7 y T47D poseen un TSI que va desde una puntuación media a una puntuación alta, mientras que la línea MDA-MB-231 tiene una puntaje alto (Sandberg y Ernberg, 2005). Por ejemplo, los grupos de polipéptidos únicos encontrados en cada tipo de línea celular de cáncer de mama pueden ser utilizados para la identificación de las diferentes clases moleculares de tumores (Figuras 34, 35, 37-40). Sin embargo, se requieren estudios exhaustivos con tejidos tumorales para validar estas suposiciones.

La línea celular de cáncer de mama MDA-MB-231 (Baja en Claudina), se caracteriza por el inmunoperfil ER⁻PR⁻HER2⁻, así como por la pérdida de algunas de las proteínas de uniones estrechas (E-caderina, claudina-3, claudina-4 y claudina-7);

además, posee un bajo nivel de expresión de Ki67 (Holliday y Speirs *et al.*, 2011). Adicionalmente, esta línea tiene una respuesta intermedia a la quimioterapia (Sturgeon *et al.*, 2008) y es una de las líneas celulares de cáncer de mama con mayor potencial de invasión (Neve *et al.*, 2006). En relación con los tumores de clase molecular Baja en claudina, estos tienen un mayor riesgo de mortalidad (Parise *et al.*, 2014) y han mostrado un enriquecimiento de marcadores relacionados con la transición epitelio-mesénquima, genes de respuesta inmune, características de células madre (Sabatier *et al.*, 2014), además de una mayor actividad en varias vías de señalización, incluyendo EGFR, SRC y TGF (Lehmann *et al.*, 2011). En este trabajo, encontramos a las proteínas sobreexpresadas CLIC4, CRK y DEK solamente en la línea celular MDA-MB-231 (Figura 38b). CLIC4 (proteína del canal intracelular de cloruro 4) está implicada en varios procesos celulares, como son: mantenimiento de las membranas intracelulares, apoptosis, diferenciación celular, inflamación y angiogénesis (Tung *et al.*, 2009; Littler *et al.*, 2010; Padmakumar *et al.*, 2014). La reducción en los niveles de expresión de CLIC4, inhibió el crecimiento de células de osteosarcoma humano *in vitro* e *in vivo*, incrementó la apoptosis y disminuyó la proliferación celular. Además, CLIC4 fue identificada en células parecidas a células madre metastásicas de cáncer colorrectal y fue asociada con un mal pronóstico (Deng *et al.*, 2014). CRK (molécula adaptadora crk) desempeña un papel clave en múltiples vías de señalización y ha sido relacionada con procesos como: proliferación celular, adhesión, migración, invasión, fagocitosis y supervivencia (Feller, 2001; Fathers *et al.*, 2012). La sobreexpresión de los polipéptidos CrkI, CrkII, CrkL se ha observada en diferentes tipos de cáncer. En cáncer de mama, se ha determinado una asociación significativa entre la expresión de la proteína CRK y tumores con proliferación alta (Fathers *et al.*, 2012). Además, se ha demostrado que las proteínas CRK son importantes para la adhesión celular, migración, invasión y proliferación de tumores de mama y en las líneas celulares basales MDA-MB-231, 1833TR y SUM1315 (Tung *et al.*, 2009). La proteína DEK está implicada en la replicación del ADN (Kappes *et al.*, 2011), en la integridad global de la heterocromatina (Alexiadis *et al.*, 2000) entre otros procesos. DEK se encuentra sobreexpresada en

adenocarcinoma gástrico (Piao *et al.*, 2014) y en cáncer de pulmón de células pequeñas (Wang *et al.*, 2014). En cáncer de mama, esta proteína se sobreexpresa en carcinoma ductal invasivo, donde los altos niveles de DEK correlacionan con un incremento en la proliferación (Liu *et al.*, 2012).

La línea celular de cáncer de mama SK-BR-3, sobreexpresa HER2 y es ER⁻ PR⁻, tiene un potencial invasivo bajo y es sensible a trastuzumab y a quimioterapia. Sin embargo, los tumores que presentan un fenotipo similar a las células que sobreexpresan HER2⁺ (ER⁻ PR⁻ HER2⁺) han demostrado un alto riesgo de mortalidad (Sabatier *et al.*, 2014). En este estudio, se identificaron algunas proteínas sobreexpresadas exclusivamente en la línea celular SK-BR-3 (Figura 38c), incluyendo a SLC27A4, DECR1, DERL1, HEXB y TPD52. La proteína SLC27A4 (transportadora de ácidos grasos de cadenas largas 4) ha sido relacionada en la translocación de cadenas de ácidos grasos largas a través de la membrana plasmática y se ha encontrado aumentada en cáncer de mama (Luo *et al.*, 2013). DECR1 (2,4-dienoil-CoA reductasa, mitocondrial) posee un papel dual en células positivas a ErbB2; disminuye la proliferación de células de cáncer de mama HER2⁺, lo cual puede proveer protección contra la apoptosis debido a la restricción de glucosa (Buzzai *et al.*, 2005). Por otro lado, DERL1 se sobreexpresa, cuando células de cáncer de mama están bajo estrés del retículo endoplásmico, lo que resulta en una resistencia a la apoptosis (Wang *et al.*, 2008). Con lo que respecta a HEXB (Beta-hexosamidasa subunidad beta), ha sido relacionada con un efecto antiangiogénico (Jones *et al.*, 2012). Finalmente, TPD52 se coexpresa en cánceres de mama Her2/Neu. Esta proteína está altamente regulada en líneas celulares de cáncer de mama positivas a ErbB2, lo cual contribuye con la supervivencia de las células de cáncer de mama (Byrne *et al.*, 1995; Wilson *et al.*, 2002; Roslan *et al.*, 2014)

Las líneas celulares de cáncer de mama Luminal A (MCF7 y T47D), poseen una morfología diferenciada y forman uniones estrechas célula-célula, a menudo responden a la quimioterapia y tienen un bajo potencial de invasión (Sandberg *et al.*, 2005; Holliday y Speirs, 2011). Los tumores que tienen características moleculares similares a las células de cáncer de mama Luminal A (ER⁺ PR⁺ HER2⁻) han demostrado tener un menor riesgo de mortalidad (Herschkowitz *et al.*, 2007). Estas líneas celulares también mostraron un grupo específico de polipéptidos sobreexpresados (Figura 38a), los cuales incluyen: UBQLN1, YBX1 y RUVBL2, entre otros. La pérdida de UBQLN1 (ubiquilina-1) resultó en un incremento en la migración celular en líneas celulares de cáncer de pulmón de células no pequeñas, una significativa subexpresión de E-caderina y marcadores epiteliales como claudina-1, así como la sobreexpresión de marcadores mesenquimales: vimentina, snail y ZEV1 (Shah *et al.*, 2014). Por lo tanto, una sobreexpresión de UBQLN1 podría ser un indicativo de un bajo potencial de invasividad de las líneas celulares Luminales A. En el caso de RUVBL2 (parecido a RuvB 2), la cual tiene un papel en la remodelación de la cromatina y la reparación de daño en el ADN. El silenciamiento de RUVBL2 a través de siRNA, causa sensibilidad a los fármacos contra el cáncer (Miyamoto-Matsubara *et al.*, 2014). RUVBL2 coopera con Ets-2 en la regulación transcripcional positiva de la telomerasa transcriptasa inversa humana (Flavin *et al.*, 2011). YBX1 (proteína de unión a la caja Y o YB-1), un miembro de la familia de proteínas de choque frío, es considerada un tipo de proteína lanzadera citoplasma-nuclear (Bargou *et al.*, 1997). Este polipéptido, está implicado en un gran número de funciones, incluyendo la regulación transcripcional y traduccional (Matsumoto *et al.*, 1998; Kohno *et al.*, 2003), crecimiento celular, empalme del RNA, reparación del ADN, proliferación, y resistencia de drogas (Eliseeva *et al.*, 2011; Lasham *et al.*, 2013). Se ha demostrado que la expresión de YBX1 en células de cáncer de próstata conduce a una transición epitelio-mesénquima, así como a una migración e invasión. Por lo tanto, la morfología mesenquimal se correlaciona con una baja regulación de marcadores epiteliales (Khan *et al.*, 2014). YB-1 está asociada con la progresión del cáncer y con un mal pronóstico para el paciente. La detección de un fragmento de

YBX1 de 18 kDa en plasma fue correlacionada con diferentes tumores malignos (Tacke *et al.*, 2014). En cáncer de mama, este fragmento se ha encontrado en un 70% de las muestras analizadas (Tacke *et al.*, 2014). Además, la expresión de YBX1 en células epiteliales mamarias humanas se asoció con una autorrenovación y transformación celular causada por una serie de eventos epigenéticos. La coexpresión de YBX1, RSK2 y hTERT en ratones, condujo a la formación de tumores con características parecidas a los cánceres tipo basales (Davies *et al.*, 2014)

Finalmente, como parte complementaria a este trabajo se decidió realizar una búsqueda de los diferentes receptores y proteínas utilizadas para la caracterización de los diferentes tipos de tumores (Tabla 6). También, se buscaron algunos genes frecuentemente alterados en cáncer en la lista de las 1020 proteínas identificadas en este trabajo (Tabla 7).

Tabla 6. Expresión diferencial de las proteínas comúnmente utilizadas para la caracterización de diferentes tipos de tumores

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
Estrogen receptor	ESR1 Sinónimos: ESR, NR3A1	P03372					
Progesterone receptor	PGR Sinónimo: NR3C3	P06401					
Receptor tyrosine-protein kinase erbB-2	ERBB2 Sinónimos: HER2, MLN19, NEU, NGL	P04626	2	1.0507	1.0641	7.4113	1.1734
Antigen KI-67	MKI67	P46013					
Epidermal growth factor receptor	EGFR Sinónimos: ERBB, ERBB1, HER1	P00533	1	0.0988	0.3827	0.2843	0.3998
Keratin, type II cytoskeletal 5	KRT5	P13647	32	0.1099	0.0584	0.0684	0.0590
Keratin, type II cytoskeletal 6A	KRT6A Sinónimos: K6A, KRT6D	P02538	34	0.0549	0.0406	0.0393	0.0356
Cadherin-1	CDH1 Sinónimos: CDHE, UVO	P12830					
Claudin-3	CLDN3 Sinónimos: C7orf1, CPETR2	O15551					
Claudin-4	CLDN4 Sinónimos: CPER, CPETR1, WBSR8	O14493					
Claudin-7	CLDN7 Sinónimos: CEPTRL2, CPETRL2	O95471					

^a Número de acceso en la base de datos de UniProt; ^b P95%, Numero de péptidos con al menos 95% de confianza en su identificación; ^{c,d,e,f} Expresión diferencial entre las diferentes líneas celulares de cáncer de mama y la línea control (113): ^c 114:113, MCF7/MCF 10A; ^d 115:113, MDA-MB-231/MCF 10A; ^e 116:113, SK-BR-3/MCF 10A; ^f 117:113, T47D/MCF 10A.

Tabla 7. Expresión diferencial de proteínas frecuentemente alteradas en cáncer

Factores de transcripción

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
Proto-oncogene c-Fos	FOS Sinónimo: G0S7	P01100					
Transcription factor AP-1	JUN	P05412					
Signal transducer and activator of transcription 1-alpha/beta	STAT1	P42224					
Transcription factor Sp1	SP1 Sinónimo: TSFP1	P08047					

Oncogenes

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
Myc proto-oncogene protein	MYC Sinónimo: BHLHE39	P01106					
GTPase KRas	KRAS Sinónimos: KRAS2, RASK2	P01116					
Epidermal growth factor receptor	EGFR Sinónimos: ERBB, ERBB1, HER1	P00533	1	0.0988	0.3827	0.2843	0.3998
E3 ubiquitin-protein ligase Mdm2	MDM2	Q00987					
Transcription factor E2F1	E2F1 Sinónimo: RBBP3	Q01094					
Vascular endothelial growth factor A	VEGFA Sinónimo: VEGF	P15692					
Cyclin-dependent kinase 4	CDK4	P11802					
Transcriptional activator Myb	MYB	P10242					
Receptor tyrosine-protein kinase erbB-2	ERBB2 Sinónimos: HER2, MLN19, NEU, NGL	P04626	2	1.0507	1.0641	7.4113	1.1734

Proto-oncogene tyrosine-protein kinase Src	SRC Sinónimo: SRC1	P12931					
GTPase HRas	HRAS Sinónimo: HRAS1	P01112					
Tyrosine-protein kinase ABL1	ABL1 Sinónimos: ABL, JTK7	P00519					
Protein Mdm4	MDM4 Sinónimo: MDMX	O15151					
RAC-alpha serine/threonine-protein kinase	AKT1 Sinónimos: PKB, RAC	P31749					
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	PIK3CA	P42336					
G1/S-specific cyclin-D1	CCND1 Sinónimos: BCL1, PRAD1	P24385					
G1/S-specific cyclin-E1	CCNE1 Sinónimo: CCNE	P24864					
Cyclin-dependent kinase 6	CDK6 Sinónimo: CDKN6	Q00534					
Fibroblast growth factor 10	FGF10	O15520					

Supresores de tumores

Proteínas	Genes	No. Acceso ^a	P95% ^b	114:113 ^c	115:113 ^d	116:113 ^e	117:113 ^f
Cellular tumor antigen p53	TP53 Sinónimo: P53	P04637					
Retinoblastoma-associated protein	RB1	P06400					
Transforming growth factor beta-1	TGFB1 Sinónimo: TGFB	P01137					
Von Hippel-Lindau disease tumor suppressor	VHL	P40337					
Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	PTEN Sinónimos: MMAC1, TEP1	P60484					

Cyclin-dependent kinase inhibitor 1 (p21)	CDKN1A Sinónimos: CAP20, CDKN1, CIP1, MDA6, PIC1, SDI1, WAF1	P38936					
Cyclin-dependent kinase inhibitor 2A (p16)	CDKN2A Sinónimos: CDKN2, MTS1	P42771					
Tumor suppressor ARF (p14 ARF)	CDKN2A Sinónimos: CDKN2, MLM	Q8N726					
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB Sinónimos: SDH, SDH1	P21912					
Adenomatous polyposis coli protein	APC Sinónimo: DP25	P25054					
Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	PTEN Sinónimos: MMAC1, TEP1	P60484					
Serine/threonine-protein kinase STK11	STK11 Sinónimos: LKB1, PJS	Q15831					

Apoptosis

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
Apoptosis regulator BAX	BAX Sinónimo: BCL2L4	Q07812					
Bcl-2 homologous antagonist/killer	BAK1 Sinónimos: BAK, BCL2L7, CDN1	Q16611					
BH3-interacting domain death agonist	BID	P55957	1				
Apoptotic protease-activating factor 1	APAF1 Sinónimo: KIAA0413	O14727					
Bcl-2-binding component 3	BBC3 Sinónimo: PUMA	Q9BXH1					
Phorbol-12-myristate-13-acetate-induced protein 1	PMAIP1 Sinónimo: NOXA	Q13794					
Apoptosis regulator Bcl-2	BCL2	P10415					

Tumor necrosis factor receptor superfamily member 6	FAS Sinónimos: APT1, FAS1, TNFRSF6	P25445					
FAS-associated death domain protein	FADD Sinónimo: MORT1	Q13158					
Caspase-8	CASP8 Sinónimo: MCH5	Q14790					
AT-rich interactive domain-containing protein 1A	ARID1A Sinónimos: BAF250, BAF250A, C1orf4, OSA1, SMARCF1	O14497					
Chromobox protein homolog 1	CBX1 Sinónimo: CBX	P83916	2	1.9012	1.4085	1.1433	1.1608

Invasión y metástasis

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
DNA replication licensing factor MCM6	MCM6	Q14566	1	2.6627	1.5206	0.9820	2.1418
Matrix metalloproteinase-9	MMP9 Sinónimo: CLG4B	P14780					
Ras-related protein Rab-5B	RAB5B	P61020	3	0.7464	0.6784	0.7042	0.9366
Endothelial cell-specific molecule 1	ESM1	Q9NQ30					
Vascular endothelial growth factor receptor 1	FLT1 Sinónimos: FLT, FRT, VEGFR1	P17948					
Vascular endothelial growth factor A	VEGFA Sinónimo: VEGF	P15692					

Metabolismo

Proteínas	Genes	No. Acceso^a	P95%^b	114:113^c	115:113^d	116:113^e	117:113^f
Fructose-2,6-bisphosphatase TIGAR	TIGAR Sinónimo: C12orf5	Q9NQ88					
Glutaminase liver isoform, mitochondrial	GLS2 Sinónimo: GA	Q9UI32					

Solute carrier family 2, facilitated glucose transporter member 3	SLC2A3 Sinónimo: GLUT3	P11169					
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1 Sinónimo: GLUT1	P11166	2	1.8310	1.7542	1.2457	1.2433
Glucose-6-phosphate 1-dehydrogenase	G6PD	P11413	12	6.7806	0.8464	4.2398	0.9151
Transketolase-like protein 1	TKTL1 Sinónimos: TKR, TKT2	P51854					
Glucose-6-phosphate isomerase	GPI	P06744	13	1.6900	1.1883	2.1083	1.6969
ATP-citrate synthase	ACLY	P53396	15	1.0810	1.5113	1.5649	1.7948
Acetyl-CoA carboxylase 1	ACACA Sinónimos: ACAC, ACC1, ACCA	Q13085					
Tumor necrosis factor receptor superfamily member 6	FAS Sinónimos: APT1, FAS1, TNFRSF6	P25445					
Ribonucleoside-diphosphate reductase subunit M2	RRM2 Sinónimo: RR2	P31350					
Ribonucleoside-diphosphate reductase subunit M2 B	RRM2B Sinónimo: P53R2	Q7LG56					
Thymidylate synthase	TYMS Sinónimo: TS	P04818					
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB	P21912					
Succinate dehydrogenase cytochrome b560 subunit, mitochondrial	SDHC Sinónimos: CYB560, SDH3	Q99643					
Succinate dehydrogenase [ubiquinone] cytochrome b small subunit, mitochondrial	SDHD Sinónimo: SDH4	O14521					
Fumarate hydratase, mitochondrial	FH	P07954	5	0.1974	0.2816	0.3784	0.2277
Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1 Sinónimo: PICD	O75874	4	1.5671	1.0143	1.9709	2.1627

^a Número de acceso en la base de datos de UniProt; ^b P95%, Numero de péptidos con al menos 95% de confianza en su identificación; ^{c,d,e,f} Expresión diferencial entre las diferentes líneas celulares de cáncer de mama y la línea control (113): ^c 114:113, MCF7/MCF 10A; ^d 115:113, MDA-MB-231/MCF 10A; ^e 116:113, SK-BR-3/MCF 10A; ^f 117:113, T47D/MCF 10A.

Conclusiones

1. Se identificaron 1020 proteínas correctamente etiquetadas, de las cuales 78 (sobrexpresadas) y 128 (subexpresadas) mostraron una expresión diferencial en todas las líneas celulares de cáncer de mama.
2. Se identificaron los diferentes procesos en los que están involucradas las proteínas sobre y subexpresadas en las líneas celulares de cáncer de mama, siendo los procesos metabólicos los más afectados en ambos casos.
3. Se encontraron proteínas sobreexpresadas involucradas en la topología del ADN, vía de las pentosas, regulación del inicio de la traducción, proteínas implicadas en la degradación por el proteosoma, corte y empalme, y glicólisis en común en todas las líneas celulares de cáncer de mama analizadas. En el caso de los polipéptidos subexpresados, estos estuvieron implicados en diferentes procesos mitocondriales, siendo el más afectado la fosforilación oxidativa.
4. Se propone a las proteínas BAG6, DDX39, ANXA8 y COX4 como probables biomarcadores en cáncer de mama. Sin embargo, se requiere de una validación exhaustiva de estos en tejidos tumorales y/o en suero de pacientes con cáncer de mama.
5. Se identificaron paneles de biomarcadores exclusivos para cada una de las líneas celulares: Luminal A, baja en claudina y HER2+.

6. Se determinaron los procesos biológicos en los que están involucradas tanto las proteínas sobre y subexpresadas en cada una de las líneas células. De igual manera, la expresión de proteínas relacionadas a procesos metabólicos se vio fuertemente afectada.

7. Se identificaron redes de interacción para los polipéptidos sobreexpresados en cada una de las líneas celulares: Luminal A (proteínas en común en MCF7 y T47D), procesamiento del RNAm. Baja en claudina, ensamble de complejos de ribonucleoproteínas, formación del complejo de inicio de la traducción, regulación del inicio de la traducción. HER2+, procesos metabólicos de pequeñas moléculas.

8. Se identificaron redes de interacción para los polipéptidos subexpresados en cada una de las líneas celulares. Luminal A (proteínas en común en MCF7 y T47D), no presento nodo de interacción significativo. Baja en claudina, transporte intracelular de proteínas y organización de membrana entre otros. HER2+, iogenesis del ribosoma, entre otros.

Perspectivas

1. Analizar la expresión diferencial de BAG6, DDX39, ANXA8 y COX4 en las líneas celulares de estadios tempranos HCC 1395 (estadio I) y HCC 1937 (estadio II) de cáncer de mama, en las líneas cancerosas (MCF7, T47D, MDA-MB-231 y SK-BR-3) así como en las líneas celulares control MCF 10A (línea proveniente de fibrosis) y MCF 12F (mama normal) a través de citometría de flujo y Western blot.
2. Determinar la localización de las proteínas BAG6, DDX39, ANXA8 y COX4 en las diferentes líneas celulares de cáncer de mama, así como en las líneas control a través de microscopia confocal.
3. Determinar la expresión diferencial de estas proteínas en tejidos tumorales y/o en suero de pacientes con cáncer de mama, así como en pacientes sanos.
4. Inhibir la expresión de algunos de estos marcadores identificados en las líneas celulares, mediante siRNA, para determinar su importancia en la biología de las células cancerosas.
5. Validar algunas proteínas sobre y subexpresadas en cada una de las líneas celulares tipo Luminal A, Baja en claudina y HER2+, en líneas celulares así como en tejidos tumorales y/o en suero de pacientes con cáncer de mama a través de citometría de flujo, Western blot y microscopia confocal.

6. Obtener proteínas de membrana plasmática y núcleo de las líneas celulares de cáncer de mama de estadios tempranos y tardíos, así como de las líneas celulares control para estudiar marcadores específicos de estos organelos.

7. Realizar análisis de expresión diferencial tanto de las proteínas de membrana plasmática y núcleo (digestión, marcaje, prefraccionamiento y análisis por espectrometría de masas).

8. Validar algunas de las proteínas diferencialmente expresadas en membrana plasmática y núcleo a través de citometría de flujo, Western blot y microscopía confocal.

CAPÍTULO VII

REFERENCIAS

Referencias

Aebersold R, Mann M: Mass spectrometry-based proteomics. *Nature* 2003, 422(6928):198-207.

Aggarwal K, Choe LH, Lee KH: Shotgun proteomics using the iTRAQ isobaric tags. *Briefings in functional genomics & proteomics* 2006, 5(2):112-120.

AJCC Cancer Staging Manual, 6th edition. In. Edited by Greene FL, D.L. Page, I.D. Fleming, A.G. Fritz, C.M. Balch, D.G. Haller, M. Morrow 6th edn. New York; 2002.

Alexiadis V, Waldmann T, Andersen J, Mann M, Knippers R, Gruss C: The protein encoded by the proto-oncogene DEK changes the topology of chromatin and reduces the efficiency of DNA replication in a chromatin-specific manner. *Genes & development* 2000, 14(11):1308-1312.

Allred DC, Mohsin SK, Fuqua SA: Histological and biological evolution of human premalignant breast disease. *Endocrine-related cancer* 2001, 8(1):47-61.

Banerjee S, Mazumdar S: Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry* 2012, 2012:282574.

Bargou RC, Jurchott K, Wagener C, Bergmann S, Metzner S, Bommert K, Mapara MY, Winzer KJ, Dietel M, Dorken B et al: Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. *Nature medicine* 1997, 3(4):447-450.

Bick U, Diekmann F: Digital mammography: what do we and what don't we know? *European radiology* 2007, 17(8):1931-1942.

Binici J, Hartmann J, Herrmann J, Schreiber C, Beyer S, Guler G, Vogel V, Tumulka F, Abele R, Mantele W et al: A soluble fragment of the tumor antigen BCL2-associated athanogene 6 (BAG-6) is essential and sufficient for inhibition of NKp30

receptor-dependent cytotoxicity of natural killer cells. *The Journal of biological chemistry* 2013, 288(48):34295-34303.

Biomarkers Definitions Working G: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology and therapeutics* 2001, 69(3):89-95.

Bogdanov B, Smith RD: Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass spectrometry reviews* 2005, 24(2):168-200.

Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 1976, 72:248-254.

Buzzai M, Bauer DE, Jones RG, Deberardinis RJ, Hatzivassiliou G, Elstrom RL, Thompson CB: The glucose dependence of Akt-transformed cells can be reversed by pharmacologic activation of fatty acid beta-oxidation. *Oncogene* 2005, 24(26):4165-4173.

Byrne JA, Tomasetto C, Garnier JM, Rouyer N, Mattei MG, Bellocq JP, Rio MC, Basset P: A screening method to identify genes commonly overexpressed in carcinomas and the identification of a novel complementary DNA sequence. *Cancer research* 1995, 55(13):2896-2903.

Cesar Bazán-Méndez (tesis de maestría, 2010)

Chandramouli K, Qian PY: Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics: HGP* 2009, 2009.

Cichon MA, Degnim AC, Visscher DW, Radisky DC: Microenvironmental influences that drive progression from benign breast disease to invasive breast cancer. *Journal of mammary gland biology and neoplasia* 2010, 15(4):389-397.

Croce CM: Oncogenes and Cancer. New England Journal of Medicine 2008, 358(5):502-511.

Dang C, Gottschling M, Manning K, O'Curraín E, Schneider S, Sterry W, Stockfleth E, Nindl I: Identification of dysregulated genes in cutaneous squamous cell carcinoma. Oncology reports 2006, 16(3):513-519.

Davies AH, Reipas KM, Pambid MR, Berns R, Stratford AL, Fotovati A, Firmino N, Astanehe A, Hu K, Maxwell C et al: YB-1 transforms human mammary epithelial cells through chromatin remodeling leading to the development of basal-like breast cancer. Stem cells 2014, 32(6):1437-1450.

De Carolis E, Vella A, Vaccaro L, Torelli R, Posteraro P, Ricciardi W, Sanguinetti M, Posteraro B: Development and validation of an in-house database for matrix-assisted laser desorption ionization-time of flight mass spectrometry-based yeast identification using a fast protein extraction procedure. Journal of clinical microbiology 2014, 52(5):1453-1458.

Deng YJ, Tang N, Liu C, Zhang JY, An SL, Peng YL, Ma LL, Li GQ, Jiang Q, Hu CT et al: CLIC4, ERp29, and Smac/DIABLO derived from metastatic cancer stem-like cells stratify prognostic risks of colorectal cancer. Clinical cancer research: an official journal of the American Association for Cancer Research 2014, 20(14):3809-3817.

Desmots F, Russell HR, Lee Y, Boyd K, McKinnon PJ: The reaper-binding protein scythe modulates apoptosis and proliferation during mammalian development. Molecular and cellular biology 2005, 25(23):10329-10337.

Diez O, Gutierrez-Enriquez S, Ramon y Cajal T: [Breast cancer susceptibility genes]. Medicina clinica 2006, 126(8):304-310.

Domon B, Aebersold R: Mass spectrometry and protein analysis. Science 2006, 312(5771):212-217.

Dowsett M, Dunbier AK: Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research* 2008, 14(24):8019-8026.

Duffy MJ, Evoy D, McDermott EW: CA 15-3: uses and limitation as a biomarker for breast cancer. *Clinica chimica acta; international journal of clinical chemistry* 2010, 411(23-24):1869-1874.

Duffy MJ: Serum tumor markers in breast cancer: are they of clinical value? *Clinical chemistry* 2006, 52(3):345-351.

El-Aneed AC, A; Banoub, J.: Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *J APPLIED SPECTROSCOPY REVIEWS* 2009, 44(3):210-230.

Eliseeva IA, Kim ER, Guryanov SG, Ovchinnikov LP, Lyabin DN: Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry Biokhimiia* 2011, 76(13):1402-1433.

Elliott MH, Smith DS, Parker CE, Borchers C: Current trends in quantitative proteomics. *Journal of mass spectrometry: JMS* 2009, 44(12):1637-1660.

Fathers KE, Bell ES, Rajadurai CV, Cory S, Zhao H, Mourskaia A, Zuo D, Madore J, Monast A, Mes-Masson AM et al: Crk adaptor proteins act as key signaling integrators for breast tumorigenesis. *Breast cancer research: BCR* 2012, 14(3):R74.

Feller SM: Crk family adaptors-signalling complex formation and biological roles. *Oncogene* 2001, 20(44):6348-6371.

Flavin P, Redmond A, McBryan J, Cocchiglia S, Tibbitts P, Fahy-Browne P, Kay E, Treumann A, Perrem K, Mcllroy M et al: RuvBl2 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS letters* 2011, 585(15):2537-2544.

Freudenberg JA, Wang Q, Katsumata M, Drebin J, Nagatomo I, Greene MI: The role of HER2 in early breast cancer metastasis and the origins of resistance to HER2-targeted therapies. *Experimental and molecular pathology* 2009, 87(1):1-11.

Gerke V, Creutz CE, Moss SE: Annexins: linking Ca²⁺ signalling to membrane dynamics. *Nature reviews Molecular cell biology* 2005, 6(6):449-461.

Gingras AC, Aebersold R, Raught B: Advances in protein complex analysis using mass spectrometry. *The Journal of physiology* 2005, 563(Pt 1):11-21.

Gingras AC, Gstaiger M, Raught B, Aebersold R: Analysis of protein complexes using mass spectrometry. *Nature reviews Molecular cell biology* 2007, 8(8):645-654.

Goncalves A, Bertucci F: Clinical application of proteomics in breast cancer: state of the art and perspectives. *Medical principles and practice: international journal of the Kuwait University, Health Science Centre* 2011, 20(1):4-18.

Gouw JW, Krijgsveld J, Heck AJ: Quantitative proteomics by metabolic labeling of model organisms. *Molecular & cellular proteomics: MCP* 2010, 9(1):11-24.

Griffiths WJ, Jonsson AP, Liu S, Rai DK, Wang Y: Electrospray and tandem mass spectrometry in biochemistry. *The Biochemical journal* 2001, 355(Pt 3):545-561.

Guerrera IC, Kleiner O: Application of mass spectrometry in proteomics. *Bioscience reports* 2005, 25(1-2):71-93.

Hamdan MH. *Cancer Biomarkers: Analytical Techniques for Discovery*. Edited by Nibbering M, Desiderio M. Wiley-Interscience Series in Mass Spectrometry. New Jersey, 2007.

Han X, Aslanian A, Yates JR, 3rd: Mass spectrometry for proteomics. *Current opinion in chemical biology* 2008, 12(5):483-490.

Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. *Cell* 2011, 144(5):646-674.

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC, Jr., American Society of Clinical O: American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 2007, 25(33):5287-5312.

Hartge P: Genes, hormones, and pathways to breast cancer. *The New England journal of medicine* 2003, 348(23):2352-2354.

Hata H, Tatemichi M, Nakadate T: Involvement of annexin A8 in the properties of pancreatic cancer. *Molecular carcinogenesis* 2014, 53(3):181-191.

Henry NL, Hayes DF: Cancer biomarkers. *Molecular oncology* 2012, 6(2):140-146.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S et al: Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology* 2007, 8(5):R76.

Ho CS, Lam CW, Chan MH, Cheung RC, Law LK, Lit LC, Ng KF, Suen MW, Tai HL: Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical biochemist Reviews / Australian Association of Clinical Biochemists* 2003, 24(1):3-12.

Holliday DL, Speirs V: Choosing the right cell line for breast cancer research. *Breast cancer research: BCR* 2011, 13(4):215.

Hondermarck H: Breast cancer: when proteomics challenges biological complexity. *Molecular & cellular proteomics : MCP* 2003, 2(5):281-291.

Horn DM, Ge Y, McLafferty FW: Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Analytical chemistry* 2000, 72(20):4778-4784.

Imyanitov EN, Hanson KP: Mechanisms of breast cancer. *Drug Discovery Today: Disease Mechanisms* 2004, 1(2):235-245.

