

Trabajo de Tesis Doctoral

Doctorado en Biomedicina  
y Biotecnología

---

# Expresión diferencial de microRNAs en tumores de mujeres jóvenes con cáncer de mama: Cribado de dianas moleculares

---

Directores:           Dra. Gloria Ribas Despuig  
                              Dra. Ana Lluch Hernández

Tutor:                 Dr. Gerardo López Rodas

Departamento de Bioquímica  
y Biología Molecular  
Facultad de Biología



*María del Carmen Peña Chilet*

*Valencia, Septiembre 2015*





VNIVERSITAT  
E VALÈNCIA



Las co-directoras de la presente tesis, la Dra. Ana Lluch Hernández, Catedrática de Medicina de Facultad de Medicina de la Universidad de Valencia y la Dra. Gloria Ribas Despuig, investigadora del Departamento de Biología del cáncer de mama del Instituto de Investigación Sanitaria - INCLIVA,

INFORMAN:

Que **María del Carmen Peña Chilet**, licenciada en Biología por la Universidad de Valencia, ha realizado la presente Tesis Doctoral “**Expresión diferencial de microRNAs en tumores de mujeres jóvenes con cáncer de mama: Cribado de dianas moleculares**” y que, a su juicio, reúne plenamente todos los requisitos necesarios para optar al **Grado de Doctora en Biomedicina y Biotecnología**, a cuyos efectos será presentada en la Universidad de Valencia. El trabajo ha sido realizado bajo su dirección, autorizando su presentación ante el Tribunal Calificador. Y para que así conste se extiende el presente certificado.

Valencia, septiembre de 2015.

Co-diretores de tesis:

**Dra. Ana Lluch Hernández**

Facultad de Medicina de la  
Universidad de Valencia

**Dra. Gloria Ribas Despuig**

Dpto. Biología del cáncer de  
mama. INCLIVA.

Tutor de la tesis:

**Dr. Gerardo López Rodas**

Departamento de Bioquímica  
y Biología Molecular, Universidad de Valencia





Esta tesis ha sido realizada en el Grupo de Genómica Funcional del Cáncer del Instituto de Investigaciones Sanitarias - INCLIVA, en colaboración con el Hospital Clínico Universitario de Valencia, bajo la supervisión de la Dra. Ana Lluch Hernández y la Dra. Gloria Ribas Despuig.

Las siguientes ayudas han permitido la realización de la presente Tesis Doctoral:

- \* Programa VALi+d para para investigadores en formación, de la Generalitat Valenciana de la Comunidad Valenciana, ACIF2011/270.
- \* Proyecto del Ministerio Instituto Carlos III, PI13/00606 y Fondos FEDER.
- \* Proyecto Intramural a investigadores emergentes. INCLIVA
- \* Donación del Proyecto Flor de Vida de la Fundación Le Cadó.





*“En la vida, no hay nada que temer,  
solamente hay que comprender.”*

*M. Curie*



*A Jose Luis y Ada*



# Agradecimientos

---

---

En primer lugar quería agradecer a La Fundación Le Cadó su generosidad, empeño y dedicación que han hecho posible hacer este trabajo.

Gracias, Ana Lluch, por dirigir un grupo tan maravilloso y por haberme permitido formar parte de él y aprender tantísimo, por hacerme sentir que formo parte de algo importante.

Gloria, gracias por haber sido mi mentora. Porque me fascina tu forma de ver y hacer las cosas y cómo puedes con todo. Te admiro. Recuerdo cuando llegué, recién licenciada, a un laboratorio sorprendentemente bien iluminado, con tan solo Maider y tú, recuerdo que pensé “este es mi sitio” y así fue. Maider, gracias por tu paciencia conmigo y por enseñarme cada una de las cosas que he aprendido a hacer, por apoyarme profesional y personalmente, y por ser mi amiga.

Maite, he aprendido muchísimo contigo, me has aportado la visión clínica, he disfrutado enseñándote que se hace en un laboratorio y aprendiendo de ti. Eres mi compi de tesis y ojalá sigamos siendo compis de trabajo, de artículos, de lo que sea, es genial trabajar contigo y tú eres genial, como profesional y como persona.

Gracias a todos los que habéis compartido vuestro tiempo conmigo, Jessi, Ariadna, Bego, Edu, Pilar, Sandra, Ray, Joan, Alberto y Chiri. Jaume, gracias por aportarme mi ración de frikismo y por tu ayuda bioinformática. Especialmente gracias a Marta y Sara, por vuestro apoyo y cariño, por vuestra ayuda y por las risas espontáneas.

Gracias, papá, por enseñarme tantos mundos, gracias por contagiarme tu pasión por tantas cosas, entre ellas la lectura y la ciencia, porque, por ti, comencé mi camino en la investigación. Gracias. Porque todas tus enseñanzas y razonamientos, tus historias, tu escepticismo, tu forma de ver la vida, me hicieron como soy y me acompañaron a cada momento. A ti te dedico este trabajo y cada uno de los que haga a continuación, con la esperanza de que, en algún plano de la existencia, estés sonriendo en este momento.

Madre, siempre me has apoyado, animándome cuando las cosas no salían, me has dado tu cariño y admiración incondicional, nunca podré agradecerte suficiente todo lo que has hecho por mí. Sé que estás orgullosa y te alegras más que nadie. Recuerdo cuando me decías que la investigación es un camino de espinas, al final las rosas también han estado ahí.

Gracias Kata, por ser la amiga más alucinante y por soportar todas mis frustraciones, desvaríos y *derivaciones* varias. Gracias a ti, Octavio, por tus escasas pero elocuentes palabras, y por esos momentos cuando, aunque no los recuerdes, me demuestras que siempre serás mi amigo. A ambos y también a ti, Lorena, gracias por acompañarme desde que, allá por 2002, empezamos Biología, gracias por los campeonatos legendarios de Mus, por las infinitas risas y por todas las que vendrán, por las noches de cervezas y las conversaciones que trascienden el tiempo y el espacio.

Rubén, a ti no puedo más que agradecerte por haber decidido formar parte de mi vida, sin ti nada tendría sentido, gracias por ser mi compañero.





# Índice

---

ÍNDICE DE FIGURAS .....	1
ÍNDICE DE TABLAS .....	5
ÍNDICE DE ANEXOS .....	7
ÍNDICE DE ABREVIATURAS .....	9
RESUMEN .....	17
ABSTRACT .....	19
INTRODUCCIÓN .....	21
CÁNCER .....	23
CÁNCER DE MAMA .....	25
<b>1. Glándula mamaria.....</b>	<b>26</b>
1.1. Anatomía.....	26
1.2. Desarrollo de la glándula mamaria.....	27
<b>2. Epidemiología y riesgos del cáncer de mama.....</b>	<b>27</b>
2.1. Factores de riesgo a padecer cáncer de mama.....	27
2.2. Incidencia del cáncer de mama .....	30
<b>3. Heterogeneidad y clasificación del cáncer de mama.....</b>	<b>31</b>
3.1. Clasificación Histopatológica .....	31
3.2. Clasificación por Grado Histológico.....	32
3.3. Clasificación según estadio .....	32
3.4. Clasificación según receptores de membrana .....	33
3.5. Clasificación por subtipos moleculares .....	33
<b>4. Tratamiento del cáncer de mama.....</b>	<b>36</b>
4.1. Cirugía.....	36
4.2. Radioterapia (RTx).....	36
4.3. Quimioterapia.....	36
4.4. Terapia dirigida.....	37
<b>5. Cáncer de mama en mujeres jóvenes .....</b>	<b>38</b>
5.1. La edad como factor de riesgo y pronóstico.....	38
5.2. Determinación del rango de edad “muy joven”.....	38
5.3. Características tumorales .....	39
5.4. Tratamiento y consideraciones específicas de las pacientes jóvenes.....	39
ESTUDIOS MOLECULARES EN CÁNCER DE MAMA.....	41
<b>1. Epigenética .....</b>	<b>41</b>
1.1. Los microRNAs y su papel en enfermedades.....	41
1.2. miRNAs en cáncer.....	42
1.3. miRNAs en cáncer de mama.....	44
1.4. Desregulación de miRNAs en células madre de cancer de mama (BCSCs).....	45
<b>2. Rutas alteradas en Cáncer de Mama .....</b>	<b>46</b>
2.1. Importancia y regulación de la ruta de las MAP cinasas.....	46
2.2. Importancia de la ruta de señalización PI3K/Akt en cáncer de mama.....	47
2.3. Importancia de la ruta de señalización de Notch.....	47
2.4. Receptores de membrana iniciadores de la cascada de señalización.....	47
2.5. Importancia de la familia HER en cáncer.....	49
<b>3. Perfiles moleculares y uso de muestras FFPE en estudios moleculares .....</b>	<b>51</b>
<b>4. Estudios moleculares en mujeres jóvenes con cáncer de mama.....</b>	<b>52</b>
4.2. Estudios de la expresión de genes en mujeres jóvenes.....	52
4.3. Estudios en miRNAs en mujeres jóvenes.....	53

<b>HIPÓTESIS Y OBJETIVOS.....</b>	<b>55</b>
HIPÓTESIS .....	57
OBJETIVOS .....	59
<b>MATERIAL Y MÉTODOS.....</b>	<b>61</b>
DISEÑO EXPERIMENTAL .....	63
1. <i>Evaluación de la idoneidad del uso de tejido incluido en parafina en estudios moleculares. ..</i>	<i>64</i>
2. <i>Estudio de la expresión de microRNAs en pacientes jóvenes de cáncer de mama.....</i>	<i>64</i>
2.1. <i>Fase I: Análisis exploratorio de la expresión de microRNAs En pacientes jóvenes de cáncer de mama.</i>	64
2.2. <i>Fase II: Validación de la expresión de los miRNAs seleccionados.....</i>	65
3. <i>Estudio de la expresión de genes diana (mRNA) y de proteína en cáncer de mama en mujeres jóvenes.....</i>	<i>66</i>
3.1 <i>Análisis de la expresión de genes mediante qRT-PCR de mRNA .....</i>	66
3.2. <i>Análisis de la proteína in situ .....</i>	66
MATERIAL EMPLEADO .....	67
1. <i>Material biológico.....</i>	<i>67</i>
1.1. <i>Muestras de tejido tumoral.....</i>	67
1.2. <i>Muestras procedentes de cultivo celular.....</i>	70
1.3. <i>Datos obtenidos a partir de bases de datos online.....</i>	71
2. <i>Material comercial .....</i>	<i>72</i>
2.1. <i>Kits comerciales empleados .....</i>	72
2.2. <i>Sondas empleadas .....</i>	73
2.3. <i>Anticuerpos empleados.....</i>	73
3. <i>Programas y herramientas empleadas.....</i>	<i>73</i>
3.1. <i>Programas empleados.....</i>	73
3.2. <i>Herramientas online y bases de datos empleadas .....</i>	74
METODOLOGÍA .....	75
1. <i>Tratamiento de las muestras de tejido biopsiado para su preparación y conservación .....</i>	<i>75</i>
1.1. <i>Tratamiento del tejido biopsiado para su congelación óptima .....</i>	75
1.2. <i>Inclusión en parafina del tejido biopsiado .....</i>	75
1.3. <i>Control histológico de las muestras .....</i>	75
2. <i>Tratamiento de cultivo celular para su crecimiento, mantenimiento y conservación .....</i>	<i>76</i>
2.1. <i>Cultivo de las líneas celulares .....</i>	76
2.2. <i>Cálculo del tiempo de doblaje.....</i>	76
3. <i>Extracción de material genético .....</i>	<i>76</i>
3.1. <i>Extracción de RNA a partir de tejido congelado .....</i>	76
3.2. <i>Extracción de RNA a partir de tejido incluido en parafina .....</i>	77
3.3. <i>Extracción de RNA a partir de pellet de células .....</i>	78
4. <i>Técnicas de cuantificación y medida de la calidad del material extraído.....</i>	<i>78</i>
4.1. <i>Cuantificación del RNA total por espectrofotometría de absorción .....</i>	78
4.2. <i>Análisis de la integridad del RNA total.....</i>	79
4.3. <i>Análisis de la integridad del RNA de pequeño tamaño .....</i>	79
5. <i>Técnicas de análisis de la expresión de material genético (RNA y miRNA) mediante PCR cuantitativa.....</i>	<i>80</i>
5.1. <i>Transcripción reversa o síntesis de cDNA.....</i>	82
5.2. <i>Preamplificación.....</i>	82
5.3. <i>PCR cuantitativa en tiempo real (qRT-PCR).....</i>	83
5.4. <i>Método <math>2^{-\Delta\Delta C_t}</math> para obtener el valor de expresión relativa .....</i>	83
6. <i>Análisis de la expresión global de miRNAs mediante la plataforma Affymetrix.....</i>	<i>85</i>
6.1. <i>Hibridación del RNA total de las muestras en chip de Affymetrix Genechip® miRNA 2.0 .....</i>	86

6.2.	Marcaje del RNA con Biotina .....	86
6.3.	Hibridación del RNA con las sondas de RNA de pequeño tamaño incluidas en el chip.....	86
6.4.	Lavados y tinción del chip .....	87
6.5.	Escaneado del chip y detección de la intensidad de fluorescencia.....	87
<b>7.</b>	<b><i>Técnicas de análisis de la expresión de proteína in situ</i></b> .....	<b>87</b>
7.1.	Construcción de los Tissue-Microarrays.....	87
7.2.	Estudio inmunohistoquímico de las muestras.....	88
7.3.	Hibridación de fluorescencia in situ para el estudio de la amplificación de HER2.....	89
7.4.	Determinación inmunohistoquímica de las muestras de cáncer de mama .....	90
ANÁLISIS ESTADÍSTICOS Y BIOINFORMÁTICOS .....		91
<b>1.</b>	<b><i>Análisis estadísticos</i></b> .....	<b>91</b>
1.1.	<i>Test de normalidad y de homogeneidad de varianzas</i> .....	91
1.2.	Comparaciones entre grupos .....	92
1.3.	Corrección por múltiples comparaciones .....	98
1.4.	<i>Curvas de supervivencia</i> .....	99
1.5.	<i>Valor de significatividad y poder estadístico</i> .....	100
1.6.	<i>Combinación de p-valores mediante el método de Fisher</i> .....	101
<b>2.</b>	<b><i>Análisis bioinformáticos</i></b> .....	<b>102</b>
2.1.	<i>Manejo, tratamiento y filtrado de datos</i> .....	102
2.2.	<i>Agrupación jerárquica</i> .....	103
2.3.	<i>Estimación del efecto acumulativo de los miRNAs sobre los genes diana</i> .....	104
2.4.	<i>Análisis de enriquecimiento de rutas</i> .....	105
2.5.	<i>Construcción de la red de interacciones</i> .....	106
2.6.	<i>Búsqueda de sitios de unión de miRNAs conservados en la zona 3'UTR de genes diana mediante TargetScan</i> .....	106
2.7.	<i>Análisis comparativo de los estudios</i> .....	107
<b>RESULTADOS</b> .....		<b>109</b>
PARTE 1: ANÁLISIS DE LA VIABILIDAD DEL TEJIDO TUMORAL BIOPSIADO		
CONSERVADO EN PARAFINA PARA EL ESTUDIO MOLECULAR .....		
1.1.	<b><i>Análisis de la integridad del RNA total</i></b> .....	<b>111</b>
1.2.	<b><i>Análisis de la integridad de los RNAs de pequeño tamaño</i></b> .....	<b>112</b>
PARTE 2: ANÁLISIS DEL PERFIL DE EXPRESIÓN DE MIRNAS EN CÁNCER DE		
MAMA DE MUJERES JÓVENES .....		
<b>1.</b>	<b><i>Estudio de la expresión de miRNAs sobre muestras tumorales de pacientes</i></b> .....	<b>117</b>
1.1.	Análisis de la intensidad de señal de los chips y control de calidad .....	117
1.2.	Análisis de la intensidad de señal de los chips y control de calidad .....	117
1.3.	Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes .....	118
1.4.	Análisis de las características tumorales como factor mediador en la significatividad obtenida....	119
1.5.	Agrupación jerárquica de muestras y miRNAs obtenidos según el grupo de edad .....	121
1.6.	Análisis individual del enriquecimiento de rutas metabólicas en genes diana de los miRNAs que constituyen los nodos seleccionados.....	124
1.7.	Análisis de enriquecimiento de rutas en los miRNAs que componen cada nodo .....	132
1.8.	Análisis de enriquecimiento de rutas en bases de datos online de dianas del total de miRNAs desregulados.....	140
1.9.	Red de interacciones de las rutas desreguladas seleccionadas .....	143
<b>2.</b>	<b><i>Estudio del perfil de expresión de miRNA en cáncer de mama de mujeres jóvenes en líneas celulares</i></b> .....	<b>148</b>
2.1.	Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes .....	148
2.2.	Agrupación jerárquica por líneas celulares y miRNAs según la edad .....	149

<b>3. Análisis del perfil de expresión de miRNAs en una serie con un tamaño muestral mayor obtenida a partir de bases de datos online (METABRIC).....</b>	<b>153</b>
3.1. Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes .....	153
3.2. Perfil de expresión mediante agrupación jerárquica de los datos analizados.....	154
3.3. Comparación y meta-análisis de los resultados obtenidos en los tres estudios.....	155
PARTE 3: VALIDACIÓN DE LA EXPRESIÓN DE LOS MIRNAS SELECCIONADOS EN UNA SERIE INDEPENDIENTE DE MUESTRAS MEDIANTE QRT-PCR .....	157
<b>1. Validación de la expresión de miRNAs en tejido tumoral mamario de mujeres jóvenes .....</b>	<b>157</b>
1.1. Análisis de la expresión de miRNAs en tumor de mama de mujeres muy jóvenes.....	157
1.2. Estudio de los miRNAs desregulados seleccionados como marcadores de la supervivencia .....	160
<b>2. Análisis de la expresión diferencial de los genes diana seleccionados y su proteína en muestras tumorales de mujeres jóvenes .....</b>	<b>162</b>
2.1. Estudio de la expresión en mujeres jóvenes del mRNA de los genes dianas seleccionados mediante qRT-PCR .....	162
2.2. Validación de la presencia diferencial de HER4 según el grupo de edad en células tumorales mediante inmunohistoquímica en tejido tumoral parafinado .....	166
DISCUSIÓN .....	169
PARTE 1: VIABILIDAD DEL TEJIDO TUMORAL BIOPSIADO CONSERVADO EN PARAFINA PARA ESTUDIOS MOLECULARES .....	171
PARTE 2: ANÁLISIS DEL PERFIL DE EXPRESIÓN DE MIRNAS EN CÁNCER DE MAMA DE MUJERES JÓVENES.....	173
2.1. <i>Las mujeres jóvenes, menores de 35 años, como entidad clínica y biológica diferenciada</i> .....	173
2.2. <i>La existencia de un perfil molecular característico de cáncer de mama.....</i>	176
2.3. <i>Validación del perfil característico de cáncer de mama en mujeres muy jóvenes .....</i>	179
PARTE 3: VALIDACIÓN DE LA EXPRESIÓN DE LOS MIRNAS SELECCIONADOS EN UNA SERIE INDEPENDIENTE DE MUESTRAS MEDIANTE QRT-PCR .....	182
3.1. <i>Selección de los miRNAs relevantes y representativos .....</i>	182
3.3. <i>Los miRNAs son validados mediante qRT-PCR.....</i>	183
3.4. <i>Relevancia de los miRNAs validados .....</i>	183
3.5. <i>miRNAs como marcadores de supervivencia .....</i>	185
PARTE 4: VALIDACIÓN DE LA IMPLICACIÓN DE LOS GENES SELECCIONADOS EN CÁNCER DE MAMA EN MUJERES JÓVENES .....	187
4.1. <i>Selección de los receptores de la ruta de NMDAR como genes candidatos.....</i>	187
4.2. <i>La expresión de los mRNAs en tejido tumoral mamario según la edad .....</i>	189
4.3. <i>Valoración de la cantidad de receptor in situ mediante IHC en mujeres muy jóvenes .....</i>	190
4.4. <i>Posibles aplicaciones terapéuticas .....</i>	193
CONCLUSIONES .....	196
BIBLIOGRAFÍA.....	199
ANEXOS .....	225

## Índice de figuras

---

---

<b>Figura 1.</b> Anatomía de la glándula mamaria. ....	26
<b>Figura 2.</b> Incidencia y mortalidad de cáncer por sexo y órgano. ....	31
<b>Figura 3.</b> Clasificación molecular del cáncer de mama.....	35
<b>Figura 4.</b> Características tumorales por grupos de edad. ....	39
<b>Figura 5.</b> Biogénesis y proceso de represión de los microRNAs.....	42
<b>Figura 6.</b> Rutas acopladas a la familia de receptores de membrana EGFR, asociadas con control del ciclo celular, evasión de la apoptosis y reparación del DNA. ....	49
<b>Figura 7.</b> Diagrama del diseño experimental llevado a cabo en este trabajo. ....	63
<b>Figura 8.</b> Diagrama de flujo del proceso de selección de pacientes para el estudio. ....	68
<b>Figura 9.</b> Método empleado por las sondas TaqMan para reportar la fluorescencia. ....	81
<b>Figura 10.</b> Cortes teñidos con Hematoxilina-eosina para el control histológico del porcentaje de celularidad tumoral ..... 88	88
<b>Figura 11.</b> Representación de la curva de Gauss característica de una distribución normal.....	91
<b>Figura 12.</b> Representación de un diagrama de Venn de dos grupos. ....	107
<b>Figura 13.</b> Análisis de la integridad del RNA total. ....	112
<b>Figura 14.</b> Valores de los Ct de las sondas estudiadas obtenidos para cada una de las muestras conservada en parafina y en congelación y la diferencia en Cts entre ambos formatos de congelación. ....	114
<b>Figura 15.</b> Correlación de los Ct entre las muestras de tejido conservado en parafina y su correspondiente par congelado.....	115
<b>Figura 16.</b> Diferencias en la recuperación de los RNA estudiados, medida por el Ct medio, en muestras congeladas y parafinadas.....	116
<b>Figura 17.</b> Histograma con los valores de intensidad obtenidos en cada uno de los chips de Affymetrix. ....	117
<b>Figura 18.</b> Distribución de la intensidad de las sondas obtenida para cada chip antes y después de la normalización. ....	118
<b>Figura 19.</b> Representación de la expresión de los microRNAs expresados en mujeres jóvenes respecto a mujeres mayores, ordenados por localización cromosómica. ....	119
<b>Figura 20.</b> Histograma de la distribución de las características tumorales de las pacientes estudiadas.....	120
<b>Figura 21.</b> Clustering jerárquico centrado en la mediana y visualización de la expresión de los miRNAs significativamente desregulados en mujeres jóvenes.....	122
<b>Figura 22.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 60. ....	124
<b>Figura 23.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 18. ....	125
<b>Figura 24.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 80. S ..... 126	126
<b>Figura 25.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 62. ....	127

<b>Figura 26.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 78. .....	128
<b>Figura 27.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 48. .....	129
<b>Figura 28.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 51. .....	130
<b>Figura 29.</b> Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 72. .....	130
<b>Figura 30.</b> Distribución de las rutas asociadas a los genes diana predichos por DIANA miRPath de los miRNAs desregulados en total y por nodo seleccionado .....	131
<b>Figura 31.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 60.....	132
<b>Figura 32.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 18.....	133
<b>Figura 33.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 80.....	134
<b>Figura 34.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs de los nodos 62 y 78.....	135
<b>Figura 35.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 48.....	136
<b>Figura 36.</b> Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 51 y 72.....	137
<b>Figura 37.</b> Heat-map del p-valor aplicando estadísticos conservadores y corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs seleccionados. ....	140
<b>Figura 38.</b> Red de interacciones de los miRNAs desregulados en cáncer de mama y sus dianas validadas en la ruta metabólica miRNAs en cáncer (hsa05206).....	146
<b>Figura 39.</b> Red de interacciones de los miRNAs desregulados en cáncer de mama y sus dianas teóricas en la ruta metabólica miRNAs en cáncer (hsa05206). ....	147
<b>Figura 40.</b> Representación de la expresión de los microRNAs en líneas celulares de cáncer de mama. Estudio realizado sobre líneas procedentes de mujeres jóvenes respecto a procedentes de mujeres mayores. ....	149
<b>Figura 41.</b> Representación del perfil de miRNAs desregulado en líneas celulares procedentes de mujeres jóvenes comparado con aquellas procedentes de mujeres mayores .....	150
<b>Figura 42.</b> Agrupación jerárquica de la expresión de los miRNAs desregulados en pacientes muy jóvenes de cáncer de mama por réplica de las líneas celulares analizadas.....	151
<b>Figura 43.</b> Representación del perfil de expresión de los miRNAs desregulados en muestras procedentes de mujeres jóvenes tras analizar las muestras de pacientes y líneas celulares conjuntamente.....	152
<b>Figura 44.</b> Representación de la expresión de los microRNAs expresados en mujeres jóvenes respecto a mujeres mayores obtenida a partir de los datos depositados en bases de datos online (METABRIC), ordenados por localización cromosómica .....	153
<b>Figura 45.</b> Clustering jerárquico supervisado centrado en la mediana y visualización de la expresión de los miRNAs significativamente desregulados en mujeres jóvenes en el estudio de los datos procedentes del estudio METABRIC. ....	154
<b>Figura 46.</b> Comparación de miRNAs significativos entre los diferentes estudios. ....	155