Jones DT, Lechertier T, Mitter R, Herbert JM, Bicknell R, Jones JL, Li JL, Buffa F, Harris AL, Hovalala-Dilke K: Gene expression analysis in human breast cancer associated blood vessels. *PloS one* 2012, 7(10):e44294.

Kappes F, Waldmann T, Mathew V, Yu J, Zhang L, Khodadoust MS, Chinnaiyan AM, Luger K, Erhardt S, Schneider R et al: The DEK oncoprotein is a Su(var) that is essential to heterochromatin integrity. *Genes & development* 2011, 25(7):673-678.

Karanjawala ZE, Illei PB, Ashfaq R, Infante JR, Murphy K, Pandey A, Schulick R, Winter J, Sharma R, Maitra A et al: New markers of pancreatic cancer identified through differential gene expression analyses: claudin 18 and annexin A8. *The American journal of surgical pathology* 2008, 32(2):188-196.

Karellas A, Vedantham S: Breast cancer imaging: a perspective for the next decade. *Medical physics* 2008, 35(11):4878-4897.

Kato M, Wei M, Yamano S, Kakehashi A, Tamada S, Nakatani T, Wanibuchi H: DDX39 acts as a suppressor of invasion for bladder cancer. *Cancer science* 2012, 103(7):1363-1369.

Kelleher NL: Top-down proteomics. *Analytical chemistry* 2004, 76(11):197A-203A.

Khan MI, Adhami VM, Lall RK, Sechi M, Joshi DC, Haidar OM, Syed DN, Siddiqui IA, Chiu SY, Mukhtar H: YB-1 expression promotes epithelial-to-mesenchymal transition in prostate cancer that is inhibited by a small molecule fisetin. *Oncotarget* 2014, 5(9):2462-2474.

Kikuta K, Kubota D, Saito T, Orita H, Yoshida A, Tsuda H, Suehara Y, Katai H, Shimada Y, Toyama Y et al: Clinical proteomics identified ATP-dependent RNA helicase DDX39 as a novel biomarker to predict poor prognosis of patients with gastrointestinal stromal tumor. *Journal of proteomics* 2012, 75(4):1089-1098.

Kim DH, Bae J, Lee JW, Kim SY, Kim YH, Bae JY, Yi JK, Yu MH, Noh DY, Lee C: Proteomic analysis of breast cancer tissue reveals upregulation of actin-remodeling proteins and its relevance to cancer invasiveness. *Proteomics Clinical applications* 2009, 3(1):30-40.

Kim S, Ruparel HD, Gilliam TC, Ju J: Digital genotyping using molecular affinity and mass spectrometry. *Nature reviews Genetics* 2003, 4(12):1001-1008.

Kohno K, Izumi H, Uchiumi T, Ashizuka M, Kuwano M: The pleiotropic functions of the Y-box-binding protein, YB-1. *BioEssays: news and reviews in molecular, cellular and developmental biology* 2003, 25(7):691-698.

Kubota D, Okubo T, Saito T, Suehara Y, Yoshida A, Kikuta K, Tsuda H, Katai H, Shimada Y, Kaneko K et al: Validation study on p16 and ATP-dependent RNA helicase DDX39 as prognostic biomarkers in gastrointestinal stromal tumour. *Japanese journal of clinical oncology* 2012, 42(8):730-741.

Kumar R, Lutz W, Frank E, Im HJ: Immediate early gene X-1 interacts with proteins that modulate apoptosis. *Biochemical and biophysical research communications* 2004, 323(4):1293-1298.

Kumar S, Mohan A, Guleria R: Biomarkers in cancer screening, research and detection: present and future: a review. *Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals* 2006, 11(5):385-405.

Kuramitsu Y, Suenaga S, Wang Y, Tokuda K, Kitagawa T, Tanaka T, Akada J, Maehara S, Maehara Y, Nakamura K: Up-regulation of DDX39 in human pancreatic cancer cells with acquired gemcitabine resistance compared to gemcitabine-sensitive parental cells. *Anticancer research* 2013, 33(8):3133-3136.

Kuramitsu Y, Tominaga W, Baron B, Tokuda K, Wang Y, Kitagawa T, Nakamura K: Up-regulation of DDX39 in human malignant pleural mesothelioma cell lines

compared to normal pleural mesothelial cells. *Anticancer research* 2013, 33(6):2557-2560.

Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, 227(5259):680-685.

Lai TC, Chou HC, Chen YW, Lee TR, Chan HT, Shen HH, Lee WT, Lin ST, Lu YC, Wu CL et al: Secretomic and proteomic analysis of potential breast cancer markers by two-dimensional differential gel electrophoresis. *Journal of proteome research* 2010, 9(3):1302-1322.

Lamb R, Harrison H, Clarke RB: Mammary development, carcinomas and progesterone: role of Wnt signalling. *Ernst Schering Foundation symposium proceedings* 2007(1):1-23.

Lane CS: Mass spectrometry-based proteomics in the life sciences. *Cellular and molecular life sciences: CMLS* 2005, 62(7-8):848-869.

Lasham A, Print CG, Woolley AG, Dunn SE, Braithwaite AW: YB-1: oncoprotein, prognostic marker and therapeutic target? *The Biochemical journal* 2013, 449(1):11-23.

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietersen JA: Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation* 2011, 121(7):2750-2767.

Leth-Larsen R, Lund R, Hansen HV, Laenkholm AV, Tarin D, Jensen ON, Ditzel HJ: Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Molecular & cellular proteomics: MCP* 2009, 8(6):1436-1449.

Liang X, Huuskonen J, Hajivandi M, Manzanedo R, Predki P, Amshey JR, Pope RM: Identification and quantification of proteins differentially secreted by a pair of normal and malignant breast-cancer cell lines. *Proteomics* 2009, 9(1):182-193.

Littler DR, Harrop SJ, Goodchild SC, Phang JM, Mynott AV, Jiang L, Valenzuela SM, Mazzanti M, Brown LJ, Breit SN et al: The enigma of the CLIC proteins: Ion channels, redox proteins, enzymes, scaffolding proteins? *FEBS letters* 2010, 584(10):2093-2101.

Liu S, Wang X, Sun F, Kong J, Li Z, Lin Z: DEK overexpression is correlated with the clinical features of breast cancer. *Pathology international* 2012, 62(3):176-181.

Loboda AV, Krutchinsky AN, Bromirski M, Ens W, Standing KG: A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid communications in mass spectrometry: RCM* 2000, 14(12):1047-1057.

Lu H, Yang Y, Allister EM, Wijesekara N, Wheeler MB: The identification of potential factors associated with the development of type 2 diabetes: a quantitative proteomics approach. *Molecular & cellular proteomics: MCP* 2008, 7(8):1434-1451.

Ludwig JA, Weinstein JN: Biomarkers in cancer staging, prognosis and treatment selection. *Nature reviews Cancer* 2005, 5(11):845-856.

Luo Y, Yang C, Ye M, Jin C, Abbruzzese JL, Lee MH, Yeung SC, McKeehan WL: Deficiency of metabolic regulator FGFR4 delays breast cancer progression through systemic and microenvironmental metabolic alterations. *Cancer & metabolism* 2013, 1(1):21.

Mann M: Functional and quantitative proteomics using SILAC. *Nature reviews Molecular cell biology* 2006, 7(12):952-958.

Marouga R, David S, Hawkins E: The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and bioanalytical chemistry* 2005, 382(3):669-678.

Matsumoto K, Wolffe AP: Gene regulation by Y-box proteins: coupling control of transcription and translation. *Trends in cell biology* 1998, 8(8):318-323.

McPherson K, Steel CM, Dixon JM: ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *Bmj* 2000, 321(7261):624-628.

Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL: Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nature biotechnology* 2001, 19(10):952-957.

Miyamoto-Matsubara M HY, Ono K, Xie MZ, Salem AMH, Shoukamy MI, Nakano T, Ide H : Depletion of RUVBL2 in human cells confers moderate sensitivity to anticancer agents. *J Cancer Sci Ther* 2014, 6:440-445.

Morrow M, Waters J, Morris E: MRI for breast cancer screening, diagnosis, and treatment. *Lancet* 2011, 378(9805):1804-1811.

Mourelatos Z: Small RNAs: The seeds of silence. *Nature* 2008, 455(7209):44-45.

Narod SA: Modifiers of risk of hereditary breast and ovarian cancer. *Nature reviews Cancer* 2002, 2(2):113-123.

Neil L. Kelleher HYL, Gary A. Valaskovic , David J. Aaserud , Einar K. Fridriksson, and Fred W. McLafferty: Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J Am Chem Soc* 1999, 121(4):806–812.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F et al: A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer cell* 2006, 10(6):515-527.

Normanno N, De Luca A, Carotenuto P, Lamura L, Oliva I, D'Alessio A: Prognostic applications of gene expression signatures in breast cancer. *Oncology* 2009, 77 Suppl 1:2-8.

Oda Y, Nagasu T, Chait BT: Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature biotechnology* 2001, 19(4):379-382.

Padmakumar V, Masiuk KE, Luger D, Lee C, Coppola V, Tessarollo L, Hoover SB, Karavanova I, Buonanno A, Simpson RM et al: Detection of differential fetal and adult expression of chloride intracellular channel 4 (CLIC4) protein by analysis of a green fluorescent protein knock-in mouse line. *BMC developmental biology* 2014, 14:24.

Parise CA, Caggiano V: Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *Journal of cancer epidemiology* 2014, 2014:469251.

Paul D, Kumar A, Gajbhiye A, Santra MK, Srikanth R: Mass spectrometry-based proteomics in molecular diagnostics: discovery of cancer biomarkers using tissue culture. *BioMed research international* 2013, 2013:783131.

Payne SJ, Bowen RL, Jones JL, Wells CA: Predictive markers in breast cancer--the present. *Histopathology* 2008, 52(1):82-90.

Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y: Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute* 2001, 93(14):1054-1061.

Piao J, Shang Y, Liu S, Piao Y, Cui X, Li Y, Lin Z: High expression of DEK predicts poor prognosis of gastric adenocarcinoma. *Diagnostic pathology* 2014, 9:67.

Pogge von Strandmann E, Simhadri VR, von Tresckow B, Sasse S, Reiners KS, Hansen HP, Rothe A, Boll B, Simhadri VL, Borchmann P et al: Human leukocyte

antigen-B-associated transcript 3 is released from tumor cells and engages the NKp30 receptor on natural killer cells. *Immunity* 2007, 27(6):965-974.

Polyak K, Kalluri R: The role of the microenvironment in mammary gland development and cancer. *Cold Spring Harbor perspectives in biology* 2010, 2(11):a003244.

Prasad SN, Houserkova D: The role of various modalities in breast imaging. *Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia* 2007, 151(2):209-218.

Radde BN, Ivanova MM, Mai HX, Salabei JK, Hill BG, Klinge CM: Bioenergetic differences between MCF-7 and T47D breast cancer cells and their regulation by estradiol and tamoxifen. *The Biochemical journal* 2014.

Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997-2015.

Rebbeck TR, Lynch HT, Neuhausen SL, Narod SA, Van't Veer L, Garber JE, Evans G, Isaacs C, Daly MB, Matloff E et al: Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. *The New England journal of medicine* 2002, 346(21):1616-1622.

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K et al: ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature genetics* 2006, 38(8):873-875.

Roslan N, Bieche I, Bright RK, Lidereau R, Chen Y, Byrne JA: TPD52 represents a survival factor in ERBB2-amplified breast cancer cells. *Molecular carcinogenesis* 2014, 53(10):807-819.

Ross JS, Hatzis C, Symmans WF, Pusztai L, Hortobagyi GN: Commercialized multigene predictors of clinical outcome for breast cancer. *The oncologist* 2008, 13(5):477-493.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S et al: Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics: MCP* 2004, 3(12):1154-1169.

Rufini A, Niklison-Chirou MV, Inoue S, Tomasini R, Harris IS, Marino A, Federici M, Dinsdale D, Knight RA, Melino G et al: TAp73 depletion accelerates aging through metabolic dysregulation. *Genes & development* 2012, 26(18):2009-2014.

Sabatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, Birnbaum D, Bertucci F: Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Molecular cancer* 2014, 13:228.

Sandberg R, Ernberg I: Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(6):2052-2057.

Sasaki T, Gan EC, Wakeham A, Kornbluth S, Mak TW, Okada H: HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53. *Genes & development* 2007, 21(7):848-861.

Shevchenko A1, Wilm M, Vorm O, Mann M: Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chemical* 1996, 68(5):850-858.

Seregni E, Coli A, Mazzucca N, Italian Group Ria-Irma Test IAoNM: Circulating tumour markers in breast cancer. *European journal of nuclear medicine and molecular imaging* 2004, 31 Suppl 1:S15-22.

Shah PP, Lockwood WW, Saurabh K, Kurlawala Z, Shannon SP, Waigel S, Zacharias W, Beverly LJ: Ubiquilin1 represses migration and epithelial-to-mesenchymal transition of human non-small cell lung cancer cells. *Oncogene* 2014.

Shaheed SU, Rustogi N, Scally A, Wilson J, Thygesen H, Loizidou MA, Hadjisavvas A, Hanby A, Speirs V, Loadman P et al: Identification of stage-specific breast markers using quantitative proteomics. *Journal of proteome research* 2013, 12(12):5696-5708.

Shiio Y, Aebersold R: Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nature protocols* 2006, 1(1):139-145.

Simhadri VR, Reiners KS, Hansen HP, Topolar D, Simhadri VL, Nohroudi K, Kufer TA, Engert A, Pogge von Strandmann E: Dendritic cells release HLA-B-associated transcript-3 positive exosomes to regulate natural killer function. *PloS one* 2008, 3(10):e3377.

Simpson KL, Whetton AD, Dive C: Quantitative mass spectrometry-based techniques for clinical use: biomarker identification and quantification. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 2009, 877(13):1240-1249.

Skoog, D. A., Holler, J. H., Nieman, T. A. *Principios de Análisis Instrumental*, 5a Edición. McGraw Hill. Madrid, España. 2001.

Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR: An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2002, 2(5):513-523.

Sontag L, Axelrod DE: Evaluation of pathways for progression of heterogeneous breast tumors. *Journal of theoretical biology* 2005, 232(2):179-189.

Stein T, Price KN, Morris JS, Heath VJ, Ferrier RK, Bell AK, Pringle MA, Villadsen R, Petersen OW, Sauter G et al: Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research* 2005, 11(19 Pt 1):6872-6879.

Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brunner N, Chan DW, Babaian R, Bast RC, Jr., Dowell B, Esteva FJ et al: National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clinical chemistry* 2008, 54(12):e11-79.

Sugiura T, Nagano Y, Noguchi Y: DDX39, upregulated in lung squamous cell cancer, displays RNA helicase activities and promotes cancer cell growth. *Cancer biology & therapy* 2007, 6(6):957-964.

Sutton CW, Rustogi N, Gurkan C, Scally A, Loizidou MA, Hadjisavvas A, Kyriacou K: Quantitative proteomic profiling of matched normal and tumor breast tissues. *Journal of proteome research* 2010, 9(8):3891-3902.

Tacke F, Galm O, Kanig N, Yagmur E, Brandt S, Lindquist JA, Eberhardt CS, Raffetseder U, Mertens PR: High prevalence of Y-box protein-1/p18 fragment in plasma of patients with malignancies of different origin. *BMC cancer* 2014, 14:33.

Throck Watson, O. David Sparkman. *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*, 4th Edition. Wiley. England. 2007.

Tsukahara T, Kimura S, Ichimiya S, Torigoe T, Kawaguchi S, Wada T, Yamashita T, Sato N: Scythe/BAT3 regulates apoptotic cell death induced by papillomavirus binding factor in human osteosarcoma. *Cancer science* 2009, 100(1):47-53.

Tung JJ, Hobert O, Berryman M, Kitajewski J: Chloride intracellular channel 4 is involved in endothelial proliferation and morphogenesis in vitro. *Angiogenesis* 2009, 12(3):209-220.

Van de Wiel MA, Costa JL, Smid K, Oudejans CB, Bergman AM, Meijer GA, Peters GJ, Ylstra B: Expression microarray analysis and oligo array comparative genomic

hybridization of acquired gemcitabine resistance in mouse colon reveals selection for chromosomal aberrations. *Cancer research* 2005, 65(22):10208-10213.

Vander Heiden MG, Cantley LC, Thompson CB: Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009, 324(5930):1029-1033.

Vermeulen M, Selbach M: Quantitative proteomics: a tool to assess cell differentiation. *Current opinion in cell biology* 2009, 21(6):761-766.

Walsh T, King MC: Ten genes for inherited breast cancer. *Cancer cell* 2007, 11(2):103-105.

Walther TC, Mann M: Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology* 2010, 190(4):491-500.

Wang J, Hua H, Ran Y, Zhang H, Liu W, Yang Z, Jiang Y: Derlin-1 is overexpressed in human breast carcinoma and protects cancer cells from endoplasmic reticulum stress-induced apoptosis. *Breast cancer research : BCR* 2008, 10(1):R7.

Wang X, Lin L, Ren X, Lin Z, Li Z, Li C, Jin T: High expression of oncoprotein DEK predicts poor prognosis of small cell lung cancer. *International journal of clinical and experimental pathology* 2014, 7(8):5016-5023.

Warburg O: On the origin of cancer cells. *Science* 1956, 123(3191):309-314.

Watson CJ, Khaled WT: Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* 2008, 135(6):995-1003.

Weigelt B, Geyer FC, Reis-Filho JS: Histological types of breast cancer: how special are they? *Molecular oncology* 2010, 4(3):192-208.

Wilson KS, Roberts H, Leek R, Harris AL, Geradts J: Differential gene expression patterns in HER2/neu-positive and -negative breast cancer cell lines and tissues. *The American journal of pathology* 2002, 161(4):1171-1185.

Wysocki VH, Resing KA, Zhang Q, Cheng G: Mass spectrometry of peptides and proteins. *Methods* 2005, 35(3):211-222.

Xu X, Qiao M, Zhang Y, Jiang Y, Wei P, Yao J, Gu B, Wang Y, Lu J, Wang Z et al: Quantitative proteomics study of breast cancer cell lines isolated from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* 2010, 10(7):1374-1390.

Yang WT: Emerging techniques and molecular imaging in breast cancer. *Seminars in ultrasound, CT, and MR* 2011, 32(4):288-299.

Yates JR, 3rd: Mass spectral analysis in proteomics. *Annual review of biophysics and biomolecular structure* 2004, 33:297-316.

Yates JR, Ruse CI, Nakorchevsky A: Proteomics by mass spectrometry: approaches, advances, and applications. *Annual review of biomedical engineering* 2009, 11:49-79.

Yong ST, Wang XF: A novel, non-apoptotic role for Scythe/BAT3: a functional switch between the pro- and anti-proliferative roles of p21 during the cell cycle. *PloS one* 2012, 7(6):e38085.

Yoo HH, Chung IK: Requirement of DDX39 DEAD box RNA helicase for genome integrity and telomere protection. *Aging cell* 2011, 10(4):557-571.

Zhang H, Ge Y: Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circulation Cardiovascular genetics* 2011, 4(6):711.

Zieske LR: A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *Journal of experimental botany* 2006, 57(7):1501-1508.

Páginas de internet

<http://www.who.int/mediacentre/factsheets/fs310/es/>

http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

<http://www.inegi.org.mx/est/contenidos/proyectos/registros/vitales/mortalidad/tabulados/ConsultaMortalidad.asp>

<http://www.cancer.gov/espanol/tipos/seno/pro/tratamiento-seno-masculino-pdq>

http://globocan.iarc.fr/Pages/bar_sex_pop_sel.aspx

<http://www.breastcancer.org/symptoms/diagnosis/staging>

http://proteomics.sysu.edu.cn/htm/list_821.html

<http://bioinfogp.cnb.csic.es/tools/venny>

<http://www.pantherdb.org/>

<http://string-db.org/>

www.qiagen.com/ingenuity

<http://www.ncbi.nlm.nih.gov/pubmed>

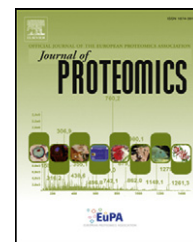
<http://www.genome.jp/kegg/>

CAPÍTULO VIII

Anexos (Publicaciones)

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry



Karla Grisel Calderón-González^a, Ma Luz Valero Rustarazo^{b,1}, Maria Luisa Labra-Barrios^c, César Isaac Bazán-Méndez^c, Alejandra Tavera-Tapia^{c,2}, Maria Esther Herrera-Aguirre^c, Manuel M. Sánchez del Pino^{b,3}, José Luis Gallegos-Pérez^d, Humberto González-Márquez^a, Jose Manuel Hernández-Hernández^c, Gloria León-Ávila^e, Sergio Rodríguez-Cuevas^f, Fernando Guisa-Hohenstein^f, Juan Pedro Luna-Arias^{c,*}

^aDoctorado en Ciencias Biológicas, División de Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Av. San Rafael Atlixco No. 186, Col. Vicentina, Iztapalapa, C.P. 09340, México, D. F., México

^bUnidad de Proteómica, Centro de Investigación Príncipe Felipe, C/Rambla del Saler 16, 46012 Valencia, España

^cDepartamento de Biología Celular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav-IPN), Av. Instituto Politécnico Nacional 2508, Col. San Pedro Zacatenco, Gustavo A. Madero, C.P. 07360, México, D. F., México

^dAB SCIE X, 500 Old Connecticut Path, Framingham, MA 01701, USA

^eDepartamento de Zoología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Prolongación de Carpio y Plan de Ayala s/n, Col. Santo Tomás, Miguel Hidalgo, C.P. 11340, México, D. F., México

^fInstituto de Enfermedades de la Mama, Fundación del Cáncer de Mama (FUCAM A.C.), Av. Bordo No. 100, Col. Viejo Ejido de Santa Ursula Coapa, Coyoacán, C.P. 04980, México, D. F., México

ARTICLE INFO

Article history:

Received 14 January 2015

Accepted 13 April 2015

Available online 24 April 2015

ABSTRACT

Breast cancer is the principal cancer in women worldwide. Although there are serum tumor markers such as CEA and HER2, they are detected in advanced stages of the disease and used as progression and recurrence markers. Therefore, there is a necessity for the identification of new markers that might lead to an early detection and also provide

* Corresponding author at: Departamento de Biología Celular, Cinvestav-IPN, Av. Instituto Politécnico Nacional No. 2508, Col. San Pedro Zacatenco, Gustavo A. Madero, C.P. 07360, México, D. F., México. Tel.: +52 55 57474016, +52 55 57473800x5567; fax: +52 55 57473393.

E-mail addresses: karla_grises@hotmail.com (K.G. Calderón-González), mluz.valero@uv.es (M.L. Valero Rustarazo), azu_car2000@yahoo.com.mx (M.L. Labra-Barrios), cesar_bazan@hotmail.com (C.I. Bazán-Méndez), taveratapia@hotmail.com (A. Tavera-Tapia), maestherha2@yahoo.com.mx (M.E. Herrera-Aguirre), sandelpi@uv.es (M.M. Sánchez del Pino), gallego7@gmail.com (J.L. Gallegos-Pérez), hgm@xanum.uam.mx (H. González-Márquez), manolo@cell.cinvestav.mx (J.M. Hernández-Hernández), leonavila60@yahoo.com.mx (G. León-Ávila), sergiorocue@gmail.com (S. Rodríguez-Cuevas), fguisa@fucam.org.mx (F. Guisa-Hohenstein), jpluna@cell.cinvestav.mx, jpluna@cinvestav.mx, jpluna2003@gmail.com (J.P. Luna-Arias).

¹ Current Address: Secció de Proteómica, Servei Central de Suport a la Investigació Experimental (SCSIE), Universitat de València, C/Doctor Moliner 50, 46100 Burjassot, València, España.

² Current Address: Centro Nacional de Investigaciones Oncológicas (CNIO), Programa de Genética del Cáncer Humano, Grupo de Genética Humana, Melchor Fernández Almagro 3, 28029 Madrid, España.

³ Current Address: Departament de Bioquímica i Biologia Molecular, Facultat de Ciències Biològiques, Universitat de València, C/ Doctor Moliner 50, 46100 Burjassot, València, España.

Keywords:

Breast cancer

Cell lines

Putative biomarkers

Quantitative proteomics

iTRAQ labelling

Tandem mass spectrometry

evidence of an effective treatment. The aim of this work was to determine the differential protein expression profiles of four breast cancer cell lines in comparison to a normal control cell line by iTRAQ labelling and tandem mass spectrometry, in order to identify putative biomarkers of the disease. We identified 1,020 iTRAQ-labelled polypeptides with at least one peptide identified with more than 95% in confidence. Overexpressed polypeptides in all cancer cell lines were 78, whilst the subexpressed were 128. We categorised them with PANTHER program into biological processes, being the metabolic pathways the most affected. We detected six groups of proteins with the STRING program involved in DNA topology, glycolysis, translation initiation, splicing, pentose pathway, and proteasome degradation. The main subexpressed protein network included mitochondrial proteins involved in oxidative phosphorylation. We propose BAG6, DDX39, ANXA8 and COX4 as putative biomarkers in breast cancer.

Biological significance

We report a set of differentially expressed proteins in the MCF7 and T47D (Luminal A), MDA-MB-231 (Claudin low) and SK-BR-3 (HER2⁺) breast cancer cell lines that have not been previously reported in breast cancer disease. From these proteins, we propose BAG6, DDX39, ANXA8 and COX4 as putative biomarkers in breast cancer. On the other hand, we propose sets of unique polypeptides in each breast cancer cell line that can be useful in the classification of different subtypes of breast cancer.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In the last twenty years, great technological advances have changed all biology search fields. These advances include high throughput technologies, such as DNA microarrays, automated DNA sequencing systems, lab-on-a-chip devices, mass spectrometry, the huge amount of information contained in genome and metagenome databases, and their easy accessibility by using personal computers through the internet. All together have raised new forefronts for searching in biology including genomics, proteomics, metabolomics, glycomics, and epigenomics, among many others. Several of these leading search fields have been involved in the discovery of new molecular biomarkers for a number of human disorders including genetic, neurodegenerative, diabetes and cancer.

Breast cancer is the most common type of cancer in women worldwide and the major cause of death by cancer in females. It has been estimated 1.67 million of new cases of breast cancer in women, corresponding to 25% of all cases of cancer diagnosed in 2012 [1]. From the total number of breast cancer deaths in women, about 60% occurred in developing countries. Current estimations for breast cancer indicate that there might be an increase of 55% and 62% in new cases and deaths worldwide by the year 2035, respectively. Moreover, estimations for this cancer in Latin America and the Caribbean indicate that the increase might be as high as 60% in the total number of new cases and approximately up to 89% of deaths by the year 2035 [1]. Therefore, breast cancer is a serious health concern in the world that needs to be diagnosed opportunely.

The diagnosis of this disease is mainly clinical examination followed by screening methods such as digital mammography, ultrasound, magnetic resonance imaging, and biopsy analysis. Some tumor markers are a complementary tool for diagnosis and prognosis. According to the Working Group and Biomarkers Consortium from the National Institutes of Health, biomarkers are molecules that are useful as indicators of normal or

pathological conditions in cells, tissues or organisms, or pharmacological responses to therapy [2]. In the case of cancer biomarkers, they are classified as: i) Diagnostic or screening biomarkers for the identification of individuals with a certain type of cancer with high specificity and sensibility; ii) Prognostic biomarkers, to predict the likely course of the disease and its recurrence, once this condition has been established; iii) Predictive biomarkers, which are principally based on mRNA expression profiles and that can be used in predicting disease outcome or the likely response to one or several drugs before starting treatment and, consequently, classifying patients as responders or non-responders to treatment [3,4]. In the case of breast cancer, some molecules have been used for the determination of recurrence of this disease, including the carcinoembryonic antigen (CEA) and carcinoma antigen 15-3 (CA 15-3) [5]. Her2/Neu/ErbB-2 has also been used. Her2/Neu is a membrane receptor characterized by three domains: the intracellular tyrosine kinase domain, the transmembrane domain and the extracellular domain (ECD). Her2-ECD can be released by shedding and consequently be found in serum. Several groups have used the presence of the Her2-ECD as marker for relapse or prognosis in breast cancer [6–9]. However, these markers are not enough to diagnose breast cancer. Therefore, there is an urgent necessity for identifying new putative biomarkers.

The mass spectrometry (MS)-based proteomics has a key role in the identification of biomarkers useful for the diagnosis and prognosis of many diseases including breast cancer. MS-based proteomics involves the separation of either intact proteins or peptides generated by trypsin treatment (or any other) using different techniques, being two-dimensional gel electrophoresis and multidimensional nano-flow HPLC the most used. Finally, peptides are identified by tandem mass spectrometry [10–12]. In the case of biomarkers, there is a necessity of quantification of the expression levels of polypeptides. Several strategies have been used in the field of breast cancer biomarkers including Stable Isotope Labelling by

Amino Acid in Cell Culture (SILAC) [13–18], 2D-Fluorescence Difference in Gel Electrophoresis (2D-DIGE) [19–21], and Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) [22–25], among others.

There are very few reports of quantitative proteomics applied to several breast cancer cell lines using iTRAQ and tandem mass spectrometry. In this paper, we determined the expression profiles of four breast cancer cell lines that are molecularly classified as Luminal A (MCF7 and T47D), Claudin-low (MDA-MB-231) and HER2⁺ (SK-BR-3) [26], in comparison to a normal cell line (MCF 10A). We identified 1,020 polypeptides labelled with the isobaric tags with at least one identified peptide with a confidence equal or higher than 95%. From this set, 206 proteins showed a change in their expression level in all breast cancer cell lines studied, 78 were overexpressed and 128 subexpressed. Additionally, we identified sets of proteins that were exclusively found in Luminal A, Claudin-low and HER2⁺ breast cancer cell lines. We discuss the pathways altered in these breast cancer cell lines and propose three panels to be used as putative biomarkers.

2. Methods

2.1. Cell lines and cell cultures

MCF7, MDA-MB-231, SK-BR-3 and T47D breast cancer cell lines, and MCF 10A control cell line were obtained from the American Type Culture Collection (Manassas, VA, USA). MCF 10A was cultured in Dulbecco's Modified Eagle Medium (DMEM)/F-12 supplemented with 10% fetal bovine serum (FBS), 10 ng/mL epidermal growth factor (EGF), 4 µg/mL insulin, 0.4 µg/mL hydrocortisone, 1X L-glutamine, and 1X penicillin-streptomycin solution. MCF7 and MDA-MB-231 cell lines were maintained in DMEM medium, while T47D was cultured in RPMI medium. The last two media were supplemented with 10% FBS, 1X sodium pyruvate, 1X L-glutamine, 1X penicillin-streptomycin, and 1X non-essential amino acids (NEAA). In addition, RPMI medium was added with 7.5% of Na₂CO₃. SK-BR-3 cell line was cultured in McCoy medium with 10% FBS, 1X L-glutamine and 1X penicillin-streptomycin solution. All cultures were incubated at 37 °C in 5% CO₂ humidified atmosphere. Cell culture media and supplements were all purchased from Invitrogen (Life Technologies, Mexico City, Mexico).

2.2. Extraction and quantification of polypeptides

Cells were grown in 27 cm² tissue culture flasks (Corning, NY, USA) until a confluence of 90% was reached. Then, culture medium was removed and discarded, and cells were washed twice in 1X phosphate buffered saline (PBS) and incubated in 5 mM EDTA in 1X PBS at 37 °C for 10 minutes. Detached cells were transferred to pre-chilled 50 mL Falcon tubes and kept in ice. EDTA incubation was repeated once and cells were centrifuged at 1,500 rpm (453 x g) for 10 minutes at 4 °C in a 5810R Eppendorf centrifuge (Germany). Cellular pellets were resuspended in 300 µL lysis buffer [6 M guanidine hydrochloride (GuHCl), 0.1% RapiGest SF surfactant (Waters, Mexico

City, Mexico) in 50 mM ammonium bicarbonate (ABC)] and sonicated three times in ice for 10 seconds at 60% amplitude in a CPX130PB Ultrasonic Processor (Cole-Parmer, Vernon Hills, IL, USA), cooling for 1 minute in ice between each sonication cycle. Then, proteins were delipidated twice with a mixture of chloroform/methanol/water (4:1:3; v:v:v), vortexed 1 minute at highest speed and centrifuged at 12,000 x g for 10 minutes at room temperature (RT) in a Heraeus Biofuge Pico (Thermo Scientific, Germany). Pellets were air-dried and solubilised in 400 µL of buffer A (8 M urea in 50 mM ABC) by sonication as described and kept at -80 °C until use. Next, proteins were precipitated with 10% trichloroacetic acid (TCA) for 30 minutes in ice, centrifuged at 12,000 x g for 5 minutes at 4 °C, washed twice with cold acetone (-20 °C) and air-dried. Finally, polypeptides were solubilised by sonication in 20 µL of a solution containing 8 M urea, 0.05% (w/v) ProteaseMAX Surfactant Trypsin Enhancer (Promega, Madison, Wisconsin, USA) and 0.5 M triethylammonium bicarbonate (TEAB) pH 8.0. Protein quantification was performed with the Bradford method [27] and samples stored at -80 °C until use.

2.3. Trypsin digestion of polypeptides and iTRAQ labelling

iTRAQ labelling was performed according to the iTRAQ Reagents – 8plex labelling manual (AB SCIEX, Madrid, Spain), with minor modifications. Briefly, before the labelling with isobaric tags, 110 µg of each sample were adjusted up to 40 µL with 0.5 M TEAB. Then, samples were reduced with 50 mM Tris-(2-carboxyethyl) phosphine (TCEP) at 37 °C for 1 hour, and alkylated with 200 mM methyl methanethiosulfonate (MMTS) at RT for 10 minutes. Urea concentration was reduced to less than 2 M with 0.5 M TEAB and polypeptides were digested with 10 µg of sequencing-grade trypsin (Promega, Madison, Wisconsin, USA) overnight at 37 °C, and dried in a Savant SpeedVac concentrator (Thermo Scientific, Germany). Labelling was performed by adding one reagent vial, containing an isobaric tag, to 110 µg of dried peptides for each cell line. The labelling reaction proceeded for 3 hours at RT after which all the samples were pooled before application of separation techniques and mass spectrometry analysis. The labelling scheme was as follows: Tag₁₁₃, MCF 10A; Tags_{114 and 118}, MCF7; Tags_{115 and 119}, MDA-MB-231; Tags_{116 and 121}, SK-BR-3; Tag₁₁₇, T47D.

2.4. Isoelectrofocusing of labelled peptides, C18 reverse phase chromatography, and MS/MS analysis

Digested and labelled peptides were solubilised in rehydration solution (7 M urea, 2 M thiourea, 2% 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 1.6% ampholytes pH 3-10, 0.002% (w/v) bromophenol blue). Isoelectrofocusing (IEF) separation procedure of pooled labelled peptides was carried out in Immobiline DryStrip gels of 13 cm in length, containing a preformed nonlinear pH 3-10 gradient on an Ettan IPGphor 3 IEF System (GE Healthcare Life Sciences, Barcelona, Spain). IEF conditions used were 75 µAmperes per strip at 20 °C, using in step 1: 5,000 Volts 30 kV hour; step 2: 500 Volts for 30 hours. Strips were washed with deionized water and divided in 12 sections (1 cm width). Peptides from each section were subsequently extracted with 50 µL of: 1) 0.1%

trifluoroacetic acid (TFA), 2) 0.1% TFA and 50% acetonitrile (ACN), and 3) 0.1% TFA in ACN, each for 10 minutes at RT, and vacuum-dried. Samples were solubilised in 0.1% TFA and cleaned with Poros R2 tips (home-made) and eluted with 40 μ L of 75% ACN and 0.1% TFA solution. Finally, samples were vacuum-concentrated to 20 μ L and volume adjusted to 125 μ L with 5% ACN and 0.1% TFA.

Before submitting samples to mass spectrometry analysis, the mass spectrometer was calibrated with 200 femtomoles of a mixture of 6 proteins previously digested with trypsin [*E. coli* β -galactosidase; *Bos taurus* serotransferrin, serum albumin, and cytochrome C; *Saccharomyces cerevisiae* alcohol dehydrogenase 1; *Gallus gallus* lysozyme C. All of them were from LC Packings (Amsterdam, NL)]. Identification of peptides for these proteins was performed with Mascot Daemon software (Matrix Science, London, UK). Once calibrated, the same conditions were used for the analysis of samples.

iTRAQ tagged peptides were loaded onto a PepMap C18 trap cartridge (300 μ m \times 5 mm, LC Packings) using as mobile phase 0.1% TFA for 10 minutes and an isocratic flow rate of 30 μ L/minute. Then, peptides were injected into a silica nanoBaume C18 reverse phase (RP) analytical chromatography column (75 μ m \times 50 cm, 5 μ m particle size, home-made) at a flow rate of 250 nL/minute. Elution of peptides was achieved using three linear gradients of ACN in 0.1% formic acid (FA) in water at a flow rate of 250 nL/minute: 1) 105 minute gradient from 1.9 to 28.5%, 2) 10 minute gradient from 28.5 to 38%, and 3) 5 min from 38 to 85%. The eluted peptides were immediately analysed in a high performance hybrid quadrupole time-of-flight mass spectrometer (QSTAR ESI XL Hybrid LC/MS/MS Mass Spectrometer System, AB SCIEX, Madrid, Spain) from the Centro de Investigación Príncipe Felipe, Valencia, Spain. The instrument was operated in IDA (Information Dependent Acquisition) mode using the Analyst QS 1.1 software (AB SCIEX, Madrid, Spain) where each 1 second MS survey scan was followed by 3 MS/MS scans of 3 seconds.

Mass spectra generated in analysis were used to identify and quantify peptides with the Paragon and ProGroup algorithms, respectively, as well as the False Discovery Rate (FDR) analysis, using the ProteinPilot software version 4.0 (AB SCIEX, Mexico City, Mexico). The database used was the UniProtKB/Swiss-Prot Human Sprot_20120120.fasta (UniProt Knowledgebase, www.uniprot.org). Parameters considered for the search were peptide label (iTRAQ 8plex), cysteine alkylation (MMTS), digestion enzyme (trypsin, allowing one missed cleavage), special factors (urea denaturation), search effort (thorough) and FDR analysis. Carbamidomethylation of Cys (C[CAM]) was used as a fixed modification; as variable modifications we selected oxidation of methionine, and deamidation of asparagine and glutamine. A tolerance on the mass measurement of 100 ppm in MS mode and \pm 0.5 Da for MS/MS ions were used in analysis. The reliability percentage of identified peptides was \geq 95% (1.30 score). In quantification settings, 113 tag was the denominator and bias correction was automatically applied. To determine significant changes in the protein expression level we performed an analysis with the PDST (ProteinPilot Descriptive Statistics Template) template v. 3.001p (AB Sciex, Mexico City, Mexico). Data were also exported to Excel for manual analysis. Proteins with a fold change greater than 1.2 were considered

overexpressed, whilst those with a fold change less than 0.8 were classified as subexpressed.

2.5. Bioinformatics analysis

Overexpressed and subexpressed polypeptides, found in common in all breast cancer cell lines and also exclusively in each cell line, were grouped in Venn-Euler diagrams using the Venny program (developed by Oliveros JC and available at <http://bioinfogp.cnb.csic.es/tools/venny/>). For the classification of identified proteins that showed changes in their expression levels according to their role in biological processes, we performed an analysis with the PANTHER classification system v 9.0 (<http://www.pantherdb.org/>) [28]. Functional interaction networks were identified with the STRING program 9.1 version, <http://string-db.org/>. Active prediction methods used in our analysis were neighbourhood, coexpression, gene fusions, experiments, co-occurrence, databases and text mining, using high confidence (0.7).

On the other hand, we searched for all the 206 differentially expressed polypeptides in common in all breast cancer cell lines in the PubMed database as well, using the recommended and alternative names assigned to each protein in UniProtKB database, and keywords such as biomarkers, cancer and breast cancer. Finally, polypeptides that showed a change in expression level were located in pathway biological maps using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>).

2.6. Western blot analysis

Proteins were electrophoresed in 12% SDS-PAGE gels and transferred to nitrocellulose membranes (GE Healthcare Life Sciences, Mexico City, Mexico). Then, membranes were blocked with 5% Skim Milk (Difco) in 0.05% (v/v) Tween 20 in 1X PBS (PBS-T) for 4 hours at RT, and incubated with 1:1,000 dilutions of primary antibodies in 5% Skim Milk in PBS-T overnight at 4 °C. The antibodies used were: anti-Ribosomal Protein S7 (E-1), anti-BAG6 (D-1), anti-COX4 (F-8), all of them mouse monoclonal antibodies; and anti-DDX39 and anti-Annexin A8 like 2 rabbit polyclonal antibodies. All experiments were performed in triplicate. Antibodies E-1, D-1 and F-8 were from Santa Cruz Biotechnology (Santa Cruz, CA, USA). Antibodies anti-DDX39 and anti-Annexin A8 like 2 were from abcam (Monterrey, NL, Mexico). Next, membranes were washed 10 times with PBS-T each for 10 minutes at RT and incubated with a 1:20,000 dilution of the corresponding secondary antibodies: anti-mouse (Zymed, Life Technologies, Mexico City, Mexico) and anti-rabbit (Zymed) polyclonal antibodies in 5% Skim Milk (Difco, BD, Biosciences, Mexico City, Mexico) in PBS-T for 2 hours at RT. After washing membranes as described, immunoreactivity was detected by chemiluminescence with a 1:8 dilution of the SuperSignal West Femto Maximum Sensitivity Substrate kit (Pierce, Mexico City, Mexico) using High Performance Chemiluminescence Films (GE Healthcare Life Sciences, Mexico City, Mexico). Digital images of western blots were obtained with the Gel Doc EZ Imager system and processed with the Image Lab software (both from Bio-Rad, Hercules, CA, USA). As

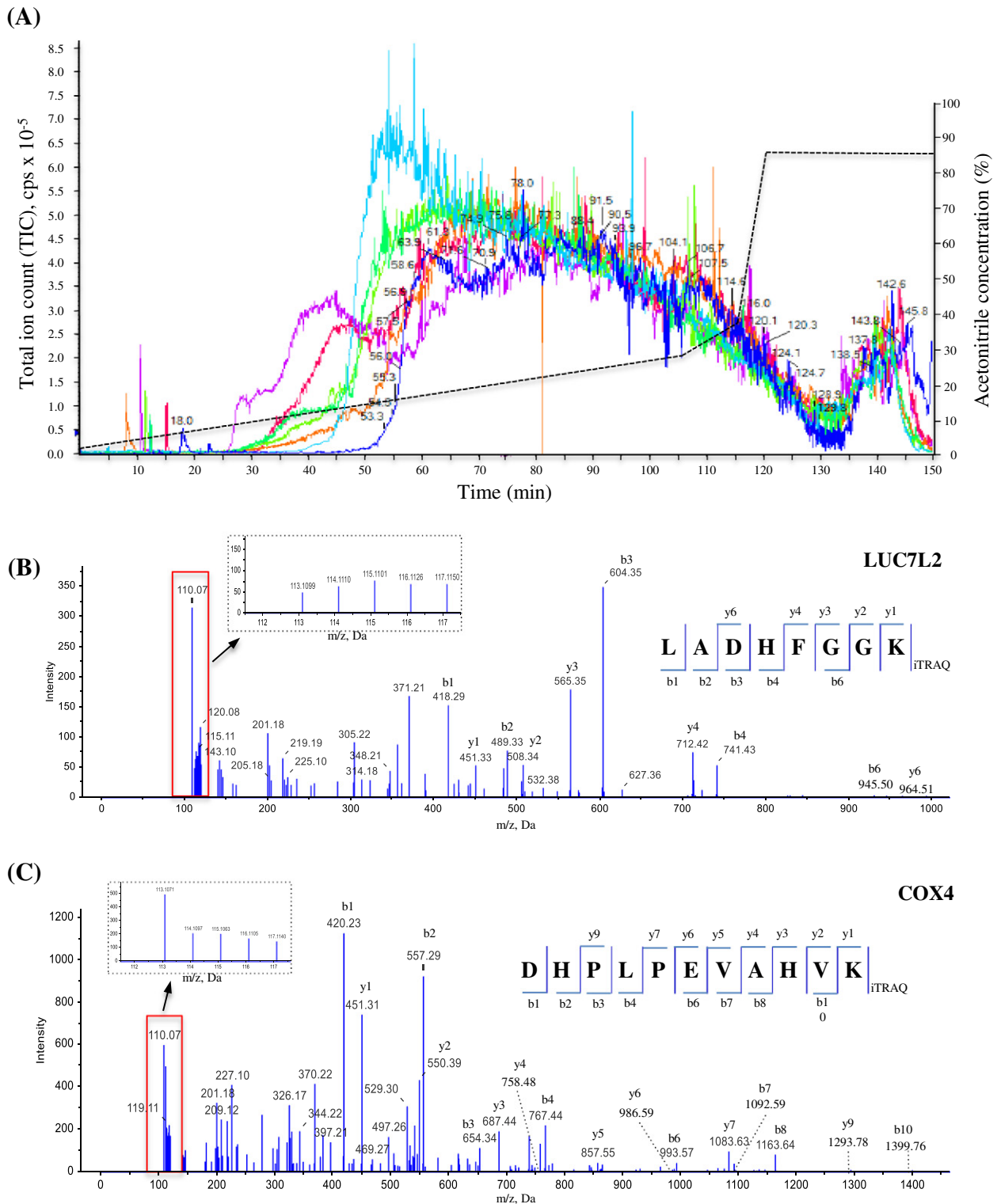


Fig. 1 – Identification and quantification of peptides from breast cancer cell lines by iTRAQ-labelling and tandem mass spectrometry. **(A)** Total ion counts as a function of the elution time from reverse phase column of sections 4-10 from the IEF strip. The different sections are indicated in colours as follow: section 4 (blue), section 5 (purple), section 6 (red), section 7 (orange), section 8 (light green), 9 (dark green) and 10 (cyan). MS/MS spectra obtained for the LADHFGGK peptide **(B)**, which belongs to the LUC7L2 polypeptide, and for the DHPLPEVAHVK peptide **(C)**, which is part of COX4. Inserts, tag relative expression levels obtained for these peptides from each cell line. Counts per second (cps). Mass-to-charge ratio (m/z). iTRAQ labels were assigned as follows: 113, MCF 10A; 114, MCF7; 115, MDA-MB-231; 116, SK-BR-3; 117, T47D. Intensity for B and C is in cps.

molecular mass markers we used Strep-tagged Precision Plus Protein All Blue Standards (Bio-Rad, Hercules, CA, USA), which were visualised by chemiluminescence with the Precision Protein StrepTactin-HRP conjugate (Bio-Rad, Hercules, CA, USA). Digital analysis of images was performed with the Quantity One software version 4.6.6 (Bio-Rad, Hercules, CA, USA).

3. Results

3.1. Differential polypeptide expression profiles of breast cancer cell lines

To determine the protein expression profiles of breast cancer cells (MCF7, MDA-MB-231, SK-BR-3 and T47D) in comparison to a normal control cell line (MCF 10A), we used the approach of iTRAQ labelling and tandem mass spectrometry as described in Methods. In the mass spectrometry analysis, most of the ions eluted between 25 and 150 minutes for each sample. We only show the elution patterns for sections 4 to 10 from the IEF strip (Fig. 1A). As examples, we include the MS/MS spectra for LADHFGGK and DHPLPEVAHVK peptides, which belong to LUC7L2 and COX4 polypeptides, respectively. We also show the relative levels of iTRAQ tags for these peptides (Fig. 1B and C). We identified 1,020 labelled proteins with at least one peptide with a minimum of 95% in confidence with the ProteinPilot software (Table 1 in Ref [29]). Peptide and protein lists were exported to PDST template to analyse the FDR and proteins plotted according to their log ratio and P-values as a volcano plot (Fig. 2). We observed a great number of polypeptides with P-values lower than 0.05. It is noteworthy that tags 114, 115 and 116 fitted very well with their replicates 118, 119 and 121, respectively.

We classified the identified proteins as overexpressed or subexpressed applying cut-off values of >1.2 or <0.8 , respectively. We found 336 overexpressed proteins in the MCF7 cell line, 326 polypeptides in T47D, 319 in MDA-MB-231, and 313 in SK-BR-3 (Fig. 3A). Remarkably, each cell line carries a set of unique polypeptides: 41, 59, 55 and 74 polypeptides for MCF7, T47D, MDA-MB-231, and SK-BR-3, respectively. Other sets are shared for different combinations of cell lines: MCF7 and T47D shared 34 proteins, T47D and MDA-MB-231 had 30 polypeptides in common, the latter and SK-BR-3 shared 24, whilst MCF7 and SK-BR-3 had 35, MCF7 and MDA-MB-231 had 35, T47D and SK-BR-3 had 19 polypeptides in common. Other sets of polypeptides are 50, 36, 20, and 27 for MCF7/T47D/MDA-MB-231, MCF7/T47D/SK-BR-3, T47D/MDA-MB-231/SK-BR-3, MCF7/MDA-MB-231/SK-BR-3, respectively. Finally, all breast cancer cell lines shared 78 polypeptides (Table 1).

In relation to the subexpressed polypeptides, we found 337 in MCF7, 321 in T47D, 310 in MDA-MB-231, and 301 in SK-BR-3 (Fig. 3B). We found specific subsets for each cell line: 39, 32, 44, and 49 polypeptides for MCF7, T47D, MDA-MB-231, and SK-BR-3, respectively. Other sets were MCF7 and T47D shared 22 proteins, T47D and MDA-MB-231 had 28 polypeptides in common, MDA-MB-231 and SK-BR-3 shared 9, MCF7 and SK-BR-3 had 33, MCF7 and MDA-MB-231 only 19, T47D and SK-BR-3 had in common 22 proteins. In addition, we found 49, 27, 13, and 20 polypeptides shared by MCF7/T47D/MDA-MB-231, MCF7/T47D/SK-BR-3, T47D/MDA-MB-231/SK-BR-3, MCF7/MDA-MB-231/SK-BR-3, respectively. Lastly, all breast cancer cell lines shared 128 subexpressed polypeptides (Table 2). For the subsequent analyses, first we will show the results obtained with the overexpressed and subexpressed polypeptides commonly found in all breast cancer cell lines. Then, we will focus on the analysis of the sets of exclusive proteins found in each type of breast cancer cell line.

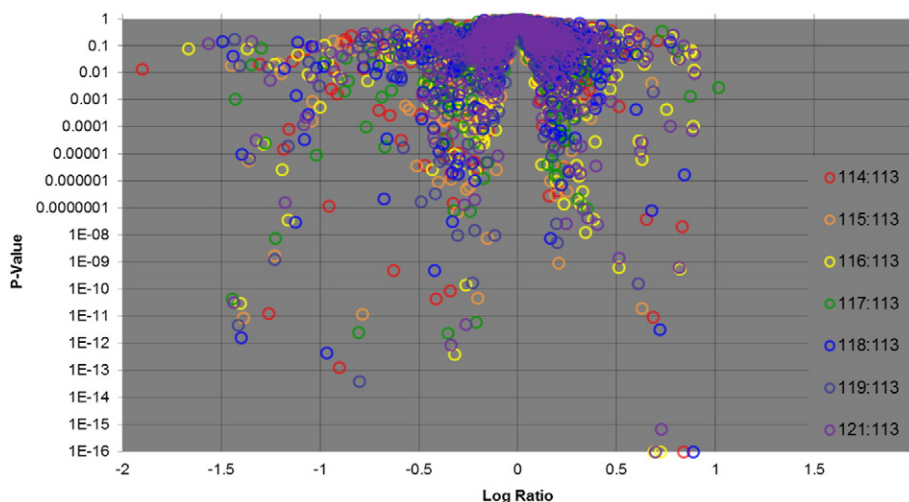


Fig. 2 – Volcano plot of proteins expressed in the breast cancer cell lines. All proteins identified harbouring an iTRAQ label were plotted. The X-axis represents the fold change of overexpressed (positive values) and subexpressed (negative values) polypeptides in breast cancer cell lines in comparison to the MCF 10A normal cell line. The Y-axis corresponds to the statistical significance (P-value) of this fold change. 113, MCF 10A; 114, MCF7; 115, MDA-MB-231; 116, SK-BR-3; 117, T47D; 118, 119 and 121 tags were duplicates of the three 114, 115 and 116 tags.

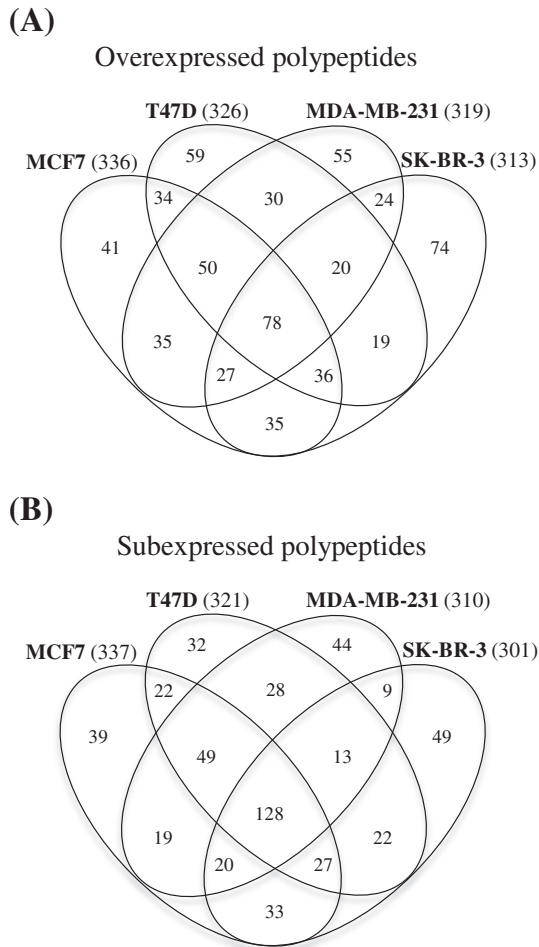


Fig. 3 – Venn-Euler diagrams of differentially expressed proteins in breast cancer cell lines. (A) Overexpressed polypeptides. (B) Subexpressed proteins. We used as cut-off values more than 1.2 for overexpressed and less than 0.8 for subexpressed polypeptides.

3.2. Functional classification of the differentially expressed proteins in common in all breast cancer cell lines

In order to obtain a global functional view of identified proteins, a classification of them with the PANTHER Classification System (version 9.0 database) was performed. The 78 overexpressed polypeptides were classified in 21 classes according to the Gene Ontology (GO) Protein Class, corresponding to 89 class hits (Fig. 4A). The main class was nucleic acid binding class (XII, with 25 polypeptides). When overexpressed proteins were grouped according to GO Biological Process, we identified 11 classes which included 132 class hits, being the class VIII the most relevant, corresponding to metabolic process (Fig. 4B). In relation to the 128 subexpressed polypeptides, the PANTHER system classified them according to the GO Protein Class in 24 groups containing 148 class hits. The most relevant classes were XV (oxidoreductase), VIII (hydrolase) and XIV (nucleic acid binding) (Fig. 4C). The classification of subexpressed proteins according to the GO Biological Process showed that they were grouped in 12 classes containing 178

class hits, being class IX (metabolic process) the main class (Fig. 4D).

3.3. Protein-protein interaction networks of the differentially expressed proteins found in common in all breast cancer cell lines

To observe the putative network of protein-protein interactions between polypeptides that changed in their expression levels in all breast cancer cell lines, we determined networks with the STRING software. In the overexpressed protein network we identified main nodes of proteins according to their reported functions: 1) DNA topology, 2) Pentose pathway, 3) Regulation of translation initiation, 4) Proteins involved in the degradation by proteasome, 5) Splicing, and 6) Glycolysis (Fig. 5). In relation to the subexpressed proteins, the network was characterized by two main nodes of proteins involved in different processes in mitochondrion: 1) Cellular respiration 2) Mitochondrial transport, and 3) Oxidative phosphorylation (Fig. 6). Polypeptides involved in the biological processes that are or not forming nodes, but having statistical significance (P -value ≤ 0.05), are shown as well.

Moreover, overexpressed and subexpressed polypeptides were submitted for their analysis to the KEGG database to localise them in the biochemical pathways. In this analysis, the most representative affected processes in the breast cancer cell lines were metabolic pathways (Fig. 7). Some of the represented pathways are metabolism of amino acids (orange), energy metabolism (purple), carbohydrate metabolism (dark blue), metabolism of cofactors and vitamins (pink), nucleotide metabolism (light red), lipids metabolism (mint green), glycan biosynthesis and metabolism (light blue), among others. Overexpressed proteins are shown in red whilst subexpressed proteins in bright green. We found 51 overexpressed proteins altered in metabolic pathways, including AGL (starch and sucrose metabolism), GALE (galactose metabolism and amino sugar and nucleotide sugar metabolism), PGLS (carbon metabolism and pentose phosphate pathway), PGD (carbon metabolism, pentose phosphate pathway and glutathione metabolism), PRPS1 (carbon metabolism, biosynthesis of amino acids, pentose phosphate pathway, purine metabolism), TPI1 (carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis, fructose and mannose metabolism and inositol phosphate metabolism), GAPDH (carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis), PTGES3 (arachidonic acid metabolism) and SEPHS1 (selenocompound metabolism). On the other hand, identified subexpressed proteins include ALDH7A1 (biosynthesis of amino acids, glycolysis/gluconeogenesis, ascorbate and aldarate metabolism, pyruvate metabolism, fatty acid degradation, glycerolipid metabolism, glycine, serine, threonine, arginine, proline, histidine, tryptophan and beta-alanine metabolism, valine, leucine, isoleucine and lysine degradation, lysine biosynthesis); Ak2 (purine metabolism), POLR2L (purine and pyrimidine metabolism), CKMT1A (arginine and proline metabolism), DNMT1 (cysteine and methionine metabolism), GLS (alanine, aspartate and glutamate metabolism, arginine and proline metabolism, D-glutamine and D-glutamate metabolism), NNMT (nicotinate and nicotinamide metabolism); HADHA (carbon metabolism, fatty acid metabolism, propanoate

Table 1 – Overexpressed polypeptides found in common in all breast cancer cell lines by iTRAQ-labelling and tandem mass spectrometry.

N	Protein name	Gene	Location	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (>95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
1	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	C	P04406	35.92	8.58	43.39	58.2	50	1.36	1.46	2.07	1.64
2	Heat shock protein HSP 90-alpha	HSP90AA1	C	P07900	84.53	4.94	32.98	28.8	37	1.71	1.69	1.56	2.02
3	Triosephosphate isomerase	TPI1	C	P60174	30.79	5.65	27.59	45.8	27	1.48	1.49	1.70	1.70
4	Cofilin-1	CFL1	N	P23528	18.37	8.26	13.75	53.6	8	2.20	1.75	1.56	1.45
5	Annexin A6	ANXA6	PM	P08133	75.74	5.41	13.08	12.5	8	1.57	4.75	3.14	2.28
6	Rab GDP dissociation inhibitor beta	GDI2	C	P50395	50.66	6.10	11.37	20.5	8	1.51	1.50	1.85	1.55
7	T-complex protein 1 subunit delta	CCT4	C	P50991	57.79	8.13	11.25	14.5	6	1.38	1.25	1.22	1.54
8	ATP-dependent RNA helicase DDX39A	DDX39A	N	O00148	49.00	5.46	11.19	12.9	8	2.29	3.25	1.27	1.41
9	Proteasome activator complex subunit 1	PSME1	C	Q06323	28.72	5.78	10.12	28.9	6	2.14	1.27	2.91	2.17
10	Eukaryotic translation initiation factor 5	EIF5	C	P55010	49.22	5.41	9.58	11.4	6	1.43	1.28	1.50	1.94
11	Phosphatidylethanolamine-binding protein 1	PEBP1	C	P30086	20.93	7.43	8.78	25.1	4	1.66	1.31	1.24	1.66
12	High mobility group protein B1	HMGB1	N	P09429	24.76	5.60	8.57	17.2	5	3.25	1.89	1.98	1.93
13	DNA topoisomerase 2-alpha	TOP2A	N	P11388	174.39	8.82	8.26	2.9	4	1.25	1.41	1.68	1.46
14	Putative RNA-binding protein Luc7-like 2	LUC7L2	Other	Q9Y383	46.51	10.02	8	8.7	5	1.54	1.63	1.35	1.54
15	6-phosphogluconate dehydrogenase, decarboxylating	PGD	C	P52209	53.01	6.88	7.62	7.0	3	1.92	1.36	1.91	1.89
16	Lactoylglutathione lyase	GLO1	C	Q04760	20.65	5.12	7.58	18.5	4	1.48	1.84	1.47	2.27
17	ATP-dependent RNA helicase DDX1	DDX1	N	Q92499	82.43	6.80	7.55	8.2	6	1.60	1.96	1.37	1.31
18	Exportin-1	XPO1	N	O14980	123.39	5.71	6.86	4.9	4	1.75	1.46	1.67	2.01
19	Proteasome activator complex subunit 2	PSME2	C	Q9UL46	27.27	5.54	6.46	16.3	4	1.63	1.41	1.84	1.43
20	Coatomer subunit gamma	COPG1	C	Q9Y678	97.72	5.32	5.7	3.5	3	1.56	1.38	1.64	1.27
21	Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	PM	P11166	54.08	8.93	4.57	1.8	2	1.83	1.75	1.25	1.24
22	Histone H1.4	HIST1H1E	N	P10412	21.73	11.03	4.42	10.1	2	1.60	1.68	1.49	2.72
23	Serine/arginine-rich splicing factor 2	SRSF2	N	Q01130	25.35	11.86	4.23	11.3	2	1.61	1.58	1.33	1.73
24	Acidic leucine-rich nuclear phosphoprotein 32 family member B	ANP32B	N	Q92688	28.79	3.93	4.22	8.4	2	1.91	1.74	1.26	1.63
25	DNA topoisomerase 2-beta	TOP2B	N	Q02880	183.14	8.16	4.05	2.6	4	1.67	1.97	1.96	1.69
26	Acetyl-CoA acetyltransferase	ACAT2	C	Q9BWD1	41.35	6.46	4.05	4.8	3	2.17	2.76	1.26	2.88
27	Thioredoxin reductase 1	TXNRD1	C	Q16881	70.91	7.16	4.05	5.2	2	2.37	1.49	5.94	5.36
28	Eukaryotic translation initiation factor 4 gamma 2	EIF4G2	C	P78344	102.36	6.69	4.03	2.1	2	1.64	1.26	1.53	2.73
29	Vesicle-fusing ATPase	NSF	C	P46459	82.59	6.52	4	4.0	2	1.26	2.33	3.56	2.32
30	WD40 repeat-containing protein SMU1	SMU1	N	Q2TAY7	57.54	6.74	4	3.9	2	1.72	1.31	1.59	1.99
31	Ribose-phosphate pyrophosphokinase 1	PRPS1	C	P60891	34.70	6.56	4	8.5	3	1.32	1.84	2.18	3.06
32	ATP-dependent RNA helicase DDX42	DDX42	C	Q86XP3	102.98	6.54	4	3.0	2	1.30	1.36	1.94	1.37
33	6-phosphogluconolactonase	PGLS	C	O95336	27.42	5.71	4	12.0	2	2.03	5.76	1.68	1.26
34	Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	C	P63244	35.08	7.60	3.94	12.6	3	1.43	1.32	1.41	1.62
35	UPF0160 protein MYG1, mitochondrial	C12orf10	N, M	Q9HB07	37.11	5.50	3.89	5.9	2	2.39	1.56	1.39	1.85
36	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	PPP2R5D	N	Q14738	69.99	8.24	3.85	4.2	2	3.20	2.10	1.34	3.93
37	UDP-glucose 4-epimerase	GALE	C	Q14376	38.28	6.26	3.6	13.2	2	1.53	1.53	4.29	1.53
38	Dnaj homolog subfamily A member 1	DNAJA1	N	P31689	44.55	6.65	3.56	2.0	1	2.16	1.95	1.47	1.48
39	Vacuolar protein sorting-associated protein 26A	VPS26A	C	O75436	38.17	6.13	3.22	6.1	2	1.56	1.96	2.33	2.28

(continued on next page)

Table 1 (continued)