<b>Figura 47.</b> Significación de la expresión de los miRNAs tras combinar los p-valores de los tres estudios realizados .....	156
<b>Figura 48.</b> Expresión relativa de los miRNAs seleccionados de los nodos según el grupo de edad mediante qRT-PCR. ....	159
<b>Figura 49.</b> Curvas de supervivencia asociada a los grupos de edad y comparación entre mujeres ER negativas y ER positivas. ....	160
<b>Figura 50.</b> Curvas de supervivencia asociadas a la expresión de miRNAs seleccionados como posibles marcadores. ....	161
<b>Figura 51.</b> Esquematización de la presencia de expresión de las subunidades GRIN2B y GRIN3A del receptor de glutamato NMDAR. Se muestran los resultados en conjunto y separados por pacientes jóvenes (>35 años) y mayores (<45 años).....	163
<b>Figura 52.</b> Expresión relativa de los mRNAs de los genes <i>ERBB4</i> , <i>GRIN2B</i> y <i>GRIN3A</i> .....	166
<b>Figura 53.</b> Tinción negativa (A) y débil (B) mediante IHC anti-HER4 de tejido mamario FFPE	166
<b>Figura 54.</b> Tinción positiva de membrana mediante IHC anti-HER4 de tejido FFPE tumoral mamario. ....	167
<b>Figura 55.</b> Intensidad y localización de la tinción del anticuerpo anti-HER4 en muestras procedentes de mujeres jóvenes comparadas con mujeres mayores. ....	168
<b>Figura 56.</b> Tinción nuclear mediante IHC anti-HER4 de tejido FFPE tumoral mamario.....	168
<b>Figura 57.</b> Receptores iniciadores de cascadas de señalización en la célula relacionados con la ruta del glutamato .....	187





## Índice de tablas

---

<b>Tabla 1.</b> Caracterización por subtipos moleculares en cáncer de mama .....	35
<b>Tabla 2.</b> MicroRNAs típicamente desregulados en cáncer. Modificado de DiLeva et al. 2013. ....	43
<b>Tabla 3.</b> miRNAs desregulados en cáncer de mama (KEGG hsa05206) .....	45
<b>Tabla 4.</b> miRNA implicados en células madre de cáncer de mama (BCSCs). .....	46
<b>Tabla 5.</b> Características tumorales de las muestras empleadas en el estudio de viabilidad. ....	67
<b>Tabla 6.</b> Características clínicas de las pacientes de cáncer de mama empleadas en el estudio. ....	70
<b>Tabla 7.</b> Características de las líneas celulares empleadas en este estudio. ....	71
<b>Tabla 8.</b> Interpretación de los resultados de la determinación del estatus hormonal y de HER2 en muestras de tejido tumoral mamario .....	90
<b>Tabla 9.</b> Valores de concentración de RNA total y RIN de cada una de las muestras analizadas. ....	111
<b>Tabla 10.</b> Valores de concentración de moléculas de RNA de pequeño tamaño. ....	113
<b>Tabla 11.</b> Diferencias en la expresión de los miRNAs asociadas a las características tumorales estudiadas. ....	121
<b>Tabla 12.</b> Nodos desregulados en mujeres jóvenes seleccionados. ....	123
<b>Tabla 13.</b> Selección de miRNAs para la validación y los criterios empleados. ....	138
<b>Tabla 14.</b> Rutas significativamente enriquecidas en los genes diana de los miRNAs seleccionados. ....	139
<b>Tabla 15.</b> Rutas KEGG enriquecidas en genes diana validados afectados por los miRNAs seleccionados por encima del 75%. ....	141
<b>Tabla 16.</b> Rutas enriquecidas en genes diana teóricos de los miRNAs desregulados. ....	142
<b>Tabla 17.</b> Rutas enriquecidas en genes diana teóricos afectados por los miRNAs .....	142
<b>Tabla 18.</b> Diferencias en la expresión de los miRNAs seleccionados en los grupos de edad. ....	158
<b>Tabla 19.</b> Valores de expresión en tumor mamario obtenidos para los receptores <i>ERBB4</i> , <i>GRIN2B</i> y <i>GRIN3A</i> . ....	164



## Índice de anexos

---

<b>ANEXO I:</b> Características de las sondas empleadas en el estudio. ....	I
<b>ANEXO II:</b> Procedimiento del análisis de enriquecimiento de rutas de dianas validadas y teóricas de los miRNAs desregulados en mujeres jóvenes. ....	II
<b>ANEXO III:</b> Perfil electroforético de las muestras descongeladas y de sus correspondientes parafinas de RNA de pequeño tamaño. ....	VIII
<b>ANEXO IV:</b> Diferencia del Ct medio de cada muestra congelada y su correspondiente muestra parafinada (DCt = Ct medio muestra parafinada - Ct medio muestra congelada). ....	IX
<b>ANEXO V:</b> Tablas con los miRNAs con expresión diferente estadísticamente significativa en pacientes jóvenes con cáncer de mama. ....	X
<b>ANEXO VI:</b> Rutas enriquecidas en genes diana por nodos de miRNAs desregulados en cáncer de mama en mujeres jóvenes mediante DIANA miRPath ( <i>genes union</i> ). ....	XIV
<b>ANEXO VII:</b> Rutas enriquecidas en miRNAs desregulados en cáncer de mama en mujeres jóvenes mediante DIANA miRPath ( <i>pathway union</i> ) ....	XXII
<b>ANEXO VIII:</b> Rutas enriquecidas en los miRNAs seleccionados como representativos del perfil de cáncer de mama en mujeres jóvenes mediante DIANA miRPath ( <i>pathway union</i> ) ....	XXVIII
<b>ANEXO IX:</b> Valores estimados de represión debida a miRNAs desregulados en pacientes jóvenes de cáncer de mama en comparación con mayores. ....	XXIX
<b>ANEXO X:</b> Rutas enriquecidas en genes diana validados por el total de miRNAs desregulados ....	XXXI
<b>ANEXO XI:</b> Rutas enriquecidas en genes diana validados más afectados por los miRNAs desregulados ....	XXXIII
<b>ANEXO XII:</b> Interacciones en las rutas seleccionadas de los genes diana validados y teóricos más afectados por los miRNAs desregulados en mujeres jóvenes. ....	XXXV
<b>ANEXO XIII:</b> Valores de significación y de expresión de los miRNAs con expresión significativamente diferente entre mujeres jóvenes y mayores en, al menos, dos de los tres estudios realizados. ....	LI
<b>ANEXO XIV:</b> Características tumorales de las muestras de las que proceden los datos analizados del consorcio METABRIC en mujeres menores de 35 años y mayores de 65. ....	LIII
<b>ANEXO XV:</b> Meta-análisis mediante Fisher de los resultados de expresión de miRNAs obtenidos a partir de muestras de pacientes, líneas celulares y datos procedentes del estudio METABRIC. ....	LIV
<b>ANEXO XVI:</b> Valores de expresión relativa (RQ) obtenidos mediante qRT-PCR de los miRNAs seleccionados para su validación ....	LV
<b>ANEXO XVII:</b> Sitios altamente conservados en vertebrados de unión a miRNAs en la región 3'UTR de los genes diana seleccionados. ....	LVII
<b>ANEXO XVIII:</b> .....	LXI



## Índice de abreviaturas

---

<i>ABCC5</i>	ATP- <i>Binding Cassette</i> , Sub-Familia C (CFTR/MRP), miembro 5.
ABI2	Proteína de interacción con Abl (Abl-Interactor) 2.
AI	Inhibidores de la aromatasa.
<i>AIB1</i>	Coactivador del receptor nuclear 3.
AML	Leucemia mieloide aguda.
AMP	Adenosín monofosfato.
AMPc	AMP cíclico.
ANOVA	Prueba de análisis de varianzas.
APAF1	Factor activador de peptidasa apoptótica 1.
AR	Receptor de andrógenos.
<i>ATM</i>	Gen ataxia telangiectasia mutado.
AU	Pares Adenina-Uracilo.
<i>AURKA</i>	Aurora cinasa A.
<i>AVEN</i>	Inhibidor de la activación de caspasa.
<i>AXL</i>	Gen del receptor Tyrosina cinasa.
<i>BACH1</i>	Homólogo 1 de BTB y CNC.
<i>BAD</i>	Agonista de muerte celular asociado a BCL2.
BC	Cáncer de mama.
BCA	Datos de expresión de miRNAs de pacientes del Hospital Clínico Universitario de Valencia obtenidos mediante Affymetrix.
<i>BCL2</i>	Homólogo de Bcl-2 (célula-B CLL/linfoma 2).
BCLXL	Proteína 1 tipo Bcl-2.
BCSC	Célula madre de cáncer de mama.
BCVY	Cáncer de mama en mujeres muy jóvenes.
BIM	Mediador de muerte celular asociado a Bcl-2.
IMC	Índice de masa corporal.
<i>BMI1</i>	Proto-oncogén del grupo Polycomb RING 1 .
<i>BRCA1</i>	Gen del cáncer de mama 1.
<i>BRCA2</i>	Gen del cáncer de mama 2.
BrEt	Bromuro de etidio.
BRIP1	Proteína de interacción con BRCA1, también FANCI.
<i>BTC</i>	Betacelulina.
<i>c-Akt</i>	Serina/treonina cinasa Akt .
<i>CaMK</i>	Proteína-cinasa activada por calcio y calmodulina.
CCND1	Ciclina D1.
CCND2	Ciclina D2.
CCNE1	Ciclina-E1 específica de G1/S.
CD44	Clúster de diferenciación 44, también HCAM.
<i>CDC25a</i>	Fosfatasa 1 inductora de la fase-M.
<i>CDC34</i>	Homólogo del gen del ciclo de división celular 34.

CDI	Carcinoma ductal infiltrante.
CDIS	Carcinoma ductal in situ.
<i>CDK6</i>	Cinasa dependiente de ciclina 6.
CEBPB	CCAAT/Proteína de unión a <i>enhancer</i> (C/EBP) $\beta$ .
<i>CHEK2</i>	Cinasa del checkpoint del ciclo celular 2.
CK	Citoquina.
<i>c-KIT</i>	Homólogo del oncogén V-Kit Hardy-Zuckerman 4 viral del sarcoma felino.
CLA	Datos de expresión de miRNAs de líneas celulares obtenidos mediante Affymetrix.
CLI	Carcinoma lobulillar infiltrante.
CLIS	Carcinoma lobular in situ.
CLL	Leucemia linfocítica crónica.
CMPI	Carcinoma micropapilar infiltrante.
<i>CO</i>	Tejido conservado en congelación.
COP	Tejido parafinado.
COX	Prostaglandina-Endoperoxida sintasa 2.
CP	Tejido parafinado.
CRC	Cáncer colorrectal.
CREB	Proteína de unión CREB.
<i>Ct</i>	Ciclo umbral.
CTGF	Factor de crecimiento de tejido conectivo.
CTNNB1	Catenina (proteína asociada a cadherina), $\beta$ 1.
CYP1B1	Citochromo P450, Polipéptido 1.
DAB	Diaminobencidina.
DABG	Detectado por encima del fondo.
<i>DGCR8</i>	Gen 8 de la región crítica del síndrome de DiGeorge.
DICER	Ribonucleasa tipo III Dicer.
DMEM	Medio de cultivo <i>Eagle</i> modificado.
DNA	Ácido desoxirribonucleico.
DNTMT1	DNA (citosina-5)-metiltransferasa 1.
DNTMT3a	DNA (citosina-5)-metiltransferasa 3A.
DNTMT3b	DNA (citosina-5)-metiltransferasa 3B.
E2F1	Factor de transcripción 1 de E2F.
E2F2	Factor de transcripción E2F 2.
ECM	Matriz extracelular.
<i>EGA</i>	Base europea de datos genómicos (European Genome-phenome Archive).
EGFR	Receptor del factor de crecimiento epidérmico, también HER1.
EGP2	Glicoproteína epitelial 2.
<i>EIS1</i>	Homólogo del sitio de integración mieloide ecotrópico viral 1.
EMT	Transición epitelio mesénquima.
<i>EPR</i>	Receptor de peptidasa de célula efectora Effector cell peptidase receptor.
ER	Receptor de Estrógenos.
<i>ERBB2</i>	Gen homólogo 2 del oncogén de la leucemia eritrobástica viral aviar, también HER2 o CD340.
<i>ERBB3</i>	Gen homólogo 3 del oncogén de la leucemia eritrobástica viral aviar, también HER3.

<i>ERBB4</i>	Gen homólogo 4 del oncogén de la leucemia eritrobástica viral aviar, también HER4.
ERK	Cinasas reguladas por señales extracelulares.
<i>ETS1</i>	Homólogo 1 del oncogén del V-Ets virus E26 de la eritroblastosis aviar.
EUSOMA	Sociedad europea de especialistas en cáncer de mama.
<i>EZH2</i>	Subunidad del complejo de represión Polycomb 2, ( <i>Enhancer Of Zeste 2</i> ).
<i>FANCD1</i>	Gen de la anemia de Fanconi D1.
<i>FANCF</i>	Proteína de interacción con BRCA1, también BRIP1.
<i>FANCN</i>	Compañero y localizador de BRCA2, también PALB2.
FDR	Tasa de falsos positivos de Benjamini & Hochberg.
FFPE	Tejido fijado en formalina y embebido en parafina.
FGF2	Factor de crecimiento fibroblástico 2.
FGFR1	Receptor del factor de crecimiento fibroblástico 1.
FISH	Hibridación in situ fluorescente.
FN1	Fibronectin 1.
FOXO3A	<i>Forkhead box O3</i> .
FWER	Tasa de error <i>Familywise</i> .
<i>FZD3</i>	Receptor tipo <i>Frizzled 3</i> .
G3	<i>Grado III</i> .
<i>GAPDH</i>	Gliceraldehído-3-fosfato deshidrogenasa.
GATA3	Proteína de unión GATA 3, factor de transcripción específico de células T.
<i>GFAP</i>	Proteína ácida fibrilar de la glía ( <i>Glial Fibrillary Acidic Protein</i> ).
GFRAS	Sistema <i>Glutamate, Rhodopsine, AMP, Fungical, Smoothened</i> de clasificación de TRKs.
GLM	Modelo lineal generalizado.
<i>GMPc</i>	Guanosín monofosfato cíclico.
GnRH	Hormona liberadora de gonadotropina.
GPCR	Receptores acoplados a proteína G.
GPX2	Glutación Peroxidasa 2.
gra	<i>Gossypium raimondii</i> .
<i>GRIN2B</i>	Subunidad 2B del receptor ionotrópico del glutamato, N-Metil-D-Aspartato.
<i>GRIN3A</i>	Subunidad 3A del receptor ionotrópico del glutamato tipo N-Metil-D-Aspartato.
<i>GRM3</i>	Receptor metabotrópico del glutamato 3.
H&E	Hematoxilina y eosina.
HER	Factor de crecimiento epidérmico.
HER2	Gen homólogo 2 del oncogén de la leucemia eritrobástica viral aviar, también ERBB2 o CD340.
HER3	Gen homólogo 3 del oncogén de la leucemia eritrobástica viral aviar, también ERBB3.
HER4	Gen homólogo 4 del oncogén de la leucemia eritrobástica viral aviar, también ERBB4.
<i>HGFR</i>	Receptor del factor de crecimiento hepático.
HIF1	Factor inducido por hipoxia 1.
HMGA2	Grupo de alta movilidad de AT-hook 2.
<i>hMLH1</i>	Homólogo humano 1 de mutL.
<i>hMSH2/6</i>	Homólogo humano 2/6 de mutS.

<i>HOXD10</i>	Homeobox D10.
<i>HPRT1</i>	Hipoxantina-guanina fosforribosil-transferasa.
HR	Riesgo, del inglés <i>Hazard's ratio</i> .
<i>H-RAS</i>	GTPasa Hras, del homólogo del oncogén viral del sarcoma de rata Harvey.
HRT	Terapia hormonal sustitutiva.
hsa	<i>Homo sapiens</i> .
<i>IBC</i>	Cáncer de mama inflamatorio.
IGF-1	Factor de crecimiento por insulina 1.
IGFR	Receptor del factor de crecimiento por insulina 1.
IHC	Tinción inmunohistoquímica
IHQ	Tinción nmunohistoquímica.
IL1RN	Antagonista del receptor de la interleucina 1.
<i>IL6</i>	Interleucina 6.
IMP-1	Subunidad 1 de la proteasa de la membrana interna mitocondrial.
<i>IRS-1</i>	Sustrato del receptor de la insulina 1.
<i>ITGB3</i>	Integrina $\beta 3$ .
<i>JAG1</i>	Homólogo de Jagged 1.
<i>JAK1</i>	Janus cinasa 1.
JNK	Cinasas c-Jun-N-terminal.
Ki67	Marcador de proliferación Ki 67.
KIP1	Inhibidor 1B de cinasa dependiente de ciclina.
KIP2	Inhibidor 2 de cinasa dependiente de ciclina.
<i>KRAS</i>	Homólogo del oncogén viral del sarcoma de rata Kirsten.
LEPR	Receptor de leptina.
LHRH	Hormona liberadora de la hormona luteinizante.
<i>LIN28</i>	Homólogo A de Lin-28.
lja	<i>Lotus japonica</i> .
<i>LUM</i>	Luminal.
MAL1	Glucosidasa de maltosa 1 Mal1.
<i>MALM3</i>	Homólogo del gen del regulón de maltosa.
MAPK	Proteína-cinasas activadas por mitógeno.
<i>MASPIN</i>	Inhibidor de la proteasa de serina mamaria.
MBD	Datos de expresión de miRNAs del consorcio METABRIC obtenidos mediante EXIQON.
Mcl-1	Proteína de diferenciación celular en leucemia mieloide inducida, También MCL1.
<i>MERTK</i>	Tirosina cinasa del proto-oncogén MER.
MET	Receptor tirosina cinasa del proto-oncogén MET.
METABRIC	<i>Molecular Taxonomy of Breast Cancer International Consortium</i> .
<i>mGluR</i>	Receptor metabotrópico del glutamato.
miRNAs	Molécula silenciadora de RNA de pequeño tamaño, también microRNA.
<i>MMP16</i>	Metalopeptidasa de la matriz 16.
MnSOD	Superóxido dismutasa 2, mitocondrial.
<i>MRE11A</i>	Homólogo A del gen de recombinación meiótica 11.
<i>M-RIP</i>	Proteína asociada a miosina fosfatasa $\rho$ .



mRNA	RNA mensajero.
mTOR	Diana mecanística de Rampamicina, también FRAP1.
MYC	Homólogo del oncogén viral de la mielocitomatosis aviar v-Myc.
MYCN	Homólogo del oncogén viral de la mielocitomatosis aviar v-Myc derivado de neuroblastoma.
Myt-1	Factor de transcripción de mielina 1.
NANOG	Nanog Homeobox.
NBN	Gen de la nibrina.
ncRNA	RNA no codificante.
NF-κB	Factor nuclear del polipéptido κ.
NGF	Factor de crecimiento neuronal.
NMDAR	Receptor de N-Methyl-D-Aspartato (glutamato).
NOR	Muestras procedentes de mujeres libres de cáncer.
NOTCH	Proteína homóloga del locus neurogénico Notch.
Nrg	Neurregulina.
OC	Contraceptivos orales.
OCT	Tejido congelado incluido en criomoldes.
ORF	Pauta de lectura abierta.
p38a	Proteína cinasa activada por mitógeno 14, también MAPK14.
p63	Proteína asociada a transformación 63.
p85	Subunidad reguladora de PI3K.
PALB2	Compañero y localizador de BRCA2, también FANCN.
PBS	Tampón fosfato salino.
PCNA	Antígeno nuclear de proliferación celular ( <i>Proliferating Cell Nuclear Antigen</i> ).
PCR	Reacción en cadena de la polimerasa.
P <sub>CT</sub>	Probabilidad de unión conservada preferente.
PDCD4	Proteína 4 de muerte celular programada ( <i>Programmed cell death protein 4</i> ).
PDCD4	Programmed Cell Death 4 (Neoplastic Transformation Inhibitor).
PDGF	Platelet-derived growth factor.
PDGFR	Platelet-derived growth factor receptor.
PDK	Dominio de la proteína cinasa.
PEBP1	Proteína de unión a Fosfatidiletanolamina 1.
PI3K	Fosfatidil inositol 3 cinasa.
PIP3	Fosfatidil-inositol-3,4,5-trifosfato .
piRNA	RNA asociado a pi-wi.
PKA	Proteína-cinasas activadas por cAMP.
PKC	Proteína-cinasas activadas por calcio y fosfolípidos.
PKD1	Cinasa 1 dependiente de fosfatidil inositol-3.
PLC	Fosfolipasa C.
PLZF	Ortólogo de dedos de zinc de la leucemia promielocítica.
Pol	Polimerasa.
PR	Receptor de Progesterona.
pta	<i>Pinus taeda</i> .
PTEN	Homólogo de la fosfatasa y tensina.

PTHrP	Proteína relacionada con la hormona paratiroidea.
PTPmu	Receptor tipo M de proteína tirosina fosfatase.
PTPRO	Receptor tipo O proteína tirosina fosfatasa.
PUMA	Modulador de la apoptorin regulado por P53.
QC	Control de calidad.
qRT-PCR	PCR cuantitativa a tiempo real.
<i>RAD50</i>	Homólogo de la proteína de reparación del DNA RAD50.
RAD51C	Parálogo C de RAD51.
<i>RAS</i>	Proteína RAS.
<i>RDX</i>	Radixina.
RECK	Reversion-Inducing-Cysteine-Rich Protein With Kazal Motifs.
<i>REDD1</i>	Proteína regulada en el desarrollo y respuesta al dño en DNA ,1.
<i>RHOA</i>	Homólogo del miembro A de familia Ras.
<i>RIN</i>	Número de integridad del RNA.
RISC	Complejo silenciador inducido por RNA.
RMA	Media robusta multimatrices.
RNA	Ácido ribonucleico.
RNAi	RNA de interferencia.
RNU6B	RNA nuclear de pequeño tamaño, también U6B.
RNU43	RNA nuclear de pequeño tamaño, también U43.
RPMI	Medio RPMI 1640.
RQ	Cantidad relativa.
rRNA	RNA ribosomal.
RTx	Radioterapia.
S/B	Señal de fondo.
<i>S100B</i>	Gen de la proteína de unión a calcio S100B.
SBF	Suero bovino fetal.
SD	Desviación típica.
SE	Error estándar.
SERM	Moodulador selectivo del receptor de estrógenos.
SET8	Dominio SET (Metiltransferasa de lisina) 8.
SHIP1	Dominio SH2 Inositol 5-fosfatasa 1.
SIRT1	Sirtuína-1 deacetilasa dependiente de NAD.
SMAD4	Miembro 4 de la familia SMAD.
SMAD5	Miembro 5 de la familia SMAD.
SNAIL	Homólogo de Snail.
snoRNA	RNA nucleolar de pequeño tamaño.
snRNA	RNA nuclear de pequeño tamaño.
SOCS1	Supresor de la señalización por citoquinas 1.
SOX2	SRY (Región Y determinante del sexo)-Box 2.
SPRY1	Sprouty Homolog 1, Antagonist Of FGF Signaling (Drosophila).
SPS	Estabilidad de la unión a semilla.
STK11/LKB	Cinasa serina-threonina 11.
SUZ12	SUZ12 subunidad del complejo de represión Polycomb 2.

<i>Suz12</i>	Subunidad del complejo de represión 2 Polycomb SUZ12.
TA	Cantidad de sitios diana.
Taq	<i>Thermus aquaticus</i> .
TBS	Tampón salino Tris.
TCL1	Gen de la leucemia de células T 1A.
TGFβII	Factor de crecimiento transformación β III.
TIMP3	Inhibidor 3 de TIMP Metallopeptidasa.
TMA	Micromatriz de tejidos.
TN	Subtipo triple negativo.
<i>TNBC</i>	Cáncer de mama triple negativo.
<i>TNC</i>	Tenascina C.
TNM	Sistema de estadificación por tamaño, afectación nodular y presencia de metástasis.
TP53	Proteína tumoral p53.
TPM1	Tropomiosina 1 α.
TRBP	Proteína de unión al RNA TAR.
<i>TRK</i>	Receptores tirosina-cinasa.
trkA	Receptor tipo 1 tirosin-cinasa neurotrófica.
tRNA	RNA de transferencia.
TRPS1	Gen del Síndrome Trichorhinophalangeal I.
TRXR2	Thioredoxina Reductasa 2.
TSP1	Trombospondina 1.
<i>UBC9</i>	Homólogo de la enzima E21 conjugada a ubiquitina (Ube9).
<i>UFO</i>	Receptor de la proteína tirosina-cinasa UFO.
UPGMA	Método de agrupamiento de pares por media aritmética.
<i>UTR</i>	Región no traducida.
UV-Vis	Espectroscopía de luz ultravioleta-visible.
VEGF	Factor de crecimiento del endotelio vascular.
VEGFR2	Receptor del dominio de cinasa.
WAF1	Inhibidor 1 de cinasa dependiente de ciclina.
WASF2	Miembro 2 de la familia de proteína WAS.
WEE1	WEE1 cinasa de checkpoint de G2.
<i>Wip1</i>	Protein fosfatasa dependiente de Mg <sup>2+</sup> /Mn <sup>2+</sup> 1D.
WNT	Proteína proto-oncogén Wnt ( <i>Wingless-type</i> MMTV).
<i>ZBTB10</i>	Proteína con dominio de unión BTB y dedos de zinc, 10.
ZEB1/2	Proteína con Homeobox E-box de unión a dedos de zinc, 1/2.



## Resumen

---

---

El cáncer de mama constituye el cáncer invasivo más frecuente en mujeres, afectando a una de cada 8 mujeres a lo largo de su vida, siendo, además, la primera causa de muerte derivada de cáncer invasivo en la población española (15,5%). Las mujeres jóvenes representan únicamente el 3% del total de casos diagnosticados; sin embargo, estas pacientes presentan tumores de mayor tamaño e índice de proliferación, estando sobrerrepresentados subtipos tumorales más agresivos y de peor pronóstico.