N	Protein name	Gene	Location	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (≥95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
40	HLA class I histocompatibility antigen, A-69 alpha chain	HLA-A	PM	P10316	40.98	6.42	3.21	5.5	2	1.71	4.04	1.96	2.96
41	Histidyl-tRNA synthetase	HARS	C	P12081	57.28	5.73	3.18	3.3	2	1.20	1.37	1.72	1.65
42	Large proline-rich protein BAG6	BAG6	N	P46379	119.41	5.40	3.17	2.1	2	2.40	2.42	1.38	1.91
43	Endoplasmic reticulum resident protein 29	ERP29	C	P30040	25.85	6.08	3.1	8.4	2	1.43	1.34	1.26	1.40
44	Eukaryotic translation initiation factor 4B	EIF4B	C	P23588	69.15	5.54	2.64	2.5	2	2.46	1.97	1.29	1.83
45	Adenylyl cyclase-associated protein 2	CAP2	PM	P40123	52.69	5.96	2.41	2.1	1	1.99	2.61	2.36	1.28
46	Activated RNA polymerase II transcriptional coactivator p15	SUB1	N	P53999	14.26	9.60	2.17	10.2	1	1.30	1.83	1.79	1.68
47	Proteasome subunit beta type-2	PSMB2	C	P49721	22.84	6.52	2.15	9.5	1	2.41	2.06	1.99	2.20
48	ATP-dependent RNA helicase DDX19A	DDX19A	N	Q9NUU7	53.84	6.21	2.14	4.0	1	1.91	1.21	1.27	1.95
49	Peptidyl-prolyl cis-trans isomerase FKBP3	FKBP3	N	Q00688	25.05	9.29	2.14	3.6	1	2.39	2.00	1.56	1.54
50	Eukaryotic initiation factor 4A-III	EIF4A3	N	P38919	46.87	6.30	2.11	6.1	2	1.60	1.39	1.46	1.35
51	Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	C	Q9NZB2	121.89	9.07	2.1	1.3	1	1.33	1.43	1.57	1.82
52	40S ribosomal protein S27	RPS27	C	P42677	9.33	9.58	2.1	14.3	2	1.71	1.28	1.21	1.27
53	Cysteine and glycine-rich protein 1	CSRP1	N	P21291	20.57	8.90	2.09	7.8	2	1.62	1.24	3.62	2.84
54	Cleavage and polyadenylation specificity factor subunit 5	NUDT21	N	O43809	26.10	8.85	2.08	10.1	3	1.76	1.79	1.64	1.69
55	Nuclear transport factor 2	NUTF2	N	P61970	14.48	5.10	2.08	6.3	1	2.55	1.68	1.22	1.21
56	Poly(rC)-binding protein 2	PCBP2	N	Q15366	38.58	6.33	2.05	14.2	4	1.51	1.42	1.43	1.42
57	Serine/arginine-rich splicing factor 3	SRSF3	N	P84103	19.33	11.64	2.03	5.5	1	2.68	1.75	1.88	2.18
58	Tropomyosin beta chain	TPM2	C	P07951	32.85	4.66	2	8.5	3	1.38	1.66	1.43	1.41
59	C-Jun-amino-terminal kinase-interacting protein 4	SPAG9	C	O60271	146.21	5.04	2	0.8	1	1.76	1.43	1.26	1.34
60	Ubiquitin-associated protein 2-like	UBAP2L	Other	Q14157	114.53	6.61	2	0.8	1	1.23	1.45	1.68	2.38
61	Selenide, water dikinase 1	SEPHS1	Other	P49903	42.78	5.64	2	3.8	1	1.65	1.26	1.41	2.16
62	Acyl-CoA-binding protein	DBI	C	P07108	9.91	6.11	2	18.4	1	3.47	1.32	2.33	1.40
63	Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	C	P50416	88.24	8.85	2	2.2	1	14.85	3.46	26.29	8.85
64	Myelin expression factor 2	MYEF2	N	Q9P2K5	64.12	8.86	2	1.7	1	6.78	3.09	3.48	2.79
65	DNA replication licensing factor MCM2	MCM2	N	P49736	101.76	5.34	2	1.9	1	7.20	5.77	4.52	4.98
66	Calcium-regulated heat stable protein 1	CARHSP1	C	Q9Y2V2	15.76	8.40	2	10.9	2	2.75	3.06	3.23	4.68
67	Prostaglandin E synthase 3	PTGES3	C	Q15185	18.70	4.32	2	6.3	1	1.69	1.44	1.57	1.44
68	Glycogen debranching enzyme	AGL	C	P35573	174.76	6.31	1.92	1.1	1	1.35	1.83	5.32	3.44
69	Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	N	P31942	36.93	6.37	1.89	2.9	1	1.33	1.46	1.30	1.44
70	ATP-binding cassette sub-family F member 3	ABCF3	Other	Q9NUQ8	79.61	5.96	1.89	1.8	1	1.89	1.73	1.58	2.04
71	N-terminal kinase-like protein	SCYL1	C	Q96KG9	89.63	5.93	1.85	1.9	1	4.28	1.74	2.30	1.78
72	Beta-catenin-like protein 1	CTNBL1	N	Q8WYA6	65.17	4.96	1.83	1.4	1	1.58	2.01	1.65	1.73
73	26S proteasome non-ATPase regulatory subunit 5	PSMD5	Other	Q16401	56.06	5.34	1.82	2.0	1	1.75	1.31	1.56	1.67
74	Nuclear pore membrane glycoprotein 210	NUP210	N	Q8TEM1	202.66	6.27	1.77	0.4	1	2.67	1.54	1.40	3.04
75	Basic leucine zipper and W2 domain-containing protein 2	BZW2	C	Q9Y6E2	48.16	6.26	1.67	2.1	1	1.65	1.36	1.44	2.05
76	High mobility group protein B2	HMGB2	N	P26583	23.90	7.77	1.47	6.2	1	4.16	3.87	2.14	2.49
77	Protein phosphatase 1G	PPM1G	N	O15355	59.27	4.27	1.44	1.5	1	2.34	1.92	1.65	2.23
78	Myosin-XVIIIa	MYO18A	C	Q92614	233.12	5.95	1.41	1.0	2	1.24	1.63	2.94	1.46

^a UniProtKB ID, UniProt Knowledgebase Identifier; ^b MW, Molecular weight; ^c pI, Isoelectric point; ^d Unused, Unused Protein Score is calculated using only peptides from spectra that have not been used by other proteins; ^e % Cov. (≥95), Percentage of coverage; ^f Pep. Ident. (>95), Number of peptides with at least 95% of confidence in their identification; ^g 114:113, MCF7/MCF 10A ratio; ^h 115:113, MDA-MB-231/MCF 10A ratio; ⁱ 116:113, SK-BR-3/MCF 10A ratio; ^j 117:113, T47D/MCF 10A ratio. C, Cytoplasm. N, Nucleus. M, Mitochondrion. PM, Plasma Membrane. Localisation of proteins was obtained with the IPA Knowledge Base.

Table 2 – Subexpressed polypeptides found in common in all breast cancer cell lines by iTRAQ-labelling and tandem mass spectrometry.

	Protein name	Gene	Location	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (≥95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
1	Plectin OS	PLEC	C	Q15149	531.79	5.74	83.39	10.4	47	0.38	0.78	0.48	0.61
2	Myosin-9	MYH9	C	P35579	226.40	5.50	52.32	13.2	37	0.23	0.70	0.77	0.57
3	Filamin-B	FLNB	C	O75369	278.16	5.47	48.96	11.8	28	0.45	0.62	0.54	0.44
4	Prelamin-A/C	LMNA	N	P02545	73.81	6.57	31.76	32.4	23	0.62	0.77	0.79	0.78
5	Stress-70 protein, mitochondrial	HSPA9	M,N	P38646	68.76	5.44	29.02	27.1	23	0.47	0.46	0.67	0.79
6	ATP synthase subunit beta, mitochondrial	ATP5B	M	P06576	51.77	5.00	28.95	37.4	28	0.43	0.31	0.53	0.52
7	Annexin A2	ANXA2	PM	P07355	38.47	7.56	28.44	42.8	21	0.32	0.70	0.35	0.38
8	ATP synthase subunit alpha, mitochondrial	ATP5A1	M	P25705	55.21	8.28	22.76	21.3	11	0.50	0.42	0.54	0.58
9	Heterogeneous nuclear ribonucleoprotein U	HNRNPU	N	Q00839	90.45	5.76	21.44	11.9	11	0.63	0.73	0.57	0.72
10	Trifunctional enzyme subunit alpha, mitochondrial	HADHA	M	P40939	79.01	8.98	19.15	18.4	10	0.34	0.38	0.61	0.42
11	4 F2 cell-surface antigen heavy chain	SLC3A2	PM	P08195	67.99	4.89	16.6	15.7	10	0.50	0.27	0.37	0.78
12	10 kDa heat shock protein, mitochondrial	HSPE1	M	P61604	10.80	8.91	14.04	69.6	8	0.46	0.39	0.51	0.60
13	Voltage-dependent anion-selective channel protein 1	VDAC1	C	P21796	30.64	8.63	13.22	27.6	7	0.23	0.38	0.34	0.40
14	Trifunctional enzyme subunit beta, mitochondrial	HADHB	M	P55084	47.48	9.24	12.59	11.2	7	0.44	0.52	0.67	0.50
15	Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	N	O60506	69.47	8.68	12.47	10.0	6	0.64	0.67	0.32	0.63
16	Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	M	P31930	49.13	5.43	12.26	19.4	8	0.49	0.47	0.63	0.43
17	EH domain-containing protein 2	EHD2	N	Q9NZN4	61.16	6.02	12.06	10.3	6	0.21	0.30	0.11	0.10
18	Core histone macro-H2A.1	H2AFY	N	O75367	39.62	9.80	11.54	18.5	7	0.43	0.57	0.50	0.46
19	Prohibitin	PHB	N	P35232	29.80	5.57	10.75	32.7	9	0.40	0.52	0.40	0.63
20	Exportin-2	CSE1L	N	P55060	110.42	5.51	10.59	8.8	8	0.68	0.49	0.62	0.45
21	Glutamate dehydrogenase 1, mitochondrial	GLUD1	M	P00367	56.01	6.71	10.57	10.9	6	0.35	0.39	0.29	0.30
22	Eukaryotic translation initiation factor 5B	EIF5B	C	O60841	138.83	5.39	10.02	6.1	6	0.50	0.51	0.41	0.41
23	Desmoplakin	DSP	PM	P15924	331.77	6.44	9.97	1.9	4	0.64	0.36	0.40	0.44
24	Junction plakoglobin	JUP	PM	P14923	81.74	5.75	9.95	7.9	9	0.33	0.37	0.40	0.52
25	Isoleucyl-tRNA synthetase, mitochondrial	IARS2	M	Q9NSE4	108.74	6.14	8.59	7.1	5	0.44	0.40	0.54	0.49
26	Fumarate hydratase, mitochondrial	FH	M	P07954	50.08	6.99	8.38	12.2	5	0.20	0.28	0.38	0.23
27	Sulfide:quinone oxidoreductase, mitochondrial	SQRDL	M	Q9Y6N5	49.96	9.18	8.14	11.3	4	0.26	0.42	0.37	0.21
28	Glutathione S-transferase P	GSTP1	C	P09211	23.22	5.44	8.13	36.7	5	0.11	0.23	0.09	0.20
29	Superoxide dismutase [Mn], mitochondrial	SOD2	M	P04179	22.20	6.86	7.84	28.4	6	0.13	0.13	0.23	0.13
30	Mitochondrial inner membrane protein	IMMT	M	Q16891	83.68	6.08	7.71	8.2	6	0.48	0.60	0.55	0.62
31	Phosphate carrier protein, mitochondrial	SLC25A3	M	Q00325	34.90	9.29	7.37	10.8	7	0.63	0.72	0.76	0.54
32	NADH-cytochrome b5 reductase 3	CYB5R3	C	P00387	34.10	7.31	7.06	17.9	6	0.45	0.78	0.59	0.48
33	Myosin-Ic	MYO1C	C	O00159	121.68	9.46	7.03	6.5	5	0.57	0.79	0.51	0.34
34	Electron transfer flavoprotein subunit alpha, mitochondrial	ETFa	M	P13804	35.07	8.62	6.59	15.9	7	0.36	0.23	0.30	0.29
35	Adenylate kinase 2, mitochondrial	AK2	M	P54819	26.48	7.67	6.58	23.0	6	0.36	0.44	0.71	0.60
36	Polymerase I and transcript release factor	PTRF	N	Q6NZI2	43.48	5.50	6.1	7.9	3	0.12	0.68	0.17	0.13
37	Cytochrome c1, heme protein, mitochondrial	CYC1	M	P08574	27.35	6.49	6.06	8.3	3	0.42	0.38	0.41	0.40
38	RuvB-like 1	RUVBL1	N	Q9Y265	50.23	6.02	6.04	8.3	3	0.73	0.73	0.60	0.71
39	Nuclear pore complex protein Nup133	NUP133	N	Q8WUM0	128.98	4.98	6	2.9	3	0.32	0.29	0.66	0.34
40	Dihydropyridyl dehydrogenase, mitochondrial	DLD	M	P09622	50.17	6.50	6	7.7	3	0.35	0.52	0.78	0.44
41	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	M	P31040	68.01	6.25	5.71	5.6	3	0.54	0.49	0.74	0.48
42	Nucleolar protein 58	NOP58	N	Q9Y2X3	59.58	9.03	5.51	8.3	4	0.71	0.61	0.60	0.68
43	Presequence protease, mitochondrial	PITRM1	M	Q5JRX3	115.64	6.26	5.38	3.2	3	0.57	0.39	0.46	0.42
44	Myosin light polypeptide 6	MYL6	C	P60660	16.80	4.56	5.35	21.9	7	0.33	0.76	0.79	0.79

(continued on next page)

Table 2 (continued)

	Protein name	Gene	Location	UniProtKB ID ^a	MW (kDa) ^b	pI ^c	Unused ^d	% Cov. (>95) ^e	Pep. Ident. (>95) ^f	114:113 ^g	115:113 ^h	116:113 ⁱ	117:113 ^j
45	Aconitate hydratase, mitochondrial	ACO2	M	Q99798	82.43	6.85	5.26	5.0	4	0.42	0.41	0.79	0.50
46	Creatine kinase U-type, mitochondrial	CKMT1A	M	P12532	43.08	7.31	5.1	4.3	4	0.28	0.15	0.51	0.16
47	ATP synthase subunit gamma, mitochondrial	ATP5C1	M	P36542	30.17	9.02	4.66	11.1	3	0.54	0.37	0.64	0.61
48	Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	M	P22695	46.78	7.74	4.5	6.2	2	0.59	0.62	0.66	0.63
49	Dihydropolyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	M	P36957	41.39	5.90	4.32	4.4	2	0.55	0.52	0.69	0.35
50	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1	M	P13073	17.20	9.16	4.23	7.1	2	0.43	0.47	0.44	0.35
51	Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	M	Q02978	33.93	9.92	4.22	9.2	3	0.36	0.71	0.73	0.59
52	Catenin alpha-1	CTNNA1	PM	P35221	99.94	5.95	4.15	3.3	2	0.50	0.34	0.42	0.51
53	DNA (cytosine-5)-methyltransferase 1	DNMT1	N	P26358	183.17	7.99	4.04	1.7	2	0.44	0.29	0.11	0.25
54	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	NDUFS8	M	O00217	20.29	5.10	4.04	9.5	2	0.31	0.50	0.50	0.44
55	Tumor-associated calcium signal transducer 2	TACSTD2	PM	P09758	33.12	8.98	4.03	8.0	3	0.42	0.23	0.47	0.34
56	Calcium-binding mitochondrial carrier protein SCaMC-1	SLC25A24	M	Q6NUK1	53.35	6.00	4.02	6.1	2	0.48	0.58	0.58	0.46
57	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB	M	P21912	28.81	8.78	4.02	8.2	2	0.44	0.53	0.77	0.44
58	Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	C	P49419	55.58	6.47	4.02	5.8	3	0.27	0.74	0.61	0.40
59	Inorganic pyrophosphatase 2, mitochondrial	PPA2	M	Q9H2U2	34.71	5.97	4	11.7	3	0.45	0.17	0.59	0.39
60	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	M	P42765	41.92	8.32	4	5.5	2	0.20	0.32	0.47	0.27
61	5'-3' exoribonuclease 2	XRN2	N	Q9H0D6	108.58	7.26	4	2.6	2	0.43	0.46	0.35	0.34
62	Cytochrome c oxidase subunit 2	MT-CO2	C	P00403	25.57	4.67	4	13.2	2	0.28	0.43	0.33	0.33
63	Ragulator complex protein LAMTOR1	LAMTOR1	PM	Q6IAA8	17.61	5.01	3.89	18.0	2	0.58	0.18	0.45	0.35
64	Polyribonucleotide nucleotidyltransferase 1, mitochondrial	PNPT1	M	Q8TCS8	80.87	6.26	3.65	2.3	2	0.77	0.47	0.71	0.80
65	SWI/SNF complex subunit SMARCC2	SMARCC2	N	Q8TAQ2	132.88	5.49	3.48	1.9	2	0.45	0.34	0.32	0.44
66	Protein S100-A2	S100A2	N	P29034	11.12	4.68	3.45	33.7	4	0.05	0.04	0.04	0.04
67	Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2	M	P43304	76.36	6.30	3.43	4.1	2	0.79	0.46	0.78	0.52
68	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	NDUFA10	M	O95299	37.15	6.87	3.42	5.6	2	0.49	0.57	0.76	0.40
69	39S ribosomal protein L38, mitochondrial	MRPL38	M	Q96DV4	41.63	6.50	3.4	4.7	2	0.47	0.43	0.74	0.50
70	ATP synthase subunit g, mitochondrial	ATP5L	M	O75964	11.30	9.65	3.33	23.3	2	0.55	0.38	0.72	0.64
71	ATP synthase subunit e, mitochondrial	ATP5I	M	P56385	7.80	9.35	3.08	30.4	2	0.32	0.22	0.36	0.39
72	ATP synthase subunit d, mitochondrial	ATP5H	M	O75947	18.36	5.22	3.01	14.3	2	0.43	0.38	0.49	0.49
73	Cytochrome c oxidase subunit 5A, mitochondrial	COX5A	M	P20674	12.50	4.88	2.99	10.0	2	0.64	0.50	0.58	0.51
74	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	NDUFA9	M	Q16795	38.90	9.68	2.92	8.2	2	0.36	0.50	0.68	0.48
75	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	NDUFS2	M	O75306	49.17	6.26	2.92	6.5	3	0.42	0.55	0.54	0.38
76	Myosin regulatory light chain 12B	MYL12B	C	O14950	19.78	4.69	2.91	9.9	2	0.68	0.80	0.71	0.61
77	Alpha-2-macroglobulin receptor-associated protein	LRPAP1	PM	P30533	37.77	6.88	2.84	4.8	2	0.69	0.65	0.49	0.72
78	Mitochondrial import inner membrane translocase subunit TIM50	TIMM50	M	Q3ZCQ8	35.08	6.01	2.49	5.1	2	0.58	0.35	0.49	0.56
79	Annexin A8-like protein 2	ANXA8L2	PM	Q5VT79	36.86	5.45	2.39	2.1	1	0.08	0.06	0.09	0.25
80	Dipeptidyl peptidase 1	CTSC	C	P53634	18.47	5.94	2.26	2.4	1	0.09	0.11	0.24	0.22
81	Neurofilament light polypeptide	NEFL	C	P07196	61.39	4.63	2.17	3.3	2	0.09	0.06	0.09	0.06
82	Glutamine synthetase	GLUL	C	P15104	41.93	6.42	2.14	2.1	1	0.19	0.04	0.54	0.03
83	ATP synthase subunit b, mitochondrial	ATP5F1	M	P24539	24.63	9.10	2.13	5.9	1	0.46	0.41	0.51	0.51
84	Serum deprivation-response protein	SDPR	PM	O95810	47.04	5.14	2.12	3.3	1	0.08	0.34	0.26	0.10
85	28S ribosomal protein S7, mitochondrial	MRPS7	M	Q9Y2R9	24.12	9.72	2.12	5.4	1	0.51	0.48	0.68	0.50

86	Mitochondrial-processing peptidase subunit alpha	PMPCA	M	Q10713	54.65	5.88	2.09	3.2	1	0.67	0.15	0.40	0.37
87	OCLIA domain-containing protein 1	OCIAD1	C	Q9NX40	27.63	7.02	2.08	7.3	1	0.57	0.52	0.32	0.43
88	Prenylcysteine oxidase 1	PCYOX1	C	Q9UHG3	53.85	5.89	2.08	3.6	1	0.66	0.78	0.60	0.74
89	Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	M	P24752	41.39	8.16	2.07	4.0	1	0.51	0.50	0.54	0.70
90	Sideroflexin-1	SFXN1	C	Q9H9B4	35.49	9.22	2.07	2.8	1	0.74	0.77	0.78	0.72
91	Serpin B5	SERPINB5	ES	P36952	42.10	5.72	2.06	3.2	1	0.13	0.18	0.17	0.22
92	Condensin complex subunit 1	NCAPD2	N	Q15021	157.18	6.19	2.04	0.7	1	0.60	0.54	0.50	0.36
93	Protein transport protein Sec31A	SEC31A	C	O94979	133.01	6.43	2.04	1.5	1	0.38	0.27	0.35	0.33
94	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	PM	P62873	37.25	5.60	2.04	2.9	1	0.33	0.32	0.33	0.24
95	Mitochondrial import inner membrane translocase subunit TIM44	TIMM44	M	O43615	51.35	8.51	2.02	3.3	1	0.49	0.65	0.25	0.55
96	Glutaminase kidney isoform, mitochondrial	GLS	C or M	O94925	71.56	7.01	2.02	2.2	2	0.62	0.57	0.51	0.55
97	Protein QIL1	C19orf70	C	Q5XKP0	13.09	9.44	2.02	16.9	1	0.28	0.49	0.38	0.37
98	Tubulin beta-2A chain	TUBB2A	C	Q13885	49.91	4.78	2	33.9	27	0.79	0.72	0.60	0.45
99	Ras-related protein Rab-10	RAB10	C	P61026	22.54	8.58	2	11.0	2	0.62	0.74	0.63	0.52
100	Protein AHNAK2	AHNAK2	Other	Q8IVF2	616.63	5.20	2	1.1	1	0.12	0.28	0.14	0.16
101	Microsomal glutathione S-transferase 1	MGST1	C	P10620	17.47	9.41	2	7.7	1	0.15	0.18	0.44	0.12
102	Quinone oxidoreductase PIG3	TP53I3	Other	Q53FA7	35.54	6.66	2	3.6	1	0.27	0.46	0.33	0.21
103	DNA-directed RNA polymerases I, II, and III subunit RPABC5	POLR2L	N	P62875	7.65	7.65	2	16.4	1	0.80	0.79	0.57	0.75
104	Electron transfer flavoprotein subunit beta	ETFB	C	P38117	27.71	8.29	2	4.7	1	0.52	0.62	0.37	0.42
105	Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	C	Q9Y6M1	66.12	8.48	2	2.3	1	0.27	0.76	0.22	0.53
106	Epidermal growth factor receptor	EGFR	PM	P00533	132.01	6.17	2	0.9	1	0.10	0.38	0.28	0.40
107	ATP synthase subunit f, mitochondrial	ATP5J2	M	P56134	10.79	9.70	2	11.7	1	0.54	0.33	0.73	0.55
108	Thioredoxin-related transmembrane protein 4	TMX4	C	Q9H1E5	36.76	4.29	2	3.7	1	0.72	0.35	0.54	0.41
109	ATP synthase subunit epsilon-like protein, mitochondrial	ATP5EP2	M	Q5VTU8	5.81	10.15	2	15.7	1	0.47	0.40	0.49	0.45
110	Acyl-coenzyme A thioesterase 13	ACOT13	C	Q9NPJ3	14.96	9.23	2	9.3	1	0.32	0.50	0.69	0.69
111	Ubiquitin fusion degradation protein 1 homolog	UFD1L	C	Q92890	34.50	6.27	2	3.3	1	0.65	0.77	0.63	0.63
112	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	NDUFB9	M	Q9Y6M9	21.70	8.59	2	8.9	1	0.37	0.19	0.61	0.35
113	Diablo homolog, mitochondrial	DIABLO	M	Q9NR28	20.77	4.78	2	4.6	1	0.49	0.56	0.56	0.44
114	Serum paraoxonase/arylesterase 2	PON2	PM	Q15165	39.40	5.33	2	2.5	1	0.24	0.21	0.63	0.34
115	Small nuclear ribonucleoprotein Sm D1	SNRPD1	N	P62314	13.28	11.56	2	10.9	1	0.42	0.39	0.34	0.38
116	Translocator protein	TSPO	C	P30536	18.83	9.43	2	4.7	1	0.27	0.29	0.48	0.35
117	Monocarboxylate transporter 4	SLC16A3	PM	O15427	49.34	8.25	2	1.5	1	0.09	0.47	0.44	0.79
118	Cytochrome c oxidase subunit 7C, mitochondrial	COX7C	M	P15954	5.36	8.98	2	14.3	1	0.40	0.66	0.68	0.51
119	Nucleoside diphosphate-linked moiety X motif 19, mitochondrial	NUDT19	M	A8MXV4	40.58	6.48	2	3.7	1	0.02	0.62	0.56	0.13
120	Metaxin-1	MTX1	C	Q13505	51.48	9.80	1.96	3.0	1	0.24	0.23	0.50	0.49
121	39S ribosomal protein L15, mitochondrial	MRPL15	M	Q9P015	31.32	9.93	1.92	3.0	1	0.59	0.65	0.61	0.67
122	Nicotinamide N-methyltransferase	NNMT	C	P40261	29.57	5.56	1.8	7.2	1	0.11	0.36	0.09	0.77
123	Interleukin-18	IL18	ES	Q14116	18.22	5.01	1.72	4.7	1	0.10	0.69	0.23	0.30
124	Dimethyladenosine transferase 2, mitochondrial	TFB2M	M	Q9H5Q4	43.26	9.11	1.69	2.3	1	0.36	0.32	0.47	0.36
125	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	NDUFB1	M	P49821	48.51	7.53	1.64	1.9	1	0.55	0.44	0.68	0.41
126	Tripartite motif-containing protein 29	TRIM29	C	Q14134	65.83	6.73	1.57	2.2	1	0.18	0.11	0.09	0.10
127	Protein ETHE1, mitochondrial	ETHE1	M	O95571	27.10	6.37	1.54	3.9	1	0.45	0.23	0.55	0.32
128	Junctional adhesion molecule A	F11R	PM	Q9Y624	29.68	6.72	1.48	3.7	1	0.56	0.24	0.77	0.48

^a UniProtKB ID, UniProt Knowledgebase Identifier; ^b MW, Molecular weight; ^c pI, Isoelectric point; ^d Unused, Unused Protein Score is calculated using only peptides from spectra that have not been used by other proteins; ^e % Cov. (≥ 95), Percentage of coverage; ^f Pep. Ident. (>95), Number of peptides with at least 95% of confidence in their identification; ^g 114:113, MCF7/MCF 10A ratio; ^h 115:113, MDA-MB-231/MCF 10A ratio; ⁱ 116:113, SK-BR-3/MCF 10A ratio; ^j 117:113, T47D/MCF 10A ratio. C, Cytoplasm. N, Nucleus. PM, Plasma Membrane. M, Mitochondrion. ES, Extracellular Space. Localisation of proteins was obtained with the IPA Knowledge Base.

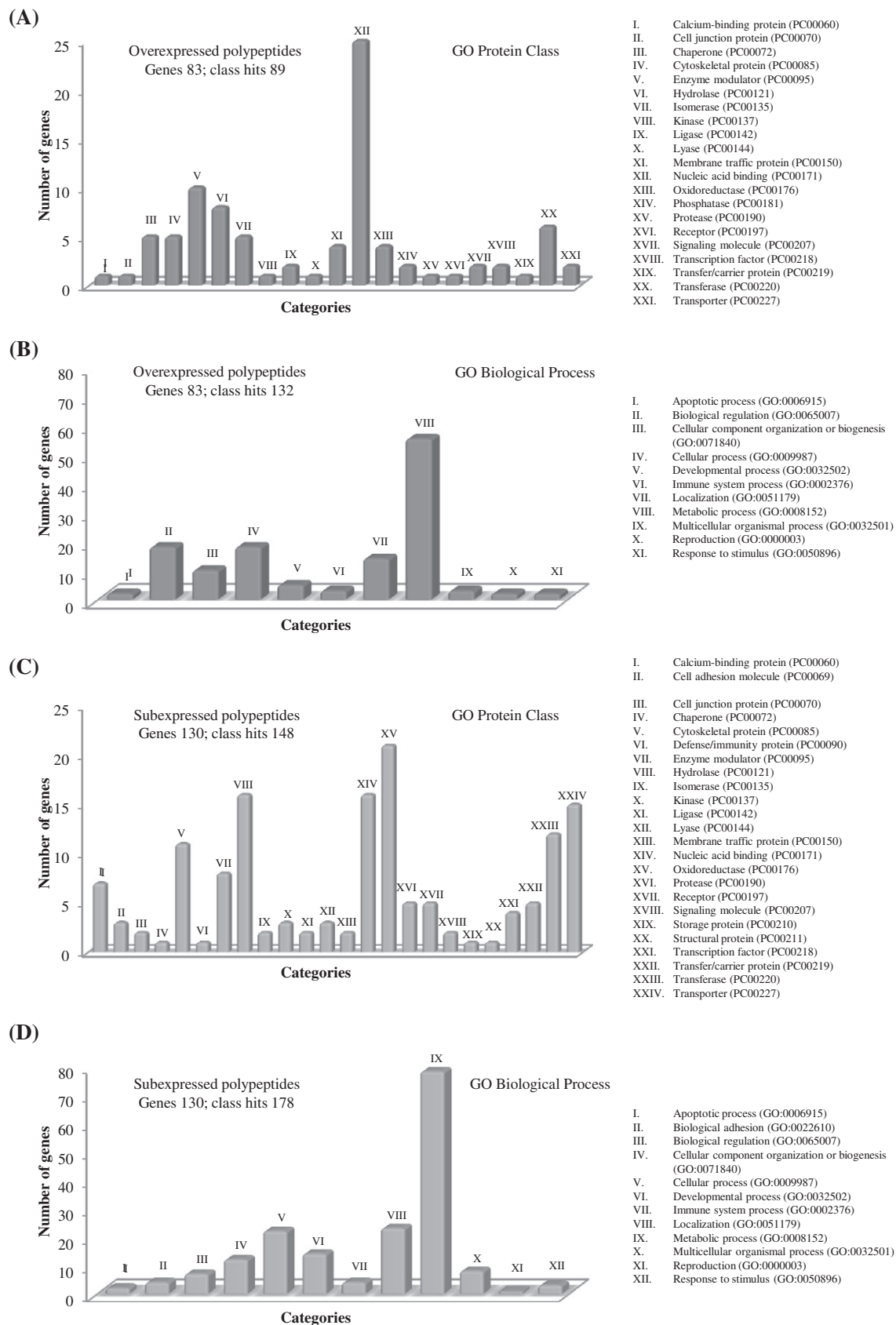


Fig. 4 – PANTHER functional classification of differentially expressed proteins found in common in the breast cancer cell lines. The proteins were classified according to Protein Class (A, C) and Biological Process (B, D). Bars show the different categories were the proteins were classified (X-axis) and the number of genes involved in each category (Y-axis). (A, B) Overexpressed polypeptides. (C, D) Subexpressed polypeptides.

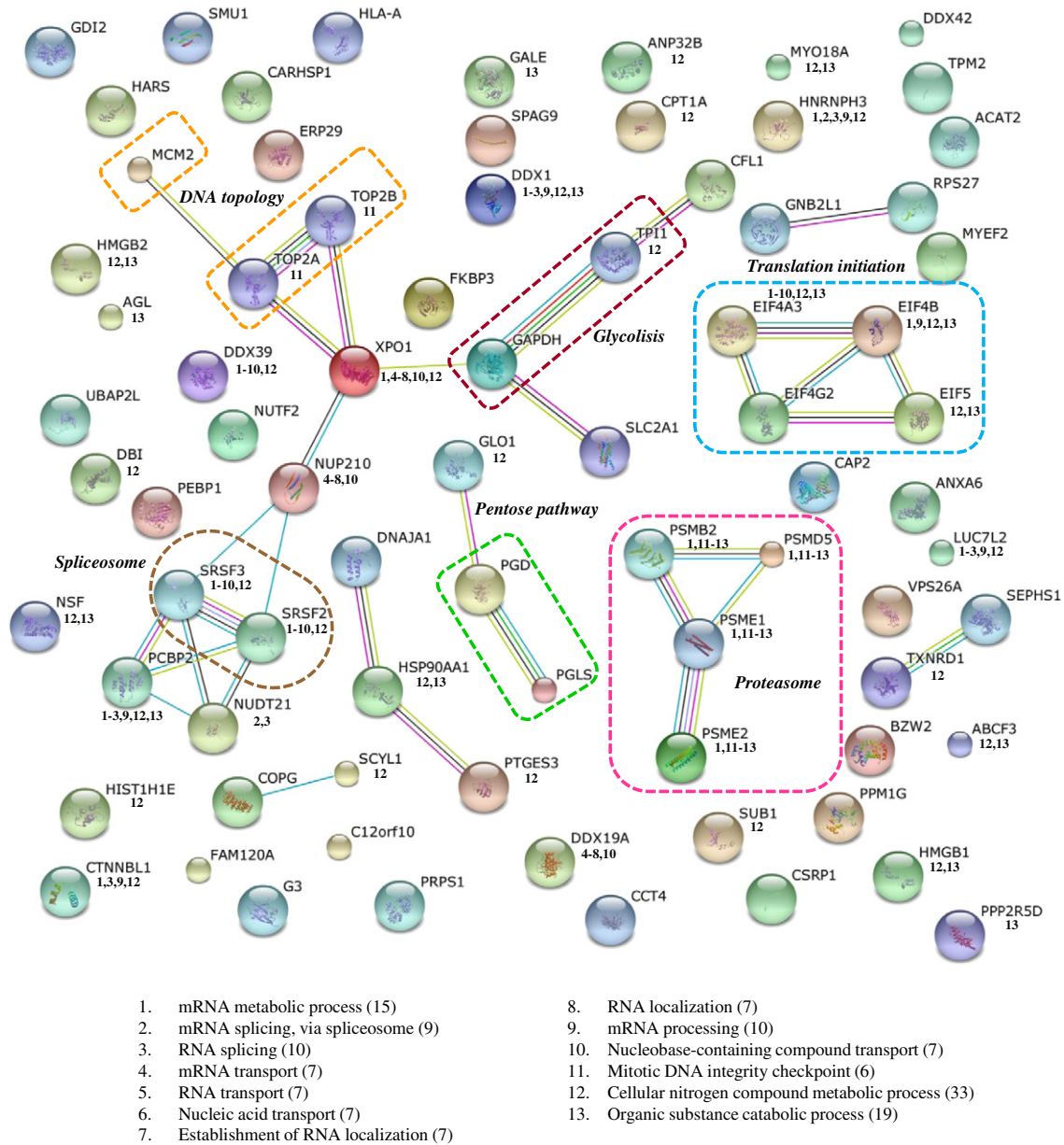
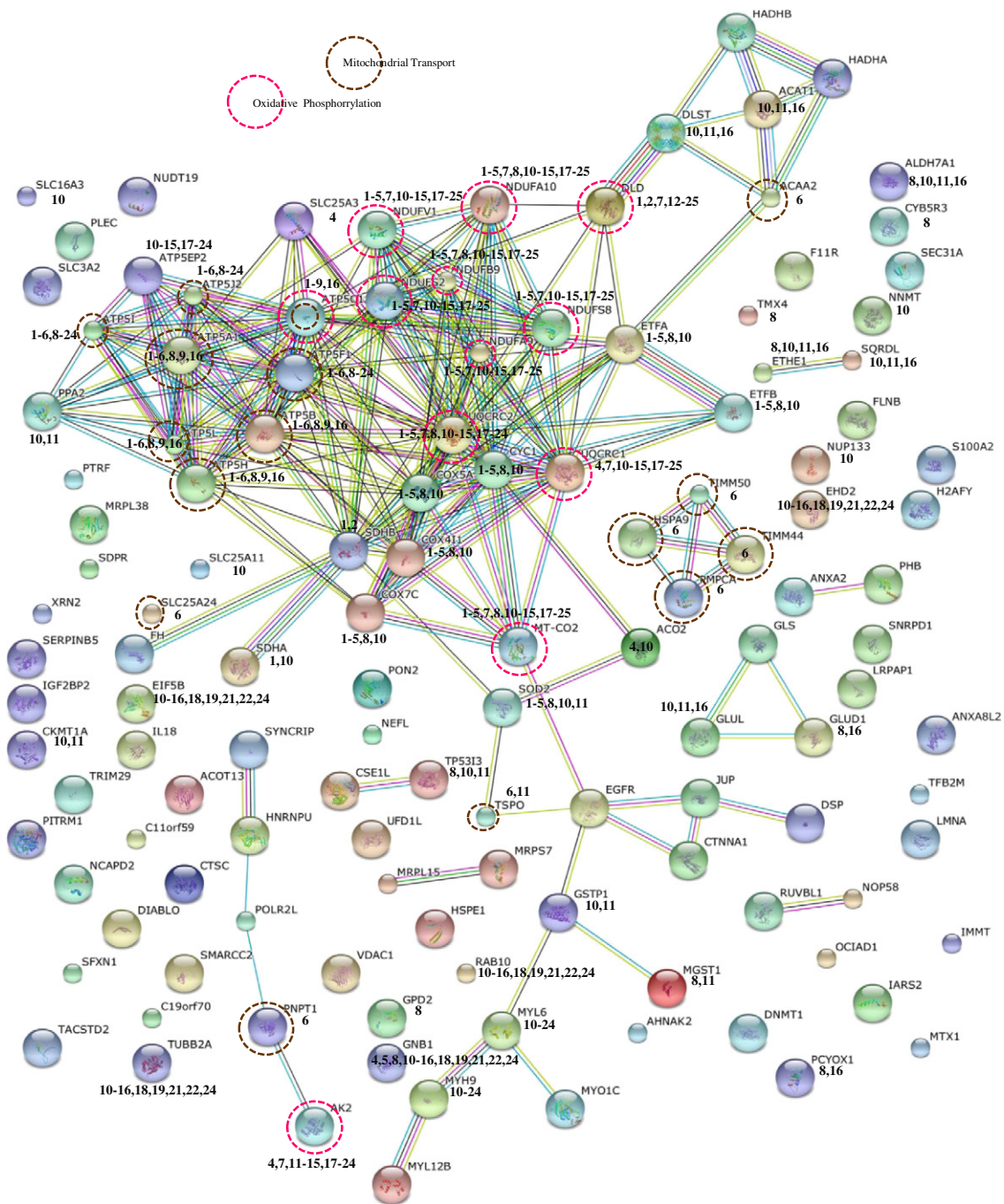


Fig. 5 – Interactome network of overexpressed polypeptides in common in all breast cancer cell lines. Coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Polypeptides involved in the same biological process were grouped as shown. DNA topology: TOP2A, TOP2B, and MCM2. Glycolysis: GAPDH and TPI1. Regulation of translation initiation: EIF5, EIF4B, EIF4G2 and EIF4A3. Spliceosome: SRSF2 and SRSF3. Pentose pathway: PGD and PGLS. Degradation by proteasome: PSMD5, PSMB2, PSME1 and PSME2. Bold numbers associated to proteins correspond to the different biological processes having statistical significance (P-value ≤ 0.05).

and butanoate metabolism, elongation and degradation of fatty acid, biosynthesis of unsaturated fatty acids, valine, leucine, isoleucine and lysine degradation, tryptophan and beta-alanine metabolism); ACAA2 and HADHB (fatty acid metabolism, elongation and degradation of fatty acid, valine, leucine and isoleucine degradation); GLUL (biosynthesis of amino acids, glyoxylate and dicarboxylate metabolism, nitrogen metabolism, alanine, aspartate, glutamate, arginine and proline metabolism); GLUD1 (nitrogen metabolism, alanine, aspartate, glutamate, arginine, proline, D-glutamine and D-glutamate

metabolism); ACO2 (carbon metabolism, 2-oxocarboxylic acid metabolism, biosynthesis of amino acids, citrate cycle, glyoxylate and dicarboxylate metabolism); DLD (carbon metabolism, glycolysis/gluconeogenesis, citrate cycle, pyruvate metabolism, glycine, serine and threonine metabolism, valine, leucine and isoleucine degradation); DLST (carbon metabolism, citrate cycle and lysine degradation); FH (carbon metabolism, citrate cycle and pyruvate metabolism); SDHA and SDHB (carbon metabolism, citrate cycle, oxidative phosphorylation); ACAT1 and ACAT2 (carbon metabolism, fatty acid metabolism,



- | | |
|---|--|
| 1. Respiratory electron transport chain (26) | 14. Ribonucleoside triphosphate metabolic process (22) |
| 2. Electron transport chain (25) | 15. Nucleoside triphosphate metabolic process (22) |
| 3. Cellular respiration (23) | 16. Organonitrogen compound catabolic process (24) |
| 4. Generation of precursor metabolites and energy (28) | 17. ATP metabolic process (17) |
| 5. Energy derivation by oxidation of organic compounds (24) | 18. Purine ribonucleoside metabolic process (22) |
| 6. Mitochondrial transport (16) | 19. Purine nucleoside metabolic process (22) |
| 7. Oxidative phosphorylation (12) | 20. Purine ribonucleoside monophosphate metabolic process (17) |
| 8. Oxidation-reduction process (29) | 21. Ribonucleoside metabolic process (22) |
| 9. Mitochondrial ATP synthesis coupled proton transport (8) | 22. Purine ribonucleotide metabolic process (22) |
| 10. Small molecule metabolic process (43) | 23. Ribonucleoside monophosphate metabolic process (17) |
| 11. Organonitrogen compound metabolic process (34) | 24. Nucleoside metabolic process (22) |
| 12. Purine ribonucleoside triphosphate metabolic process (22) | 25. ATP synthesis coupled electron transport (9) |
| 13. Purine nucleoside triphosphate metabolic process (22) | |

Fig. 6 – Interactome network of subexpressed polypeptides commonly found in all breast cancer cell lines. Coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Bold numbers associated to proteins correspond to the different biological processes having statistical significance (P -value ≤ 0.05). Proteins involved in oxidative phosphorylation were labelled with a pink circle, whilst those involved in mitochondrial transport were labelled with a brown circle.

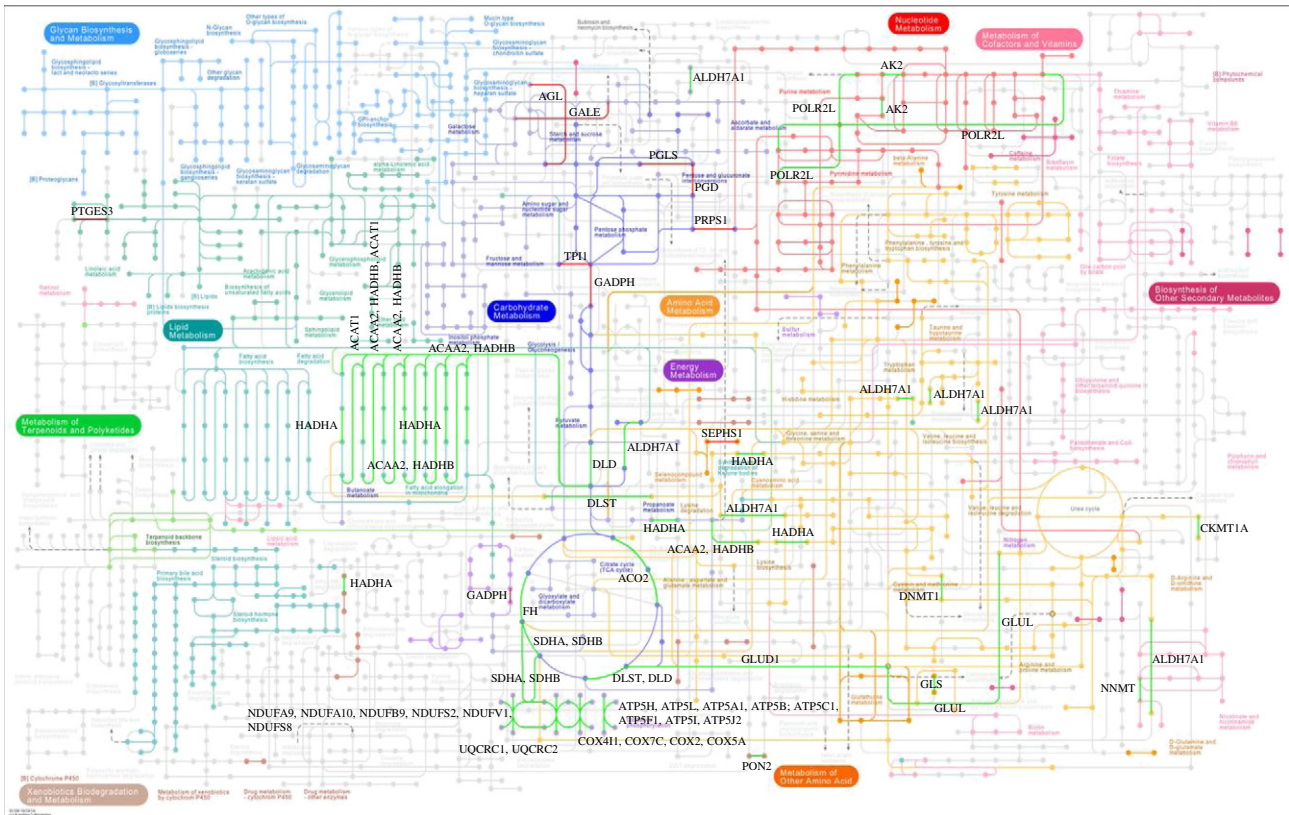


Fig. 7 – Global view of the metabolic pathways affected in all breast cancer cell lines. The 206 polypeptides that showed a change in their expression levels in all breast cancer cell lines were represented in KEGG pathways. The overexpressed proteins are shown in red and subexpressed polypeptides are shown in dark green.

pyruvate metabolism, glyoxylate and dicarboxylate metabolism, propanoate and butanoate metabolism, fatty acid degradation, synthesis and degradation of ketone bodies, valine, leucine, isoleucine and lysine degradation, tryptophan metabolism and terpenoid backbone biosynthesis) and PON2.

Remarkably, we found 24 subexpressed polypeptides involved in oxidative phosphorylation (Fig. 7). Their specific localisations in each complex of the electron transport of the respiratory chain are shown in Fig. 8. NDUFA9, NDUFA10, NDUFB9, NDUFS2, NDUFV1, and NDUFS8 were identified as elements of complex I (NADH dehydrogenase). In the case of complex II (succinate dehydrogenase) we found SDHA and SDHB polypeptides. UQCRC1 (COR1), UQCRC2 (QCR2) and CYC1 (Cyt1) were located in complex III (ubiquinol-cytochrome c reductase). As part of complex IV (cytochrome c oxidase) we found COX2, COX4I1 (COX4), COX5A, and COX7C. Finally, we localised 9 polypeptides that belong to complex V (ATP synthase), ATP5H (d subunit), ATP5L (g subunit), ATP5C1 (gamma subunit), ATP5I (subunit e), ATP5J2 (f2 subunit), ATP5A1 (alpha subunit), ATP5F1 (subunit b1), ATP5B (beta subunit) and PPA2 (pyrophosphatase).

3.4. IPA analysis of the differentially expressed proteins found in common in all breast cancer cell lines

To determine whether over- or subexpressed polypeptides have been involved in diseases and functions, or localised in

networks, we used the IPA software to perform mainly three different core analyses: a) Core I considered all tissues, primary cells and all cell lines. b) Core II was performed for breast cancer cell lines and mammary gland. c) Core III, for all cancer cell lines excluding breast cancer cell lines. They were obtained for different categories of diseases and biofunctions, using a p-value equal or lower than 0.05, which was obtained with Fisher’s Exact Test (see Table 2 in Ref [29]). Furthermore, we obtained the two top networks for both the over- and subexpressed proteins (Fig. 1 in Ref [29]). Networks were related to cancer, reproductive system disease, dermatological diseases and conditions, haematological diseases, cellular development, cellular growth and proliferation, cell cycle, hereditary disorders, and haematological diseases, among other processes.

3.5. Selection of putative biomarkers

The final step in our analysis was the identification of potential biomarkers using two different approaches: 1) Filtering candidate biomarkers with the IPA software through three filters, and 2) Searching of proteins in the PubMed database. In the former case, the first filter, named Biomarkers I, was carried out in all tissues, primary cells and all cell lines information contained in the Ingenuity Knowledge database. The second filter was through breast cancer cell lines and mammary gland (Biomarkers II). The third one consisted in

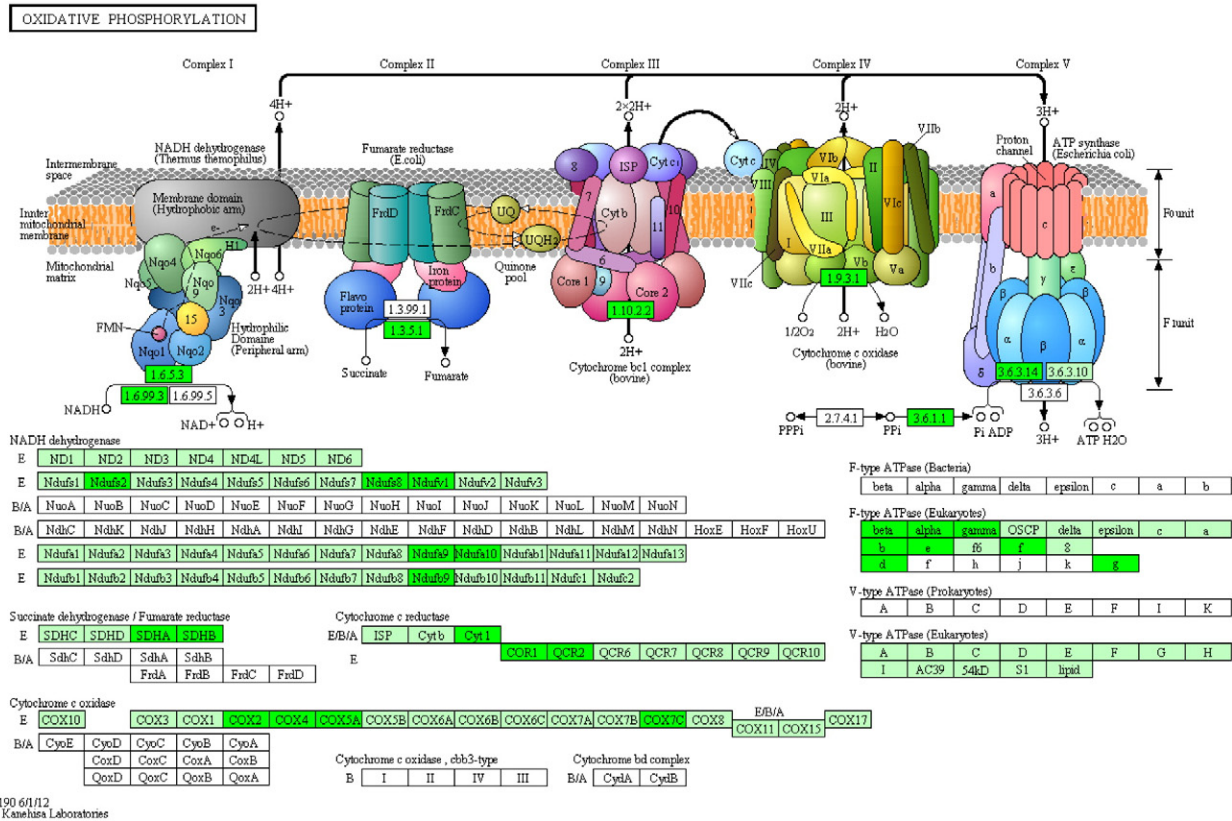


Fig. 8 – Localization of the subexpressed polypeptides in the respiratory chain obtained from the KEGG database. The subexpressed proteins (bright green) in all breast cancer cell lines involved in oxidative phosphorylation are shown in each complex of the mitochondrion.

filtering through cancer cell lines, excluding breast cancer cell lines and was named Biomarkers III. Results obtained in Biomarkers I revealed 39 potential candidates (15 over- and 24 subexpressed polypeptides), Biomarkers II included 9 putative targets (2 for over- and 7 for subexpressed proteins) and Biomarkers III selected 24 polypeptides (11 overexpressed and 13 subexpressed) (Fig. 2 in Ref [29]). All these potential biomarkers were in common in all breast cancer cell lines in the present study.

In the case of PubMed search, we obtained 35 overexpressed (Fig. 2A in Ref [29]) and 53 subexpressed polypeptides (Fig. 2B in Ref [29]) not previously reported in breast cancer, with one exception, ANXA8, which has been reported once [30]. These analyses revealed only 3 specific overexpressed polypeptides in Biomarkers I and 34 in the PubMed search (Fig. 2A in Ref [29]). For subexpressed proteins, we obtained 6 polypeptides for Biomarkers I and 51 in PubMed (Fig. 2B in Ref [29]). Biomarkers II and III modules found non-exclusive markers in both cases.

3.6. Validation of selected polypeptides by Western blot

With all this information, we decided to validate the BAG6 and DDX39 as overexpressed polypeptides and ANXA8 and COX4 as subexpressed proteins by western blot, as they have not been reported in breast cancer (Fig. 9). The overexpression level for BAG6 and DDX39 obtained with iTRAQ labelling and mass spectrometry were corroborated by western blot.

Although these values did not match perfectly with those obtained by iTRAQ, they displayed a tendency to be overexpressed (Fig. 9A and B). The same behaviour was also obtained with ANXA8 and COX4 in western blot (Fig. 9C and D). Remarkably, in the western blots of BAG6, DDX39 and ANXA8, we detected additional bands of different molecular masses, including in extracts from control cell lines as recommended by suppliers. These new bands could be due to posttranslational modifications of polypeptides.

3.7. Functional classification of the proteins differentially expressed in Luminal A, Claudin-low and HER2⁺ breast cancer cell lines

The overexpressed and subexpressed specific proteins found in each breast cancer cell line (Figs. 3 and 4 in Ref [29]) were classified according to their molecular classification into Protein Class and Biological Processes with PANTHER. For overexpressed polypeptides we obtained 25 categories, being nucleic acid binding class (XIV) the most representative category for Luminal A and Claudin-low cell lines, whilst hydrolase (VIII) and oxidoreductase (XV) classes were the most representatives in the HER2⁺ cell line (Fig. 10A). In the case of the Biological Processes, we found 12 categories, being the metabolic process the category with more genes implicated in all breast cancer cell lines (Fig. 10B). On the other hand, the categorization of subexpressed proteins specific for

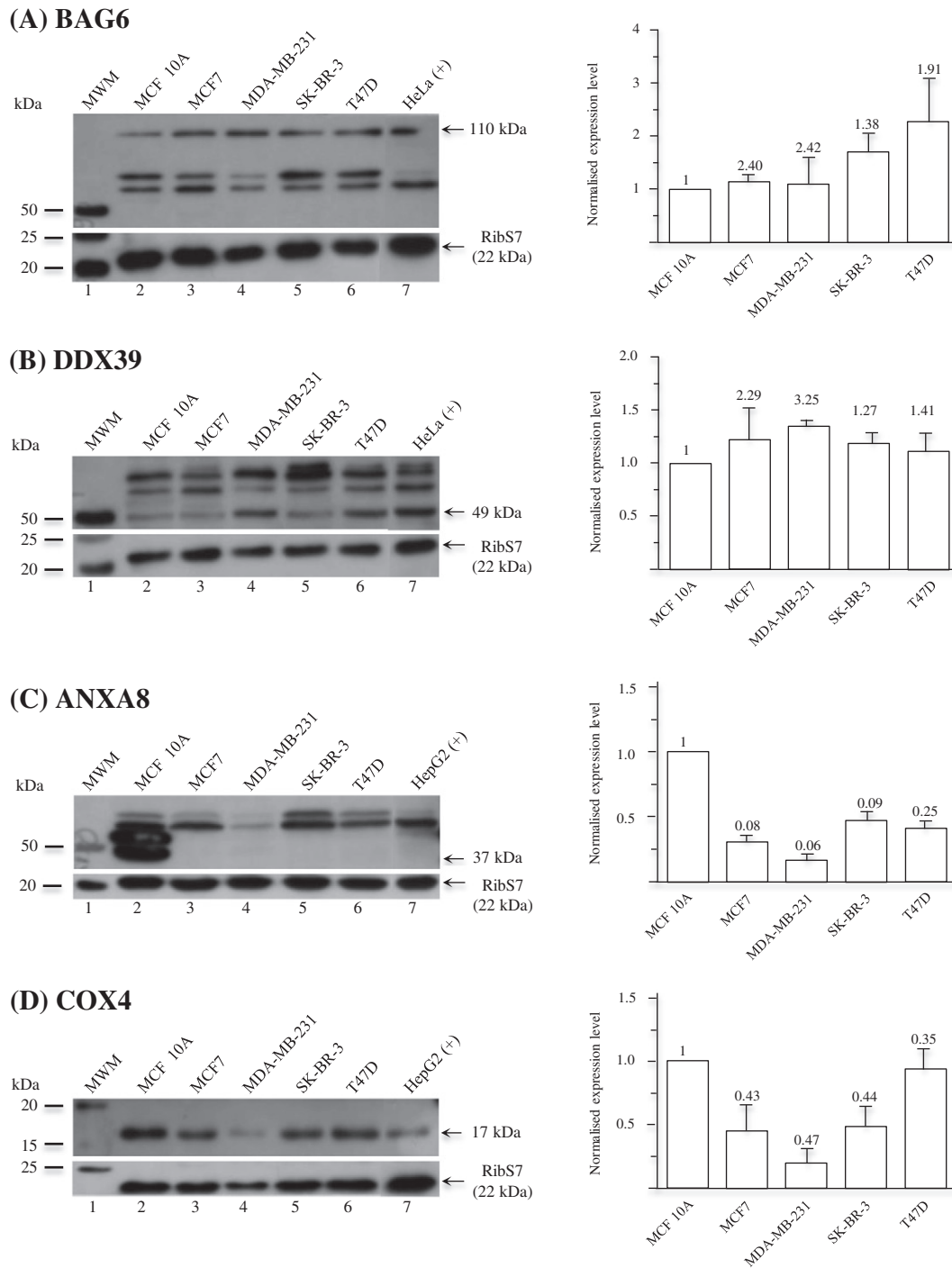


Fig. 9 – Western blot analyses of candidate protein biomarkers in breast cancer. Western blots and densitometry analysis of the overexpressed proteins BAG6 (A) and DDX39 (B) and subexpressed polypeptides ANXA8 (C) and COX4 (D) for the breast cancer cell lines compared to a control cell line. Densitometry values were first normalised to the loading control RibS7 protein and then to the values obtained in the control cell line. Bars correspond to standard deviation. Numbers above bars are values obtained with iTRAQ.

the different molecular types showed 25 categories in Protein Class and 11 in Biological Processes. Protein Classes with the mayor number of genes were nucleic acid binding (XIV) for Claudin-low and HER2⁺, whilst hydrolase (VIII) and cytoskeletal protein (V) classes were found for Luminal A cell lines

(Fig. 10C). Finally, the most affected Biological Process was the metabolic process (VIII) in all types of breast cancer cell lines (Fig. 10D). Furthermore, we also performed a PANTHER analysis of the exclusive sets of proteins found in each breast cancer cell line (Figs. 5 and 6 in Ref [29]).

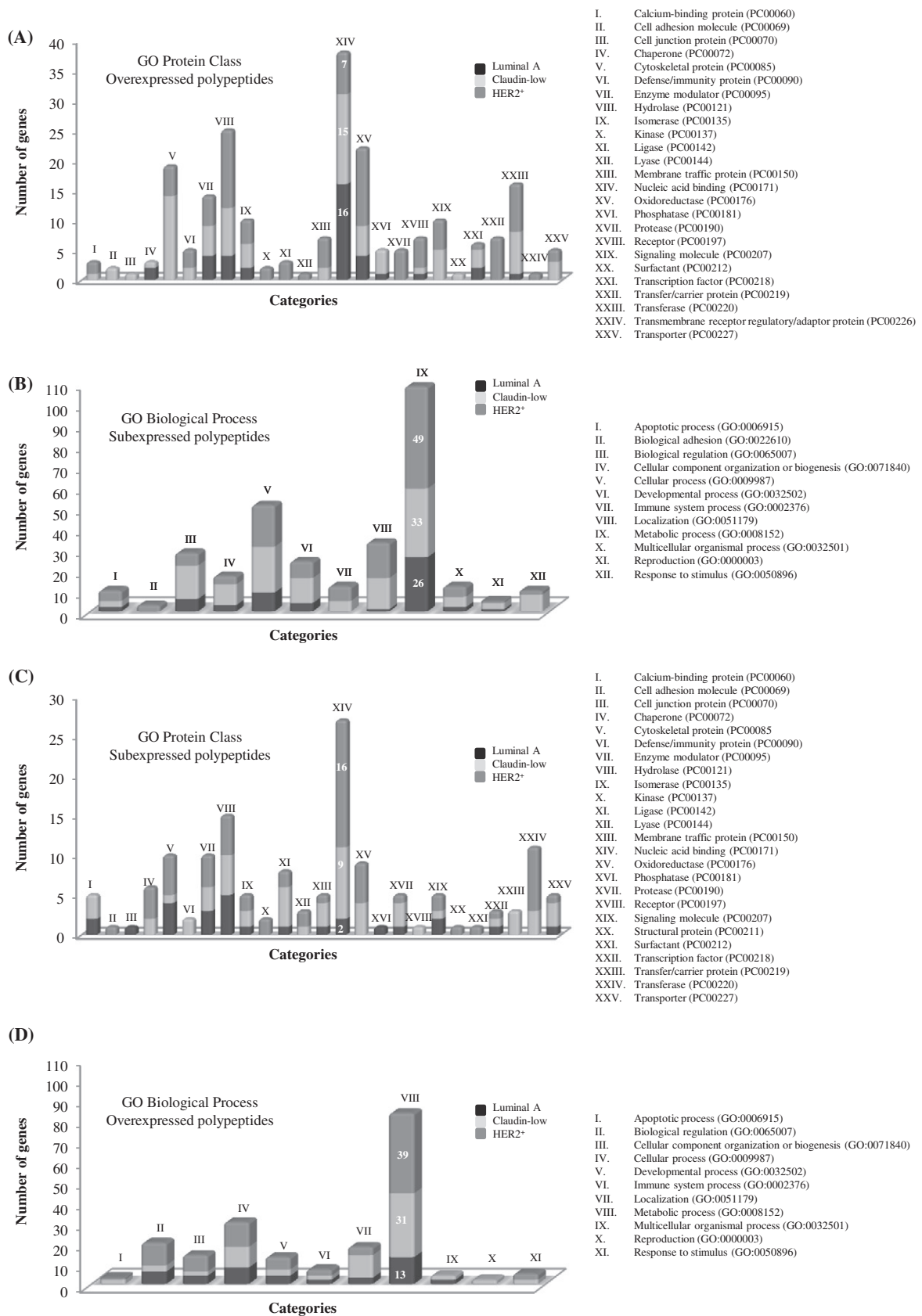
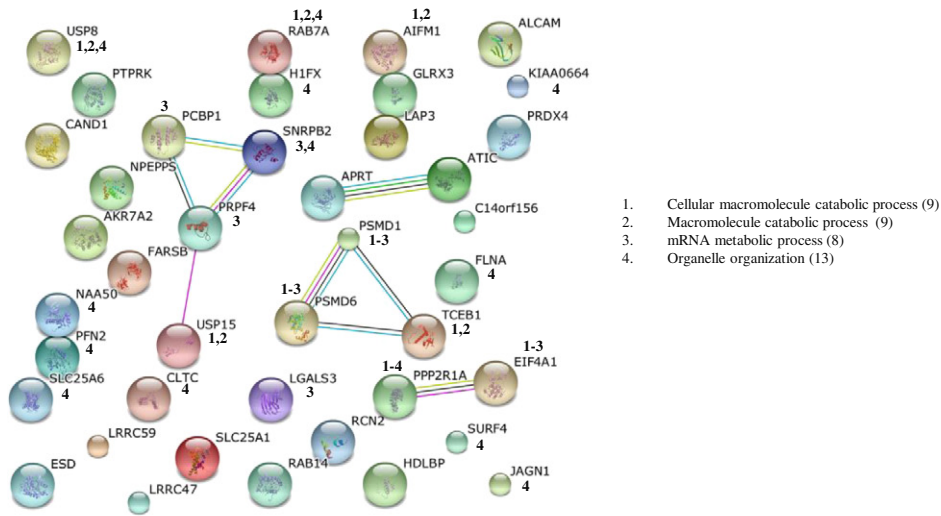


Fig. 10 – PANTHER functional classification of overexpressed and subexpressed proteins exclusively found in Luminal A, Claudin-low and HER2⁺ breast cancer cell lines. GO Protein Class for overexpressed (A) and subexpressed (C) polypeptides. GO Biological Process for overexpressed (B) and subexpressed (D) polypeptides. Luminal A (MCF7 and T47D); Claudin-low (MDA-MB-231); HER2⁺ (SK-BR-3).