Los microRNAs (miRNAs) son pequeñas moléculas de RNA no codificante, cuyo papel en la célula es regular la expresión de genes, bloqueando el paso de RNA mensajero a proteína. Su desregulación es típica en numerosas enfermedades, incluyendo cáncer, constituyen una herramienta muy útil como biomarcador diagnóstico y pronóstico debido a su estabilidad y reproducibilidad.

El principal objetivo del trabajo de esta tesis ha sido establecer un perfil molecular característico del cáncer de mama en mujeres jóvenes. Para ello se seleccionaron 22 pacientes de cáncer de mama menores de 35 años y 12 pacientes mayores de 65 que no hubiesen recibido tratamiento neoadyuvante, no metastásicas y sin mutación en BRCA1/2, además de 3 mujeres sanas. Mediante un chip de hibridación de miRNAs se obtuvo la expresión en tejido tumoral de 1100 miRNAs humanos, de los cuales 91 mostraron una expresión diferente en mujeres menores de 35 años. Se obtuvo un perfil de miRNAs específico de mujeres jóvenes con cáncer de mama. Este perfil identificativo fue validado en líneas celulares procedentes de mujeres jóvenes (HBL100, HCC1500, HCC1937). Se estudió, además, en una serie de datos de 750 pacientes de cáncer de mama obtenidas a partir de datos depositados online de un estudio previo (METABRIC). Los miRNAs desregulados en mujeres jóvenes apuntaban a genes diana de rutas relacionadas con proliferación, supervivencia celular, migración y movilidad celular, y relacionadas directamente con procesos cancerosos.

Se seleccionaron 18 miRNAs representativos, de los cuales, seis fueron validados mediante qRT-PCR en una nueva serie de 55 pacientes: hsa-miR-30c-5p, hsa-miR-92b, hsa-miR-139, hsa-miR-1228, hsa-miR-1275, hsa-miR-3196. Los miRNAs hsa-miR-92b y hsa-miR-139 mostraron una peor supervivencia asociada a su represión y sobre-expresión, respectivamente. Entre sus genes diana, se seleccionaron receptores de membrana relevantes desde un punto de vista traslacional: *HER4*, *GRIN2B*, *GRIN3A* y *GRM3* y se evaluó su expresión mediante qRT-PCR, mostrando *GRIN2B* una expresión mayor en mujeres jóvenes. Se midió la cantidad de proteína HER4 en 31 cortes de tejido tumoral de pacientes jóvenes y 29 de mayores de 65 años, mediante tinción inmunohistoquímica in situ, lo que confirmó que existe una mayor expresión de HER4, localizada en membrana, en mujeres jóvenes.

Los resultados respaldan la idea de que el cáncer de mama en mujeres jóvenes constituye una entidad clínica y biológica diferenciada dentro del cáncer de mama general, con una etiología particular que responde a una desregulación de procesos celulares relacionados con proliferación, supervivencia celular, migración y movilidad celular, que merece una mayor atención de cara a ofrecer un tratamiento personalizado más acorde a las necesidades particulares de este grupo de pacientes.



## Abstract

---

---

Breast cancer is the most frequent cancer in women, affecting one in eight women in their lifetime. It is also the first invasive cancer derived cause of death in Spanish population (15.5%). Young women only account for the 3% of total diagnosed cases, however, these patients usually present larger tumors with higher proliferation and overrepresented in more aggressive tumor subtypes.

MicroRNAs (miRNAs) are small non coding RNA molecules that regulate gene expression by silencing the protein translation. MiRNA deregulation typically affects in various diseases, including cancer, and constitute a useful tool as a diagnosis and prognosis biomarker, due to their stability and reproducibility.

The aim of this study was to establish a molecular profile unique of breast cancer in very young women. With that purpose, 22 breast cancer patients younger than 35 years and 12 patients older than 65 years were selected, besides of 3 healthy women. None of the patients received neoadjuvant treatment, were metastatic nor/or have known BRCA1/2 mutation. Using a miRNA Affymetrix microarray, the expression of 1100 miRNAs was assessed in tumoral tissue. 91 of those miRNAs showed differential expression in patients younger than 35 years. The analyses performed revealed a molecular profile distinctive of young women with breast cancer. The profile was validated on cell lines derived from mammary tumor of young women (HBL100, HCC1500, HCC1937). The analyses were also carried out on a dataset from 750 breast cancer patients obtained from a published study (METABRIC repository). Deregulated miRNAs in young women pointed out to target genes related with proliferation, cell survival, migration and cell motility pathways, and with cancer-related events.

Eighteen representative miRNAs were selected, six of those were validated by qRT-PCR in an independent series of 55 patients: hsa-miR-30c-5p, hsa-miR-92b, hsa-miR-139, hsa-miR-1228, hsa-miR-1275, hsa-miR-3196. MiRNAs hsa-miR92b and hsa-miR-139 show differential survival prognosis when downregulated and upregulated, respectively. Among the target genes, relevant receptors were selected: *HER4*, *GRIN2B*, *GRIN3A* y *GRM3* and their mRNA expression was measured using qRT-PCR assays. The results showed a higher expression of *GRIN2B* in tissue from young women. The *HER4* protein was quantified by IHC in tumoral tissue from 31 young patients and 29 older patients, confirming a higher membrane expression of *HER4* in young women.

The results of this work support the idea that breast cancer in very young women constitutes a clinical and biological entity inside breast cancer, which particular etiology is due to the deregulation of cellular processes related to proliferation, cell survival, migration and cell motility. Breast cancer in very young women, though, deserves utter attention towards offering more focussed treatment that grants the needs of this particular group of patients.





# INTRODUCCIÓN



---

---

# CÁNCER

---

---

Parte

# 1

Lo que se entiende comúnmente por cáncer se trata en realidad de un neoplasma maligno. El término neoplasma se acuña del griego *νεο- neo*, "nuevo" y *πλάσμα plasma*, "formación", se puede deducir que se trata de una masa anormal de tejido de nueva formación, producida como resultado de una neoplasia. Se llama neoplasia a la proliferación anormal de células que crecen de manera no coordinada y a mayor velocidad que la usual del tejido adyacente. Antiguamente el término tumor se utilizaba para cualquier tipo de hinchazón, no en vano tumor era el vocablo en latín para hinchazón, bulto,... pero actualmente en medicina moderna se llama tumor a un neoplasma que, encapsulado o no, excede los límites del tejido formando una protuberancia. No todos los neoplasmas acabarán abultándose, ni siendo malignos. Los neoplasmas pueden ser benignos, pre-malignos (carcinoma *in-situ*) o malignos (carcinoma invasivo o cáncer).

- \* **Neoplasma benigno:** están localizados y limitados y no se transforman en cáncer. Se incluyen en esta denominación fibromas uterinos y nevi melanocíticos, por ejemplo.
- \* **Neoplasma potencialmente maligno o pre-maligno:** incluyen carcinomas *in situ*, a diferencia de los malignos no invaden y dañan los tejidos, pero si no se tratan y tras pasar un tiempo suficiente pueden o no evolucionar a neoplasma maligno.
- \* **Neoplasma maligno:** comúnmente llamado cáncer, invaden y dañan el tejido adyacente, pudiendo tener o desarrollar capacidad metastática.
- \* **Neoplasma secundario:** se refiere a neoplasma de cualquier tipo que puede tratarse tanto de un tumor producto de la metástasis derivada de un tumor primario como de un tumor aparentemente no relacionado.

Neoplasma maligno es, en realidad, un término general para un grupo heterogéneo de varias enfermedades, existen más de 200 tipos de cáncer descritos, que tienen como características comunes el crecimiento celular incontrolado y la capacidad de invadir localmente a los tejidos adyacentes o a tejidos distantes, denominándose a este proceso metástasis. Las señas de identidad de un proceso canceroso están bien definidas, pudiendo tener una célula parte o la totalidad de ellas, siendo los más relevantes la evasión de la muerte celular, inmortalización celular, evasión de los sistemas de supresión del crecimiento, activación de la capacidad de invasión y metástasis, inducción de la angiogénesis y mantenimiento aberrante de las señas de proliferación (1, 2).

Al tratarse de una enfermedad compleja de penetrancia variable, el riesgo de padecer cáncer depende de un gran número de factores, siendo estos tanto ambientales como genéticos, que, a su vez, varían dependiendo del tipo de cáncer.

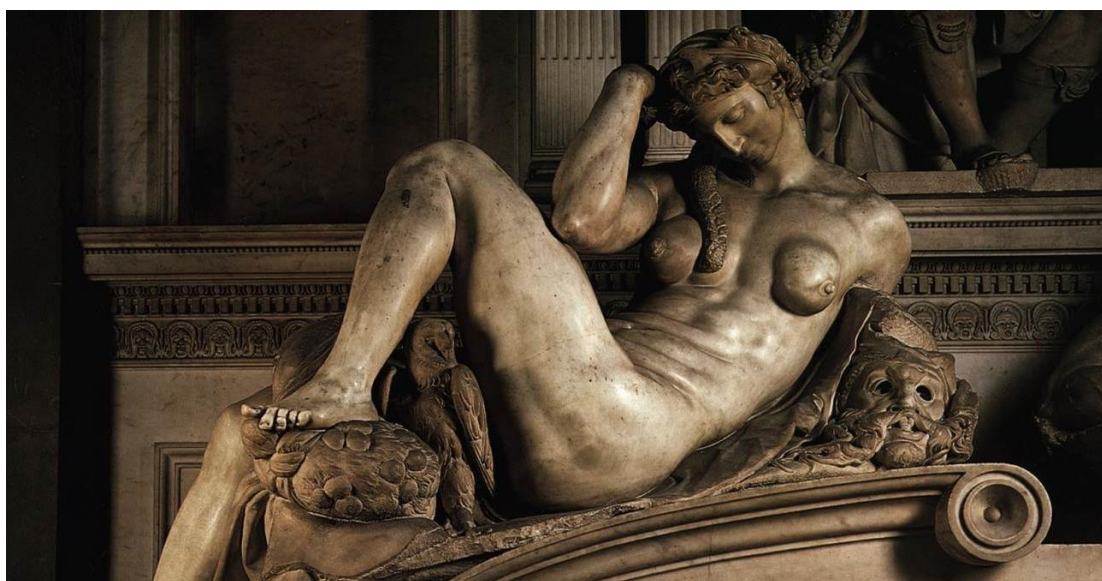


# CÁNCER DE MAMA

Parte

# 2

El cáncer de mama es uno de los cánceres mejor conocidos desde la antigüedad y del que se tienen más registros documentados debido a su visibilidad, pudiéndose observar en el arte a lo largo de la historia.



*Noche*, Miguel Ángel 1526/31, escultura en mármol. Nueva sacristía de San Lorenzo, Florencia.



De izquierda a derecha: *Retrato de una joven (La Fornarina)*, Raffaello Sanzio 1518/19, óleo sobre tabla, Galería Borghese, Roma. *Las tres gracias*, Peter Paul Rubens 1636/39, óleo sobre tabla, Museo del Prado, Madrid. *Betsabé con la carta de David*, Rembrandt Harmenszoon van Rijn 1645, óleo sobre lienzo, Museo del Louvre, París.

El primer registro del que se tiene constancia proviene del Antiguo Egipto y data del 1500 a.C. El *Papiro Edwin Smith* es un tratado de medicina que, entre 48 historias clínicas, describe ocho casos de tumor mamario, para los que “*no existe tratamiento una vez el tumor es sangrante y duro*”. En la antigua Grecia, Hipócrates describe varios casos y apunta a que las pacientes con tumores infiltrantes no deben ser tratadas, ya que así aumenta su supervivencia.

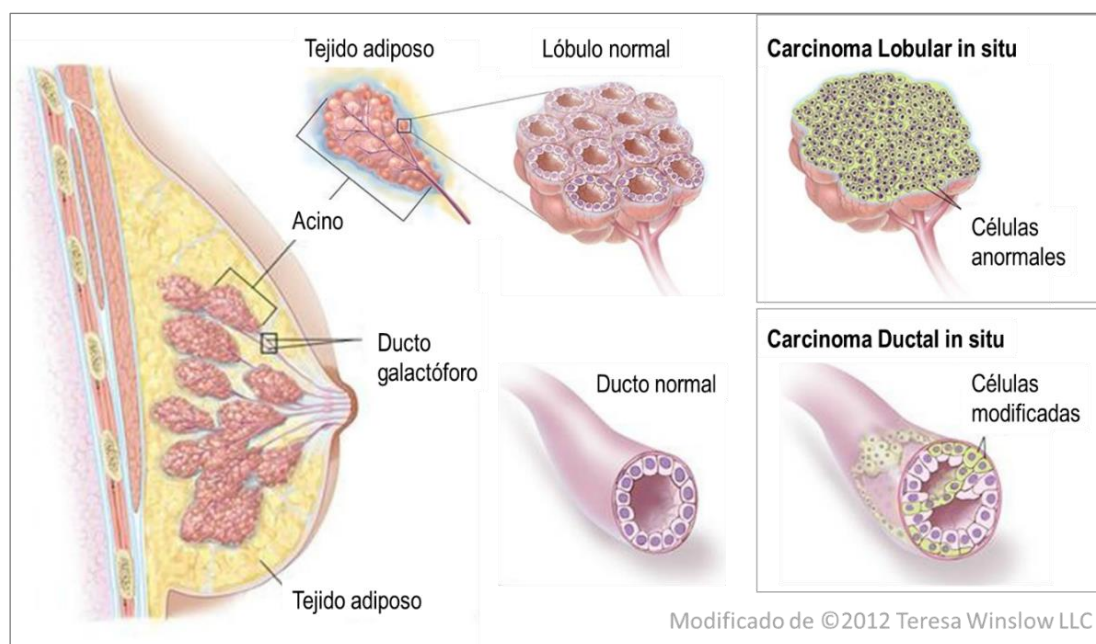
Durante la historia varios médicos han descrito casos con una triste conclusión, no fue hasta que la medicina amplió su conocimiento en el sistema circulatorio, en el siglo XVII, cuando se obtuvieron los primeros resultados favorables, al poder relacionar la enfermedad con los nódulos linfáticos axilares y ser capaces de extraer el tumor junto con los nódulos linfáticos y los músculos pectorales; aunque la mortalidad se mantuvo debido a la cantidad de muertes por sepsis derivada, abrió el camino a la mastectomía moderna.

El cáncer de mama es un neoplasma maligno que se origina en tejido mamario. Se le denomina carcinoma debido a su origen epitelial, aunque en sentido estricto se trata en algunos casos de adenocarcinomas por su origen glandular, sarcomas que derivan de tejido conectivo o linfomas de los ganglios linfáticos, estos dos últimos en menor frecuencia. Dado que en primera instancia derivan de tejido epitelial, el término carcinoma es correcto médicamente.

## 1. GLÁNDULA MAMARIA

### 1.1. Anatomía

La glándula mamaria es un órgano presente en mamíferos cuya función es producir la leche que alimentará a la descendencia. Se trata de un tipo específico de glándula apocrina, cuya secreción se produce por escisión de parte de la célula junto con las vesículas secretadas (Figura 1).



**Figura 1. Anatomía de la glándula mamaria.** Representa un proceso tumoral que origina bien un carcinoma lobular in situ (CLIS) o un carcinoma ductal in situ (CDIS) según las células de procedencia.

El componente principal de la glándula mamaria son los alvéolos, cavidades cuboidales tapizadas por células productoras de leche y células mioepiteliares. Estos alvéolos se agrupan formando los llamados acinos glandulares o lóbulos, desembocando cada lóbulo en un conducto galactóforo, que es el encargado de dirigir la leche producida hasta el exterior por el pezón. Las células mioepiteliares se contraen por el efecto de la oxitocina, dirigiendo la leche secretada por el ducto. Los componentes se encuentran embebidos en una matriz extracelular (ECM) llamada estroma mamario, compuesto principalmente por adipocitos, fibroblastos y células inflamatorias, entre otros.

## 1.2. Desarrollo de la glándula mamaria

La glándula mamaria es el único órgano cuyo desarrollo se produce en diferentes etapas, y alcanza su madurez mucho después del desarrollo embrionario y del nacimiento (3-5).

### Desarrollo embrionario y post-natal

Durante el desarrollo embrionario se forma un sistema ductal rudimentario, dependiente de hormonas sistémicas y maternas, así como de la comunicación paracrina entre las células epiteliales y mesenquimales por medio de la proteína relacionada con la hormona paratiroidea (PTHrP). En las primeras semanas tras el nacimiento se produce un breve desarrollo de los conductos galactóforos, pero en su mayoría este desarrollo se produce en la pubertad (6). El estrógeno promueve la diferenciación de los ductos rudimentarios, y su crecimiento es controlado por el factor de crecimiento epidérmico EGF y por componentes de la matriz extracelular (ECM) (7).

### Embarazo y lactancia

En las primeras etapas del embarazo, los aumentos en los estrógenos y progesterona promueven el desarrollo del árbol de los conductos galactóforos, que alcanzan su máximo durante la lactancia. Tras la lactancia se desencadena la involución de la glándula mamaria proceso en el cual se pierde drásticamente la mayoría de tejido mamario epitelial y glandular, que sufre muerte celular programada o apoptosis (3, 8, 9).

## 2. EPIDEMIOLOGÍA Y RIESGOS DEL CÁNCER DE MAMA

---

### 2.1. Factores de riesgo a padecer cáncer de mama

#### Factores hormonales externos

La terapia hormonal sustitutiva (HRT, del inglés *hormone replacement therapy*) se emplea para tratar los síntomas provocados por la menopausia. Las evidencias, procedentes tanto de estudios epidemiológicos como de ensayos clínicos, indican un elevado riesgo entre la población de mujeres receptoras de HRT, dependiente de la duración del tratamiento, reduciéndose por cada año llegando al nivel del riesgo normal tras 5 años de cese en el tratamiento (10, 11)

El riesgo también es un 15% mayor en usuarias actuales de contraceptivos orales (OC), sin embargo tras 10 años sin usarlos el riesgo se iguala con el de no usuarias, se considera únicamente un factor de riesgo en mujeres jóvenes, que tienen una incidencia de cáncer de mama basal, por

lo que el uso de OC no es relevante como factor de riesgo a padecer cáncer de mama en la población general (10, 12).

### **Factores hormonales internos**

La edad de la menarquía y la menopausia influye en el riesgo de padecer cáncer de mama, ya que éste es mayor conforme aumenta el número de ciclos menstruales. Este riesgo disminuye un 15% por cada año de retraso en la menarquía y aumenta un 3% por cada año más en la edad de menopausia. Ambos factores están relacionados en tanto que determinan los años totales de carga hormonal acumulados por una mujer (11) (13).

Tanto el embarazo como la lactancia aumentan el riesgo a padecer cáncer de mama a corto plazo, probablemente debido al aumento en los niveles de estrógenos en los primeros meses de la gestación. Se ha propuesto la existencia de una asociación entre la involución de la glándula mamaria tras la lactancia y el mayor riesgo a desarrollar cáncer de mama en los 5 años posteriores al embarazo, este hecho es debido, según algunos autores, a la reordenación de la matriz extracelular y modificación del microambiente, facilitando la proliferación de las posibles células pre-malignas existentes (14, 15).

Sin embargo, embarazo y lactancia ejercen un papel protector en el riesgo a largo plazo. La nuliparidad actúa como factor de riesgo conocido, debido a que las mujeres con al menos un embarazo llevado a término alcanzan niveles menores de globulinas receptoras de hormonas sexuales y un aumento en la prolactina. Adicionalmente la lactancia tiene un efecto beneficioso al suprimir la ovulación durante la misma, reduciendo el riesgo en un 4,3% por cada año de lactancia (16).

Por lo tanto, el primer embarazo a edad temprana, así como un efecto protector ligero propiciado por el número de embarazos, junto a la lactancia constituyen los principales factores protectores frente al cáncer de mama (17) (13).

Síntomas fibroquísticos y fibroadenomas, las enfermedades de mama más comunes de carácter benigno, se asocian con un riesgo de 2 a 3 veces más elevado de padecer cáncer de mama, no se trata de condiciones pre-neoplásicas, pero sí denotan una elevada proliferación epitelial, ligada a alteraciones hormonales intrínsecas.

### **Edad**

La edad se considera un factor directamente relacionado con riesgo de padecer cáncer de mama, siendo edades más avanzadas de mayor riesgo. Probablemente esto está asociado al mayor número de ciclos menstruales acumulados por la mujer y una mayor carga hormonal recibida. Debido a que el cáncer de mama es una enfermedad post-menopáusica en su mayoría, este factor se vería influenciado por la edad de aparición de la menopausia y por la administración a este grupo de mujeres de HRT, demostrado factor de riesgo (13, 18-21).

Basándonos en estos supuestos la relación entre edad y aparición de cáncer de mama debería ser lineal, siendo más elevada a más años, pero en la práctica se observa una distribución bimodal, con un pico de aparición a los 50 años y otro a los 65 años de edad (18). Por esto la comunidad se refiere a ambos tipos de cáncer de mama como pre-menopáusicos y post-menopáusicos, con algunas diferencias en cuanto a características tumorales, incidencia y pronóstico.



## Factores relacionados con hábitos de vida

La ingesta de alcohol aumenta hasta en un 30% el riesgo a desarrollar un cáncer de mama en consumidoras de 3 o más bebidas alcohólicas diarias (22).

Existen varios estudios que evalúan la relación entre la dieta y el riesgo de padecer cáncer de mama, sin embargo no se ha hallado ninguna asociación remarcable con cáncer de mama (23, 24). De todos los aspectos considerados, únicamente parece haber un ligero papel protector en la ingesta de soja y folatos (por el efecto neutralizador que ejercen sobre la toma de alcohol).

Es posible, sin embargo, que los niveles hormonales y la nutrición durante el período intrauterino e infancia sí afecten en la carcinogénesis de mama. De hecho, algunos trabajos relacionan el tamaño y peso al nacer (derivado de la ingesta nutricional de la madre durante el embarazo, entre otros factores) con un 6% mayor riesgo de padecer cáncer de mama por cada 500gr de sobrepeso y un 9% por cada dos centímetros de aumento en el tamaño relativo (25, 26).

Aparte del tamaño como componente de riesgo, pero consistente con él, otros factores antropomórficos han demostrado tener relación con la etiología del cáncer de mama. Se ha establecido el índice de masa corporal (IMC) como factor de riesgo, relacionándose valores de IMC superiores a 30 (es decir correspondientes con obesidad) con mayor probabilidad a desarrollar cáncer de mama post-menopáusico (por encima de los 50 años de edad) (23). Sorprendentemente, en mujeres pre-menopáusicas (menores de 50 años) parecen ser valores de IMC bajos los que aumentan este riesgo (18).

Probablemente, tanto la obesidad como la ingesta de alcohol, tengan relación con mecanismos que impliquen aspectos hormonales o metabólicos.

## Predisposición genética

La existencia de antecedentes familiares aporta un riesgo de padecer cáncer de mama de 2 a 3 veces mayor que el de la población normal, y este riesgo aumenta según el número de parientes de primer grado afectados, la edad de diagnóstico y si en algún caso se trata de cáncer de mama bilateral o en hombres (27). Alrededor de un 10% de los casos de cáncer de mama tienen un componente familiar, pudiéndose atribuir a factores genéticos y genes de alta, media y baja penetrancia, en su mayoría relacionados con el metabolismo y regulación de hormonas, y daño y reparación del DNA.

**Genes de alta penetrancia:** Del porcentaje de cáncer de mama familiar, un 20% corresponde a familias portadoras de mutaciones en los genes de alta penetrancia *BRCA1* y *BRCA2 (FANCD1)*. Mutaciones en estos genes aumentan el riesgo de padecer cáncer de mama hasta en un 80%, además de un 15-40% de padecer cáncer de ovario y un 65% de desarrollar un segundo tumor de mama. En estas pacientes el cáncer de mama se presenta a edades tempranas y su evolución tiende a mal pronóstico. Ambos son genes supresores de tumores en la ruta de reparación del DNA.

La frecuencia de mutaciones en los genes *BRCA1* y *BRCA2* varía dependiendo de la población, pero en general se mantiene baja, con una prevalencia de 1 en 860, y 1 en 740, respectivamente, explicando hasta un 25% de los casos familiares. Existen otros genes que pueden explicar este componente familiar, como son *TP53*, *PTEN*, *STK11/LKB*, *RAD51C*.

**Genes de media y baja penetrancia:** Además de los genes de alta penetrancia, mutaciones en muchos otros genes provocan un aumento del riesgo de cáncer de mama. En esta

categoría se incluyen mutaciones en *CHEK2*, *ATM*, *NBN*, *MRE11A*, *RAD50*, *BRIP1*, *PALB2* (28-31).

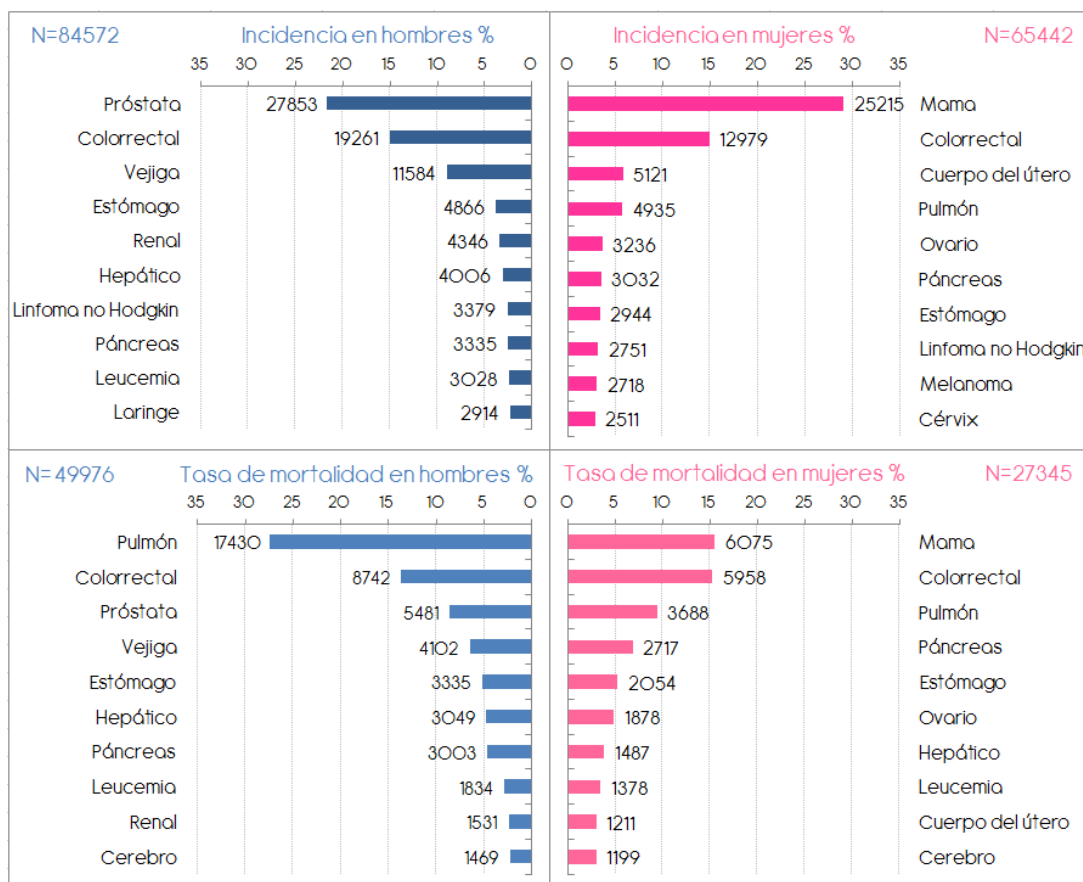
- \* *CHEK2* (Cinasa del checkpoint del ciclo celular 2) es un gen supresor de tumores. Codifica para una cinasa que interviene en el control del ciclo celular, regulando la división celular e impidiendo el crecimiento descontrolado.
- \* *ATM*, o gen ataxia telangiectasia mutado, es el gen que codifica para la enzima fosfatidil inositol cinasa (PI3K), implicada en la regulación de los procesos de control de la división celular y en la reparación del DNA. Junto con la proteína mTOR, entre otras, conforman el grupo de las fosfatidil inositol-3-cinasas de clase V. Las mutaciones en este gen predisponen a la aparición de diferentes tipos de cáncer.
- \* *NBN* codifica la proteína Nbs1 o nibrina, que forma complejo con *MRE11A* y *RAD50* y se unen a *ATM*. Este complejo desempeña un papel esencial en la detección de cadenas de DNA rotas y su reparación, manteniendo la estabilidad y regulando la división celular. La acumulación de mutaciones en *NBN* provoca un riesgo aumentado de desarrollar cáncer, observándose un riesgo hasta tres veces mayor de desarrollar cáncer de mama. Así mismo, se ha propuesto que algunas mutaciones hereditarias en *RAD50* aumentan el riesgo de padecer cáncer de mama, aunque los resultados son contradictorios y algunos autores concluyen que *RAD50* no es un factor de riesgo.
- \* *BRIP1* y *PALB2*, también conocidos como *FANCF* y *FANCG*, respectivamente, forman parte del grupo de genes que conforman la ruta metabólica asociada con la anemia de Fanconi. Se han descubierto varios genes de esta misma ruta asociados a diversos cánceres, entre ellos a cáncer de mama.

Del total de casos familiares diagnosticados, resta un 65% de casos sin mutación conocida en ninguno de estos genes, pero que cuentan con un claro componente genético o epigenético.

## 2.2. Incidencia del cáncer de mama

Se trata del cáncer invasivo más frecuente en mujeres, afecta al 12% de la población femenina, es decir una de cada 8 mujeres lo padecerá a lo largo de su vida. Esta probabilidad varía según el rango de edad, siendo del 1,9% hasta los 49 años, 2,3% de los 50 a los 59 años, entre 60 años y 69 aumenta hasta un 3,5% y alcanza el máximo a partir de los 70 años de edad, siendo de un 6,7% (32). Aunque su tasa de supervivencia es muy alta (cerca del 83% de media en Europa), como se muestra en la Figura 2, debido a su elevada incidencia es la primera causa de muerte derivada de cáncer invasivo en la población española, con un 15,5% (33).

La incidencia de cáncer de mama ha sufrido un ascenso considerable desde el año 1984, estabilizándose en la última década o incluso disminuyendo en algunos países (34). Este punto de inflexión puede atribuirse al aumento y mejora del sistema de cribado de tumores mamarios por mamografías, de forma rutinaria e intensiva en la población a partir de cierta edad, que permite la detección precoz de carcinomas in situ y el tratamiento antes de su transformación maligna. Por otro lado, existe un mayor control sobre el empleo de terapia hormonal sustitutiva (HTR) (35, 36).



**Figura 2. Incidencia y mortalidad de cáncer por sexo y órgano.** Número total y porcentaje de nuevos casos diagnosticados de cáncer y muertes relacionadas, segmentado por zona de aparición del tumor, en hombres y mujeres por separado en España según el informe Globocan de 2012.

### 3. HETEROGENEIDAD Y CLASIFICACIÓN DEL CÁNCER DE MAMA

El cáncer de mama es objeto de varias clasificaciones atendiendo a diferentes aspectos del mismo, así se puede clasificar según su procedencia histológica, grado de diferenciación de sus células, estadio, presencia o ausencia de ciertos receptores hormonales y subtipo molecular. Esto da una idea de la heterogeneidad que existe dentro de esta enfermedad, siendo objeto de múltiples estratificaciones y agrupaciones, cada una con diferente pronóstico y aproximación terapéutica, por ello es vital llevar a cabo una clasificación lo más precisa posible y caracterizar el tumor de manera completa.

#### 3.1. Clasificación Histopatológica

Se determina el tejido de procedencia de las células del carcinoma y su histología, además pueden ser *in situ*, si no han atravesado la barrera de ese tejido y en cuyo caso no podemos llamar cáncer de mama, o *invasivo*, si se han infiltrado a otros tejidos.

Los tipos de cáncer de mama más frecuentes en la práctica clínica son:

- \* **Carcinoma ductal invasivo o infiltrante (CDI):** Representa más del 80% de los casos diagnosticados, las células cancerosas provienen de las células que tapizan el ducto galactóforo. Aunque este carcinoma puede afectar a mujeres de cualquier edad, resulta más común a medida que la mujer envejece. Según la Sociedad Americana del Cáncer, aproximadamente dos tercios de las mujeres que son diagnosticadas con cáncer de mama invasivo tienen 55 años o más.
- \* **Carcinoma lobulillar invasivo o infiltrante (CLI):** Es menos frecuente que el carcinoma ductal, pero aun así bastante común, dándose en un 10% de las pacientes de cáncer de mama. Se origina en las células que conforman los lóbulos o acinos glandulares. Aunque, al igual que los CDI, este tipo de carcinoma puede aparecer a cualquier edad, los CLI tienden a aparecer a edades algo más avanzadas que los carcinomas ductales invasivos, alrededor de los 60 años.
- \* **Carcinoma medular:** Representa cerca del 3 al 5 % de todos los casos de cáncer de mama. Se denomina carcinoma “medular” porque el tumor es una masa suave y pulposa que recuerda al bulbo raquídeo o médula. Es más frecuente en mujeres de 45 a 55 años. El carcinoma medular afecta con más frecuencia a mujeres que tienen una mutación del gen *BRCA1*. La morfología de sus células es similar a células cancerosas agresivas y muy anómalas, pero no actúan como tales. El carcinoma medular no crece rápidamente y por lo general no se propaga fuera de la mama hacia los ganglios linfáticos.
- \* **Carcinoma mucinoso:** Representa cerca del 2-3 % de todos los casos de cáncer de mama. En este tipo de cáncer, el tumor se forma a partir de células anómalas embebidas en acumulaciones de mucina. El carcinoma mucinoso suele afectar a las mujeres postmenopáusicas, con una edad promedio al momento del diagnóstico de 60 años o más. Tiene menos probabilidad de propagarse a los ganglios linfáticos que otros tipos de cáncer de mama y, además, es más fácil de tratar.

### 3.2. Clasificación por Grado Histológico

Corresponde a la clasificación del grado Bloom-Richardson-Elston o clasificación de Nottingham. Las células cancerosas pierden diferenciación, siendo usualmente cánceres más agresivos y de peor pronóstico aquellos cuyas células están más indiferenciadas. Se asigna un valor de la escala de I a III, siendo III el valor más indiferenciado y de peor pronóstico.

### 3.3. Clasificación según estadio

El estadio se determina siguiendo el sistema de estadificación TNM, un método estandarizado en el que se describe cuán extendido está el cáncer y que atiende al tamaño del tumor primario (T), a si afecta o no a los ganglios o nódulos linfáticos (N), y a la aparición o no de metástasis distante (M). Los estadios se basan en el manual de estadificación de 2010 de la AJCC (*American Joint Committee on Cancer*) (37).

### 3.4 Clasificación según receptores de membrana

Las células expresan receptores de membrana para su interacción con el entorno y para recibir estímulos químicos que causan cambios y consecuencias en la célula, como son las hormonas y factores de crecimiento. En cuanto a las células de cáncer de mama, se ha establecido una caracterización de tres receptores importantes para determinar el pronóstico y progresión del cáncer de mama (38).

- \* **Receptores hormonales:** Éstos son los receptores hormonales de estrógenos (ER), y de progesterona (PR). Unos niveles anormalmente altos de estrógenos o de su receptor favorecen el crecimiento tumoral.  
Un tumor de mama es ER positivo, cuando expresa ER en su membrana, detectándose mediante tinción inmunohistoquímica en el tejido. Su tasa de supervivencia es elevada, en torno al 95% a los 5 años cuando el cáncer se ha diagnosticado a tiempo-
- \* **Receptor ERBB2:** expresado por el gen homólogo 2 del oncogén de la leucemia eritrobástica viral aviar (*ERBB2*), también llamado *CD340* (clúster de diferenciación 340), o *HER2* (del inglés *Human epidermal growth factor receptor 2*). Este oncogén de la familia de receptores del factor de crecimiento epidérmico (EGFR/ERBB) se encuentra amplificado, o sobreexpresado en aproximadamente del 10% al 30% de cánceres de mama diagnosticados. Los niveles de expresión son bajos en epitelio de mama normal, por lo que tumores con sobreexpresión de HER2 presentan un fenotipo muy proliferativo y agresivo.

### 3.5 Clasificación por subtipos moleculares

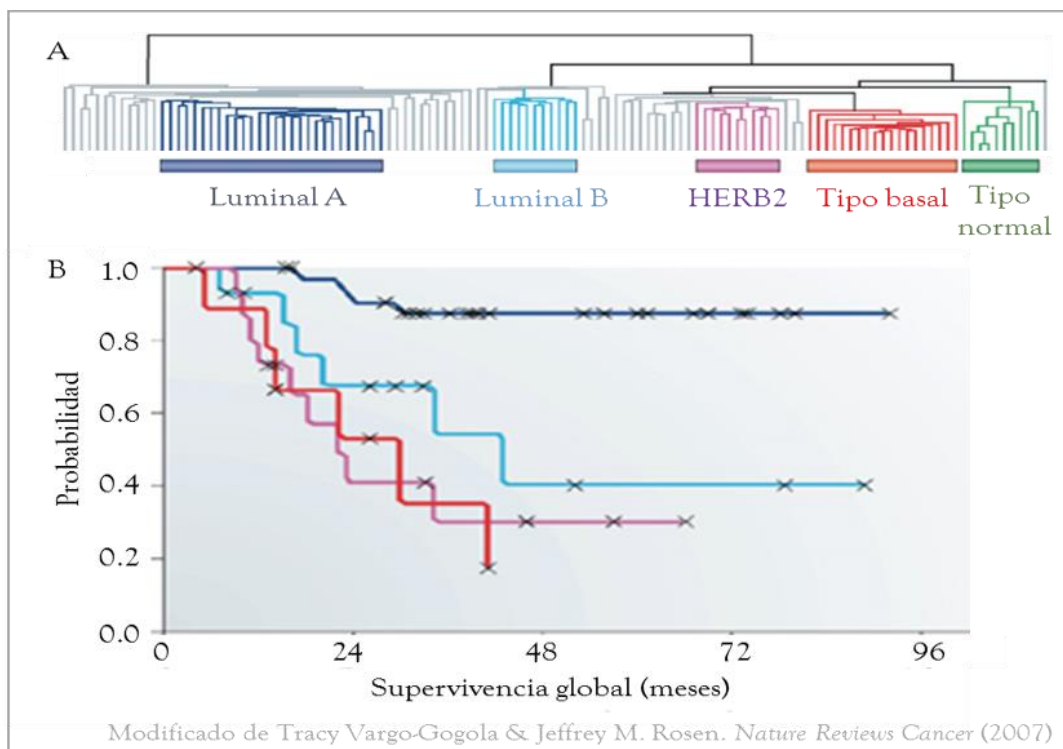
Perou y colaboradores propusieron la primera clasificación molecular del cáncer de mama, basada en el análisis de la expresión de determinados genes mediante técnicas de alto rendimiento (39). Se obtuvo así una clasificación en cinco subtipos, con un perfil de expresión característico y representativo de cada uno y con valor pronóstico significativo, ya que, además se observó que cada subtipo molecular estaba asociado con una supervivencia global determinada (Figura 3). Estos grupos podían separarse a su vez en ER+ y ER-, manteniendo por tanto las clasificaciones moleculares anteriores (38, 40).

- \* **Subtipos luminales:** Se subdivide en luminal A y luminal B. Llamados así por expresar grupos de genes de manera similar a los expresados normalmente por el epitelio luminal de la mama, es decir, el que está en contacto con el lumen del conducto galactóforo. Tanto el subtipo luminal A como el luminal B expresan citoquinas CK8/18, GATA3, y receptores de estrógenos, por lo que ambos formarían parte del grupo ER+. Sin embargo, los tumores clasificados/caracterizados como luminal A tienen una expresión muy elevada de ER, con niveles bajos de genes de proliferación, dando por tanto fenotipos menos agresivos, de grado histológico bajo y muy buen pronóstico. Por otro lado los denominados como luminal B expresan ER en menor cantidad y altos niveles de genes de proliferación, dando como resultado tumores con mayor grado histológico y peor pronóstico que los luminales A. Una proporción significativa de tumores con sobreexpresión de HER2, pero positivos para receptores endocrinos, se agrupa en este subtipo, luminal B.

- \* **Subtipo HER2:** Este subtipo se caracteriza por la presencia de tumores con HER2 sobreexpresado, aparte de otros genes de la ruta de HER, localizados en la banda cromosómica 17q12. Son completamente negativos para receptores hormonales. Usualmente estos tumores son agresivos, presentando un peor pronóstico.
- \* **Subtipo *Basal-like*:** Este subtipo tumoral expresa principalmente genes típicos de células basales o mioepiteliares, como serían EGFR, citoquinas como CK5, CK14, CK17, CD44, caderinas, caveolina 1 y 2 y nestina. Dan, además, resultados negativos para ER, PR y HER2, por lo que también se les conoce como tumores triple negativos, cuyas características incluyen un comportamiento muy agresivo y proliferativo, con grado histológico alto, áreas metaplásicas, necrosis central, e infiltración nodular. Sus características histopatológicas se asemejan a los tumores típicos de mujeres con mutaciones en BRCA1.
- \* **Subtipo *Breast-like*:** No se encuentra establecido como subtipo representativo, ya que su importancia clínica no ha sido determinada, además, algunos autores han propuesto que se trate de una falsa categoría, debida a la calidad del tejido empleado, pudiendo mostrar una mayor representación de células adiposas. Esta categoría se agrupa con las muestras de tejido mamario sano, expresando genes típicos de adipocitos y otras células no epiteliares, incluyendo las células basales (Cornejo 2014).
- \* **Subtipo *Claudin-Low*:** No se definió como subtipo hasta más adelante, podría definirse como un subtipo especial dentro de los *basal-like*, ya que son ER, PR y HER2 negativos. Se caracterizan por bajos niveles de expresión de genes relacionados con adhesión celular, como son claudinas, ocludinas y E-caderina. Aunque su expresión se asemeja a la del subtipo *basal-like*, es inconsistente, muestra bajos niveles de expresión de genes luminales y altos de marcadores de células endoteliales y linfocitos. Son tumores con características agresivas del tipo *basal-like*, contando con un alto número de carcinomas con grado histológico II o III, que comparten caracteres con la expresión mesenquimal de tumores de tipo escamoso.

#### Otros marcadores en cáncer de mama

- \* **Ki67:** es la proteína codificada por el gen *MKI67*, antígeno identificado por el anticuerpo monoclonal ki-67, cuya expresión se evalúa como indicador de proliferación tumoral, sirviendo como marcador pronóstico y de respuesta, debido a que la supresión temprana de Ki67 tras haber administrado terapia hormonal predice el posterior fracaso la misma.
- \* **P53:** proteína codificada por el gen supresor de tumores *TP53*. Dicho gen se encuentra comúnmente mutado en cáncer y se estudian sus niveles como rutina en la práctica clínica.
- \* **EGFR:** es un receptor transmembrana del factor de crecimiento epidérmico, se ha encontrado sobreexpresado en el 10% de los tumores de mama, acompañando a un resultado negativo para receptores endocrinos, siendo usualmente *basal-like*.



**Figura 3. Clasificación molecular del cáncer de mama.** A, representa los diferentes grupos en los que se agrupan las muestras de cáncer de mama en luminal A (azul), luminal B (cian), HER2 (rosa), basal (rojo) y tipo normal (verde). B, representa la supervivencia global en meses de los diferentes subtipos descritos en A.

**Tabla 1.** Caracterización por subtipos moleculares en cáncer de mama

Subtipo	ER/PR/HER2	Otros marcadores	Rutas	Grado	Pronóstico
Luminal A	ER <sup>3+</sup> /PR <sup>3+</sup> /HER2 <sup>-</sup>	CK8,CK18,GATA	ES	Bajo	Bueno
Luminal B	ER <sup>+</sup> /PR <sup>+/</sup> /HER2 <sup>+/</sup>	CK8,CK18,GATA	ES, TP53	Interm.	Interm.
HER2	ER <sup>-</sup> /PR <sup>-</sup> /HER2 <sup>+</sup>		HER, TP53	Alto	Malo
Basal-like	ER <sup>-</sup> /PR <sup>-</sup> /HER2 <sup>-</sup>	CK5/14/17, EGFR, c-KIT, CD44, nestina, caveolina2,	TP53, BRCA1	Alto	Malo
Claudin-low	ER <sup>-</sup> /PR <sup>-</sup> /HER2 <sup>-</sup>	Claudinas <sup>-</sup> , E-cadherinas, ocludina	Adhesión celular	Alto	Malo
Breast-like	ER <sup>+/</sup> /PR <sup>+/</sup> /HER2 <sup>-</sup>			Bajo	

ER: receptor de estrógenos, PR: receptor de progesterona, HER2: receptor epidérmico humano 2, CK: citoqueratinas, GATA3: factor de transcripción tipo GATA 3, EGFR: receptor del factor de crecimiento epidérmico, TP53: gen de la proteína tumoral p53, CD44: molécula antigénica CD44, BRCA1: gen del cáncer de mama 1.

### Clasificación del riesgo según el perfil molecular

El estudio de la expresión de genes mediante técnicas de alto rendimiento se ha convertido en una técnica ampliamente empleada en el estudio de enfermedades, incluyendo cáncer de mama. Además, se ha visto su valor en la identificación de biomarcadores, como valor diagnóstico y pronóstico y para una correcta clasificación de los pacientes.

En la práctica clínica se emplean paneles de genes cuya expresión clasifica los tumores de mama según su riesgo, facilitando la decisión del tratamiento y un pronóstico correcto. Los más empleados son PAM50, que analiza la expresión de 58 genes, clasificando el tumor en uno de los cinco subtipos y dando, además, una medida del riesgo de recurrencia a 5 años; ONCOTYPE, de

modo similar al PAM50, analiza 21 genes en casos de cáncer de mama ER positivos, determinando la probabilidad de recurrencia y la utilidad de la quimioterapia para reducir ese riesgo; por último, MAMMAPRINT estudia un panel de 70 genes e indica el riesgo de recurrencia a diez años(41-43).

Existen múltiples clasificaciones según perfiles moleculares, pudiendo o no estar establecidas como marcadores pronóstico o de terapia, pero está clara la heterogeneidad y cantidad de procesos subyacentes que pueden ser objeto de un proceso canceroso o verse modificados por este (44).

## 4. TRATAMIENTO DEL CÁNCER DE MAMA

---

### 4.1. Cirugía

La extirpación quirúrgica del tumor es el tratamiento más empleado y suele ir acompañado de otros tratamientos atendiendo a las características del tumor y a las de la paciente, como pueden ser radioterapia, quimioterapia o terapias dirigidas.

El tratamiento por cirugía se puede clasificar, atendiendo al tejido y cantidad de éste que es extirpado, en cirugía conservativa (aquella en la que siempre se mantiene parte de la glándula mamaria, como la lumpectomía); y en cirugía no conservativa o mastectomía, en la que se elimina por completo la glándula mamaria y el tejido adyacente, incluyendo la piel y pezón (45).

### 4.2. Radioterapia (RTx)

La terapia mediante radiación ionizante emplea partículas de alta energía para destruir las células cancerosas provocando daños en el DNA que provocan la muerte celular.

En cáncer de mama se emplea tras cirugía conservativa de la mama, para disminuir la probabilidad de recurrencia o de invasión a los nódulos linfáticos adyacentes; tras una mastectomía en pacientes con tumores de más de 5cm; o cuando existe infiltración en los nódulos.

### 4.3. Quimioterapia

Los agentes de quimioterapia actúan destruyendo las células que se dividen rápidamente, una de las propiedades principales de la mayoría de las células cancerosas. Se puede administrar tras una cirugía para eliminar células cancerosas residuales (tratamiento adyuvante) o antes de la cirugía, en cuyo caso su finalidad es la de reducir el tamaño del tumor previamente a la cirugía para facilitar la intervención (neoadyuvante).

En la mayoría de los casos, la quimioterapia es más efectiva cuando se administra en combinación con otras drogas, las más empleadas en el tratamiento de cáncer de mama incluyen antraciclinas y taxanos.



#### 4.4. *Terapia dirigida*

La terapia dirigida es aquella que tiene una diana concreta, usualmente un receptor, o un elemento de una ruta metabólica específica. Es más específica y efectiva, pero requiere de una selección correcta y exhaustiva de los pacientes.

##### **Terapia endocrina**

El cáncer de mama es una enfermedad extremadamente compleja, pero simplificando se puede separar en ER positiva y ER negativa, según expresa o no receptores de estrógenos (46). En el primer caso el tumor se vale del estrógeno para su crecimiento, estableciendo una diana clara para su tratamiento, ya que bloquear la producción de estrógeno o su disponibilidad reducirá el crecimiento tumoral. Las terapias endocrinas están dirigidas al grupo ER positivo, que constituye el 64% de todos los tumores (47).

El estrógeno es una de las principales hormonas femeninas reguladoras de la reproducción, es producido por los ovarios hasta la menopausia, cesando su actividad. Las mujeres post-menopáusicas producen una limitada cantidad de estrógenos, ya que las glándulas adrenales producen andrógenos, hormona que es metabolizada en estrógeno por la enzima aromatasa.

- \* Ooforectomía: es la terapia endocrina más simple, consiste en la extirpación quirúrgica de los ovarios, reduciendo la producción de estrógeno al mínimo. Se trata de un procedimiento drástico e invasivo, en el que se induce la menopausia irreversible a una mujer, privándole de su capacidad reproductora y forzándole todos los síntomas asociados a la menopausia prematuramente.
- \* Moduladores selectivos del receptor de estrógenos (SERMs): actúan bloqueando el estrógeno en el tejido, son el método más empleado en mujeres pre-menopáusicas, de los cuales el más conocido es el tamoxifeno.
- \* Análogos de LHRH: actúan a nivel hipotálamo-hipófisis bloqueando la comunicación hormonal e inhibiendo la síntesis de hormonas sexuales, como los estrógenos y andrógenos. La hormona liberadora de hormona luteinizante (LHRH), es secretada por neuronas reguladas mediante diferentes neurotransmisores; controla en última instancia los niveles de estrógenos y progesterona.
- \* Inhibidores de la aromatasa (AIs): bloquean la producción de estrógeno por parte de la enzima aromatasa, se consideran la primera línea de terapia para cáncer de mama metastático en mujeres post-menopáusicas, ER positivas.

##### **Terapia anti-HER2**

Hasta la aparición del Trastuzumab, un anticuerpo monoclonal humanizado con unión al dominio extracelular de los receptores de HER2, los tumores HER2 positivos constituían los casos de cáncer de mama con menor supervivencia, la implementación como primera línea de tratamiento ha mejorado considerablemente el pronóstico, aumentando la supervivencia a 5 años hasta el 77,1% actual (48, 49).

## 5. CÁNCER DE MAMA EN MUJERES JÓVENES

---

### 5.1. La edad como factor de riesgo y pronóstico

La edad se considera *per se* un factor de riesgo asociado a cáncer de mama, sin embargo ya hemos comentado el hecho de que al menos existen dos grupos de cáncer de mama atendiendo a la edad de diagnóstico: post y pre-menopáusicos.

Dentro de las pre-menopáusicas la edad se sigue considerando un factor de riesgo, sin embargo, la edad al diagnóstico se trata además como un factor pronóstico, siendo en este caso de peor pronóstico en mujeres jóvenes, aumenta con cada año de adelanto en la edad de diagnóstico.