(A) MCF7



1. Cellular macromolecule catabolic process (9)
2. Macromolecule catabolic process (9)
3. mRNA metabolic process (8)
4. Organelle organization (13)

(B) T47D

1. Protein localization to endoplasmic reticulum (9)
2. Viral transcription (8)
3. Translational termination (8)
4. Viral gene expression (8)
5. Multi-organism metabolic process (8)
6. SRP-dependent cotranslational protein targeting to membrane (8)
7. Cotranslational protein targeting to membrane (8)
8. Protein targeting to ER (8)
9. Establishment of protein localization to endoplasmic reticulum (8)
10. Cellular protein complex disassembly (8)
11. Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (8)
12. Protein complex disassembly (8)
13. Translational initiation (8)
14. Macromolecular complex disassembly (8)
15. Protein targeting to membrane (8)
16. Viral life cycle (8)
17. Nuclear-transcribed mRNA catabolic process (8)
18. mRNA catabolic process (8)
19. Cytoplasmic transport (12)
20. Translational elongation (7)
21. RNA catabolic process (8)
22. Establishment of protein localization to membrane (8)
23. Nucleobase-containing compound catabolic process (13)
24. Intracellular protein transport (11)
25. Protein localization to organelle (10)
26. Aromatic compound catabolic process (13)
27. Heterocycle catabolic process (13)
28. Cellular nitrogen compound catabolic process (13)
29. mRNA metabolic process (11)
30. Intracellular transport (14)
31. Organic cyclic compound catabolic process (13)
32. Establishment of localization in cell (16)
33. Protein localization to membrane (8)
34. Cellular component disassembly (9)
35. Single-organism membrane organization (10)
36. Macromolecule catabolic process (12)
37. Catabolic process (18)
38. Membrane organization (11)
39. Organic substance catabolic process (17)
40. Protein complex subunit organization (14)
41. Establishment of protein localization to organelle (8)
42. Cellular protein localization (12)
43. Single-organism cellular localization (8)
44. Single-organism localization (8)
45. Cellular macromolecule localization (12)
46. Cellular localization (16)
47. Protein targeting (8)
48. Ribosome biogenesis (6)
49. Regulation of translation (7)

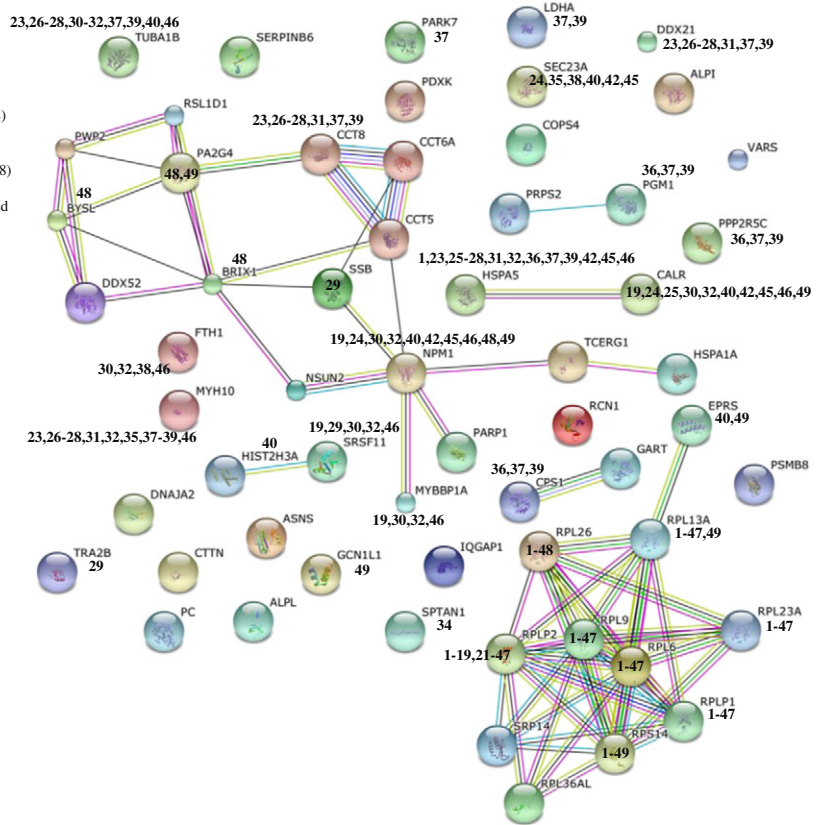


Fig. 11 – Interactome of overexpressed unique polypeptides in each Luminal A breast cancer cell lines. Interactomes show the predicted functional interaction networks of MCF7 (A) and T47D (B) breast cancer cell lines. The coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Networks also show some proteins, with statistical significance (P-value ≤ 0.05, bold numbers), involved in different biological processes.

3.8. Protein-protein interaction networks of the differentially expressed proteins only found in Luminal A, Claudin-low or HER2⁺ breast cancer cell lines

Interaction network maps revealed that in the case of the exclusively overexpressed polypeptides found in MCF7 cells, the main node was involved in mRNA metabolic processes (Fig. 11A), whilst those of T47D showed two main nodes, the first one mainly related to ribosome biogenesis, whilst the second had a relationship to a number of different functions, including intracellular protein transport, protein targeting to endoplasmic reticulum, among others (Fig. 11B). When we analysed the interactome of the overexpressed proteins in common in Luminal A cell lines, the main node was mainly involved with mRNA processing (Fig. 12A). In the case of the unique overexpressed proteins found in the Claudin-low MDA-MB-231 cell line, the principal node was related to several processes, including the ribonucleoprotein complex assembly, formation of the translation preinitiation complex, and regulation of translational initiation (Fig. 12B). Finally, The network of the unique set of overexpressed polypeptides found in the HER2⁺ cell line was mainly involved in small molecule metabolic processes (Fig. 12C).

The analysis of the interactome of the exclusive subexpressed proteins of MCF7, did not reveal any significant node (Fig. 13A). On the contrary, T47D had a node related to different processes, such as regulation of actin polymerization/depolymerization, actin filament length and actin cytoskeleton organisation, and activation of immune process (Fig. 13B). The subexpressed proteins shared by Luminal A cells did not show any significant node (Fig. 14A). However, the Claudin-low interactome revealed a small node related to 10 biological processes, including intracellular protein transport and membrane organization (Fig. 14B). Finally, the HER2⁺ network showed two main nodes, one implicated in almost all biological processes and other mainly related to ribosome biogenesis (Fig. 14C).

4. Discussion

We found 78 overexpressed proteins and 128 subexpressed polypeptides in common for all breast cancer cell lines. These proteins were analysed with the IPA software in the module of Biomarker filters for the selection of some putative biomarkers in breast cancer. Through IPA, we established three Biomarker filters: Biomarkers I (all tissues, primary cells and all cell lines), Biomarkers II (mammary gland and breast cancer cell lines) and Biomarkers III (all cancer cell lines excluding breast cancer cell lines). The Biomarkers filter I was the only one in identifying proteins unshared with the other filters (Fig. 2 in Ref [29]). Some of the selected proteins with this filter are detectable in tissues like heart, lung and kidney, as well as in cell lines derived from central nervous system, colon, breast, kidney, lung, myeloma, ovary and prostate, among others.

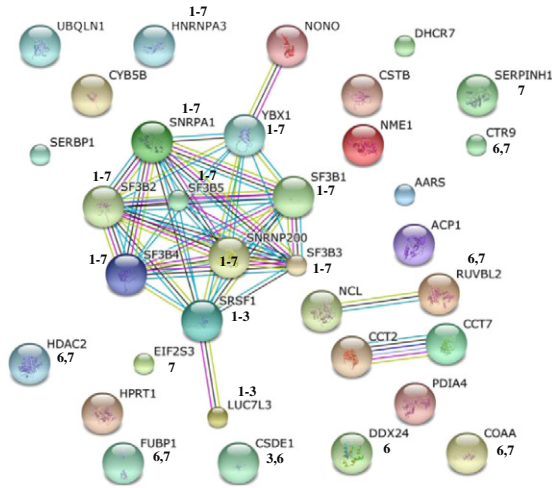
Furthermore, there have been developed some specific inhibitory drugs for some of these polypeptides proposed in this work by IPA as putative biomarkers, including: CPT1A (perhexiline), DNMT1 (5-azacytidine, decitabine), HSP90AA1

(17-dimethylaminoethylamino-17-demethoxygeldanamycin, IPI-504, cisplatin), NNMT (nicotinic acid). Noteworthy that we did not find exclusive proteins belonging to the Biomarkers filter II and that some of the putative biomarkers proposed by IPA have been extensively studied. For this reason, we aimed to find proteins that have not been previously reported in breast cancer through a search in PubMed database. In this analysis, we found 35 overexpressed and 53 subexpressed polypeptides (Fig. 2 in Ref [29]). We selected the BAG6 and DDX39 overexpressed proteins and the ANXA8 and COX4 subexpressed polypeptides for validation of their expression levels in all cell lines by western blot (Fig. 9).

The BCL2-Associated Athanogene 6 (BAG6) is a large proline-rich protein of 1,132 residues in length, which contains an N-terminal ubiquitin-like domain, two proline-rich domains, a domain of unknown function (DUF3538), a caspase-3 cleavage site, a nuclear localization signal and a conserved BAG domain in the C-terminus [31,32]. Several reports have suggested a role of BAG6 (also called Scythe/BAT3) in apoptosis [33-35]. The interaction of the Immediate Early gene X-1 (IEX-1) with BAG6 and other proteins modulate apoptosis and cellular growth [33]. The inactivation of BAG6 in mice showed strong developmental defects in the lung, kidney and brain, leading to embryonic lethality. These defects were related to increased apoptosis and cellular proliferation [34]. A high level of Papillomavirus Binding Factor (PBF) is associated with poor prognosis in osteosarcoma. The BAG6-PBF complex inhibits apoptosis through a caspase-9 independent pathway. Interestingly, BAG6 was localised to the cytoplasm in normal tissue, in contrast to its nuclear colocalisation with PBF in osteosarcoma cells [35]. Another role of BAG6 is related to the cell cycle progression. In this case, BAG6 is capable of switching on pro- and antiproliferative roles of p21 during the cell cycle, thus regulating its activity, abundance and intracellular localisation to facilitate G1/S transition and G2/M progression [36]. Moreover, in the absence of BAG6, the complex p300-p53 is not formed, and acetylation of p53 through p300 is abrogated, affecting the functions of p53 target genes, including p21 and PUMA [37]. Remarkably, BAG6 has been involved in the immunosurveillance of tumor cells through its release from them in response to cellular stress. It interacts with NKp30 receptor present in NK cells through a binding site located in residues 686-936 of BAG6. This novel domain is sufficient for docking to NKp30 receptor and inhibition of NK cell cytotoxicity [38-40].

DDX39 belongs to the DEAD box RNA helicase family and is upregulated in human pancreatic and pleural mesothelial malignant cells [41,42], as well as in lung squamous cell cancer and urinary bladder cancer [43,44]. There has been revealed that DDX39 interacts with components of the RNA splicing and transcription machineries such as ALY, CIP29 and FUS/TLS [45]. Overexpression of DDX39 accelerates cell proliferation of HeLa cells, which suggests a significant role in cancer pathogenesis [44]. However, DDX39 may play a role as a tumor suppressor gene as well. When DDX39 was knocked down in bladder cancer cells, an increased ability of invasion was observed in them in comparison to control cells [43]. Studies on patients have shown that DDX39 expression was inversely proportional to disease-free survival rates in gastrointestinal stromal tumors [46,47], indicating a bad prognosis in patients with an elevated level of DDX39.

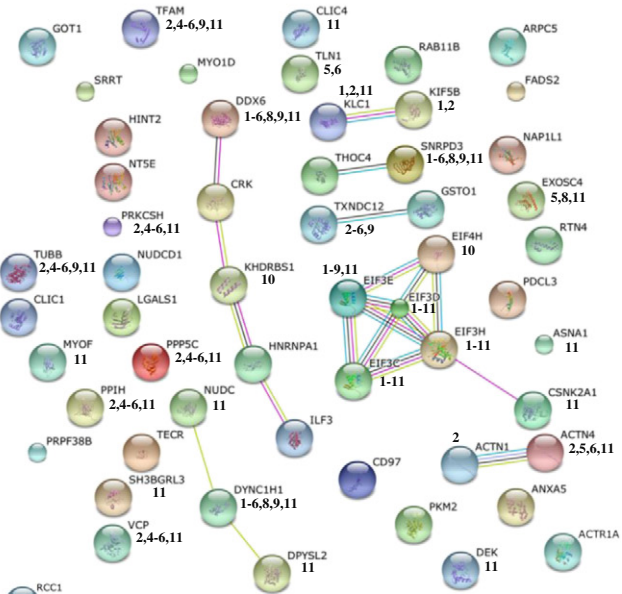
(A) Luminal A (MCF7 and T47D)



1. mRNA splicing, via spliceosome (11)
2. mRNA processing (11)
3. mRNA metabolic process (12)
4. RNA splicing (9)
5. RNA processing (9)
6. RNA metabolic process (16)
7. Gene expression (16)

1. Ribonucleoprotein complex subunit organization (9)
2. Macromolecular complex subunit organization (18)
3. Ribonucleoprotein complex assembly (7)
4. Macromolecular complex assembly (14)
5. Cellular component biogenesis (17)
6. Cellular component assembly (16)
7. Formation of translation preinitiation complex (4)
8. Ribonucleoprotein complex biogenesis (8)
9. Cellular macromolecular complex assembly (10)
10. Regulation of translational initiation (5)
11. Cellular component organization or biogenesis (24)

(B) Claudin-low (MDA-MB-231)



1. Oxoacid metabolic process (15)
2. Organic acid metabolic process (15)
3. Cellular lipid catabolic process (7)
4. Small molecule metabolic process (23)
5. Carboxylic acid metabolic process (13)

(C) HER2+ (SK-BR-3)

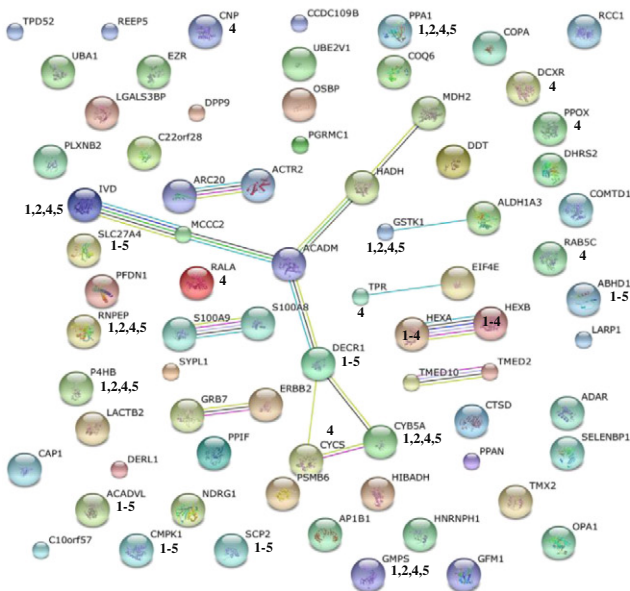
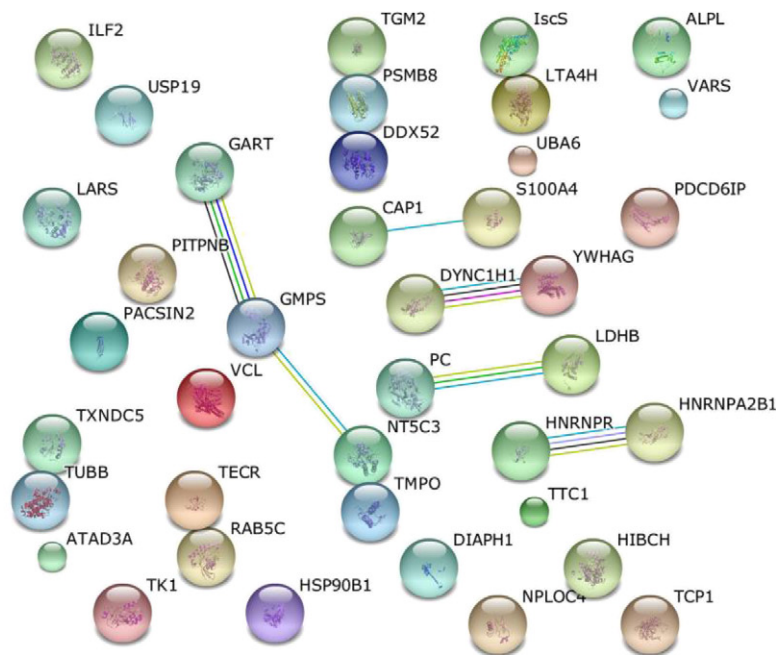
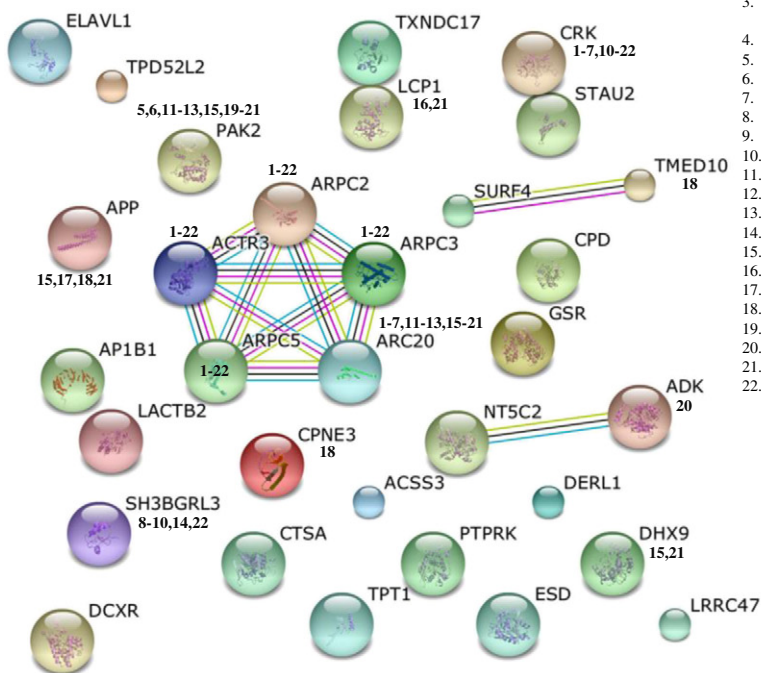


Fig. 12 – Interactomes of specific overexpressed proteins found in Luminal A, Claudin-low and HER2+ breast cancer cell lines. (A) Polypeptide interaction map of overexpressed polypeptides shared by Luminal A cell lines (MCF7 and T47D). (B) Interactome of exclusive overexpressed polypeptides found in Claudin-low breast cancer cell line (MDA-MB-231). (C) Interaction network of exclusively overexpressed HER2+ breast cancer cell line (SK-BR-3). Coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Networks also show proteins with statistical significance (P -value ≤ 0.05 , bold numbers) involved in different biological processes.

(A) MCF7



(B) T47D



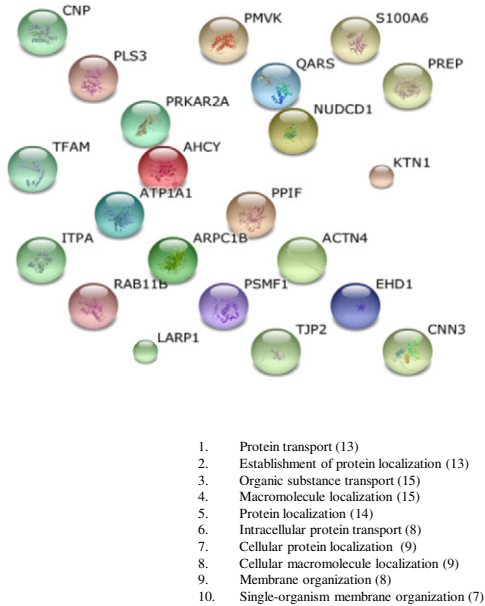
1. Fc-gamma receptor signaling pathway involved in phagocytosis (6)
2. Fc receptor mediated stimulatory signaling pathway (6)
3. Immune response-regulating cell surface receptor signaling pathway involved in phagocytosis (6)
4. Fc-gamma receptor signaling pathway (6)
5. Immune response-activating cell surface receptor signaling pathway (7)
6. Fc receptor signaling pathway (7)
7. Phagocytosis (6)
8. Regulation of actin polymerization or depolymerization (5)
9. Regulation of actin filament length (5)
10. Regulation of actin cytoskeleton organization (6)
11. Immune response-activating signal transduction (7)
12. Immune response-regulating cell surface receptor signaling pathway (7)
13. Activation of immune response (7)
14. Regulation of actin filament-based process (6)
15. Innate immune response (9)
16. Immune effector process (7)
17. Endocytosis (7)
18. Vesicle-mediated transport (9)
19. Immune response-regulating signaling pathway (7)
20. Positive regulation of immune system process (8)
21. Immune response (10)
22. Regulation of cytoskeleton organization (6)

Fig. 13 – Networks of subexpressed specific polypeptides found in Luminal A breast cancer cell lines. The interactome shows the predicted functional interaction of specific proteins of MCF7 (A) and T47D (B) cell lines. The coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Networks also show some proteins, with statistical significance (P -value ≤ 0.05 , bold numbers), involved in different biological processes.

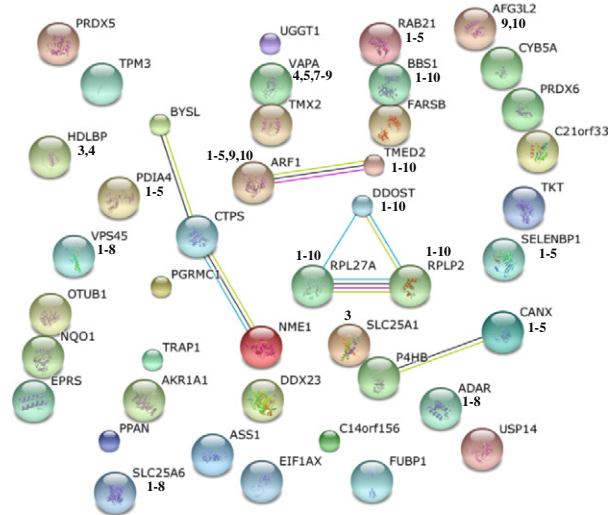
Spotlighting, DDX39 interacts with hTERT and TRF2 subunit of shelterin, where TRF2 has been implicated in the protection of chromosome ends [48].

ANXA8 belongs to the family of Ca^{2+} -dependent phospholipid binding proteins. The ANXA proteins can bind to some membrane phospholipids and form networks. Through these

(A) Luminal A (MCF7 and T47D)



(B) Claudin-low (MDA-MB-231)



(C) HER2+ (SK-BR-3)

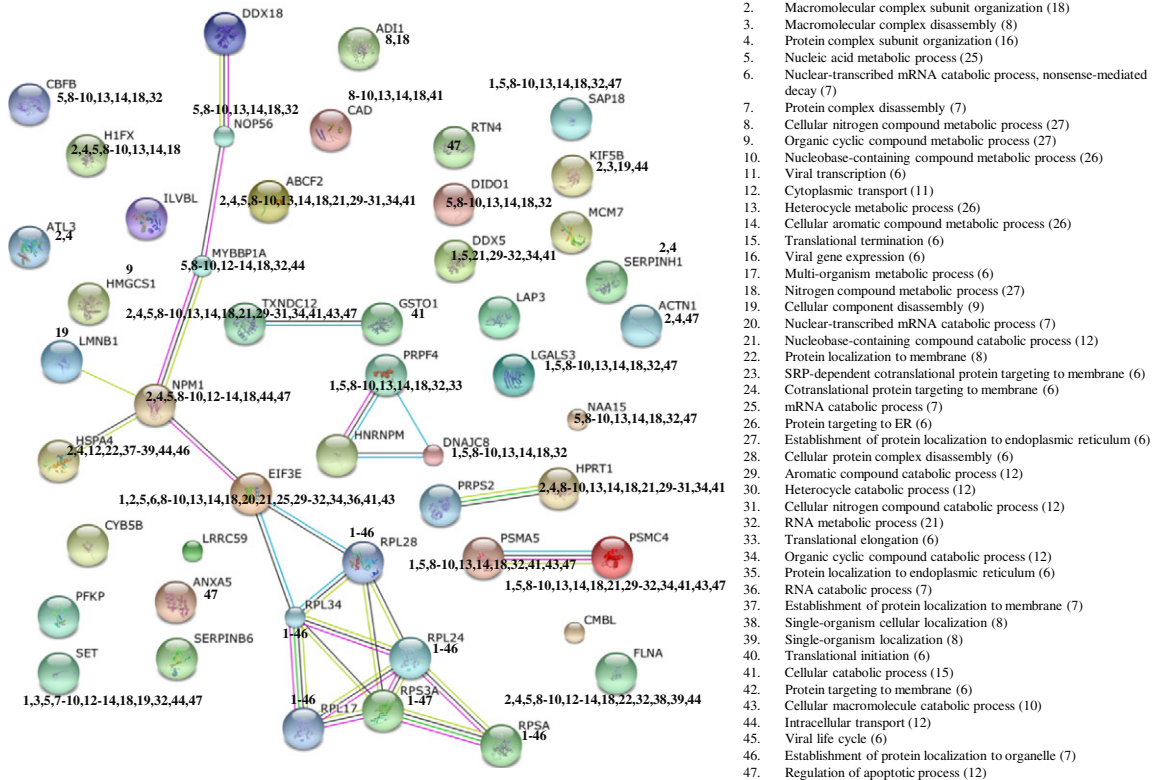


Fig. 14 – Networks of subexpressed proteins exclusively found in Luminal A, Claudin-low and HER2+ breast cancer cell lines. (A) Interactome of proteins shared by Luminal A breast cancer cell lines (MCF7 and T47D). (B) Interactome of exclusively subexpressed polypeptides found in Claudin-low breast cancer cell line (MDA-MB-231). (C) Interaction network of exclusively subexpressed HER2+ breast cancer cell line (SK-BR-3). Coloured lines represent different evidences for each interaction: red line, fusion; green line, neighbourhood; blue line, cooccurrence; purple line, experimental; yellow line, text mining; light blue line, database; black line, coexpression. Networks also show some proteins, with statistical significance (P-value ≤ 0.05, bold numbers), involved in different biological processes.

networks, annexins can function as organizers of membrane domains and as scaffolds for recruitment of proteins [49]. ANXA8 mRNA levels have been found increased in human pancreatic cells [50], and the inhibition of ANXA8 using siRNA in pancreatic cells reduced the number of migrating cells and also diminished the HIF-1 α mRNA levels [51]. Nevertheless, these findings contrast with our results that showed that ANXA8 was subexpressed in breast cancer cell lines (Fig. 9C). In agreement with our findings, an immunohistochemical analysis revealed ANXA8 in low expression levels in the epithelium of lobules and ducts, and in some basal cells as well as in normal human breast tissues. However, ANXA8 was found overexpressed at very low frequency in tumor cells [30].

Cytochrome c oxidase (COX) is the terminal enzyme of the mitochondrial respiratory chain. It is a multisubunit complex conformed by 13 subunits and its main function is to couple electron transfer from cytochrome c to molecular oxygen. Regulation of the cytochrome c oxidase subunit IV isoform 1 (COX4i1) has been associated to p73, a member of the p53 family. The control of the COX4i1 gene expression relies on two promoters, which allows the expression of TAp73 and Δ Np73, harbouring or not the transactivation domain, respectively. Therefore only TAp73 has the ability to transactivate COX4i1. Remarkably, Tap73-deficient mice, which had a reduction in the COX4i1 expression levels, showed mitochondrial dysfunction and increased levels of ROS, with consequently premature aging [52]. The role of COX4i1 in disease has been mainly associated to metabolic disorders such as type 2 diabetes and cancer. In the transgenic mice MKR that develops type 2 diabetes, a downregulation of multiple proteins involved in oxidative metabolism were observed in the pancreatic islets compared to the non-diabetic control mice [53]. Some of the affected proteins (COX2, COX4i1 and COX5A) are part of complex IV. In cancer, the mRNAs of COX4i1, COX5B and COX7C have been found upregulated mainly in cutaneous squamous cell carcinoma [54]. Besides, the subunit COX4i1 was found decreased in response to antitumoral treatments of adrenocortical carcinoma and other solid tumors by using Mitotane and Gemcitabine [55,56]. Other study showed that T47D cells exhibited a notably higher mitochondrial reserve capacity than MCF7 cells in the absence of serum hormones, which augmented after treatment with 4-hidroxytamoxifen. Under these conditions, a low level of COX4i1 was observed [57].

To our knowledge, there are no studies showing a relationship of BAG6 and DDX39 with breast cancer. In the case of ANXA8, this was mentioned only in one article so far. In this report, as well as in our work, ANXA8 is proposed as part of a markers panel in breast cancer. In accordance with the other report, we proposed ANXA8 as a putative biomarker in breast cancer disease. Regarding to COX4, it was recently observed that its expression is oestrogen sensitive in T47D cells [57], but has not been reported as a putative biomarker in breast cancer. In resume, BAG6, DDX39, ANXA8 and COX4 might be putative biomarkers in breast cancer, but the validation of these polypeptides in a wide cohort of tumors and serum from patients with breast cancer disease is required. We do not discard the other molecules identified in our study, but we need to validate them first as well.

Our global analysis showed that some proteins were involved in metabolic processes. The overexpressed proteins were altered in starch and sucrose metabolism, biosynthesis of amino acids, pentose phosphate pathway, purine metabolism, and glycolysis, among others (Fig. 7). On the other hand, most of the subexpressed polypeptides were implicated in different mitochondrial processes. The tricarboxylic acid cycle (TCA) generates vital products such as NADH and FADH for energy production through oxidative phosphorylation. Remarkably, we found subexpressed polypeptides that are critical for the generation of NADH and FADH (Fig. 7). Additionally to these findings, we determined that 24 polypeptides that are forming part of complexes I, II, III, IV and V of the respiratory chain, were subexpressed (Fig. 8). In consequence, the oxidative phosphorylation was severely affected. In summary, the altered processes are related with the Warburg effect, which is characterized by the generation of ATP through aerobic glycolysis instead of oxidative phosphorylation [58,59].

All analyses were based on the overexpressed and subexpressed polypeptides in common in all breast cancer cell lines in this study. Nevertheless, we identified sets of polypeptides that were unique for each breast cancer cell line (Figs. 11-14; Figs. 3 and 4 in Ref [29]). This result showed important differences in expression of proteins among cell lines. The similarity between some tumoral tissues and their corresponding cell lines has been established through gene expression profiles. A tissue similarity index (TSI) measures the similarity in gene expression between cell lines and tumor tissues. A high TSI score indicates a close similarity in their gene expression levels. MCF7 and T47D cell lines have high to medium TSI score whilst MDA-MB-231 has a high score [60]. For example, the sets of exclusive polypeptides found in each type of breast cancer cell line might be useful for the identification of each specific molecular class in tumors (Figs. 3, 11-14). However, thorough studies with tumor tissues are required to validate these assumptions.

The MDA-MB-231 Claudin-low breast cancer cell line is characterized by the ER⁺ PR⁺ HER2⁻ immunoprofile, lacks some tight junction proteins (E-cadherin, claudin-3, claudin-4 and claudin-7) and has a low expression level of Ki67 [26,61]. In addition, it has an intermediate response to chemotherapy [26] and is one of the breast cancer cell lines with the higher potential of invasion [62]. In relation to the tumors of the Claudin-low molecular class, they have the highest risk of mortality [63] and have shown an enrichment of markers enrolled in the epithelial-to-mesenchymal transition, genes of the immune response, and characteristics of stem cells [64], besides a higher activity in several signalling pathways, including EGFR, SRC and TGF β [65]. In this work, we found CLIC4, CRK and DEK as overexpressed proteins only in MDA-MB-231 cells (Fig. 12B). CLIC4 (Chloride intracellular channel protein 4) is involved in different cellular processes, including maintenance of intracellular membranes, apoptosis, cellular differentiation, inflammation and angiogenesis [66-68]. The reduction in the CLIC4 expression levels inhibited the growth of human osteosarcoma cells in vitro and in vivo, increased apoptosis, and decreased cell proliferation. CLIC4 was identified in metastatic cancer stem-like cells of colorectal cancer and was associated with a poor prognosis [69]. CRK (Adapter molecule crk) plays key roles in multiple signalling pathways, and is related to cell proliferation,

adhesion, migration, invasion, phagocytosis, and survival [70,71]. Overexpression of CrkI, CrkII, CrkL polypeptides have been observed in different types of cancer. A significant association between CRK protein expression and high proliferative tumors was observed in breast cancer [71]. In addition, there has been demonstrated that CRK proteins are important for cellular adhesion, migration, invasion and *in vivo* proliferative breast tumor outgrowth in the basal cell lines MDA-MB-231, 1833TR and SUM1315 [66]. The DEK protein is involved in DNA replication [72] and global heterochromatin integrity [73], among other processes. DEK protein has been found overexpressed in gastric adenocarcinoma [74] and in small cell lung cancer [75]. In breast cancer, this protein was overexpressed in invasive ductal carcinoma, where the high DEK levels correlated with an increased proliferation [76].

SK-BR-3 breast cancer cell line, which is classified as overexpressing HER2⁺, and ER⁻ PR⁻, has a low invasive potential [62], and is both responsive to trastuzumab and chemotherapy [26]. However, tumors displaying a phenotype similar to overexpressing HER2⁺ cells (ER⁻ PR⁻ HER2⁺) have shown a high risk of mortality [63]. We identified some overexpressed proteins that were exclusively overexpressed in the SK-BR-3 HER2⁺ cells (Fig. 12C), including SLC27A4, DECR1, DERL1, HEXB and TPD52. SLC27A4 (Long-Chain fatty acid transport 4) has been involved in the translocation of long chain fatty acids through the plasma membrane and has also been found upregulated in breast cancer [77]. DECR1 (2,4-dienoyl-CoA.reductase, mitochondrial) has a dual role in ErbB-2 positive cells; it restricts proliferation on Her2⁺ breast cancer cells and it might provide protection against apoptosis due to glucose withdrawal [78]. DERL1 is overexpressed when breast cancer cells are under endoplasmic reticulum stress, resulting in resistance to apoptosis [79]. In the case of HEXB (Beta-hexosamidase subunit beta), it has been involved in an antiangiogenic effect [80]. Finally, TPD52 has been found coexpressed with Her2/Neu breast cancer. It is upregulated in ERBB2-positive breast cancer cell lines and it contributes to survival of breast cancer cells [81–83].

The Luminal A breast cancer cell lines MCF7 and T47D have a differentiated morphology and form tight cell-cell junctions, they often are responsive to chemotherapy and have low potential of invasion [26,60]. Tumors having molecular characteristics similar to Luminal A breast cancer cells (ER⁺ PR⁺ HER2⁻) have shown a lower risk of mortality [61]. These cell lines also showed a group of exclusively overexpressed polypeptides (Fig. 12A), including UBQLN1, YBX1 and RUVBL2, among others. The loss of UBQLN1 (Ubiquilin-1) resulted in an increase in cell migration in non-small cell lung cancer cell lines, a significant subexpression of E-cadherin and claudin-1 epithelial markers, and the overexpression of vimentin, Snail and ZEV1 mesenchymal markers [84]. Thus, an overexpression of UBQLN1 might be indicative of the low potential of invasion of the Luminal A cell lines. In the case of RUVBL2 (RuvB-like 2), it has a role in chromatin remodelling and DNA damage repair. The siRNA knockdown of RUVBL2 causes sensibility to anticancer drugs [85]. Outstanding, RUVBL2 cooperates with Ets-2 in the positive transcriptional regulation of the human telomerase reverse transcriptase [86].

YBX1 (Y box binding protein 1 or YB-1), a member of the cold-shock protein family, is considered a type of nuclear-

cytoplasm shuttle protein [85]. It is involved in a number of functions, including transcriptional and translational regulation [86,87], cell growth, RNA splicing, DNA repair, proliferation, and drug resistance [88,89]. There has been shown that the YBX1 expression in prostate cancer cells leads to epithelial-to-mesenchymal transition, as well as migration and invasion. Thus, the mesenchymal morphology was correlated with a down-regulation of epithelial markers [90]. YB-1 is associated with cancer progression and poor patient prognosis. The detection of an 18 kDa YBX1 fragment in plasma was correlated with different malignancies [91]. In breast cancer, this fragment was found in 70% of the samples analysed [91]. In addition, the YBX1 expression in human mammary epithelial cells was associated with a self-renewal and cellular transformation caused by a series of epigenetic events. Remarkably, the coexpression of YBX1, RSK2 and hTERT in mice, led to tumour formation with characteristics related to basal-like cancer [92].

5. Conclusions

We found overexpressed proteins implicated in DNA topology, pentose pathway, regulation of translation initiation, proteins involved in degradation by proteasome, splicing and glycolysis in common in all breast cancer cell lines analysed. In relation to the subexpressed polypeptides found in common in all cells studied, they were involved in different mitochondrial processes, being the most affected oxidative phosphorylation. BAG6, DDX39, ANXA8 and COX4 might be putative biomarkers in breast cancer disease. However, they require an exhaustive validation in breast cancer tumor tissues or sera.

Authors declare that there is no interest conflict.

List of abbreviations

DMEM	Dulbecco's Modified Eagle Medium
RPMI	Roswell Park Memorial Institute
SFB	fetal bovine serum
EGF	epidermal growth factor
NEAA	non-essential amino acids
PBS	phosphate buffered saline
GuHCl	Guanidine hydrochloride
ABC	Ammonium bicarbonate
TEAB	Triethylammonium bicarbonate
TCEP	Tris-(2-carboxyethyl) phosphine
MMTS	Methyl methanethiosulfonate
TFA	Trifluoroacetic acid
FA	Formic acid
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate

Acknowledgements

We thank to Consejo Nacional de Ciencia y Tecnología (Conacyt) and the Instituto de Ciencia y Tecnología del

Distrito Federal (ICyTDF) (now transformed in Secretaría de Ciencia Tecnología e Innovación (SECITI) del Distrito Federal), both from Mexico, for the financial support to Dr. Juan Pedro Luna Arias for doing this work. We thank Conacyt for the scholarship to Karla G. Calderón-González during her Ph.D. studies, and the ECOS Mexico-France program for the financial support to Karla G. Calderón-González for a short academic stay in Lille, France. The Unidad de Proteómica from the Centro de Investigación Príncipe Felipe belongs to ProteoRed, PRB²-ISCIII.

REFERENCES

- [1] GLOBOCAN. Estimated cancer incidence, mortality and prevalence worldwide in 2012. International Agency for Research on Cancer World Health Organization; 2012 [<http://globocan.iarc.fr/Default.aspx>].
- [2] Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
- [3] Hamdan M. In: Hamdan MH, editor. *Cancer biomarkers: Analytical techniques for discovery*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2007. p. 1–20.
- [4] van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- [5] Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value? *Clin Chem* 2006;52:345–51.
- [6] Carney WP, Bernhardt D, Jasani B. Circulating HER2 extracellular domain: A specific and quantitative biomarker of prognostic value in all breast cancer patients? *Biomark Cancer* 2013;5:31–9.
- [7] Ha JH, Seong MK, Kim EK, Lee JK, Seol H, Lee JY, et al. Serial serum HER2 measurements for the detection of breast cancer recurrence in HER2-positive patients. *J Breast Cancer* 2014;17: 33–9.
- [8] Lam L, Czerniecki BJ, Fitzpatrick E, Xu S, Schuchter L, Xu X, et al. Interference-free HER2 ECD as a Serum Biomarker in Breast Cancer. *J Mol Biomark Diagn* 2014;4:151.
- [9] Di Gioia D, Dresse M, Mayr D, Nagel D, Heinemann V, Stieber P. Serum HER2 in combination with CA 15-3 as a parameter for prognosis in patients with early breast cancer. *Clin Chim Acta* 2015;440:16–22.
- [10] Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004;5:699–711.
- [11] Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. *J Cell Biol* 2010;190:491–500.
- [12] Goncalves A, Bertucci F. Clinical application of proteomics in breast cancer: state of the art and perspectives. *Med Princ Pract* 2011;20:4–18.
- [13] Liang X, Zhao J, Hajivandi M, Wu R, Tao J, Amshey JW, et al. Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma. *J Proteome Res* 2006;5:2632–41.
- [14] Leth-Larsen R, Lund R, Hansen HV, Laenkholm AV, Tarin D, Jensen ON, et al. Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Mol Cell Proteomics* 2009;8:1436–49.
- [15] Lund R, Leth-Larsen R, Jensen ON, Ditzel HJ. Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. *J Proteome Res* 2009;8: 3078–90.
- [16] Xu X, Qiao M, Zhang Y, Jiang Y, Wei P, Yao J, et al. Quantitative proteomics study of breast cancer cell lines isolated from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* 2010;10:1374–90.
- [17] Geiger T, Madden SF, Gallagher WM, Cox J, Mann M. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Res* 2012;72:2428–39.
- [18] Villarreal L, Mendez O, Salvans C, Gregori J, Baselga J, Villanueva J. Unconventional secretion is a major contributor of cancer cell line secretomes. *Mol Cell Proteomics* 2013;12: 1046–60.
- [19] Lai TC, Chou HC, Chen YW, Lee TR, Chan HT, Shen HH, et al. Secretomic and proteomic analysis of potential breast cancer markers by two-dimensional differential gel electrophoresis. *J Proteome Res* 2010;9:1302–22.
- [20] Schulz DM, Bollner C, Thomas G, Atkinson M, Esposito I, Hofler H, et al. Identification of differentially expressed proteins in triple-negative breast carcinomas using DIGE and mass spectrometry. *J Proteome Res* 2009;8:3430–8.
- [21] Huang HL, Stasyk T, Morandell S, Dieplinger H, Falkensammer G, Griesmacher A, et al. Biomarker discovery in breast cancer serum using 2-D differential gel electrophoresis/ MALDI-TOF/TOF and data validation by routine clinical assays. *Electrophoresis* 2006; 27:1641–50.
- [22] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004;3:1154–69.
- [23] Ho J, Kong JW, Choong LY, Loh MC, Toy W, Chong PK, et al. Novel breast cancer metastasis-associated proteins. *J Proteome Res* 2009;8:583–94.
- [24] Bouchal P, Roumeliotis T, Hrstka R, Nenutil R, Vojtesek B, Garbis SD. Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J Proteome Res* 2009;8:362–73.
- [25] Shaheed SU, Rustogi N, Scally A, Wilson J, Thygesen H, Loizidou MA, et al. Identification of stage-specific breast markers using quantitative proteomics. *J Proteome Res* 2013; 12:5696–708.
- [26] Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. *Breast Cancer Res* 2011;13:215.
- [27] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;72:248–54.
- [28] Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2013;41:D377–86.
- [29] Calderón-González KG, Valero Rustarazo ML, Labra-Barrios ML, Bazán-Méndez CI, Tavera-Tapia A, Herrera-Aguirre ME, Sánchez del Pino MM, et al. Data set of the protein expression profiles of Luminal A, Claudin-low and overexpressing HER2⁺ breast cancer cell lines by iTRAQ labelling and tandem mass spectrometry. *Data in Brief* 2015 [in press].
- [30] Stein T, Price KN, Morris JS, Heath VJ, Ferrier RK, Bell AK, et al. Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer. *Clin Cancer Res* 2005;11:6872–9.
- [31] Kumar R, Lutz W, Frank E, Im HJ. Immediate early gene X-1 interacts with proteins that modulate apoptosis. *Biochem Biophys Res Commun* 2004;323:1293–8.
- [32] Desmots F, Russell HR, Lee Y, Boyd K, McKinnon PJ. The reaper-binding protein scythe modulates apoptosis and proliferation during mammalian development. *Mol Cell Biol* 2005;25:10329–37.
- [33] Tsukahara T, Kimura S, Ichimiya S, Torigoe T, Kawaguchi S, Wada T, et al. Scythe/BAT3 regulates apoptotic cell death

- induced by papillomavirus binding factor in human osteosarcoma. *Cancer Sci* 2009;100:47–53.
- [34] Yong ST, Wang XF. A novel, non-apoptotic role for Scythe/BAT3: a functional switch between the pro- and anti-proliferative roles of p21 during the cell cycle. *PLoS One* 2012;7:e38085.
- [35] Sasaki T, Gan EC, Wakeham A, Kornbluth S, Mak TW, Okada H. HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53. *Genes Dev* 2007;21:848–61.
- [36] Pogge von Strandmann E, Simhadri VR, von Tresckow B, Sasse S, Reiners KS, Hansen HP, et al. Human leukocyte antigen-B-associated transcript 3 is released from tumor cells and engages the Nkp30 receptor on natural killer cells. *Immunity* 2007;27:965–74.
- [37] Simhadri VR, Reiners KS, Hansen HP, Topolar D, Simhadri VL, Nohroudi K, et al. Dendritic cells release HLA-B-associated transcript-3 positive exosomes to regulate natural killer function. *PLoS One* 2008;3:e3377.
- [38] Binici J, Hartmann J, Herrmann J, Schreiber C, Beyer S, Guler G, et al. A soluble fragment of the tumor antigen BCL2-associated athanogene 6 (BAG-6) is essential and sufficient for inhibition of Nkp30 receptor-dependent cytotoxicity of natural killer cells. *J Biol Chem* 2013;288:34295–303.
- [39] Kuramitsu Y, Tominaga W, Baron B, Tokuda K, Wang Y, Kitagawa T, et al. Up-regulation of DDX39 in human malignant pleural mesothelioma cell lines compared to normal pleural mesothelial cells. *Anticancer Res* 2013;33:2557–60.
- [40] Kuramitsu Y, Suenaga S, Wang Y, Tokuda K, Kitagawa T, Tanaka T, et al. Up-regulation of DDX39 in human pancreatic cancer cells with acquired gemcitabine resistance compared to gemcitabine-sensitive parental cells. *Anticancer Res* 2013;33:3133–6.
- [41] Kato M, Wei M, Yamano S, Kakehashi A, Tamada S, Nakatani T, et al. DDX39 acts as a suppressor of invasion for bladder cancer. *Cancer Sci* 2012;103:1363–9.
- [42] Sugiura T, Nagano Y, Noguchi Y. DDX39, upregulated in lung squamous cell cancer, displays RNA helicase activities and promotes cancer cell growth. *Cancer Biol Ther* 2007;6:957–64.
- [43] Sugiura T, Sakurai K, Nagano Y. Intracellular characterization of DDX39, a novel growth-associated RNA helicase. *Exp Cell Res* 2007;313:782–90.
- [44] Kikuta K, Kubota D, Saito T, Orita H, Yoshida A, Tsuda H, et al. Clinical proteomics identified ATP-dependent RNA helicase DDX39 as a novel biomarker to predict poor prognosis of patients with gastrointestinal stromal tumor. *J Proteomics* 2012;75:1089–98.
- [45] Kubota D, Okubo T, Saito T, Suehara Y, Yoshida A, Kikuta K, et al. Validation study on pftin and ATP-dependent RNA helicase DDX39 as prognostic biomarkers in gastrointestinal stromal tumour. *Jpn J Clin Oncol* 2012;42:730–41.
- [46] Yoo HH, Chung IK. Requirement of DDX39 DEAD box RNA helicase for genome integrity and telomere protection. *Aging Cell* 2011;10:557–71.
- [47] Gerke V, Creutz CE, Moss SE. Annexins: linking Ca²⁺ signalling to membrane dynamics. *Nat Rev Mol Cell Biol* 2005;6:449–61.
- [48] Karanjawala ZE, Illei PB, Ashfaq R, Infante JR, Murphy K, Pandey A, et al. New markers of pancreatic cancer identified through differential gene expression analyses: claudin 18 and annexin A8. *Am J Surg Pathol* 2008;32:188–96.
- [49] Hata H, Tatemichi M, Nakadate T. Involvement of annexin A8 in the properties of pancreatic cancer. *Mol Carcinog* 2014;53:181–91.
- [50] Rufini A, Niklison-Chirou MV, Inoue S, Tomasini R, Harris IS, Marino A, et al. TAp73 depletion accelerates aging through metabolic dysregulation. *Genes Dev* 2012;26:2009–14.
- [51] Lu H, Yang Y, Allister EM, Wijesekera N, Wheeler MB. The identification of potential factors associated with the development of type 2 diabetes: a quantitative proteomics approach. *Mol Cell Proteomics* 2008;7:1434–51.
- [52] Dang C, Gottschling M, Manning K, O'Curra E, Schneider S, Sterry W, et al. Identification of dysregulated genes in cutaneous squamous cell carcinoma. *Oncol Rep* 2006;16:513–9.
- [53] Hescot S, Slama A, Lombes A, Paci A, Remy H, Leboulleux S, et al. Mitotane alters mitochondrial respiratory chain activity by inducing cytochrome c oxidase defect in human adrenocortical cells. *Endocr Relat Cancer* 2013;20:371–81.
- [54] van de Wiel MA, Costa JL, Smid K, Oudejans CB, Bergman AM, Meijer GA, et al. Expression microarray analysis and oligo array comparative genomic hybridization of acquired gemcitabine resistance in mouse colon reveals selection for chromosomal aberrations. *Cancer Res* 2005;65:10208–13.
- [55] Radde BN, Ivanova MM, Mai HX, Salabei JK, Hill BG, Klinge CM. Bioenergetic differences between MCF-7 and T47D breast cancer cells and their regulation by oestradiol and tamoxifen. *Biochem J* 2014;465:49–61.
- [56] Warburg O. On the origin of cancer cells. *Science* 1956;123:309–14.
- [57] Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009;324:1029–33.
- [58] Sandberg R, Ernberg I. Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). *Proc Natl Acad Sci U S A* 2005;102:2052–7.
- [59] Herschkowitz JJ, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 2007;8:R76.
- [60] Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 2006;10:515–27.
- [61] Parise CA, Caggiano V. Breast cancer survival defined by the ER/PR/HER2 subtypes and a surrogate classification according to tumor grade and immunohistochemical biomarkers. *J Cancer Epidemiol* 2014;2014:469251.
- [62] Sabatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, et al. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Mol Cancer* 2014;13:228.
- [63] Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 2011;121:2750–67.
- [64] Littler DR, Harrop SJ, Goodchild SC, Phang JM, Mynott AV, Jiang L, et al. The enigma of the CLIC proteins: Ion channels, redox proteins, enzymes, scaffolding proteins? *FEBS Lett* 2010;584:2093–101.
- [65] Padmakumar V, Masiuk KE, Luger D, Lee C, Coppola V, Tessarollo L, et al. Detection of differential fetal and adult expression of chloride intracellular channel 4 (CLIC4) protein by analysis of a green fluorescent protein knock-in mouse line. *BMC Dev Biol* 2014;14:24.
- [66] Tung JJ, Hobert O, Berryman M, Kitajewski J. Chloride intracellular channel 4 is involved in endothelial proliferation and morphogenesis in vitro. *Angiogenesis* 2009;12:209–20.
- [67] Deng YJ, Tang N, Liu C, Zhang JY, An SL, Peng YL, et al. CLIC4, ERp29, and Smac/DIABLO derived from metastatic cancer stem-like cells stratify prognostic risks of colorectal cancer. *Clin Cancer Res* 2014;20:3809–17.
- [68] Feller SM. Crk family adaptors-signalling complex formation and biological roles. *Oncogene* 2001;20:6348–71.
- [69] Fathers KE, Bell ES, Rajadurai CV, Cory S, Zhao H, Mourskaia A, et al. Crk adaptor proteins act as key signaling integrators for breast tumorigenesis. *Breast Cancer Res* 2012;14:R74.

- [70] Kappes F, Waldmann T, Mathew V, Yu J, Zhang L, Khodadoust MS, et al. The DEK oncoprotein is a Su(var) that is essential to heterochromatin integrity. *Genes Dev* 2011;25:673–8.
- [71] Alexiadis V, Waldmann T, Andersen J, Mann M, Knippers R, Gruss C. The protein encoded by the proto-oncogene DEK changes the topology of chromatin and reduces the efficiency of DNA replication in a chromatin-specific manner. *Genes Dev* 2000;14:1308–12.
- [72] Piao J, Shang Y, Liu S, Piao Y, Cui X, Li Y, et al. High expression of DEK predicts poor prognosis of gastric adenocarcinoma. *Diagn Pathol* 2014;9:67.
- [73] Wang X, Lin L, Ren X, Lin Z, Li Z, Li C, et al. High expression of oncoprotein DEK predicts poor prognosis of small cell lung cancer. *Int J Clin Exp Pathol* 2014;7:5016–23.
- [74] Liu S, Wang X, Sun F, Kong J, Li Z, Lin Z. DEK overexpression is correlated with the clinical features of breast cancer. *Pathol Int* 2012;62:176–81.
- [75] Luo Y, Yang C, Ye M, Jin C, Abbruzzese JL, Lee MH, et al. Deficiency of metabolic regulator FGFR4 delays breast cancer progression through systemic and microenvironmental metabolic alterations. *Cancer Metab* 2013;1:21.
- [76] Buzzai M, Bauer DE, Jones RG, Deberardinis RJ, Hatzivassiliou G, Elstrom RL, et al. The glucose dependence of Akt-transformed cells can be reversed by pharmacologic activation of fatty acid beta-oxidation. *Oncogene* 2005;24:4165–73.
- [77] Wang J, Hua H, Ran Y, Zhang H, Liu W, Yang Z, et al. Derlin-1 is overexpressed in human breast carcinoma and protects cancer cells from endoplasmic reticulum stress-induced apoptosis. *Breast Cancer Res* 2008;10:R7.
- [78] Jones DT, Lechertier T, Mitter R, Herbert JM, Bicknell R, Jones JL, et al. Gene expression analysis in human breast cancer associated blood vessels. *PLoS One* 2012;7:e44294.
- [79] Byrne JA, Tomasetto C, Garnier JM, Rouyer N, Mattei MG, Bellocq JP, et al. A screening method to identify genes commonly overexpressed in carcinomas and the identification of a novel complementary DNA sequence. *Cancer Res* 1995;55:2896–903.
- [80] Wilson KS, Roberts H, Leek R, Harris AL, Geradts J. Differential gene expression patterns in HER2/neu-positive and -negative breast cancer cell lines and tissues. *Am J Pathol* 2002;161:1171–85.
- [81] Roslan N, Bieche I, Bright RK, Lidereau R, Chen Y, Byrne JA. TPD52 represents a survival factor in ERBB2-amplified breast cancer cells. *Mol Carcinog* 2014;53:807–19.
- [82] Shah PP, Lockwood WW, Saurabh K, Kurlawala Z, Shannon SP, Waigel S, et al. Ubiquilin1 represses migration and epithelial-to-mesenchymal transition of human non-small cell lung cancer cells. *Oncogene* 2014;34:1709–17.
- [83] Miyamoto-Matsubara MHY, Ono K, Xie MZ, Salem AMH, Shoulkamy MI, Nakano T, et al. Depletion of RUVBL2 in human cells confers moderate sensitivity to anticancer agents. *J Cancer Sci Ther* 2014;6:440–5.
- [84] Flavin P, Redmond A, McBryan J, Cocchiglia S, Tibbitts P, Fahy-Browne P, et al. RuvBl2 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS Lett* 2011;585:2537–44.
- [85] Bargou RC, Jurchott K, Wagener C, Bergmann S, Metzner S, Bommert K, et al. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. *Nat Med* 1997;3:447–50.
- [86] Matsumoto K, Wolffe AP. Gene regulation by Y-box proteins: coupling control of transcription and translation. *Trends Cell Biol* 1998;8:318–23.
- [87] Kohno K, Izumi H, Uchiumi T, Ashizuka M, Kuwano M. The pleiotropic functions of the Y-box-binding protein, YB-1. *Bioessays* 2003;25:691–8.
- [88] Lasham A, Print CG, Woolley AG, Dunn SE, Braithwaite AW. YB-1: oncoprotein, prognostic marker and therapeutic target? *Biochem J* 2013;449:11–23.
- [89] Eliseeva IA, Kim ER, Guryanov SG, Ovchinnikov LP, Lyabin DN. Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry (Mosc)* 2011;76:1402–33.
- [90] Khan MI, Adhami VM, Lall RK, Sechi M, Joshi DC, Haidar OM, et al. YB-1 expression promotes epithelial-to-mesenchymal transition in prostate cancer that is inhibited by a small molecule fisetin. *Oncotarget* 2014;5:2462–74.
- [91] Tacke F, Galm O, Kanig N, Yagmur E, Brandt S, Lindquist JA, et al. High prevalence of Y-box protein-1/p18 fragment in plasma of patients with malignancies of different origin. *BMC Cancer* 2014;14:33.
- [92] Davies AH, Reipas KM, Pambid MR, Berns R, Stratford AL, Fotovati A, et al. YB-1 transforms human mammary epithelial cells through chromatin remodeling leading to the development of basal-like breast cancer. *Stem Cells* 2014;32:1437–50.



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data set of the protein expression profiles of Luminal A, Claudin-low and overexpressing HER2⁺ breast cancer cell lines by iTRAQ labelling and tandem mass spectrometry



Karla Grisel Calderón-González^a, Ma Luz Valero Rustarazo^{b,1},
 Maria Luisa Labra-Barrios^c, César Isaac Bazán-Méndez^c,
 Alejandra Tavera-Tapia^{c,2}, MaríaEsther Herrera-Aguirre^c,
 Manuel M. Sánchez del Pino^{b,3}, José Luis Gallegos-Pérez^d,
 Humberto González-Márquez^e, Jose Manuel Hernández-Hernández^c,
 Gloria León-Ávila^f, Sergio Rodríguez-Cuevas^g,
 Fernando Guisa-Hohenstein^{f,g}, Juan Pedro Luna-Arias^{c,*}

^a Doctorado en Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana,

Unidad Iztapalapa, Av. San Rafael Atlixco No. 186, Col. Vicentina, Iztapalapa, C.P. 09340 México DF, Mexico

^b Unidad de Proteómica, Centro de Investigación Príncipe Felipe, C/Rambla del Saler 16, 46012 Valencia, Spain

^c Departamento de Biología Celular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav-IPN), Av. Instituto Politécnico Nacional 2508, Col. San Pedro Zacatenco, Gustavo A. Madero, C.P. 07360 México DF, Mexico

^d AB SCIEX, 500 Old Connecticut Path, Framingham, MA 01701, USA

^e Departamento de Ciencias de la Salud, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Av. San Rafael Atlixco No. 186, Col. Vicentina, Iztapalapa, C.P. 09340 México DF, Mexico

^f Departamento de Zoología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Prolongación de Carpio y Plan de Ayala s/n, Col. Santo Tomás, Miguel Hidalgo, C.P. 11340 México DF, Mexico

^g Instituto de Enfermedades de la Mama, Fundación del Cáncer de Mama (FUCAM A.C.), Av. Bordo No. 100, Col. Viejo Ejido de Santa Ursula Coapa, Coyoacán, C.P. 04980 México DF, Mexico

DOI of original article: <http://dx.doi.org/10.1016/j.jprot.2015.04.018>

* Corresponding author. Tel.: +52 55 57474016, +52 55 57473800x5567; fax: +52 55 57473393.

¹ Current address: Secció de Proteòmica, Servei Central de Suport a la Investigació Experimental (SCSIE), Universitat de València, C/Doctor Moliner 50, 46100 Burjassot, València, Spain.

² Current address: Centro Nacional de Investigaciones Oncológicas (CNIO), Programa de Genética del Cáncer Humano, Grupo de Genética Humana, Melchor Fernández Almagro 3, 28029 Madrid, Spain.

³ Current address: Departament de Bioquímica i Biologia Molecular, Facultat de Ciències Biològiques, Universitat de València, C/Doctor Moliner 50, 46100 Burjassot, València, Spain.

<http://dx.doi.org/10.1016/j.dib.2015.04.025>

2352-3409/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ARTICLE INFO

Article history:

Received 21 April 2015

Received in revised form

28 April 2015

Accepted 29 April 2015

Available online 17 June 2015

ABSTRACT

Breast cancer is the most common and the leading cause of mortality in women worldwide. There is a dire necessity of the identification of novel molecules useful in diagnosis and prognosis. In this work we determined the differentially expression profiles of four breast cancer cell lines compared to a control cell line. We identified 1020 polypeptides labelled with iTRAQ with more than 95% in confidence. We analysed the common proteins in all breast cancer cell lines through IPA software (IPA core and Biomarkers). In addition, we selected the specific overexpressed and subexpressed proteins of the different molecular classes of breast cancer cell lines, and classified them according to protein class and biological process. Data in this article is related to the research article “Determination of the protein expression profiles of breast cancer cell lines by Quantitative Proteomics using iTRAQ Labelling and Tandem Mass Spectrometry” (Calderón-González et al. [1] in press).

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Specifications table

Subject area	Cell Biology
More specific subject area	Breast cancer
Type of data	Tables, figures
How data was acquired	Isobaric labelling, preparative isoelectrofocusing, reverse phase chromatography and tandem mass spectrometry using an AB SCIEX high performance hybrid quadrupole time-of-flight QSTAR ESI XL Hybrid LC/MS/MS Mass Spectrometer System
Data format	Analysed and filtered
Experimental factors	Samples were reduced with 50 mM TCEP, alkylated with 200 mM MMTS, digested with sequencing grade trypsin and labelled according to the manufacturer’s protocol of iTRAQ 8-plex kit, with minor modifications
Experimental features	An amount of 110 µg of total protein from each normal and breast cancer cell line were trypsin-digested, iTRAQ-labelled and pooled. Labelled-peptides were separated by isoelectrofocusing on a non-linear pH 3–10 gradient. Strips were divided in sections and peptides extracted and chromatographed in a C18 reverse phase column. Elution was done using several linear gradients of ACN and eluted peptides were analysed in a QSTAR ESI XL Hybrid LC/MS/MS Mass Spectrometer System. Collected mass spectra were analysed with the 4.0 ProteinPilot program and a cut-off of 20% fold change in the relative quantitation ratio was set. Identified proteins were grouped and classified with PANTHER (v 9.0) and IPA software
Data source location	Mexico City, Mexico
Data accessibility	Data is within this article

Value of the data

- Differential protein expression profiles of breast cancer cell lines.
- Diseases and biofunctions analysis by IPA of over- and subexpressed polypeptides found in common in all breast cancer cell lines.
- Comparative analysis of putative biomarkers found in common in all breast cancer cell lines.
- Identification of specific over- and subexpressed proteins for each breast cancer cell line.
- Classification of specific proteins in each breast cancer cell lines according to protein class and biological process with PANTHER.