Aunque la media de edad de aparición del cáncer de mama es a los 61 años, 1 de cada 40 mujeres diagnosticadas es muy joven, constituyendo del 5 al 7% de muertes por cáncer en este grupo de edad (50, 51). Esto puede ser debido a que los tumores en estas pacientes se detectan en estadios más avanzados, ya que la mayor densidad mamaria imposibilita el diagnóstico por técnicas de imagen y que, además no se realizan técnicas de cribado para el diagnóstico rutinario, permitiendo la detección del tumor cuando éste es palpable y de tamaño considerable (52).

La mamografía, herramienta muy efectiva para el cribado de cáncer de mama en mujeres mayores, no es una herramienta adecuada para la detección de cáncer en las mujeres jóvenes puesto que éstas tienden a tener un tejido mamario denso y que previene la detección de tumores o calcificaciones. Debido a la falta de herramientas de detección adecuadas, las mujeres más jóvenes son a menudo diagnosticadas en una etapa tardía de la enfermedad si comparamos con las mujeres de mayor edad (35, 52).

### 5.2. Determinación del rango de edad “muy joven”

Qué se considera joven a la hora de desarrollar cáncer de mama ha sido un tema muy discutido en la comunidad científica, y de gran heterogeneidad, considerándose joven a partir de los 30, 35, 40 e incluso 45 años, complicando su estudio y la reproducibilidad del mismo.

La *European Society of Breast Cancer Specialists* (EUSOMA) propuso la definición de mujeres jóvenes, como aquel grupo de mujeres con una edad por debajo de los 40 años. Aunque la biología y el estado endocrino son un continuo, la definición de grupo de edad es arbitraria (50). Sin embargo, las mujeres de menos de 40 años presentan características propias como el mantenimiento de la fertilidad, el embarazo y la lactancia que no mantienen las mujeres mayores de 40 años y que por ello implica un enfoque distinto y establecer diferencias entre mujeres mayores, peri- y pre-menopáusicas (18).

Un trabajo en 1703 pacientes de un mismo centro determinó que el riesgo de recurrencia asociado a la edad era un parámetro lineal continuo, mostrando un 4% de descenso en la recurrencia y un 2% en la probabilidad de muerte por cada año de aumento en la edad (53). Sin embargo, estos resultados no han sido reproducidos hasta la fecha.

Han y colaboradores, 2010, analizaron 9885 casos de cáncer de mama en mujeres menores de 50 años. Los estratificaron en grupos atendiendo a la edad de las pacientes a partir de 30 años (<30, 30-35, 35-40, 40-50), con un total de 1443 pacientes por debajo de los 35 años (54).

Observaron que el grupo de mujeres menor de 30 años mostró peor supervivencia que el de 30-35 años, y este a su vez, peor que el de 35-40 años, sin embargo los grupos de edades 35-40 y 40-50 no mostraron diferencias, pudiendo considerarse como un solo grupo en términos de pronóstico. Estas correlaciones las observaron solo en el grupo de pacientes ER+ y no en el grupo de pacientes negativo para ER. Los autores concluyen que 35 años es un límite razonable para definir al cáncer de mama en mujeres muy jóvenes, coincidiendo con otros estudios (55).

### 5.3. Características tumorales

Los tumores de mujeres muy jóvenes se caracterizan por ser típicamente más agresivos, teniendo una mayor representación de tumores de mayor tamaño, más indiferenciados, con mayor afectación ganglionar, sobreexpresión de HER2 y ausencia de receptores hormonales, todos ellos factores de mal pronóstico, tal y como se observa en la Figura 4 (56-59). La tasa de supervivencia a 5 años es de 83% para las mujeres más jóvenes, más baja que la tasa de supervivencia para las mujeres post-menopáusicas. Tal es la asociación que la Conferencia de St. Gallen de 1998 estableció que edades por debajo de 35 años actúan como variable con valor pronóstico (60) sin embargo, tiene mayor peso como indicador de pronóstico la respuesta a terapia endocrina (es decir la presencia o no de receptores hormonales) (61).

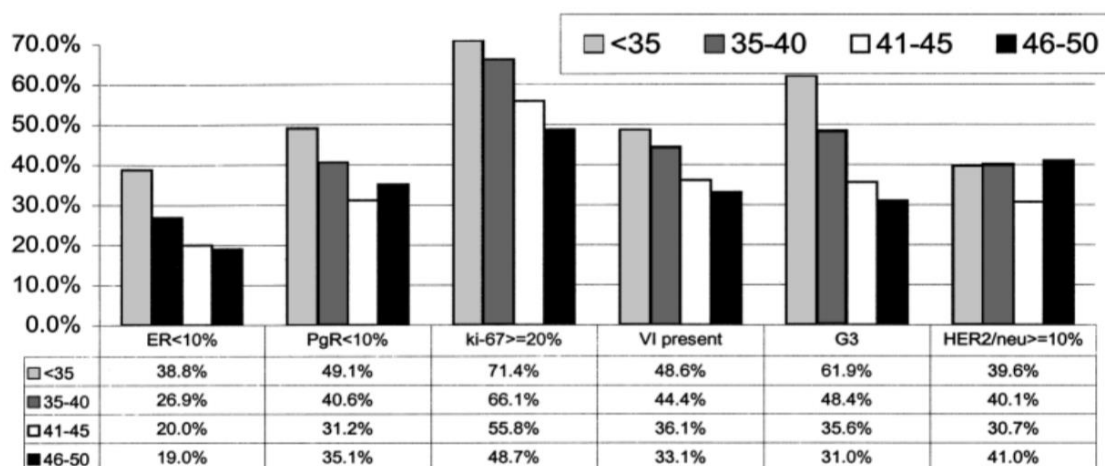


Figura 4. Características tumorales por grupos de edad. Expresión de RE, RP, ki67, presencia peritumoral de VI, porcentaje de tumores de grado III y sobreexpresión de HER2. Sacada de Colleoni y colaboradores 2001.

### 5.4. Tratamiento y consideraciones específicas de las pacientes jóvenes

Un área adicional de preocupación para las mujeres más jóvenes con cáncer de mama incluyen cuestiones relacionadas con la aparición temprana de la menopausia o dificultades con su fertilidad, uno de los principales efectos adversos de la quimioterapia o terapias hormonales, ya que muchas mujeres jóvenes que son diagnosticadas con cáncer de mama aún no han comenzado una familia. Por otro lado las secuelas psicológicas derivadas del tratamiento o de la enfermedad en sí son más acusadas en pacientes jóvenes, debido a la presión social, laboral o una disminución en la autoestima (62).

La mastectomía, empleada como tratamiento del cáncer de mama, provoca en pacientes jóvenes trastornos hormonales y secuelas psicológicas, especialmente en aquellas que no han satisfecho sus necesidades reproductivas. Se debe tener en cuenta la situación particular de cada mujer y tener en cuenta los estudios para sensibilizar acerca de este hecho, que afirman que la

supervivencia global en mujeres muy jóvenes (menores de 40 años) con cáncer de mama es similar tras una mastectomía o una lumpectomía (63).

Aunque se tienen consideraciones específicas a la hora de recomendar un tratamiento en mujeres jóvenes, estas atienden principalmente a los deseos reproductivos de las pacientes y suelen ser más conservadores (55, 64-66). No reflejan la necesidad de las pacientes de un tratamiento específico acorde con la agresividad del tumor en este grupo de edad. Por ello se pone de manifiesto en varios trabajos la necesidad de un tratamiento dirigido a este grupo y de una mejor estratificación de los casos para alcanzar una aproximación más efectiva que contemple tanto las necesidades específicas de las pacientes como la biología del tumor (18).

Como se ha dicho, las mujeres jóvenes afrontan problemas específicos que son menos relevantes para las mujeres mayores pre-menopáusicas, además de que presentan un peor desenlace de la enfermedad que las mujeres mayores, por ello, definir correctamente el término “muy joven” es vital para una correcta estratificación de las pacientes en base a su edad de diagnóstico, que conlleva un importante valor pronóstico (65).

---

---

# ESTUDIOS MOLECULARES

---

---

Capítulo

# 3

---

---

## 1. EPIGENÉTICA

---

---

Una célula tiene a lo largo de su vida el mismo componente genético, al igual que todas las células que integran nuestro cuerpo; sin embargo, el comportamiento que siguen no es el mismo tanto entre diferentes tipos celulares como en diferentes momentos de la vida. La carga genética transmisible es la misma, sin embargo, el comportamiento diferente de la célula depende de más factores.

Se denomina epigenética al conjunto de procesos químicos que suceden en la célula, que modifican la actividad del DNA sin modificar su estructura, ampliando el abanico de posibilidades y diferencias celulares con un mismo genoma. Los últimos estudios en este campo destacan el papel de la epigenética en procesos como el envejecimiento.

Desde que la secuenciación del genoma humano fue completada, quedó patente que existen enormes regiones intergénicas cuya función era desconocida o subestimada. Sólo el 2% del genoma corresponde a genes conocidos con una función clara, sin embargo cerca del 50% del DNA se transcribe a RNA sin dar lugar a una proteína, este RNA se denomina RNA no codificante (ncRNA). Dentro de la categoría de ncRNA se encuentran los microRNAs, interruptores endógenos pertenecientes a la ruta de ribointerferencia (RNAi), descubierta por Andrew Fire y col en 1998 (67).

### 1.1. *Los microRNAs y su papel en enfermedades*

Los microRNAs (miRNAs) son pequeñas moléculas de RNA no codificante, de unos 22 nucleótidos, cuya función es reprimir la expresión de genes secuestrando el paso de RNA mensajero (mRNA) a proteína. Descubiertos en los 90s, no fue hasta 2001 que se propuso una función para ellos, desde ese momento apareció un nuevo nivel de complejidad a la ya de por sí compleja red de regulación de la expresión, junto con las modificaciones post-traduccionales y la metilación del DNA (68, 69).

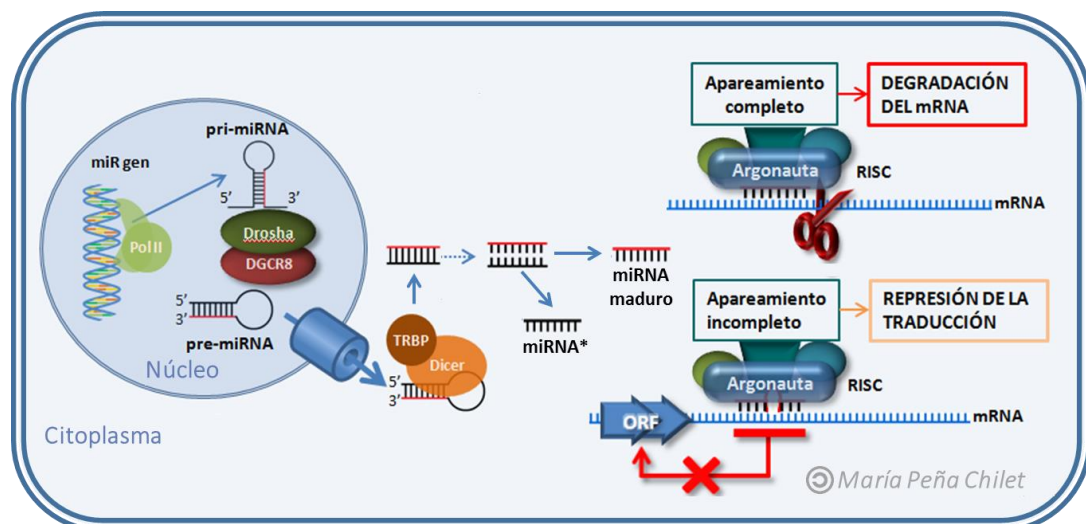
La síntesis de miRNAs tiene lugar en el núcleo, dónde la enzima polimerasa II transcribe el gen del miRNA formando el pri-miRNA, este adopta una estructura en horquilla al emparejarse las secuencias complementarias que lo conforman. El complejo enzimático Drosha junto a DGCR8 (del inglés *DiGeorge syndrome critical region gene 8*), reconocen el extremo de cadena simple seguido por secuencia de cadena doble y se une a él, cortándolo y formando el precursor del miRNA, el pre-miRNA, que es exportado al citoplasma por una exportina. Ya en el citoplasma, la enzima Dicer, que necesita a la proteína TRBP (proteína de unión al RNA TAR, del inglés *HIV-1 TAR RNA binding protein*) para funcionar, reconoce secuencias de doble cadena asociadas a un bucle, estructura típica de un pre-miRNA, y corta este bucle, dejando libre

la secuencia de doble cadena. Esta secuencia es reconocida y desapareada por una helicasa del complejo proteico RISC (*RNA-induced silencing complex*) dejando libres las dos secuencias complementarias, por un lado, el microRNA maduro (usualmente se trata de la hebra anti-sentido) y un anti-miRNA, o hebra pasajera, con su secuencia complementaria (miRNA\*). El miRNA maduro es incorporado por la endonucleasa argonata, también del complejo proteico RISC, que se encarga de unirse a cualquier mRNA que contenga la secuencia complementaria del miRNA asociado, provocando la represión del mRNA (Figura 5).

En el caso de que el apareamiento entre secuencias sea completo Argonata corta la secuencia del mRNA complementaria al miRNA y es degradado; si, por el contrario, la complementariedad es parcial, tiene lugar la represión de la traducción mediante el bloqueo de la pauta de lectura, pero el mRNA queda intacto en la célula, pudiendo ser recuperado si los niveles de miRNA libre cambian (70, 71).

Un mismo miRNA puede actuar sobre múltiples genes (del orden de cientos), a su vez, los niveles de miRNAs pueden ser regulados, bien por regulación de la transcripción del gen del miRNA, o por la transcripción de pseudogenes conservados evolutivamente, que contengan la secuencia complementaria y que actuarían como sumideros de miRNAs, disminuyendo la cantidad efectiva de miRNAs (72, 73).

Los miRNAs representan la clase principal de ncRNA de pequeño tamaño, controlando a casi un tercio de todos los genes humanos, estando implicados en diversas funciones vitales de la célula. La desregulación de uno o varios miRNAs se ha comprobado que está asociada a un gran número de desórdenes y enfermedades humanas (68).



**Figura 5. Biogénesis y proceso de represión de los microRNAs.** Pol II, Polimerasa II. DGCR8, subunidad del complejo microprocesador, del inglés *DiGeorge syndrome critical region gene 8*. TRBP, proteína de unión al RNA TAR, del inglés *HIV-1 TAR RNA binding protein*. miRNA, microRNA. mRNA, RNA mensajero. ORF, Pauta de lectura abierta, del inglés *open reading frame*. RISC, complejo de silenciamiento génico, del inglés *RNA induced silencing complex*.

## 1.2. miRNAs en cáncer

Conforme se acumulan los datos acerca de los miRNAs, se les atribuyen cada vez más papeles en el control de proliferación, diferenciación, supervivencia y apoptosis (74). Como se ha destacado antes, la alteración de cualquiera de estas rutas, entre muchas otras, sería suficiente para producir en sí misma un proceso canceroso (75).

Un gran número de estudios recientes se centran en el papel de los miRNAs en el riesgo a padecer cáncer, o en su desarrollo, aportando una herramienta de estudio útil a la hora de establecer perfiles moleculares que puedan servir bien como marcadores de diagnóstico o pronóstico, bien para proporcionar indicios sobre rutas alteradas o mecanismos subyacentes que puedan estar contribuyendo al desarrollo del cáncer, aún desconocidos para nosotros, pudiendo indicarnos posibles dianas terapéuticas y ayudarnos a comprender mejor la etiología del cáncer (76).

De manera similar a los genes codificantes, los miRNAs pueden estar sobreexpresados o reprimidos, por lo tanto pueden actuar como supresores de tumores o como oncogenes (denominándoseles oncomir), dependiendo del gen o genes que estén controlando. Por ejemplo, los primeros miRNAs descritos como supresores de tumores fueron miR-15 y miR-16, ya que su pérdida impide la inhibición de oncogenes como *BCL2*, *BMI1*, *CCND1* y *CCND2*, promoviendo el crecimiento y progresión tumoral. Por otro lado, miR-21 se encuentra altamente expresado en la mayoría de tejidos tumorales, actuando como represor de genes pro-apoptóticos como *PTEN* y *PDCD4*, estimulando así la proliferación celular (75).

**Tabla 2.** MicroRNAs típicamente desregulados en cáncer. Modificado de DiLeva et al. 2013.

miRNA	Localización	Expresión	Cáncer y tejido	Mecanismo	Genes diana
<b>miRNAs supresores de tumores</b>					
Let-7a2	11q24	↓	Mama, pulmón, CRC, ovario y estómago	Reprimido por MYC	KRAS, HMGA2, MYC, DICER, BCL-XL, IMP-1, CDC34, IL6
miR-15/16	13q31	↓	CLL, próstata y adenomas pituitarios	Pérdida genómica, mutación, activado por p53	BCL2, COX, CHECK1, CCNE1, CCND1, CCND2, BMI1, FGF2, FGFR1, VEGF, VEGFR2, CDC25a
Familia miR-29	7q32 1q30	↓	AML, CLL, pulmón, mama, linfoma, hepatocarcinoma	Pérdida genómica, activado por p53, reprimido por MYC	CDK6, MCL1, TCL1, DNMT1, DNMT3a, DNMT3b
Familia miR-34	1p36 11q23	↓	CRC, pulmón, mama, riñón y vejiga	Reprimido por MYC	SIRT1, BCL2, NOTCH, HMGA2, MYC, MET, AXL, NANOG, SOX2, MYCN, SNAIL
miR-26a	3p22	↓	Hígado	Reprimido por MYC	CCND2, CCNE2
Familia miR-200	1p36 12p13	↓	Cáncer de mama agresivo	Reprimido por ZEB1/2	ZEB1, ZEB2, BMI-1, SUZ-12, FN1, LEPR, CTNBN1, JAG1, MALM2, MALM3, p38α
<b>Oncomirs</b>					
miR-155	21q21	↑	CLL, AML, mama, pulmón, CRC y linfoma	Activado por NF-κB	SOCS1, BACH1, MEIS1, ETS1, FOXO3A, hMSH2/6, hMLH1, SMAD5, WEE1, SHIP1, CEBPB
miR-21	17q23	↑	Pulmón, mama, próstata, páncreas, estómago, CLL, AML, glioblastoma	Activado por IL6, GF1α	PTEN, TPM1, PDCD4, SPRY1, TIMP3, RECK
miR-221/222	Xp11	↑	CDI, pulmón, carcinoma hepatocelular, tiroides papilar	Activado por MET, PLZF, NF-κB o cJun; reprimido por ERα en mama	P27(Kip1), p57(Kip2), PTEN, TIMP3, FOXO3A, ERα, KIT, TRSP1, DICER, APAF1, PUMA, PTPμ
miR-17/92	13q22	↑	Pulmón, mama y CRC	Activado por E2F1 y MYC	PTEN, BIM, HIF1, PTPRO, p63, E2F2, E2F3, TSP-1, CTGF, p21(WAF1), JAK1, SMAD4, TGFβII, MnSOD, GPX2, TRXR2

CRC: cáncer colorrectal; CLL: leucemia linfocítica crónica; AML: leucemia mieloide aguda; CDI: carcinoma ductal infiltrante.

El desarrollo de nuevas técnicas de alto rendimiento para la detección de miRNAs han permitido establecer perfiles de expresión de miRNAs en diversos tumores malignos, como son leucemia crónica linfocítica, cáncer de mama, cáncer de pulmón, de estómago, de próstata, carcinoma papilar tiroideo, tumores pancreáticos, glioblastoma o carcinoma hepatocelular (77, 78).

Algunos estudios sugieren que el estudio de estos perfiles de miRNAs se alza como una herramienta más eficaz y menos costosa que las técnicas actuales que emplean la expresión de los genes, en el ámbito del diagnóstico, clasificación y pronóstico. Esto sería debido principalmente a la gran estabilidad de los miRNAs frente a la degradación, la gran especificidad y reproducibilidad de los perfiles y por último, la habilidad de éstos perfiles para identificar el tejido de origen en células metastásicas, lo que reduciría el coste total del procedimiento (75, 79, 80).

Por otro lado, numerosos estudios han remarcado la utilidad de los miRNAs como biomarcadores tumorales, llegando a establecer un “miRNoma” común en cáncer (81, 82). Además se ha demostrado que el perfil de expresión de microRNAs es capaz de clasificar los cánceres humanos, proponiéndose incluso como marcadores para la respuesta clínica de los pacientes y siendo investigados en ensayos clínicos (83, 84). Recientemente ha surgido el papel terapéutico de los miRNAs para el tratamiento del cáncer, debido a que su función puede ser reprimida mediante antagomirs artificiales, apoyando su potencial como una nueva estrategia terapéutica para tratar el cáncer, a parte de otras enfermedades (83, 85, 86).

### *1.3. miRNAs en cáncer de mama*

A lo largo de décadas los mecanismos implicados en el cáncer de mama han sido exhaustivamente investigados, sin embargo, todavía existen retos que debemos enfrentar a la hora de establecer un diagnóstico precoz y específico de cada paciente, de cara a predecir su respuesta y resistencias adquiridas al tratamiento.

Como hemos visto, los miRNAs han mostrado un potencial valor pronóstico, además de como dianas terapéuticas. Es por eso que, desde que en 2005 se informó por primera vez una desregulación de miRNAs en cáncer de mama, son muchos los estudios que han dirigido sus esfuerzos a investigar la expresión de miRNAs en cáncer de mama y su papel en el desarrollo de la enfermedad (87).

Gasparini y colaboradores aseguran que la detección combinada de los miRNAs miR-155, miR-493, miR-30e y miR-27a, junto con las técnicas rudimentarias de detección inmunohistoquímica, separa las pacientes de cáncer de mama triple negativo en un grupo de alto riesgo y otro de bajo riesgo (88). Este trabajo deja clara la utilidad y posibles aplicaciones en el campo.



**Tabla 3.** MicroRNAs desregulados en cáncer de mama (KEGG hsa05206)

miRNAs	Genes diana	Función	Referencias
<b>miRNAs supresores de tumores</b>			
miR-206	<i>ESR1</i>	Señalización ER	Iorio2008(89), Adams2007(90)
miR-17-5p	<i>AIB1, CCND1, E2F1</i>	Proliferación	Hosain 2006 (91), Yu 2008 (92)
miR-125a,b	<i>HER2, HER3</i>	Crecimiento dependiente de anclaje	Mattie 2006(93), Scott 2007(94)
miR-200c	<i>BMI1, ZEB1, ZEB2</i>	Señalización TGF- $\beta$	Gregory 2008(95), Shimono 2009(96)
let-7	<i>HRAS, HMGA2, LIN28, PEBP1</i>	Proliferación y diferenciación	Johnson 2005 (97), Yu 2007(98)
miR-34a	<i>CCND1, CDK6, E2F3, MYC</i>	Proliferación, reparación del DNA	Christophersen 2010, (99) Sun 2008, (100) Welch 2007 (101)
miR-30e	<i>ITGB3, UBC9</i>	Supresor del crecimiento celular	Wu 2009 (102)
miR-31	<i>FZD3, ITGA5, MIRIP, MMP16, RDX, RHOA</i>	Metástasis	Valastyan 2009(103)
miR-335	<i>SOX4, PTPRN2, MERTK, TNC</i>	Metástasis	Tavazoie 2008 (104)
miR-27b	<i>CYP1B1</i>	Respuesta del tumor a drogas anti-cancerígenas	Tsuchiya 2006 (105)
miR-126	<i>IRS-1</i>	Progresión fase G1/G0 a S	Zhang 2008 (106)
miR-101	<i>EZH2</i>	Actividad oncogénica y metastásica	Varambally 2008 (107)
miR-145	Represión de <i>cMyc</i> mediada por <i>p53</i>	Supresor de la invasión celular y metástasis	Sachdeva 2009 (108)
miR-146a/b	<i>NF-<math>\kappa</math>B</i>	Regulador de NF- $\kappa$ B, y de la invasión	Bhaunik 2008 (109)
miR-205	<i>ERBB3, VEGFA</i>	Inhibe invasión y crecimiento tumoral	Wu 2009(110)
<b>Oncomirs</b>			
miR-21	<i>BCL-2, TPM1, PDCD4, PTEN, MASPIN</i>	Apoptosis	Frankel 2008 (111), Qi 2009(112), Zhu 2008(113)
miR-155	<i>RHOA</i>	Señalización TGF- $\beta$	Kong2008(114)
miR-10b	<i>HOXD10</i>	Metástasis	Ma 2007 (115)
miR-373	<i>CD44</i>	Metástasis	Huang 2008(116)
miR-27a	<i>ZBTB10, Myt-1</i>	Progresión del ciclo celular. G2-M	Mertens 2007 (117)
miR221/222	<i>p27Kip1</i>	Resistencia a Tamoxifeno	Miller 2008 (118)

#### 1.4. Desregulación de miRNAs en células madre de cancer de mama (BCSCs)

El proceso de transición epitelio-mesénquima (EMT) es esencial durante el desarrollo embrionario de los mamíferos; sin embargo, este proceso se ha relacionado con el inicio de procesos cancerosos y su progresión. Está caracterizado por la pérdida del marcador epitelial E-caderina, de proteínas de adhesión celular y de la polaridad de la célula (119).

El estudio de EMT es importante también en el cáncer de mama para establecer el inicio de la carcinogénesis, ya que las células cancerosas reproducen el comportamiento de las células madre normales. Yu y colaboradores fueron los primeros en estudiar la expresión de miRNAs en células madre de cáncer de mama (BCSCs del inglés *breast cancer stem cells*), para ello compararon la expresión de miRNAs en BCSCs autorrenovables y células diferenciadas provenientes tanto de líneas celulares como de tumores primarios de mama (120).

**Tabla 4.** miRNA implicados en células madre de cáncer de mama (BCSCs).

miRNA	Dianas	Reguladores	Procesos relacionados
<b>Reprimidos en BCSCs</b>			
Familia let-7	<i>RAS</i> <i>HMGGA2</i>	<i>Lin28</i> bloquea su biogénesis	Propicia la actividad tumorigénica e influencia la formación de mamoesferas in vitro. Aumenta la formación de tumores y su potencial metastático in vivo. Inhibe la diferenciación celular, mantiene la proliferación y promueve EMT.
Familia miR200	<i>Bmi-1</i> <i>Suz12</i>	---	Regula EMT. Relevante para las funciones de células madre en células cancerosas in vitro Induce propiedades tipo células madre.
Familia miR30	<i>Ubc9</i> <i>ITGB3</i> <i>AVEN</i>	---	Capacidad de mantenimiento celular y propiedades anti-apoptóticas. Modulador de las propiedades tipo células madre de BCSCs. Regula la formación y el crecimiento no adherente de mamoesferas. Controla genes implicados en apoptosis y proliferación de BCSCs.
miR128	<i>Bmi-1</i> <i>ABCC5</i>	---	Asociado a resistencia a quimioterapia y tasas de supervivencia de los pacientes de Influye en el número y tamaño de las mamoesferas in vitro. Reduce el crecimiento tumoral e induce la apoptosis in vivo.
miR34c	<i>Notch4</i>	regulación mediante metilación	Influye en el auto mantenimiento celular y EMT. Actúa sobre la formación de mamoesferas in vitro. Controla la migración de células tumorales.
miR16	<i>Wip1</i>	---	Aumenta el número y tamaño de las mamoesferas. Responsable de la sensibilidad a doxorubicina.
<b>Sobreexpresados en BCSCs</b>			
miR181	<i>ATM</i>	TGF- $\beta$	Induce la formación de mamoesferas.
mi495	<i>REDD1</i>	---	Conlleva un descenso de los niveles de E-caderina. Promueve la formación de colonias in vitro. Aumenta la formación de tumores in vivo.

BCSCs: Células madre de cáncer de mama (del inglés *Breast cancer stem cells*).

EMT: Transformación epitelio-mesénquima.

## 2. RUTAS ALTERADAS EN CÁNCER DE MAMA

### 2.1. Importancia y regulación de la ruta de las MAP cinasas

La estimulación de la mayor parte de los receptores de membrana conduce, de una manera más o menos directa, a la activación de proteína cinasas y a la fosforilación de proteínas diana como parte del mecanismo de regulación de procesos celulares por agonistas. Existen rutas que conducen a la activación de diversas proteína-cinasas activadas por segundos mensajeros, como las dependientes de AMP cíclico (PKA), de calcio y calmodulina (CaMK) o de calcio y fosfolípidos (PKC). Existe otro grupo de proteína cinasas, conocidas con el nombre de MAP cinasas (MAPK, del inglés *mitogen activated protein kinases*), que forman parte de un importante grupo de vías de transducción de señales extracelulares (121). Existen al menos una docena de MAPK, entre ellas destacan ERK (del inglés *extracellular signal regulated kinases*), JNK (del inglés *c-Jun-N-terminal kinase*) y p38.

Estas vías se organizan en módulos que contienen una proteína G y tres proteína cinasas que actúan en serie. Las MAPK son activadas por fosforilación catalizada por MAPK cinasas (MAPKK, MAP2K o MEK), a su vez fosforiladas y activadas por MAPKK cinasas (MAPKKK), reguladas por proteínas G (122).

Las vías de las MAPK son activadas por estímulos extracelulares que incluyen hormonas, mitógenos, factores tróficos y de crecimiento, citoquinas inflamatorias y diversas formas de estrés celular (osmótico, oxidativo, radiación, etc.); participan en el control de la expresión génica por receptores de membrana, regulando procesos tan importantes como proliferación, diferenciación

y transformación celulares, desarrollo, apoptosis, respuesta a estrés inducido y choque térmico u osmótico e inflamación crónica (123).

## 2.2. Importancia de la ruta de señalización PI3K/Akt en cáncer de mama

La ruta PI3K/Akt/mTOR es una ruta metabólica que concluye en la transcripción de genes anti-apoptóticos, favoreciendo la supervivencia celular y asociándose con procesos tumorigénicos.

La cascada de señalización se inicia a través del reclutamiento de la proteína-cinasa PI3K a la membrana plasmática a través de la unión con la subunidad reguladora de PI3K, p85. La actividad cinasa de PI3K produce el fosfatidil-inositol-3,4,5-trifosfato o PIP<sub>3</sub>, que sirve como ligando para reclutar la serina/treonina cinasa Akt (*c-Akt* o PKB) a la membrana plasmática. Una vez en la cara interna de la membrana, Akt es fosforilado por otra serina/treonina cinasa, la cinasa 1 dependiente de fosfatidil inositol-3 (PKD1), resultando en la activación de Akt.

La activación de Akt controla la supervivencia celular a través de la fosforilación de proteína-cinasas como mTOR, inactivación de proteínas apoptóticas (como BAD, caspasa 9, p53 o FOX), y activación de factores de transcripción (NF- $\kappa$ B, HIF-1 $\alpha$ , CREB) que tiene como consecuencia un incremento en la transcripción de genes anti-apoptóticos, como Mcl-1, o factores de crecimiento, como VEGF (124, 125). El resultado es un incremento en la supervivencia celular, proliferación, crecimiento y metabolismo de la célula (126, 127).

Las actividades cinasas son reguladas por las fosfatasas que actúan eliminando los fosfatos de las proteínas diana. Existen evidencias de que PTEN desfosforila PIP<sub>3</sub>, actuando por tanto como regulador de la ruta de señalización de PI3K, PTEN suprime el crecimiento celular tumoral ejerciendo un efecto antagonista al de las proteínas tirosina cinasas, regulando la invasión de las células tumorales y la metástasis a través de las interacciones con las adhesiones focales. La inactivación de PTEN debida a alteraciones epigenéticas o genéticas puede resultar en la pérdida de la expresión de la proteína en varios tumores esporádicos, encontrándose en el 28% de pacientes con cáncer de mama en un estudio (128).

## 2.3. Importancia de la ruta de señalización de Notch

La ruta de señalización de Notch es un sistema de diferenciación celular altamente conservado entre las especies de organismos (129). En mamíferos existen cuatro receptores de Notch, NOTCH1, NOTCH2, NOTCH3 y NOTCH4. Entre las funciones que son controladas por la ruta de señalización Notch destaca su importancia en la comunicación intercelular implicada en la regulación de procesos que controlan la diferenciación celular durante el desarrollo embrionario o la vida adulta, estando implicada en numerosos tipos de cáncer (130). Además la inhibición de la expresión de Notch tiene efectos antiproliferativos en cultivos celulares y en modelos murinos (131, 132).

## 2.4. Receptores de membrana iniciadores de la cascada de señalización.

Entre los receptores de membrana que activan las principales rutas de transducción de señales, mediante activación directa de PI3K, encontramos a los receptores acoplados a proteínas G (GPCR del inglés *G-protein-coupled receptor*) y receptores tirosina cinasa (RTK, o TKR del inglés *tyrosine kinase receptor*).

Los GPCR, también conocidos como receptores de 7 dominios transmembrana, constituyen una de las mayores superfamilias de proteínas conocida (con cerca de 800 genes humanos conocidos), responsable de la iniciación, tras un estímulo extracelular, de la transducción de señales, bien mediada por AMPc (adenosín monofosfato cíclico), GMPc (guanósín monofosfato cíclico) o por PI3K. Entre sus papeles destacan la transducción de la percepción sensorial, la regulación del comportamiento y del sistema inmune, la modulación de la homeostasis celular y están implicada en la proliferación celular, así como en la capacidad de metástasis y el crecimiento tumoral (133, 134). Se han descrito mutaciones en las proteínas G que las activan como oncogenes y que han sido encontradas en diversos tipos de tumores (135).

Se clasifican mediante el sistema GRAFS (*Glutamate, Rhodopsine, AMP, Fungal, Smoothened*):

- \* **Clase A:** Receptores semejantes a rodopsina (incluye receptores de quemoquinas, interleucinas, angiotensinas, somatostatinas, y melatoninas entre otros).
- \* **Clase B:** Receptores de secretina (incluye los receptores hormonales clásicos y los regulados por secretina o glucagón).
- \* **Clase C:** Receptores metabotrópicos de Glutamato (mGluR).
- \* **Clase D:** Receptores fúngicos involucrados en la determinación sexual.
- \* **Clase E:** Receptores de AMPc.
- \* **Clase F:** *Frizzled/Smoothened*, de las vías Wnt y Hedgehog, respectivamente.

Los RTK son receptores de membrana de alta afinidad a factores de crecimiento, citoquinas y hormonas (incluyendo estrógenos y progesterona) (136). La activación de GPCR puede dar lugar a la fosforilación y activación de algunos RTK en ausencia de sus agonistas, como los receptores de EGF (EGFR), PDGF (PDGFR) e IGF-1 (IGFR), lo que inicia la activación de la vía Ras/MAPK. Los RTK han demostrado su papel en la transducción de señales de la regulación de procesos de la célula, por lo que desempeñan un papel crítico en el desarrollo y progresión de numerosos tipos de cáncer (137). Entre las 20 clases de RTK conocidas destacamos:

- \* **RTK clase I:** Familia de los receptores para el factor de crecimiento epidérmico (EGFR del inglés *Epidermal Growth Factor Receptor*).
- \* **RTK clase II:** Familia de los receptores de insulina (IGFR, del inglés *Insulin Growth Factor Receptor*).
- \* **RTK clase III:** Familia de los receptores del factor de crecimiento derivado de plaquetas (PDGFR, del inglés *Platelet-derived Growth Factor Receptor*).
- \* **RTK clase IV:** Familia de los receptores del factor de crecimiento de fibroblastos (FGFR, del inglés *Fibroblast Growth Factor Receptor*).
- \* **RTK clase V:** Familia de los receptores del factor de crecimiento del endotelio vascular (VEGFR, del inglés *Vascular Endoteliar Growth Factor Receptor*).
- \* **RTK clase VI:** Familia de los receptores del factor de crecimiento de hepatocitos (HGFR, del inglés *Hepatocyte Growth Factor Receptor*), conocidos como cMET.
- \* **RTK clase VII:** Familia de receptores de Trk.
- \* **RTK clase VIII:** Familia de receptores de membrana activados en respuesta a efrina.
- \* **RTK clase IX:** Familia de receptores tirosina cinasa UFO, codificados por *AKL*.

Como hemos visto, las proteína-cinasas juegan un papel fundamental en casi todos los procesos biológicos, apoptosis, ciclo celular, progresión, ordenación del citoesqueleto, diferenciación celular, desarrollo, respuesta inmune, función del sistema nervioso o transcripción. La desregulación de alguna proteína cinasa tiene lugar en numerosas

enfermedades, incluyendo cáncer, diabetes, enfermedades autoinmunes, cardiovasculares o trastornos del sistema nervioso.

## 2.5. Importancia de la familia HER en cáncer

Cohen descubrió la importancia de la familia ERBB cuando describió por primera vez la función del factor de crecimiento epidérmico (EGFR), revelando su acción como proteína-tirosina cinasa (138), además fue el primer receptor del que se encontró relación directa entre su sobreexpresión y el desarrollo de cáncer, es por ello uno de los receptores más estudiados por su papel en transducción de señales y oncogénesis (139). Pero no sólo EGFR, toda la familia ERBB, formada por EGFR o HER1, ERBB2 o HER2, ERBB3 o HER3 y ERBB4 o HER4, se encuentra entre las familias de señalización más estudiadas en biología (Figura 6), (140).

Los receptores HER se expresan de manera ubicua en epitelio, mesénquima y células neuronales. Entre sus ligandos encontramos factores de crecimiento como EGF o TGF- $\alpha$ ; neuregulinas 1, 2, 3 y 4; EPR (*effector cell protease receptor-1*); o betacelulina (BTC), entre otros. No se conoce ningún ligando para HER2, por lo que se supone que su activación ocurre por la unión a otros miembros de la familia mediante heterodimerización. Tras la activación del receptor por parte del ligando, pasa de su forma monomérica inactiva a un dímero con actividad cinasa (141).

Consisten en un dominio extracelular al que se une el ligando, un segmento hidrofóbico transmembrana simple y un dominio intracelular con actividad cinasa. El dominio II del fragmento extracelular participa en la formación de homo- y hetero-dímeros. Los homodímeros formados por HER2/HER2 y HER3/HER3 no son funcionales, ya que HER2 no posee dominio extracelular y HER3 no tiene dominio catalítico, sin embargo una sobreexpresión de HER2 fuera de lo normal puede dar lugar a un homodímero de HER2 funcional. Los heterodímeros de HER2 con HER1 y HER3, sin embargo, son completamente funcionales (Figura 6B) (142).

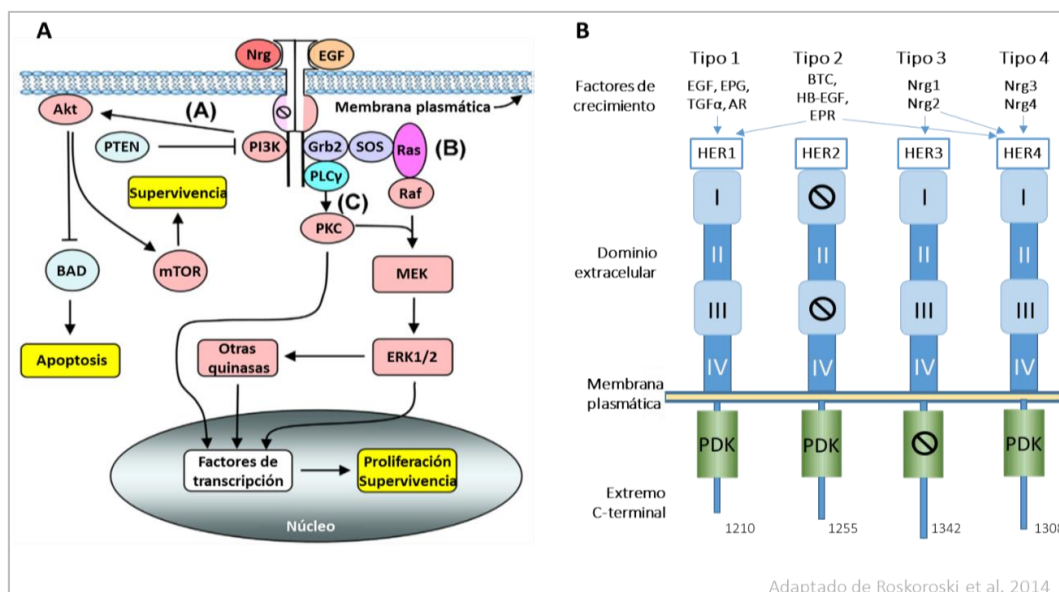


Figura 6 A. Rutas acopladas a la familia de receptores de membrana EGFR, asociadas con control del ciclo celular, evasión de la apoptosis y reparación del DNA. Nrg, Neuregulina. EGF, factor de crecimiento epidérmico. PI3K, fosfoinositol-3-cinasa. (A) Ruta Akt/mTOR. (B) Ruta RAS/Raf. (C) Ruta MAPK. B. Representación esquemática de la estructura secundaria de los diferentes receptores de membrana de la familia EGFR. Se representan los receptores HER1, HER2, HER3 y HER4, sus dominios y los factores de crecimiento que conforman sus ligandos de tipo 1, 2 y 3 y 4.

Al tratarse de receptores tipo RTK, su activación inicia la transducción de señales de las rutas MAPK, PI3K/Akt/mTOR, Ras/Raf, JNK o la ruta de la fosfolipasa C (PLC), entre otras, responsables de la internalización al núcleo de factores de transcripción que favorecen la proliferación y supervivencia celular. Su desregulación, por tanto, contribuye en la iniciación de procesos oncogénicos en cáncer de pulmón, cáncer de colon y cáncer gástrico, cáncer de mama, cáncer de cabeza y cuello, glioblastoma y cáncer de páncreas (Figura 6A) (141).

## EGFR

Aproximadamente la mitad de los casos de cáncer de mama triple negativos (TNBC) e inflamatorios (IBC) sobreexpresan EGFR, por lo que el tratamiento con inhibidores de EGFR ha sido evaluado en numerosos estudios para tratar estos tipos de cáncer de mama tan agresivos. Recientes trabajos han revelado que la ruta de señalización activada por EGFR regula el EMT, migración e invasión tumoral, indicando que la elevada expresión de EGFR actúa como predictor de mal pronóstico en sí mismo. Aunque los recientes estudios indican que la terapia dirigida a EGFR puede tener un papel prometedor en TNBC y IBC, se necesita de unos biomarcadores adecuados que determinen qué pacientes responden mejor a dicho tratamiento. (143).

## HER2

HER2 se encuentra sobreexpresado en el 20 - 30% de todos los cánceres de mama, usualmente ocurre por amplificación del proto-oncogén *ERBB2*, el número de copias se puede determinar mediante FISH. Además, se estima que un 1,6% de las pacientes son portadoras de mutación activadora en *ERBB2* (144). La sobreexpresión de HER2 constituye un factor pronóstico, siendo de mal pronóstico hasta la implantación de la terapia anti-HER2.

## HER3

HER3 juega un importante papel en promover la supervivencia celular, ya que la subunidad reguladora de PI3K (p85), se une preferentemente a HER3, debido a su cantidad de residuos de fosfo-tirosinas, lo que lleva a la activación de PI3K y de la ruta Akt/mTOR. La sobreexpresión de HER3 es un pronóstico de menor supervivencia en pacientes de cáncer de mama (145). De hecho, la coexpresión del heterodímero formado por HER2 y HER3 constituye una importante unidad oncogénica asociada a peor supervivencia (146).

## HER4

La expresión de HER4 es considerado como factor de buen pronóstico en cáncer de mama, ya que las evidencias sugieren que promueve la diferenciación celular e inhibe el crecimiento de las células de cáncer de mama (147). Sin embargo, trabajos recientes demuestran que ciertas isoformas de HER4, promueven la carcinogénesis y aumentan la capacidad de migración de la célula (148-150). Por otro lado algunos trabajos sugieren que la localización subcelular de su dominio intracelular determina el pronóstico; además, parece que la localización nuclear de HER4 media en la adquisición de resistencia a trastuzumab (terapia anti-HER2) en pacientes HER2 positivas (151, 152).

## Receptores HER en el núcleo

Además de su papel más que conocido como receptores de membrana y su acción en el citoplasma, los receptores de la familia HER pueden internalizarse y traslocarse al núcleo, donde participan en la señalización celular por sí mismos (153). Entre las funciones de estos receptores

en el núcleo se incluyen proliferación celular, reparación del DNA, transcripción de genes, desarrollo y expansión tumoral (154). La región C-terminal de los receptores HER contiene una secuencia transactivadora de genes, entre sus dianas destacan *CCND1*, *NOS2*, *AURKA*, *PCNA*, *CDK2*, genes de la glía, *GFAP*, *S100B* (141). EGFR se ha encontrado localizado en el núcleo en numerosos tipos de cáncer, incluyendo cáncer de mama, por lo que se discute el papel de la localización nuclear de los receptores HER en los procesos de metástasis, progresión del tumor y resistencia a radioterapia (154, 155).

### 3. PERFILES MOLECULARES Y USO DE MUESTRAS FFPE EN ESTUDIOS MOLECULARES

---

Las técnicas empleadas para la caracterización molecular de tumores suelen requerir de una gran calidad de las muestras, tanto a la hora de conservarlas como a la hora de extraer el material, en este caso RNA, por lo que se emplean muestras frescas o material congelado en condiciones óptimas. Sin embargo, cada vez son más los estudios moleculares llevados a cabo utilizando material parafinado (FFPE del *inglés formaline fixed paraffine embedded*), por su mayor disponibilidad con carácter retrospectivo. Sin embargo, los análisis de perfiles transcriptómicos (mRNA) empleando muestras parafinadas son técnicamente un reto ya que, en estas condiciones, los ácidos nucleicos no se conservan de manera adecuada (156) Esto debido a la fuerte degradación enzimática y química del RNA durante el proceso de fijación, a su continuo deterioro a lo largo del tiempo, y a la interferencia en la reacción de retrotranscripción, necesaria en técnicas como la PCR (157-161).

El material parafinado es ideal para estudios histológicos, ya que el proceso de fijación e inclusión en parafina conduce a un mantenimiento idóneo de la estructura tisular, evitando la putrefacción del tejido. Sin embargo, Por un lado, el proceso de fijación e inclusión en parafina produce entrecruzamientos entre las moléculas de RNA y las proteínas, haciendo al RNA más resistente a la extracción. Además, se obtiene un peor rendimiento de la extracción y una peor calidad e integridad del RNA, debido a la degradación enzimática y química, que ocurren antes y durante el proceso de fijación. Sin embargo, los miRNAs no se ven afectados de manera significativa durante el proceso de fijación, son una diana menos susceptible de degradación química y enzimática, y pueden extraerse fácilmente a partir de muestras parafinadas dado su pequeño tamaño (20-22 nucleótidos) y su alta estabilidad (162, 163).

Se ha propuesto que la protección de los miRNA frente a la degradación se debe a su fuerte asociación con proteínas del complejo RISC, siendo tan sólo un bajo porcentaje de miRNAs maduros los que se encuentran libres en las células (164). En comparación con los mRNAs, parece que los miRNAs no son tan susceptibles a los efectos negativos de la fijación asociados a muestras parafinadas, por lo que resultan candidatos idóneos para ser estudiados empleando material parafinado, lo que les da un valor añadido a la hora de su estudio como potenciales biomarcadores en cáncer (165).

## 4. ESTUDIOS MOLECULARES EN MUJERES JÓVENES CON CÁNCER DE MAMA

---

Es aún tema de debate controvertido si el cáncer de mama en mujeres jóvenes se trata de una entidad biológica diferenciada del cáncer de mama en mujeres mayores, o si su comportamiento y peor pronóstico son un reflejo de la mayor frecuencia de subtipos agresivos que presenta este grupo de edad. Existen estudios contradictorios tanto a favor como en contra de que su comportamiento tenga una base biológica diferente (56, 57, 59).

### 4.1. Biomarcadores pronóstico en cáncer de mama de mujeres jóvenes

Lin y colaboradores, 2011, determinan que en mujeres menores de 35 años, las mutaciones en TP53, el porcentaje de Ki67 y la expresión de HER2 tienen mayor peso como factores pronóstico que el estatus hormonal. Sugieren que, en vista del limitado valor pronóstico que muestran los receptores hormonales en este grupo de edad, los biomarcadores empleados para cáncer de mama en general, sin tener en cuenta la edad de la paciente, podrían no ser del todo adecuados para mujeres muy jóvenes con cáncer de mama (166).

Anderson y Matsuno, 2006, basándose en la distribución de la edad de diagnóstico y en su tasa de mortalidad, definieron el cáncer de mama como una mezcla de, al menos, dos tipos de cáncer (167). El primero tendría su edad de aparición alrededor de los 50 años y el segundo se daría en mujeres de más de 65 años, pudiendo ser definidos como pre y post-menopáusicos, atendiendo al principal evento hormonal que diferencia ambos grupos. Aun así, los autores destacan en su trabajo que “*Como mínimo, no se debe ver el cáncer de mama como una entidad biológica única nunca más. El concepto de un modelo de mezcla debe conformar la base para el nuevo marco conceptual. Por ello, si el cáncer de mama consiste en una mezcla de al menos dos tipos, necesitamos una aproximación estratificada más que unificada para la investigación, prevención y tratamiento del cáncer de mama.*” Su aseveración pone de manifiesto la necesidad de un enfoque diferente tanto en la investigación como en el tratamiento del cáncer de mama. El hecho de que el cáncer de mama en mujeres muy jóvenes no forme un pico de frecuencia más elevado en la edad de aparición puede deberse, simplemente, al hecho de que su proporción es mucho menor si la comparamos con el cáncer de mama pre y post-menopáusico. Esto no implica necesariamente que la biología subyacente no sea una entidad diferenciada dentro del cáncer de mama.

### 4.2. Estudios de la expresión de genes en mujeres jóvenes

Como se ha destacado antes, existen trabajos que apoyan esta hipótesis, así como otros que la desmienten, sugiriendo que el diferente comportamiento biológico es una consecuencia natural de la acumulación de subtipos agresivos, y que la existencia de un perfil molecular distintivo de mujeres muy jóvenes es un artefacto producto de este hecho (56, 57, 59, 64, 168, 169).

Anders y colaboradores, 2008, describieron en dos trabajos una expresión diferencial de genes en mujeres menores de 45 años en comparación con mujeres mayores de 65 años. Estos genes únicos de tumores mamarios de mujeres jóvenes se relacionan con función inmune, hipoxia, BRCA1, apoptosis, deacetilación de histonas, así como con múltiples rutas oncogénicas,



incluyendo PI3K/Akt/mTOR, Myc, Ras, E2F1 y B-cateninas, p53, PTEN y MAP cinasas. Los autores, además, asocian este perfil de expresión génica a un peor pronóstico (56, 57).

Recientemente, Colak 2013, revelaron una expresión de genes diferente en tejido tumoral mamario según la edad, separando los perfiles en mujeres muy jóvenes (menores de 35 años), mujeres jóvenes (de 35 a 45 años), pre-menopáusicas (de 45 a 55 años) y mujeres mayores (de 55 años en adelante). Analizando por separado los perfiles distintivos de cada grupo, destacan que los procesos biológicos enriquecidos asociados con genes desregulados en mujeres muy jóvenes (menores de 35 años) incluían control del ciclo celular, morfogénesis, proliferación y muerte celular, entre otros. De manera similar al perfil obtenido para mujeres jóvenes (de 35 a 45 años), los análisis de redes biológicas indicaron alteraciones en las rutas de señalización de MAPK, PI3K/Akt y NFkB, además de revelar como potencialmente importantes los genes IL1RN, ESR1 y la familia de ERBB2 (169).