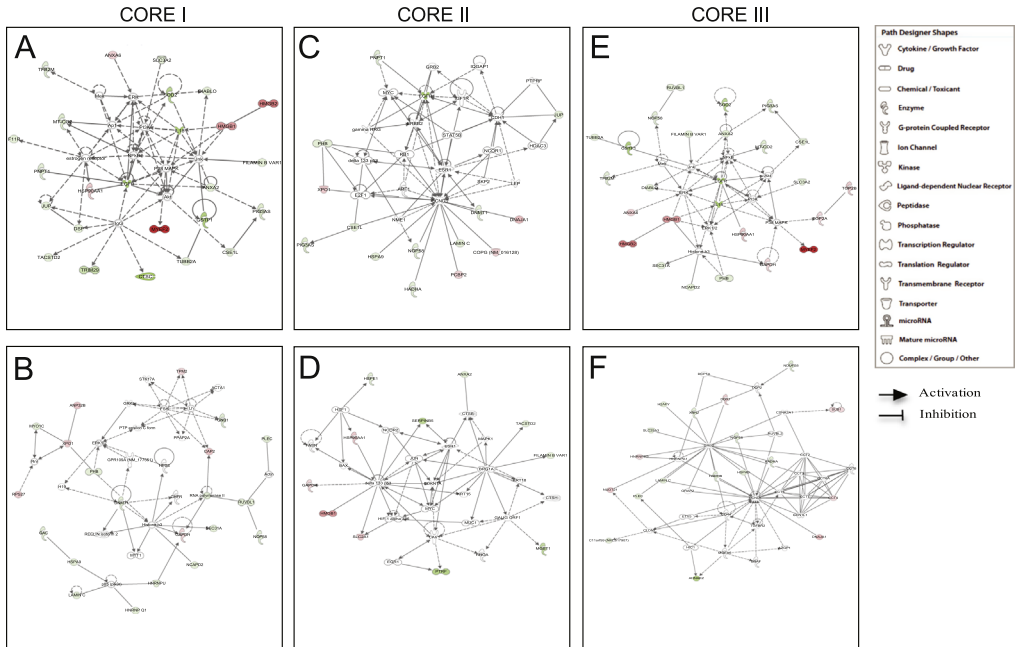


Fig. 1. IPA functional networks of differentially expressed proteins through different core analyses. There are only shown the two top networks of each core analysis. Core I networks are related with: (A) Cancer; Reproductive System Diseases; Dermatological Diseases and Conditions; Haematological Diseases. (B) Connective Tissue Disorders; Neurological Diseases; Skeletal and Muscular Disorders; DNA replication, Recombination and Repair; Cancer; Gastrointestinal Diseases. Core II networks are implicated in: (C) Cellular Development; Cellular Growth and Proliferation; Cell Cycle. (D) Cell Death and Survival; Tumour Morphology; Cellular Development. Core III analysis was performed for proteins involved in: (E) Cell Cycle; Cancer; Hereditary Disorders; Haematological Diseases. (F) Developmental Disorders, Hereditary Disorders; Metabolic Diseases; Cell Death and Survival. Different shapes, indicating the functional class to which they belong, represent proteins. Molecules in red are overexpressed polypeptides, whilst those green molecules correspond to subexpressed proteins. Molecules in grey, which are not specified, were incorporated into the networks by IPA through relationships with other molecules. Molecular relationships between polypeptides are indicated with lines. A continuous line illustrates a direct interaction and a dotted line is used for indirect interactions.

1. Data

We identified 1020 iTRAQ-labelled proteins with at least one peptide with a minimum of 95% in confidence with the ProteinPilot software (Table 1).

1.1. IPA analysis of the differentially expressed proteins found in common in all breast cancer cell lines

To determine whether over- or subexpressed polypeptides have been involved in diseases and biofunctions, or localised in networks, we used the IPA software to perform mainly three different core analyses: (a) Core I considered all tissues, primary cells and all cell lines. (b) Core II was performed for breast cancer cell lines and mammary gland. (c) Core III, for all cancer cell lines excluding breast cancer cell lines (Table 2). They were obtained for different categories of diseases and biofunctions, using a p -value equal or lower than 0.05, which was obtained with Fisher's Exact Test. Core analyses I and III shared similarities in the results obtained. Core I included 78 categories of diseases and biofunctions, whilst 11 categories were found in Core II and 72 in Core III. The most representative categories in Core I were Nucleic Acid Metabolism (16 proteins and p -values between 4.46×10^{-10} and 2.53×10^{-2}), Small Molecule Biochemistry (36 polypeptides and p -values ranging from 4.46×10^{-10} to 3.77×10^{-2}), DNA Replication, Recombination and Repair (19 proteins, p -values

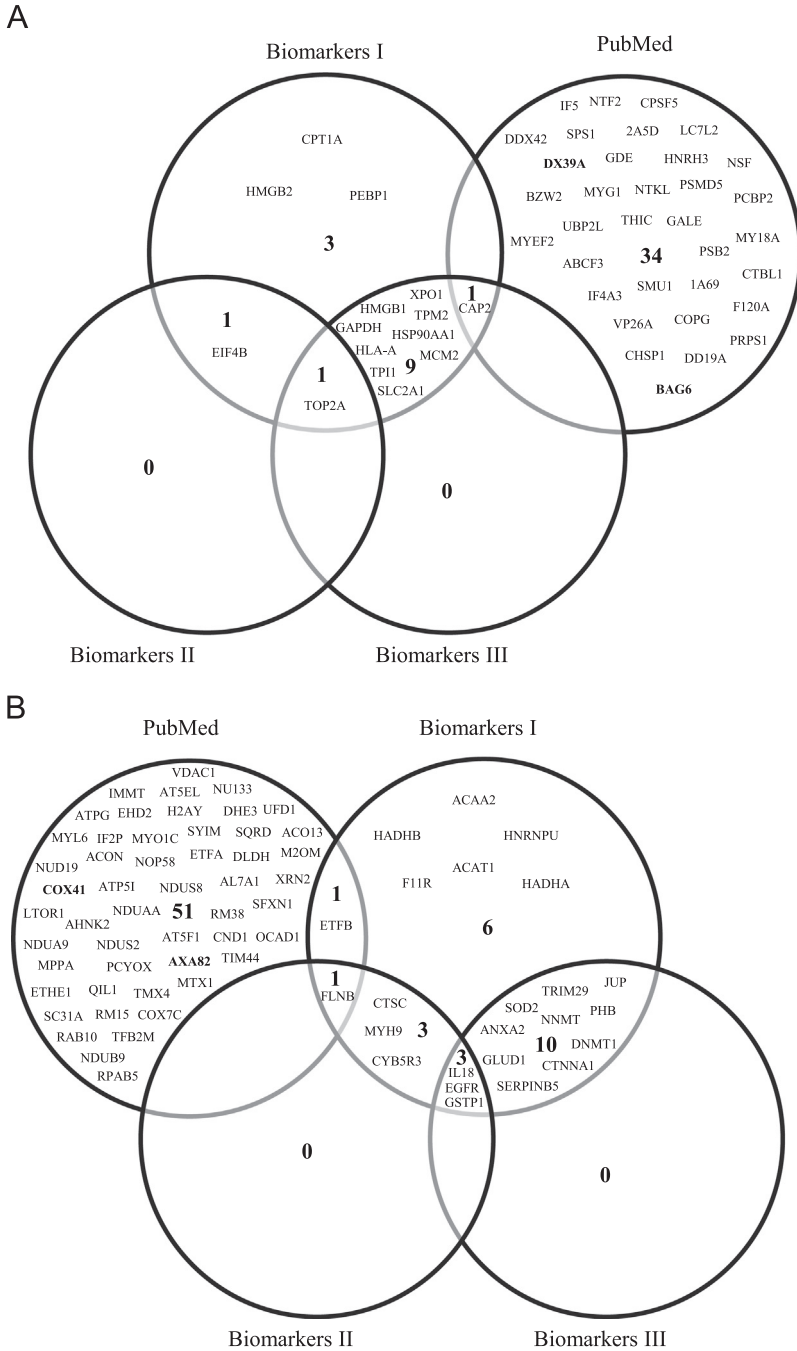


Fig. 2. Venn-Euler diagrams of putative biomarkers obtained through IPA and PubMed analyses. Diagrams show the putative biomarkers found with IPA Biomarkers filter I (all tissues, primary cells and all cell lines), II (mammary gland and breast cancer) and III (cancer cell lines excluding breast cancer cell lines), and those obtained with PubMed database search. (A) Overexpressed proteins and (B) Subexpressed polypeptides common in all breast cancer cell lines.

Overexpressed proteins

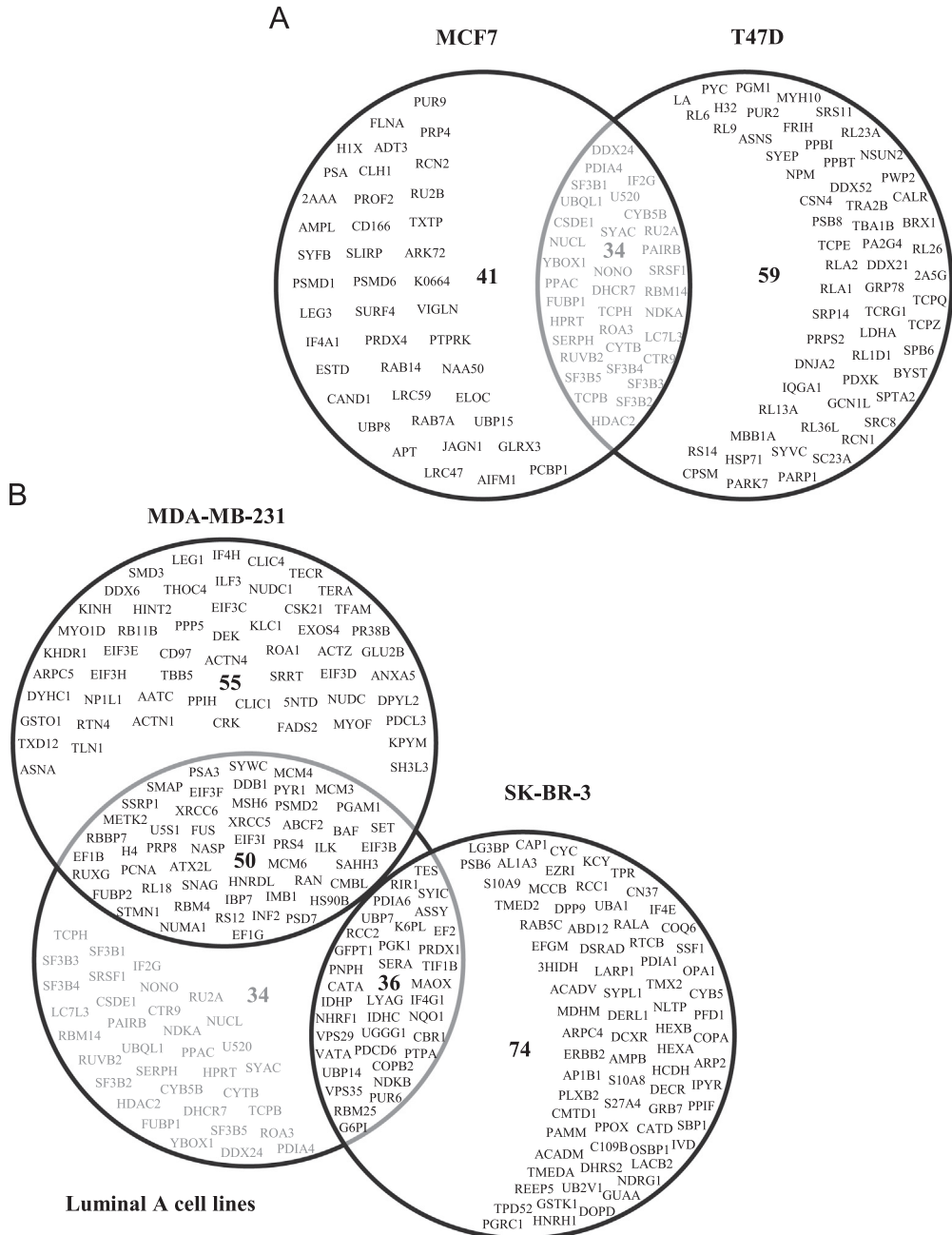


Fig. 3. Venn-Euler diagrams of overexpressed proteins in the different molecular types of breast cancer cell lines. (A) The diagram shows the specific proteins in the Luminal A cell lines (MCF7 and T47D) as well as common proteins in these lines. (B) Diagram shows the unique proteins for Claudin-low (MDA-MB-231) and HER2⁺ (SK-BR-3) cell lines, and the comparison of the common proteins in Luminal A cell lines with the other types of cell lines.

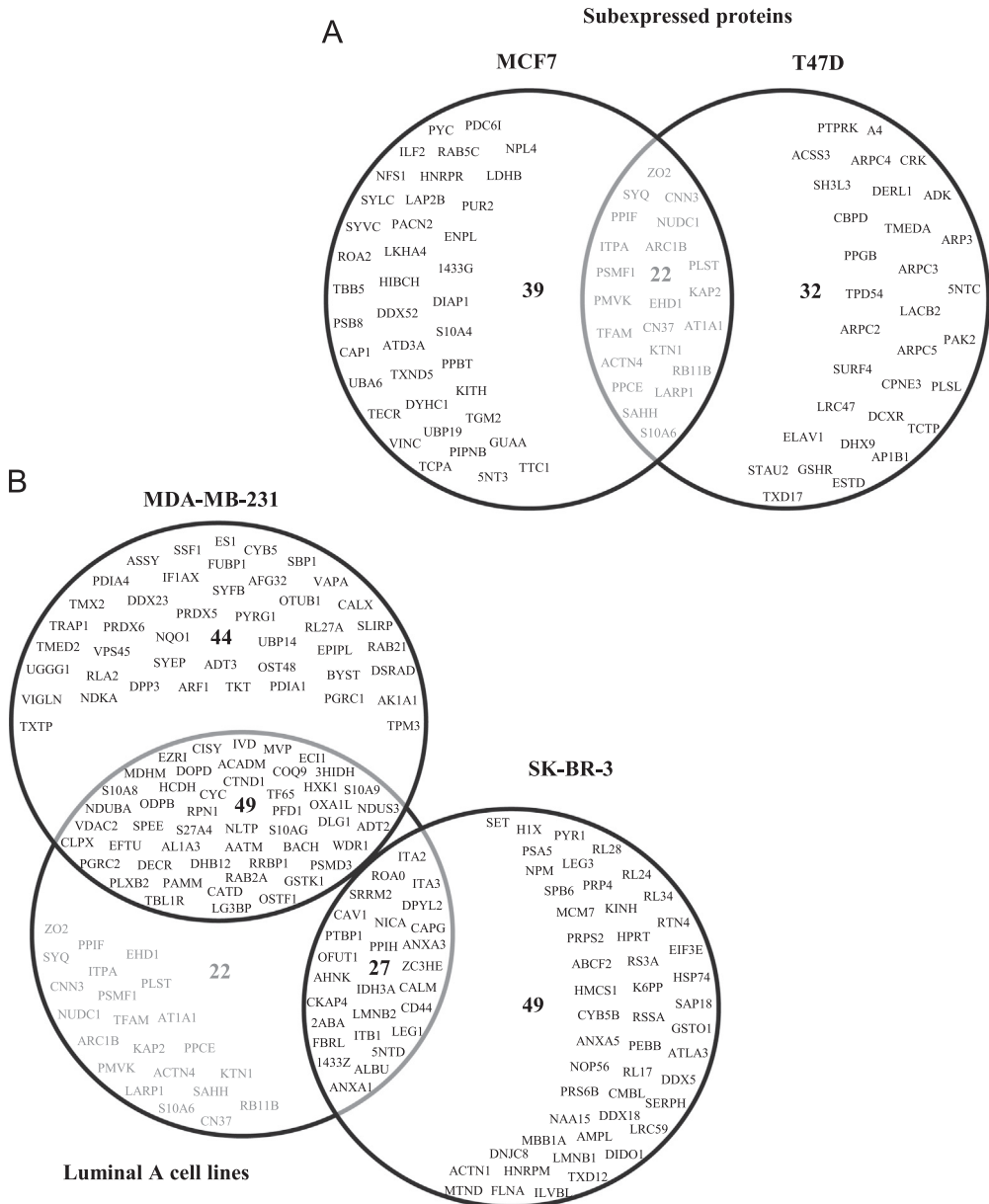


Fig. 4. Venn-Euler diagrams of subexpressed proteins in the different molecular types of breast cancer cell lines. (A) The diagram shows the specific proteins in the Luminal A cell lines (MCF7 and T47D) as well as common proteins in these lines. (B) Diagram show the unique proteins for Claudin low (MDA-MB-231) and HER2⁺ (SK-BR-3) cell lines, and the comparison of the common proteins in Luminal A cell lines with the other types of cell lines.

between 7.24×10^{-10} and 3.77×10^{-2} , and Energy Production (10 polypeptides and a p -values of 7.24×10^{-10}). In Core II we got the following categories: Cellular Response to Therapeutics (only 2 proteins with a p -value of 2×10^{-3}), Cellular Movement (9 polypeptides and p -values from 1.72×10^{-2} to 3.67×10^{-2}), Cell Morphology and Cellular Assembly and Organisation (2 polypeptides, p -value of 1.85×10^{-2}). Finally, the most representative categories in Core III were the same as in Core

I: Nucleic Acid Metabolism (16 polypeptides, p -values ranging from 4.6×10^{-10} to 4.02×10^{-2}), Small Molecule Biochemistry (34 proteins, p -values between 4.6×10^{-10} and 4.02×10^{-2}), DNA Replication, Recombination and Repair (21 proteins, p -values from 7.46×10^{-10} to 3.78×10^{-2}), and Energy Production (12 proteins, p -values from 7.46×10^{-10} to 4.02×10^{-2}).

The two top networks obtained in Core I analysis were related with: (1) Cancer; Reproductive System Disease; Dermatological Diseases and Conditions; Haematological Diseases, which showed a score of 38 and contained 25 target molecules. (2) Connective Tissue Disorders; Neurological Disease; Skeletal and Muscular Disorders; DNA Replication, Recombination and Repair; Cancer; Gastro-intestinal Diseases, that had a score of 27 and a network focused on 20 molecules (Fig. 1A and B). In Core II analysis, the two most representative networks were involved in: (1) Cellular Development, Cellular Growth and Proliferation, and Cell Cycle, with a score of 15 and a network of 15 polypeptides. (2) Cell Death and Survival, Tumour Morphology, and Cellular Development. In this case, the score was 9 and the network generated was focused on 11 molecules (Fig. 1C and D). Finally, the Core III analysis contained polypeptides playing a role in: (1) Cell Cycle, Cancer, Hereditary Disorders, and Haematological Diseases. In this case, it showed a score of 41 and considered 26 proteins (Fig. 1E). (2) Developmental Disorders, Hereditary Disorders, Metabolic Diseases, and Cell Death and Survival. This network had a score of 24 and contained 18 polypeptides (Fig. 1F).

1.2. Selection of putative biomarkers found in common in all breast cancer cell lines

For the identification of putative biomarkers we used two different approaches: (1) Filtering candidate biomarkers with the IPA software, and (2) Searching of proteins in the PubMed database. The first filter, named Biomarkers I, was carried out in all tissues, primary cells and all cell lines information contained in the Ingenuity Knowledge database; the second one was through breast cancer cell lines and mammary gland (Biomarkers II), and the third consisted in filtering through cancer cell lines, excluding breast cancer cell lines and was named Biomarkers III. Results obtained in Biomarkers I revealed 39 potential candidates: ACAA2, ACAT1, ANXA2, CAP2, CPT1A, CTNNA1, CTSC, CYB5R3, DNMT1, EGFR, EIF4B, ETVB, F11R, FLNB, GAPDH, GLUD1, GSTP1, HADHA, HADHB, HLA-A, HMGB1, HMGB2, HNRNPU, HSP90AA1, IL18, JUP, MCM2, MYH9, NNMT, PEBP1, PHB, SERPINB5, SLC2A1, SOD2, TOP2A, TPI1, TPM2, TRIM29, XPO1. Biomarkers II included 9 putative targets: CTSC, CYB5R3, EGFR, EIF4B, FLNB, GSTP1, IL18, MYH9, TOP2A. Finally, Biomarkers III selected 24 polypeptides, which included ANXA2, CAP2, CTNNA1, DNMT1, EGFR, GAPDH, GLUD1, GSTP1, HLA-A, HMGB1, HSP90AA1, IL18, JUP, MCM2, NNMT, PHB, SERPINB5, SLC2A1, SOD2, TOP2A, TPI1, TPM2, TRIM29 and XPO1. All of these potential biomarkers were in common in all studied breast cancer cell lines.

In relation to the search for candidates in the PubMed database, we obtained 35 overexpressed and 53 subexpressed polypeptides not previously reported in breast cancer, with one exception, ANXA8, which has only one report [2]. Finally, we performed a comparison between the biomarkers found with IPA and those in the PubMed search. These analyses revealed only 3 specific overexpressed polypeptides in Biomarkers I, and 34 in the PubMed search (Fig. 2A). For subexpressed proteins, we obtained 6 polypeptides for Biomarkers I, and 51 in PubMed (Fig. 2B). Biomarkers II and III modules did not find exclusive markers in both cases.

1.3. Panel of putative biomarkers exclusively found in Luminal A, Claudin-low and HER2⁺ breast cancer cell lines

We identified sets of specific proteins for Luminal A, Claudin-low and HER2⁺ breast cancer cell lines. We obtained a set of proteins shared by Luminal A cell lines containing 34 overexpressed (Fig. 3A) and 22 subexpressed polypeptides (Fig. 4A). In the case of Claudin-low, we identified a set with 55 and 44 overexpressed and subexpressed polypeptides, respectively (Figs. 3B and 4B). The HER2⁺ cell line showed 74 overexpressed (Fig. 3B) and 49 subexpressed proteins (Fig. 4B). However, we need to determine the expression levels of these proteins in tumour tissues and or sera to validate them as biomarkers.

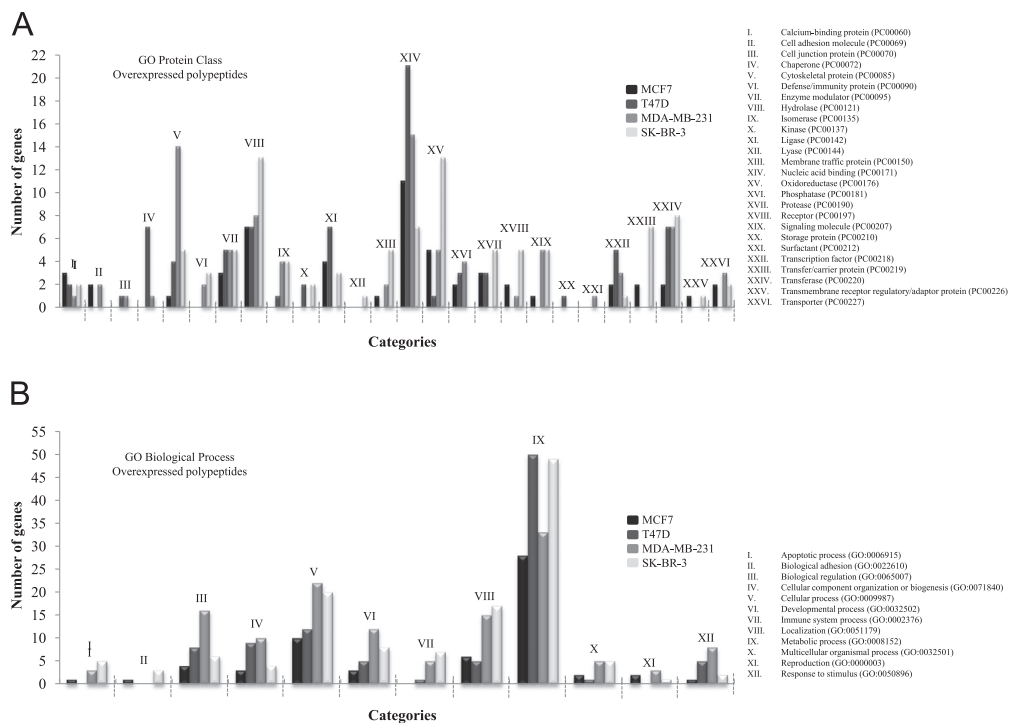


Fig. 5. PANTHER functional classification of overexpressed proteins found exclusively in each breast cancer cell line. The graph shows the number of genes involved in the different Protein Class (A) and Biological Process (B) of the specific proteins in each breast cancer cell line. The total numbers of class hits were 54, 77, 84 and 97 for Protein Class and 61, 96, 132, and 127 for Biological Process in MCF7, T47D, MDA-MB-231 and SK-BR-3 cell lines.

1.4. Functional classification of the exclusive proteins differentially expressed in each breast cancer cell line

The overexpressed and subexpressed specific proteins of each breast cancer cell line (MCF7, T47D, MDA-MB-231, SK-BR-3) were classified with PANTHER. For Protein Class of overexpressed proteins, we obtained 26 categories, being the Nucleic Acid Binding (XIV) the most representative category for MCF7, T47D and MDA-MB-231, and Hydrolase (VIII) and Oxidoreductase (XV) for SK-BR-3 (Fig. 5A). In the case of the Biological Processes, we found 12 categories, being the Metabolic process (IX) the category with more genes implicated (Fig. 5B). On the other hand, the categorisation of subexpressed proteins specific for each cell line had 25 categories in Protein Class and 12 in Biological Processes. The categories of Protein Class with mayor number of genes were Hydrolase (VII) for MCF7, and T47D, Nucleic Acid Binding (XIII) for MDA-MB-231 and SK-BR-3 (Fig. 6A). The category of Biological Process most affected was Metabolic process (IX) in each of cell line (Fig. 6B).

2. Experimental design, materials and methods

2.1. IPA analysis of proteins in common in all breast cancer cell lines

To identify diseases and biological functions, canonical pathways, molecular networks, and putative candidates for biomarkers from the proteins that are commonly express in all breast cancer cell lines, we used the complete functional Ingenuity Pathway Analysis software (IPA, QIAGENs Redwood City, www.qiagen.com/ingenuity). We performed three different analyses named Cores I, II and III. In all cases

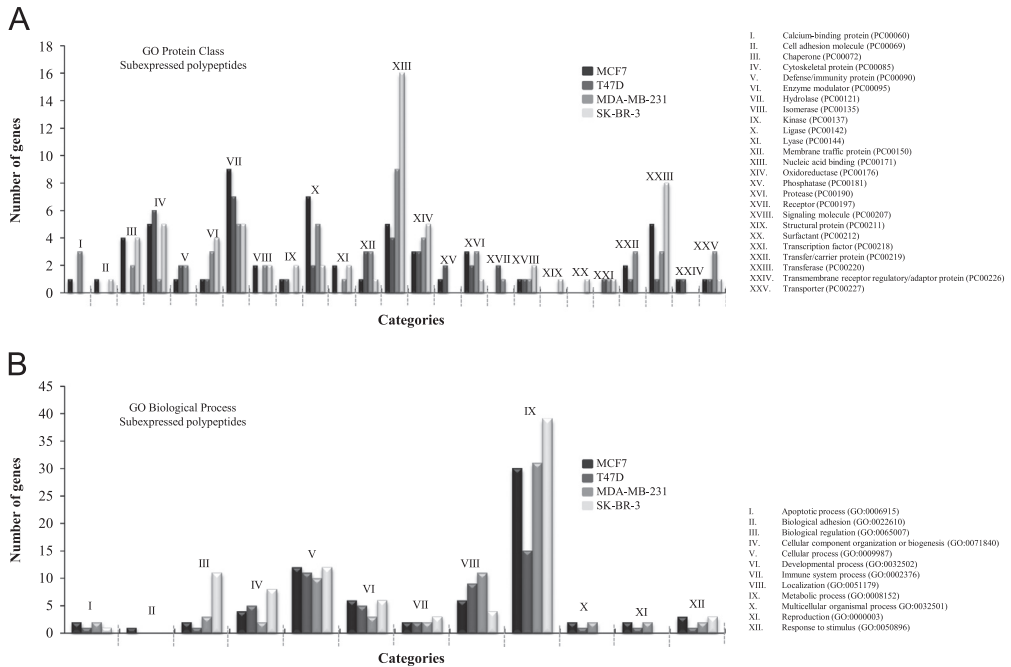


Fig. 6. PANTHER functional classification of subexpressed proteins found in each breast cancer cell line. The graph shows the number of genes involved in the different Protein Class (A) and Biological Process (B) of the specific proteins in each breast cancer cell line. The total numbers of class hits were 57, 41, 55 and 64 for Protein Class and 72, 52, 70, and 87 for Biological Process in MCF7, T47D, MDA-MB-231 and SK-BR-3 cell lines.

we used the stringent filter. The Fisher's Exact Test was used to determine the p -value, which was considered as significant with values ≤ 0.05 . Core I was done to classify identified proteins in all tissues, primary cells and all cell lines including cervical, central nervous system, colon, hepatoma, immune, kidney, leukaemia, lung, lymphoma, macrophage, melanoma, myeloma, neuroblastoma, osteosarcoma, ovarian, pancreatic, prostate, teratocarcinoma, breast cancer and other cell lines. Core II was used in mammary gland and breast cancer cell lines. Core III in all cancer cell lines, excluding breast cancer cell lines. Parameters used in the complete analysis were: (i) ingenuity knowledge base (genes only), considering direct and indirect relationships. (ii) Interaction networks including endogenous chemicals, default value of 35 molecules per network and 25 networks per analysis. (iii) All data source. (iv) Confidence: experimentally observed, highly and moderately predicted. (v) Human species with stringent filter. (vi) All mutations.

2.2. Selection of putative biomarkers common in all breast cancer cell lines

We used the module IPA biomarker to prioritise protein biomarker candidates in our data. Three analyses were carried out, Biomarkers I included a filter for all tissues, primary cells and all cell lines, Biomarkers II for breast cancer cell lines and mammary gland, and Biomarkers III for cancer cell lines, excluding breast cancer. Parameters used were: (1) human species. (2) All molecule types. (3) All biofluids. (4) All diseases for Biomarkers I, and cancer disease for Biomarkers II and III. (5) Biomarkers I, all biomarkers and disease application. For Biomarkers II, all biomarkers application and breast cancer disease (including breast cancer, breast carcinoma, ductal carcinoma, ductal carcinoma in situ, infiltrating ductal breast carcinoma, invasive ductal breast cancer, lobular breast carcinoma, and metastatic breast cancer). In the case of Biomarkers III, all biomarkers application and cancer disease excluding breast cancer.

Moreover, we searched every one of the 206 identified polypeptides that changed in their expression levels in common in all breast cancer cell lines in PubMed database using key words such as biomarkers, cancer and breast cancer. Finally, we performed a comparison analysis of the results obtained with IPA Biomarker modules and PubMed with the Venny program (developed by Oliveros JC and available at <http://bioinfogp.cnb.csic.es/tools/venny/>).

2.3. Selection of biomarkers panel in the different molecular classification of breast cancer cell lines

For the selection of the exclusive sets of overexpressed and subexpressed polypeptides for Luminal A, Claudin-low and HER2⁺ cell lines, we performed a comparative analysis of between them with the Venny program.

2.4. Functional classification of the exclusive sets of proteins for each breast cancer cell lines according to their corresponding molecular classification with PANTHER

The specific sets of polypeptides for each breast cancer cell line according to their corresponding molecular classification were analysed with the PANTHER classification system v 9.0 (<http://www.pantherdb.org/>) [3]

Acknowledgements

We thank to Consejo Nacional de Ciencia y Tecnología (Conacyt) and the Instituto de Ciencia y Tecnología del Distrito Federal (ICyTDF) (now transformed in Secretaría de Ciencia Tecnología e Innovación (SECITI) del Distrito Federal), both from Mexico, for the financial support to Dr. Juan Pedro Luna Arias for doing this work. We thank Conacyt for the scholarship to Karla G. Calderón-González during her Ph.D. studies, and the ECOS Mexico–France program for the financial support to Karla G. Calderón-González for a short academic stay in Lille, France. The Unidad de Proteómica from the Centro de Investigación Príncipe Felipe belongs to ProteoRed, PRB2-ISCIIL.

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2015.04.025>.

References

- [1] K.G. Calderón-González, M.L. Valero Rustarazo, M.L. Labra-Barrios, C.I. Bazán-Méndez, A. Tavera-Tapia, M.E. Herrera-Aguirre, et al., Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry, *J. Proteomics* 124 (2015) 50–78, <http://dx.doi.org/10.1016/j.jprot.2015.04.018>.
- [2] T. Stein, K.N. Price, J.S. Morris, V.J. Heath, R.K. Ferrier, A.K. Bell, et al., Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 11 (2005) 6872–6879.
- [3] H. Mi, A. Muruganujan, P.D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees, *Nucleic Acids Res.* 41 (2013) D377–D386.