### 4.3. Estudios en miRNAs en mujeres jóvenes

En cuanto al papel de los miRNAs en la aparición de cáncer de mama en mujeres muy jóvenes, poco se conoce. Dos trabajos similares relacionan la presencia de un polimorfismo funcional en el gen del miR-146a con un adelanto en la edad de aparición de cáncer de mama (170, 171). Del mismo modo, se relaciona un polimorfismo en la zona de unión del miR-502, localizada en la región 3'UTR del gen SET8, con un adelanto en la edad de aparición de cáncer de mama (172) (Song 2009). Recientemente, un estudio en mujeres libanesas con cáncer de mama reveló una sobreexpresión del miR-155 en pacientes menores de 40 años, al compararlo con mayores de 40 años (173) (Nassar 2014). Existe también asociación descrita entre el perfil de miRNAs y el cáncer de mama familiar no debido a BRCA1/2, un tipo de cáncer que se origina principalmente en mujeres jóvenes (174, 175).

La identificación de un grupo de pacientes de cáncer de mama según su pronóstico y basado en el perfil de rutas oncogénicas ofrece la oportunidad de un tratamiento personalizado según el riesgo de recurrencia, y, además, permite incorporar terapias según la biología del tumor de manera individual, con el fin último de mejorar el pronóstico de la enfermedad en los pacientes.



# HIPÓTESIS Y OBJETIVOS



---

## Hipótesis

---

Los tumores de mama en mujeres muy jóvenes, menores de 35 años, aunque representen menos del 3% del total son de especial relevancia. Se trata de mujeres que acaban de incorporarse a la vida productiva profesional y personalmente, ya que se encuentran en plena edad reproductiva, en muchos casos sin haber satisfecho las necesidades reproductivas. Estas mujeres cuentan con un periodo largo de vida laboral siendo seguidas clínicamente durante largos periodos de tiempo.

El diagnóstico en este grupo de mujeres es menos eficaz y, debido a las actuales técnicas de diagnóstico, sufre un retraso en el tiempo respecto a mujeres de más edad (principalmente por la mayor densidad mamaria de las mujeres jóvenes), detectándose cuando el tumor ha alcanzado un tamaño considerable.

Los tumores de cáncer de mama en estas mujeres, que se identifican y se tratan en la práctica clínica según su respuesta a receptores hormonales y sobreexpresión de HER2, presentan sin embargo, características clínicas diferenciadas, puesto que son más agresivos, presentan una mayor capacidad metastática, tienen tamaños más grandes, suelen ser menos diferenciados y tener una mayor capacidad proliferativa. Como consecuencia, estos tumores responden peor al tratamiento y se asocian con mal pronóstico.

Todo ello nos lleva a pensar que estos tumores podrían constituir un subgrupo como entidad propia, el estudio de la etiología y biología de éstos tumores permitiría mejorar el diagnóstico, y quizá cambiar las pautas terapéuticas asignadas por otras mejor adaptadas a su perfil, considerando las necesidades específicas de las pacientes de este grupo de edad.



---

---

## Objetivos

---

---

El objetivo fundamental de este trabajo de tesis ha sido la caracterización molecular del cáncer de mama en mujeres muy jóvenes, consideradas así a las pacientes menores de 35 años, Para ello, nos hemos planteado los siguientes objetivos específicos:

1. Evaluar la utilidad de muestras de tejido conservadas en parafina (FFPE) para estudios moleculares basados en microRNA
2. Identificar un perfil molecular en tumores de pacientes muy jóvenes con cáncer de mama utilizando la expresión de miRNAs mediante plataformas moleculares de alto rendimiento.
3. Evaluar a nivel molecular las líneas celulares procedentes de mujeres jóvenes para establecer su utilidad como modelo en estudios funcionales.
4. Validar el perfil de expresión de miRNAs empleando bases de datos públicas.
5. Determinar aquellas rutas metabólicas desreguladas por los miRNAs mediante la utilización de bases de datos anotadas públicas.
6. Seleccionar miRNAs representativos y validar su expresión con otra tecnología en una muestra de pacientes independiente.
7. Detectar la expresión de genes diana seleccionados en muestras de tejido tumoral de pacientes:
  - \* Mediante el análisis de la expresión de mRNA.
  - \* Mediante la cuantificación de proteína por tinción inmunohistoquímica.



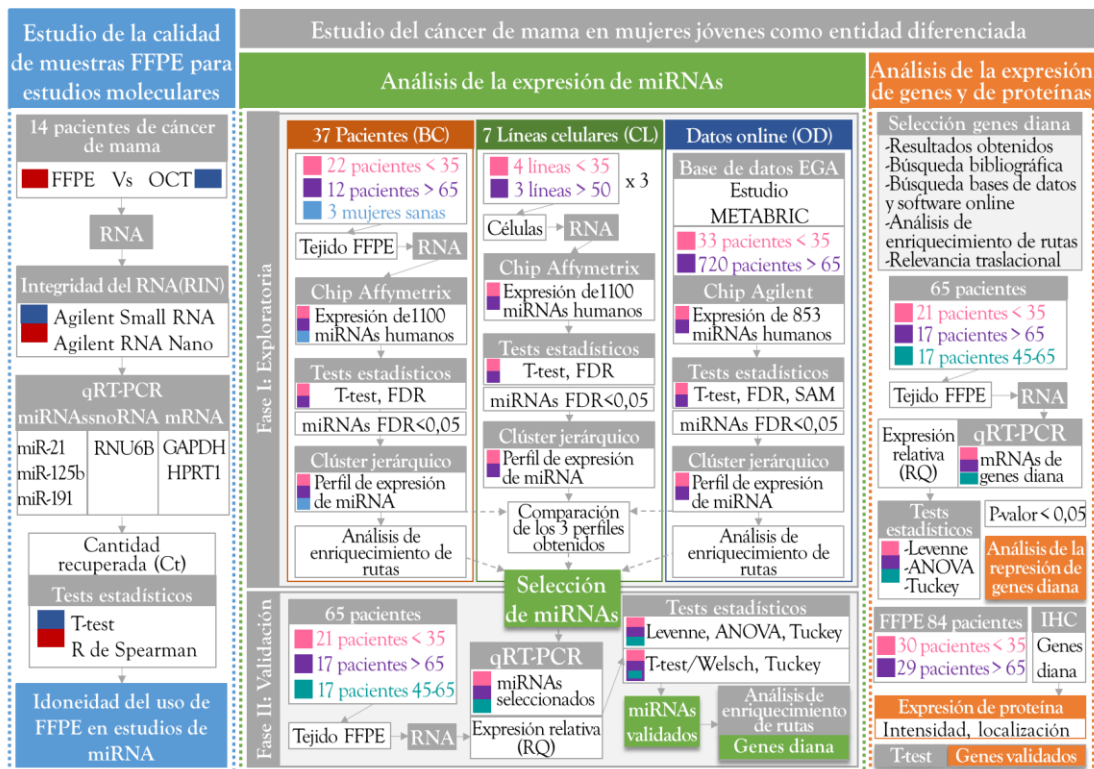


# MATERIAL Y MÉTODOS



# DISEÑO EXPERIMENTAL

El trabajo realizado en esta tesis doctoral puede separarse en tres grandes bloques; en el primero se evalúa la viabilidad de las muestras de tejido conservadas en parafina y su idoneidad para el empleo en estudios moleculares. En un segundo bloque se estudia la expresión de miRNAs en cáncer de mama en mujeres jóvenes y, en el último bloque se estudia la expresión de genes y proteína en tejido tumoral mamario de pacientes jóvenes. Estos dos últimos bloques constituyen el grueso del trabajo de esta tesis, cuyo tema se centra en evaluar si el cáncer de mama en mujeres jóvenes se trata de una entidad biológica diferente dentro del cáncer de mama general (Figura 7).



**Figura 7. Diagrama del diseño experimental llevado a cabo en este trabajo.** Los cuadros de color en cada análisis y test realizado corresponden al grupo de muestras sobre el que ha sido realizado, el código de colores se detalla en cada apartado individualmente. FFPE, tejido parafinado (*formaline-fixed paraffin-embedded*); OCT, tejido congelado incluido en criomoldes OCT (*optimal cutting temperature*); RIN, medida de la integridad del RNA (*RNA integrity number*); qRT-PCR, PCR cuantitativa en tiempo real; Ct, ciclo umbral de detección; FDR, test de corrección por múltiples comparaciones de Benjamini & Hochberg (*false discovery rate*); EGA, base europea de datos genómicos (*European Genome-phenome Archive*); SAM, Análisis de la significatividad sobre microchips de expresión (*significance analysis of microarray*); RQ, Cantidad relativa de material (*relative quantification*).

# 1. EVALUACIÓN DE LA IDONEIDAD DEL USO DE TEJIDO INCLUIDO EN PARAFINA EN ESTUDIOS MOLECULARES

---

Durante el año 2011 se recogieron en un Biobanco hospitalario un total de 14 muestras tisulares para ser criopreservadas procedentes de pacientes afectas de cáncer de mama que no habían sido sometidas a ningún tratamiento previo.

En este trabajo se plantea el análisis de la integridad del RNA total y del RNA de pequeño tamaño obtenido de estas 14 muestras de tejido congelado conservado en OCT y de sus correspondientes bloques de parafina procedentes del mismo tumor mamario.

Para este estudio se partió de dos formatos de conservación de muestras: congelados óptimos (CO) y parafinas (FFPE). El estudio se basó en la extracción del RNA total, incluyendo los RNAs de pequeño tamaño, de todas las muestras (14 parejas congelado óptimo/parafina) y su análisis comparativo de integridad. Para ello, la extracción de RNA se realizó con kits comerciales y se analizó tanto la integridad del RNA total como la de los RNAs de pequeño tamaño.

Finalmente, se analizaron mediante PCR cuantitativa en tiempo real los valores de Ct medios de estas muestras para tres microRNAs, un snoRNA y dos mRNAs con el objetivo de evaluar la recuperación de estas moléculas a partir de los distintos formatos de conservación de la muestra de partida. La finalidad última es determinar la calidad e idoneidad de las muestras de tejido incluido en parafina para futuros estudios moleculares basados en RNA.

# 2. ESTUDIO DE LA EXPRESIÓN DE MICRORNAS EN PACIENTES JÓVENES DE CÁNCER DE MAMA

---

El estudio del cáncer de mama en mujeres jóvenes como entidad biológica diferenciada del cáncer de mama se basó en la obtención de un perfil molecular mediante el análisis de la expresión de microRNAs. Para este trabajo se han empleado un total de 89 muestras de tejido mamario procedente de pacientes de cáncer de mama, 3 muestras de tejido normal de mujeres sanas, 7 líneas celulares y los datos depositados online de 745 pacientes de cáncer de mama. El trabajo consta de varias partes diferenciadas, que se detallan a continuación.

## 2.1. *Fase I: Análisis exploratorio de la expresión de microRNAs en pacientes jóvenes de cáncer de mama.*

### **Análisis de la expresión en muestras procedentes de pacientes**

En esta fase se realizó, en primer lugar, una selección exhaustiva de pacientes procedentes del Hospital clínico universitario de Valencia, atendiendo a criterios demográficos y clínicos. Una vez seleccionada la muestra, se obtuvieron cortes de tejido tumoral mamario FFPE, de los que se extrajo el RNA total, incluyendo la fracción micro, que contiene los miRNAs. Posteriormente se analizó la expresión de 1100 miRNAs humanos, incluidos en el chip de Affymetrix microArray 2.0 empleando esta plataforma disponible. Se analizaron 37 muestras de RNA; 22 muestras procedentes de pacientes menores de 35 años, 12 procedentes de pacientes mayores de 65 años y

3 muestras de tejido mamario normal procedente de mujeres sanas. Estas últimas muestras se emplearon como control para discernir aquellas diferencias debidas a la edad de las pacientes.

Una vez obtenida la expresión de los miRNAs en las muestras, se analizaron estadísticamente las diferencias entre mujeres jóvenes y mujeres mayores, seleccionándose los miRNAs significativamente diferentes para realizar, mediante agrupamiento jerárquico, el perfil molecular distintivo de cáncer de mama en mujeres muy jóvenes.

Los miRNAs ejercen su función reprimiendo el mRNA de genes, de modo que, para evaluar su posible implicación en los procesos que le confieren mayor agresividad al cáncer de mama en mujeres jóvenes o en el desarrollo de la enfermedad, debemos estudiar los genes diana que son reprimidos por estos miRNAs desregulados. Para ello se analizó el enriquecimiento de rutas, de manera que se obtuvieron aquellas rutas en las que los miRNAs tienen una mayor representación, obteniendo, además, los posibles genes diana.

Con los resultados de este análisis se seleccionaron las rutas más relevantes, así como miRNAs y los genes diana más representativos o interesantes desde un punto de vista traslacional.

### **Análisis de la expresión en líneas celulares**

Con el fin de comprobar si la desregulación de miRNAs observada en pacientes de cáncer de mama muy jóvenes se reproduce en líneas celulares, se analizó, de manera similar al análisis realizado sobre muestras de pacientes, la expresión de los miRNAs en líneas celulares de cáncer de mama procedentes de mujeres jóvenes, y se comparó con la expresión en líneas de mujeres de más edad. Las líneas celulares se crecieron en cultivo por triplicado, para asegurar la reproducibilidad del estudio, recogiendo las células de cada una de las réplicas y tratándolas como muestras separadas. Se analizaron los miRNAs diferencialmente expresados en ambos grupos de líneas celulares. Por último, se trató de reproducir el perfil de expresión de miRNAs obtenido en muestras de pacientes y característico de mujeres jóvenes, para evaluar la adecuación de estas líneas celulares como modelo para el estudio del cáncer de mama en mujeres jóvenes.

### **Análisis de la expresión de miRNAs en una serie independiente mayor de pacientes, obtenida a partir de bases de datos online**

Con el fin de reproducir los resultados de perfil de expresión obtenidos en el estudio realizado con pacientes, se emplearon datos de estudios moleculares depositados en bases de datos online, concretamente EGA. Se seleccionaron los datos de expresión de miRNA obtenidos mediante la plataforma Agilent y se filtraron los datos de miRNAs humanos, así como los datos clínicos relevantes para el estudio. De las 1460 muestras depositadas en el estudio, se analizaron un total de 33 mujeres menores de 35 años y 720 mayores de 65. Se evaluaron los miRNAs diferencialmente expresados y se realizó un análisis de enriquecimiento de rutas para, en última instancia, comparar los resultados con los obtenidos en los otros análisis.

## *2.2. Fase II: Validación de la expresión de los miRNAs seleccionados*

Para validar los resultados obtenidos en la primera fase de este trabajo, se seleccionó una nueva serie de muestras de tejido tumoral mamario procedente de pacientes del Hospital Clínico Universitario de Valencia, con características similares a las empleadas en la primera fase. Se obtuvo muestra de 21 pacientes jóvenes, 17 mayores de 65 años y 17 pacientes con edades comprendidas entre los 45 y los 65 años. Este último grupo fue introducido en el análisis para

asegurarnos de que las diferencias observadas no eran un reflejo de los dos grupos de cáncer de mama descritos hasta la fecha, postmenopáusico y pre-menopáusico.

La validación se llevó a cabo, pues, en una serie de mujeres independiente de la de la Fase I, mediante una técnica diferente. En esta fase se analizó la expresión de miRNAs en el tejido tumoral mediante PCR cuantitativa (qRT-PCR). Mediante análisis estadísticos se comparó la expresión obtenida en los tres grupos, primero como grupos independientes y después considerando el grupo de mujeres mayores e intermedias como un único grupo para añadir robustez a los resultados, en el caso de que ambos grupos fuesen estadísticamente iguales.

Por último, se realizó un análisis de enriquecimiento de rutas con los miRNAs validados, con el fin de recrear el perfil de cáncer de mama en mujeres jóvenes obtenido, empleando únicamente dichos miRNAs. Esta reproducción permitiría aislar los biomarcadores, aportando relevancia traslacional al estudio.

### 3. ESTUDIO DE LA EXPRESIÓN DE GENES DIANA (MRNA) Y DE PROTEÍNA EN CÁNCER DE MAMA EN MUJERES JÓVENES

---

Con el fin de confirmar la acción de los miRNAs seleccionados sobre los genes diana relevantes, se analizó la expresión de los genes diana, por un lado mediante la cuantificación del mRNA y por otro lado mediante la observación de la cantidad de proteína *in situ*.

#### 3.1 *Análisis de la expresión de genes mediante qRT-PCR de mRNA*

El análisis de la expresión de los genes seleccionados se llevó a cabo en la misma serie de pacientes empleada para el análisis de miRNAs mediante qRT-PCR (Fase II del apartado anterior). Un total de 56 muestras de tejido tumoral FFPE procedente de pacientes de cáncer de mama fueron analizadas. Para ello se extrajo el RNA total de las muestras de tejido FFPE y se cuantificó la cantidad de mRNA mediante sondas específicas empleando la técnica qRT-PCR. Con los datos obtenidos se realizaron tests estadísticos para evaluar la expresión diferencial en mujeres jóvenes con cáncer de mama.

#### 3.2 *Análisis de la proteína in situ*

Para este estudio se emplearon 60 muestras tumorales de mama procedente de pacientes del Hospital Clínico Universitario de Valencia. Se seleccionaron los genes más relevantes y se analizó la cantidad de proteína en el tejido tumoral mamario de pacientes menores de 35 años y mayores de 45. Para este análisis se construyeron 5 TMAs a partir del tejido FFPE, que incluían muestra de tejido normal de 60 pacientes además de tejido normal procedente de 3 mujeres sanas. Se realizaron cortes y se incubaron con anticuerpos específicos para la proteína mediante tinción inmunohistoquímica; por último, un patólogo experto evaluó la cantidad de proteína, asignándole un valor relativo, así como la localización de la tinción y, por tanto, de la proteína. Posteriormente se analizaron estadísticamente los datos y se valoró la diferencia o no en la expresión de proteína entre ambos grupos de mujeres, atendiendo a aspectos como la intensidad y la localización de la tinción.

# MATERIAL EMPLEADO

## 1. MATERIAL BIOLÓGICO

### 1.1. Muestras de tejido tumoral

Todas las muestras de tejido proceden de pacientes de cáncer de mama o mujeres sanas del Hospital Clínico Universitario de Valencia que fueron sometidas a cirugía. El trabajo ha sido aprobado por el comité de ética.

#### Pacientes para el estudio de la viabilidad de muestras FFPE en estudios con miRNA

Las 28 muestras analizadas en este trabajo son muestras tisulares procedentes de 14 pacientes afectas de cáncer de mama durante el año 2011, que dieron su consentimiento para la obtención y conservación de su muestra en un Biobanco hospitalario y para que ésta pudiera ser utilizada en futuros estudios biomédicos.

Tabla 5. Características tumorales de las muestras empleadas en el estudio de viabilidad.

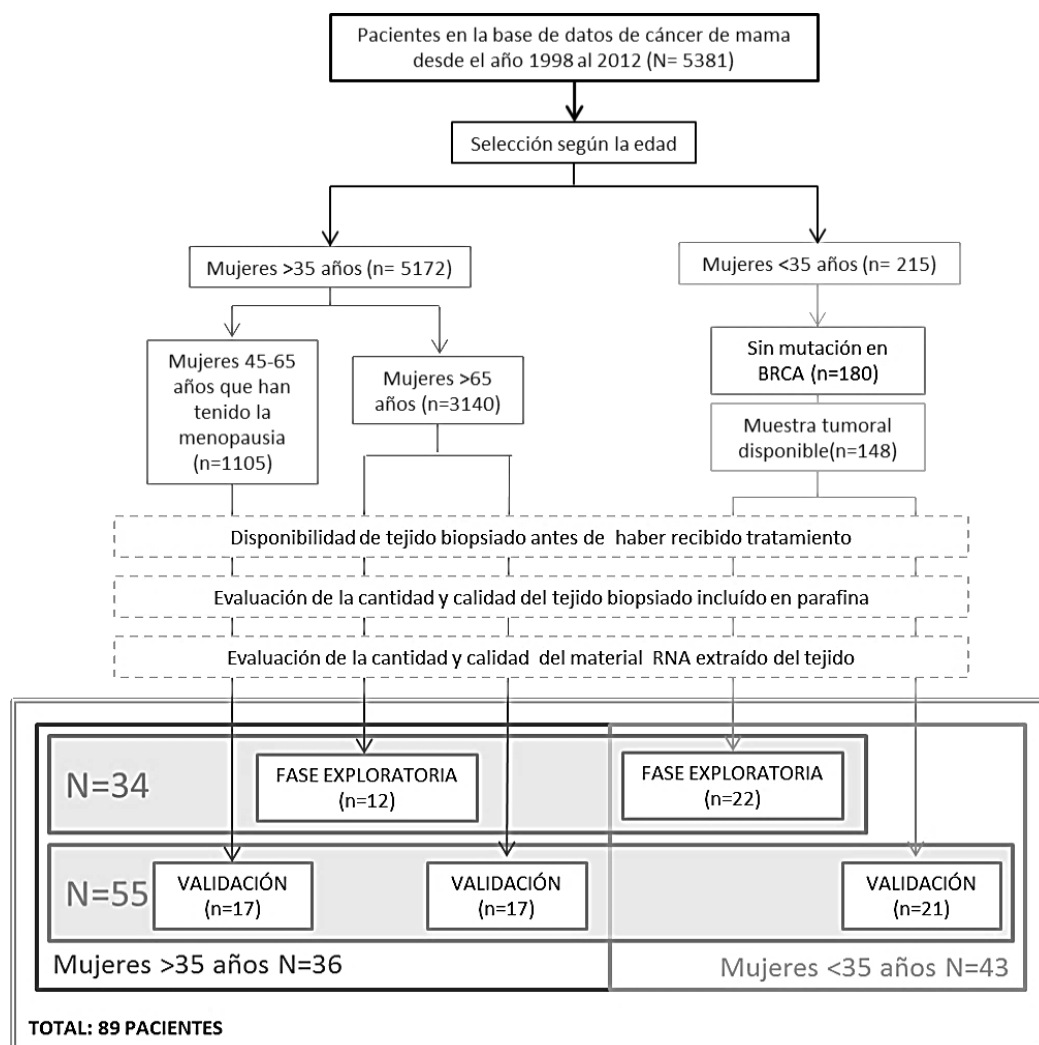
ID Muestra	Edad	Dx	Tipo	Grado	Tamaño (cm)	pT	pN	ER (%)	PR (%)	HER2 (IHQ/FISH)	Ki67 %	CtC %	CtP %
BC1	52	CDI	LUM	II	1,9	T1c	N1(mi)	+(100)	+(100)	-(1+)	5	70	60
BC2	52	CMPI	LUM	II	1,5	T1c	NO(sn)	+(100)	+(100)	-(1+)	10	50	65
BC3	54	CDI	LUM	II	1,5	T1c	N1a	+(100)	+(60)	-(1+)	30	55	60
BC4	66	CDI	LUM	I	1,8	T1c	NO(sn)	+(100)	+(30)	-(1+)	5	80	60
BC5	50	CDI	LUM	II	1,3	T1c	NO(sn)	+(80)	+(40)	-(2+/na)	30	70	55
BC6	70	CDI	LUM	I	2,2	T2	N1(mi)	+(100)	+(80)	-(1+)	10	60	60
BC7	55	CDI	LUM	I	1,1	T1c	NO(sn)	+(100)	+(100)	-(1+)	5	70	60
BC8	84	CDI	LUM	II	3,5	T2	NO(sn)	+(100)	+(80)	-(0)	25	70	70
BC9	47	CDI	LUM	I	1,2	T1c	NO(sn)	+(100)	+(100)	-(0)	5	80	60
BC10	58	CDI	LUM	I	0,9	T1b	NO(sn)	+(100)	+(100)	-(2+/na)	20	90	70
BC11	85	CDI	TN	II	0,9	T1b	NO	-	-	-(0)	60	50	50
BC12	50	CDI	LUM	I	1,1	T1c	NO	+(100)	+(80)	-(2+/na)	13	60	50
BC13	73	CDI	LUM	I	0,8	T1b	NO(sn)	+(100)	+(100)	-(2+/na)	10	60	35
BC14	52	CDI	TN	III	3,0	T2	Nx	-	-	-(0)	70	50	50

CO=congelado óptimo; COP=parafina del congelado óptimo. Dx: diagnóstico. CDI=carcinoma ductal infiltrante; CLI=carcinoma lobulillar infiltrante; CMPI=carcinoma micropapilar infiltrante. LUM= luminal; TN=triple negativo; (sn)= (sentinel node) ganglio centinela; (mi)=microinvasor; ER=receptor de estrógenos; PR=receptor de progesterona; IHQ=inmunohistoquímica; na=no amplificado; CtC=porcentaje de celularidad tumoral de muestras congeladas; CtP=porcentaje de celularidad tumoral de muestras fijadas e incluidas en parafina.

Para el análisis de la viabilidad del tejido incluido en parafina como objeto de estudio en técnicas moleculares, se emplearon muestras de tejido de tumor de mama procedente de pacientes que no habían recibido tratamiento previo, de las que se disponía tanto de tejido almacenado en condiciones óptimas de congelación, como de tejido fijado en formalina e incluido en parafina (FFPE). Las características clínicas e histológicas de los tumores empleados se encuentran detalladas en la Tabla 5.

### Pacientes para el estudio de la expresión de miRNAs en mujeres jóvenes con cáncer de mama

Para llevar a cabo este estudio, se seleccionaron un total de 37 muestras, 22 mujeres menores de 35 años (“grupo de estudio”) y 12 mayores de 65 años (“control mayores”), empleamos 3 mujeres sanas (“control sanas”) como control para el análisis global de la expresión de miRNAs.



**Figura 8. Diagrama de flujo del proceso de selección de pacientes para el estudio.** Se representa el proceso de selección de pacientes y muestras empleadas y los criterios de selección. Se muestra en cada paso el número de casos seleccionados (n). Se muestran los grupos finales (mujeres menores de 35 años y mujeres mayores de 65 años) en cada fase (exploratoria y fase de validación) junto con el número de muestras empleadas de cada grupo (N). BRCA: gen del cáncer de mama 1 ó 2 (del inglés *breast cancer gene*).

Para minimizar el sesgo introducido, se procuró emparejar los grupos según su subtipo molecular, teniendo en cuenta también características tumorales como el grado de diferenciación,



estadio, tamaño tumoral, afectación de los nodos linfáticos y sobreexpresión del factor de proliferación Ki67.

Para la fase de validación, se seleccionaron 21 mujeres menores de 35 años, 17 mayores de 65 años y 17 pacientes de un nuevo grupo con edades comprendidas entre 45 y 65 años (“control intermedias”).

El proceso de selección de las integrantes del estudio se llevó a cabo sobre la base de datos del Hospital Clínico Universitario de Valencia, que cuenta con registro de 5381 pacientes (ver la Figura 8). Las características clínicas de las pacientes de cáncer de mama empleadas en este estudio se detallan en la tabla 6.

Los criterios para la selección de los grupos de estudio fueron los siguientes:

#### Grupo de estudio

- \* Pacientes con cáncer de mama diagnosticado a los 35 años o antes.
- \* Sin mutación conocida en BRCA1/2.
- \* Sin haber recibido tratamiento neoadyuvante.
- \* Pacientes no metastásicas.
- \* Distribución de la muestra representativa en cuanto a características clínicas y tumorales.

#### Grupo control mayores

- \* Pacientes con cáncer de mama diagnosticado a los 65 años o en adelante.
- \* Sin mutación conocida en BRCA1/2.
- \* Sin haber recibido tratamiento neoadyuvante.
- \* Pacientes no metastásicas.
- \* Muestra representativa con características clínicas y tumorales similares al grupo de estudio.

#### Grupo control intermedias

- \* Pacientes con cáncer de mama diagnosticado entre los 45 y los 65 años de edad.
- \* Sin mutación conocida en BRCA1/2.
- \* Sin haber recibido tratamiento neoadyuvante.
- \* Pacientes no metastásicas.

#### Grupo control sanas

- \* Mujeres libres de cáncer.
- \* Con edades similares al grupo de estudio de mujeres menores de 35 años.
- \* Con parámetros demográficos y características similares al grupo de estudio.

**Tabla 6.** Características clínicas de las pacientes de cáncer de mama empleadas en el estudio.

Grupo de edad	Fase I N=34 (%)		Fase II N=55 (%)		
	<35 (n=22)*	>65 (n=12)	<35 (n=21)	45-65 (n=17)	>65 (n=17)
Media de edad (SD)	31.41 (4.04)	73.3 (10.18)	31.4 (2.87)	56.94 (5.58)	69.4 (4.48)
Media de IMC(SD)	21.98 (3.53)	29.44 (6.66)	23.95 (5.50)	30.15 (8.04)	30.23 (6.55)
<b>Grado histológico</b>					
I	1 (4.55)	3 (25.00)	3 (14.29)	8 (47.06)	3 (17.65)
II	8 (36.36)	3 (25.00)	7 (33.33)	5 (29.41)	12 (70.59)
III	13 (59.09)	6 (50.00)	11 (52.38)	4 (23.53)	2 (11.76)
<b>Tipo histopatológico</b>					
CDI	19 (86.36)	11 (91.67)	19 (90.48)	13 (76.47)	12 (70.59)
CLI	0 (0.00)	0 (0.00)	1 (4.76)	0 (0.00)	2 (11.76)
Otros	3 (13.64)	1 (8.33)	1 (4.76)	4 (23.53)	3 (17.65)
<b>Tamaño tumoral</b>					
< 2 cm	3 (13.64)	9 (75.00)	3 (14.29)	11 (64.71)	14 (82.35)
2-5 cm	15 (68.18)	2 (16.67)	15 (71.43)	6 (35.29)	3 (17.65)
> 5 cm	4 (18.18)	1 (8.33)	3 (14.29)	0 (0.00)	0 (0.00)
<b>Afectación ganglionar</b>					
Positiva	10 (45.45)	4 (33.33)	7 (33.33)	6 (35.29)	3 (17.65)
Negativa	12 (54.55)	8 (66.67)	14 (66.67)	11 (64.71)	14 (82.35)
<b>Receptores</b>					
ER+	17 (77.27)	9 (75.00)	14 (66.67)	14 (82.35)	14 (82.35)
ER-	5 (22.73)	3 (25.00)	7 (33.33)	3 (17.65)	3 (17.65)
PR+	15 (68.18)	7 (58.33)	14 (66.67)	13 (76.47)	13 (76.47)
PR-	7 (31.82)	5 (41.67)	7 (33.33)	4 (23.53)	4 (23.53)
HER2+	10 (45.45)	3 (25.00)	5 (23.81)	2 (11.76)	2 (11.76)
HER2-	12 (54.55)	9 (75.00)	16 (76.19)	15 (88.24)	15 (88.24)
<b>Ki67 (%)</b>					
1-14%	5 (22.73)	3 (25.00)	5 (23.81)	8 (47.06)	5 (29.41)
14-30%	8 (36.36)	5 (41.67)	11 (52.38)	7 (41.18)	10 (58.82)
>30%	9 (40.91)	4 (33.33)	5 (23.81)	2 (11.76)	2 (11.76)
<b>Subtipo molecular</b>					
Luminal A	5 (22.73)	3 (25.00)	3 (14.29)	8 (47.06)	5 (29.41)
Luminal B	6 (27.27)	4 (33.33)	10 (47.62)	4 (23.53)	9 (52.94)
TN	2 (9.09)	2 (16.67)	3 (14.29)	3 (17.65)	2 (11.76)
Luminal/HER2	5 (22.73)	2 (16.67)	2 (9.52)	0 (0.00)	1 (5.88)
HER2	4 (18.18)	1 (8.33)	3 (14.29)	2 (11.76)	0 (0.00)

IMC: índice de masa corporal, expresado en kg/m<sup>2</sup>. ER: receptor de estrógenos; PR: receptor de progesterone; HER2: receptor ERBB2; +/-: presencia (+) y ausencia (-) de sobreexpresión del receptor. HER2 se considera positivo (+) cuando los análisis inmunohistoquímicos muestran +++/+++ or ++/+++ (y el análisis FISH muestra amplificación). \*Una de las 22 muestras iniciales menores de 35 años fue eliminada del estudio debido a una pobre calidad en los controles de calidad (QC) metodológicos.

## 1.2. Muestras procedentes de cultivo celular

Para el estudio de la expresión de miRNAs en líneas celulares se emplearon 7 líneas celulares de tumor mamario procedentes de la *American Type Culture Collection* (ATCC, Rockville, MD), Se seleccionaron 4 líneas de mujeres jóvenes, HBL100, HCC1500, HCC1937 y Hs 566(B).T. Como control se emplearon 3 líneas celulares de cáncer de mama, típicas y ampliamente usadas: MDA-MB-231, MDA-MB-468 y MCF7. Las características de las líneas se detallan en la tabla 7.

**Tabla 7.** Características de las líneas celulares empleadas en este estudio.

Línea celular	Subtipo	Receptores expresados	Tipo de tumor	Edad	Medio de cultivo	Condiciones	Suplementos
HCC 1500	Basal	EGP2	CDI	32	PRMI	37 °C 5%CO <sub>2</sub>	10% SBF
HCC1937	Basal	EGP2	CDI	24	RPMI	37 °C 5%CO <sub>2</sub>	1% L-glu 1% Pir-Sod 10% SBF
HBL100	Basal	---	Normal	27	DMEM	37 °C 5%CO <sub>2</sub>	1% L-glu 10% SBF
Hs566.B(T)	---	---	Carcinoma	35	DMEM	37 °C 5%CO <sub>2</sub>	1% Pir-Sod 10% SBF
MDAMB-231	Basal	EGFR, TGFβ	Carcinoma	51	Leibovitz L-15	37 °C	10% SBF
MDAMB468	---	EGFR, TGFα	Carcinoma	51	Leibovitz L-15	37 °C	10% SBF
MCF7	Luminal	ER, IGF1P	CDI	69	DMEM	37 °C 5%CO <sub>2</sub>	1% Insulina 10% SBF

EGP2: Glicoproteína epitelial 2; EGFR: Receptor de factor de crecimiento epidérmico; TGFβ/α: Factor de crecimiento transformante β/α; ER: receptor de estrógenos; IGF1P: Proteína de unión al factor de crecimiento insulínico; RPMI: medio RPMI 1640; DMEM: *Modified Eagle's Medium*; SBF: suero bovino fetal; L-glu: L-glutamina; Pir-Sod: piruvato sódico.

Las líneas celulares procedentes de mujeres jóvenes, excepto Hs566.B(T), crecieron formando capas adheridas al frasco de cultivo. Todas ellas mostraron morfología epitelial a excepción de la línea Hs566.B(T), cuyas células mostraban una morfología alargada similar a un fibroblasto. Las líneas HBL100 y HCC1937 iniciaron su crecimiento formando agregados celulares en monocapa mientras que las líneas HCC1500 y Hs566.B(T) crecieron inicialmente como células aisladas, obteniendo un crecimiento en monocapa más tardío en el caso de HCC1500.

### 1.3. Datos obtenidos a partir de bases de datos online

Mediante búsqueda bibliográfica se detectaron varios conjuntos de datos de expresión de microRNAs y tumores de cáncer de mama; sin embargo, existen pocos estudios lo bastante grandes como para proporcionar suficiente número de mujeres jóvenes. Se decidió, por tanto, emplear los datos del estudio más grande con datos clínicos completos encontrado. Los datos usados son originarios del grupo *Cambridge Breast Unit* del Doctor Carlos Caldas (Dvinge et al., 2013). En este trabajo se estudia la desregulación de microRNAs en 1302 tumores de cáncer primario e invasivo de mama, además de 116 muestras de tejido adyacente tumoral y 28 líneas de cáncer de mama. Los datos están depositados en *European Genome Archive* ([www.ebi.ac.uk/ega](http://www.ebi.ac.uk/ega)) en el EBI bajo el código de acceso EGAD00010000438.

Así pues, se solicitaron los datos crudos a este organismo, obteniéndose:

- \* Los datos demográficos, fenotípicos y patológicos.
- \* Los datos de la expresión ya normalizados de los microRNAs (plataforma Agilent).

Estas muestras se obtuvieron con el consentimiento del comité de ética de Cambridge y Vancouver, donde se realizó el análisis molecular de las muestras, depositadas en los bancos de tumores respectivos a raíz de un estudio inicial METABRIC (*Molecular Taxonomy of Breast Cancer International Consortium*) (176).

Los datos de expresión se obtuvieron mediante matrices donde hibridan microRNAs de la compañía Agilent Technologies Inc. La plataforma consta de ocho matrices individuales por

portaobjetos. Cada serie contiene un total de 10.226 sondas únicas, de las cuales 3.388 son dirigidas contra secuencias de miRNAs maduros canónicas (2.645) o putativas (743) identificados a partir de secuenciación de alto rendimiento, dirigidas a un total de 1406 miRNAs o estrechamente relacionados. Las sondas restantes fueron dirigidas contra otros RNAs no codificantes, incluyendo rRNAs, snRNAs, snoRNAs, piRNAs y otros (177). Sin embargo, finalmente, los datos depositados del estudio contemplan solamente datos de expresión de 853 microRNAs, ya filtrados por su buena señal y normalizados, de 1.440 muestras (1.302 tumores de cáncer de mama, 116 muestras normales, 28 muestras de líneas celulares de cáncer de mama).

Para nuestro estudio, se emplearon los datos de expresión de microRNAs de pacientes de cáncer de mama, para evaluar las diferencias en su expresión según la edad al diagnóstico de las pacientes.

## 2. MATERIAL COMERCIAL

---

### 2.1. *Kits comerciales empleados*

Se utilizaron los siguientes kits comerciales siempre siguiendo las especificaciones del fabricante:

- \* CoolSafe™ 110, ScanVac (LaboGene™ ApS, Lyngø, Denmark).
- \* RecoverAll™ Total Nucleic Acid Isolation Kit de ambion® (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* mirVana™ miRNA Isolation Kit de ambion® (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* High-Capacity cDNA Reverse Transcription kit (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® PreAmp Master Mix (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® Universal Master Mix no UNG (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® MicroRNA Assays (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® Gene Expression Assays (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* LabChip® Agilent RNA 6000 Nano Assay (Agilent Technologies Inc., Santa Clara, CA, USA).

- \* LabChip® Agilent Small RNA Assay (Agilent Technologies Inc., Santa Clara, CA, USA).
- \* FlashTag™ Biotin HSR RNA Labeling Kit (Affymetrix Ltd., Santa Clara, CA, USA).
- \* Affymetrix GeneChip® miRNA 2.0 (Affymetrix Ltd., Santa Clara, CA, USA).
- \* GeneChip® *Eukaryotic Hybridization Control Kit* (Affymetrix Ltd., Santa Clara, CA, USA).
- \* PAM50-Based Prosigna Breast Cancer Assay (Nanostring technologies, Seattle, WA, USA).
- \* HER2 IQFISH pharmaDx™ (Dako, Glostrup, Denmark).
- \* Dako EnVision™ System-HRP (DAB) (Agilent Technologies Inc., Santa Clara, CA, USA).

## 2.2. Sondas empleadas

Los cebadores y las sondas empleados en este trabajo están basados en los ensayos de expresión de TaqMan®. La información completa de las sondas empleadas está detallada en el Anexo I.

- \* TaqMan® MicroRNA Assays (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).
- \* TaqMan® Gene Expression Assays (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).

## 2.3. Anticuerpos empleados

- \* Anti-HER2. Anticuerpo policlonal de PharmaDX (Dako Corp.) Clon A0485.
- \* Anti-HER4p180. Anticuerpo monoclonal LS-C50208 - LSBio. (LifeSpan BioSciences, Inc.) Clon EP2270Y.

# 3. PROGRAMAS Y HERRAMIENTAS EMPLEADAS

---

## 3.1. Programas empleados

- \* Cluster Eisen software. [http://www.eisenlab.org/eisen/?page\\_id=7](http://www.eisenlab.org/eisen/?page_id=7)
- \* Cytoscape 3.2.1. (Cytoscape Consortium). [www.cytoscape.org](http://www.cytoscape.org)
- \* *Expert Software* v.B.01.02.SI136.(Agilent Technologies Inc., Santa Clara, CA, USA).
- \* Expression Console. (Affymetrix Ltd., Santa Clara, CA, USA).

- \* GeneChip® *Eukaryotic Hybridization Control Kit* (Affymetrix Ltd., Santa Clara, CA, USA).
- \* GraphPad Prism 6.0. (Graphpad Software Inc., La Jolla, CA, USA).
- \* Notepad++. <https://notepad-plus-plus.org>
- \* Paquete de Microsoft Office 2010.
- \* R-studio. <https://www.rstudio.com>
- \* R versiones i386, 3.0, 3.1.1 y 3.2.0. <https://cran.r-project.org/>
- \* Treeview Eisen software. [http://www.eisenlab.org/eisen/?page\\_id=7](http://www.eisenlab.org/eisen/?page_id=7)

### 3.2. *Herramientas online y bases de datos empleadas*

- \* **Diana TOOLS:** <http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index>
- \* **Doubling-Time software online:** <http://www.doubling-time.com>
- \* **EGA** (European Genome Archive): <http://www.ebi.ac.uk/ega>
- \* **EnrichR:** <http://amp.pharm.mssm.edu/Enrichr/enrich>
- \* **Ensembl:** <http://www.ensembl.org>
- \* **Gene Cards:** <http://www.genecards.org>
- \* **GEO** (Gene Expression Omnibus): <http://www.ncbi.nlm.nih.gov/geo>
- \* **KEGG:** Kyoto Encyclopedia of Genes and Genomes: <http://www.genome.jp/kegg>
- \* **miRBase:** The microRNA database: <http://www.mirbase.org>
- \* **miRTarBase:** <http://mirtarbase.mbc.nctu.edu.tw>
- \* **microT-CDS:** <http://www.microna.gr/microT-CDS>
- \* **PomeloII:** herramienta para el cálculo de diferencias estadísticas en archivos de expresión. <http://pomelo2.iib.uam.es>
- \* **REACTOME:** A curated pathway database: <http://www.reactome.org>
- \* **Statpages:** <http://statpages.org>
- \* **TargetScanHuman:** Prediction of microRNA targets (version 5.2) [http://www.targetscan.org/vert\\_50](http://www.targetscan.org/vert_50)
- \* **UNIPROT:** <http://www.uniprot.org/uniprot>

---

## 1. TRATAMIENTO DE LAS MUESTRAS DE TEJIDO BIOPSIADO PARA SU PREPARACIÓN Y CONSERVACIÓN

---

### *1.1. Tratamiento del tejido biopsiado para su congelación óptima*

Las muestras de tejido sólido se congelan en criomoldes de plástico incluidas en OCT (del inglés *optimal cutting temperature*), mediante congelación rápida por inmersión en isopentano previamente enfriado a  $-110^{\circ}\text{C}$  en congelador portátil (CoolSafe™ 110, ScanVac, LaboGene™ ApS, Lynge, Denmark). Finalmente las muestras se almacenan en congeladores mecánicos de  $-80^{\circ}\text{C}$  durante un tiempo indefinido.

### *1.2. Inclusión en parafina del tejido biopsiado*

En primer lugar y tras recibir de cirugía la pieza de tejido, se procede a su tallado. Este proceso consiste en hacer secciones de la muestra para su mejor fijación. Tras esto, se fijan las muestras por inmersión en un baño de formol tamponado al 4%. A continuación, las muestras se someten a un progresivo proceso de deshidratación que permite eliminar el agua existente en el tejido. Esta operación se lleva a cabo empleando baños sucesivos de 3 horas de duración, en alcoholes de graduación creciente 50%, 70%, 80%, 95% y alcohol absoluto. Tras la deshidratación las muestras se someten a un proceso de aclarado, en que se sustituye el agente deshidratante (etanol en nuestro caso) por una sustancia miscible en el medio de inclusión, el xileno. Para ello, se sumergen las muestras tisulares en 3 baños sucesivos en xileno de 30 minutos cada uno. Una vez completado este proceso, se procede a la infiltración de la muestra, que consiste en infiltrar completamente la muestra histopatológica en parafina líquida a una temperatura entre  $58^{\circ}$  y  $65^{\circ}\text{C}$ , dependiendo del punto de fusión de la misma. Para facilitar la completa impregnación y eliminación total de los restos de xilol se realizan dos baños sucesivos en diferentes recipientes.

### *1.3. Control histológico de las muestras*

Se realizó un corte adicional de  $3\text{-}5\ \mu\text{m}$  de cada una de las muestras, que se colocó sobre portas pretratados con poli-L-lisina. A estos cortes histológicos se les realizó una tinción hematoxilina-eosina (H&E), sirviendo así como control histológico de los casos a estudiar. De este modo, se determinó el porcentaje de celularidad tumoral presente en cada muestra analizada.

## 2. TRATAMIENTO DE CULTIVO CELULAR PARA SU CRECIMIENTO, MANTENIMIENTO Y CONSERVACIÓN

---

### 2.1. Cultivo de las líneas celulares

Las características de las líneas y las condiciones de cultivo se indican en la tabla 8. Las células se crecieron en cultivo siguiendo las condiciones recomendadas del proveedor. Cuando los cultivos llegaban a una confluencia del 80-90%, las células fueron tripsinizadas con tripsina al 0,05% y se dividieron en un ratio de 1:2 en medio fresco para permitir la expansión de los cultivos. Para la recogida del *pellet*, las células se pasaron a placas Petri de 10 mL y al alcanzar una confluencia del 95%, se recogieron las células mediante raspado con *Cell Scraper*. Seguidamente, las células recogidas se centrifugaron a 1.500 rpm durante 10 minutos y se guardaron a -80 °C en tanque de isopropanol hasta próximo uso.

### 2.2. Cálculo del tiempo de doblaje

Para calcular el tiempo de doblaje de las líneas celulares se hicieron recuentos de células a intervalos regulares en el pase 8. En el día 0 las células se cultivaron por triplicado en frascos de 25 cm<sup>2</sup> en sus correspondientes medios de cultivo con una población inicial de  $1 \times 10^5$  células. Se tiñeron las células con azul de Trypan y se hicieron recuentos en cada triplicado en cámara de Neubauer en intervalos de 24 h durante 10 días seguidos. A partir de los datos obtenidos se calculó el tiempo de doblaje mediante la herramienta online <http://www.doubling-time.com>.

## 3. EXTRACCIÓN DE MATERIAL GENÉTICO

---

### 3.1. Extracción de RNA a partir de tejido congelado

En el caso de las muestras que sufrieron el proceso de descongelación, éstas se incluyeron de nuevo en OCT, almacenándose de nuevo a -80 °C hasta su utilización. Posteriormente, se trataron exactamente igual que el material congelado óptimo.

Para la extracción de RNA total de tejido congelado en OCT almacenado a -80°C se partió de 10 criosecciones de 10 µm, realizadas en criostato a -20°C, y recogidas en un microtubo de 2 ml conservado en hielo seco, empleándose el kit comercial mirVana™ miRNA Isolation Kit de ambion® (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA). Este kit se basa en una extracción orgánica seguida de una inmovilización del RNA sobre filtros de fibra de vidrio (extracción en fase sólida) para purificar el RNA total, incluidos los miRNAs, de muestras tisulares.

En resumen, a cada tubo de 2 ml que contiene las criosecciones de tejido se le añadió una bolita de acero de 5 mm (QIAGEN, Hilden, Germany) y se dejó enfriar en hielo seco 5 minutos. La interrupción y homogenización del tejido se realizó empleando el equipo *TissueLyser* LT (QIAGEN, Hilden, Germany). Para ello, se colocaron los microtubos de 2 ml que contenían las



muestras con la bola de acero en el rotor del equipo de disrupción y se les añadieron 600µl de tampón de lisis del kit, iniciándose inmediatamente la disrupción y homogenización de las muestras durante 30 segundos a 50Hz. Seguidamente, se descartó la bola de acero, pasándose el lisado a un microtubo de 1,5 ml nuevo, al que se le añadieron 60µl de la solución miRNA Homogenate Additive del kit, dejándose la mezcla en hielo durante 10 minutos tras haberse agitado vigorosamente. A continuación, se añadieron 600µl de fenol ácido:cloroformo a pH 4,5 (premezclado con alcohol isoamílico en proporción 125:24:1 fenol:cloroformo:alcohol isoamílico), agitándose la mezcla durante 30 segundos y centrifugándose posteriormente durante 10 minutos a 10.000 x g a temperatura ambiente. Se transfirió la fase acuosa a un nuevo microtubo al que se le añadieron 750µl de etanol 100%. La mezcla se pasó por la columna, que contenía el filtro de fibra de vidrio (extracción en fase sólida), mediante centrifugación durante 15 segundos a 10.000 x g. Seguidamente se realizó un lavado con 700µl de la solución de lavado 1 y otros dos lavados de 500µl con la solución de lavado 2/3 suministradas en el kit. Finalmente se eluyó el RNA total (que incluía miRNAs) con 80µl de agua libre de RNasa precalentada a 95°C, almacenándose el RNA extraído a -80°C para su conservación, previamente se alicuotaron 2 µl para medir la concentración con NanoDrop, como control de la extracción y para evitar descongelaciones innecesarias que degradarían el RNA.

### 3.2. Extracción de RNA a partir de tejido incluido en parafina

Para la extracción de RNA total del tejido fijado en formaldehído e incluido en parafina, conservado a temperatura ambiente, se partió de 4 secciones de 20 µm, realizadas en microtomo a temperatura ambiente y recogidas en un microtubo de 1,5 ml, empleándose el kit comercial RecoverAll™ Total Nucleic Acid Isolation Kit de ambion® (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).

En resumen, a cada tubo de 1,5 ml que contiene las secciones de tejido se le añadieron 1000 µl de xilol 100%, agitándose vigorosamente y dejándose la mezcla durante 3 minutos a 50°C. A continuación se centrifugó la mezcla durante 3 minutos a 10.000 x g y se eliminó el xilol. Se repitió este paso con xilol para conseguir desparafinar mejor la muestra. Seguidamente, se realizaron 2 lavados de la muestra con etanol 100%, lo que permitió eliminar los restos de xilol y acelerar el secado de la muestra, que se dejó secar al aire durante 30 minutos a temperatura ambiente. A continuación, se añadieron 200 µl de tampón de digestión y 4 µl de proteasa, suministrados en el kit, incubándose la mezcla a 50°C durante una hora y, posteriormente, durante 15 minutos a 80°C. En este momento, se añadieron a cada muestra 240 µl de la solución del kit *Isolation Additive* y 500 µl de etanol 100%. La mezcla se pasó por columna, mediante centrifugación durante 30 segundos a 10.000 x g. Seguidamente se realizó un lavado con 700 µl de la solución de lavado 1 y otros dos lavados de 500 µl con la solución de lavado 2/3 suministradas en el kit. Tras los lavados, se añadieron 60 µl de la mezcla de DNasa en el centro del filtro de la columna y se dejó a temperatura ambiente durante 30 minutos. De nuevo, se realizó un lavado con 700 µl de la solución de lavado 1 y otros dos lavados de 500 µl con la solución de lavado 2/3 suministradas en el kit. Finalmente, se eluyó el RNA total (que incluía miRNAs) con 70 µl (2 x 35 µl) de agua libre de RNasa, almacenándose el RNA extraído a -80°C para su conservación tras alicuotar 2 µl para la medida de concentración.

### 3.3. Extracción de RNA a partir de pellet de células

Para la extracción de RNA total a partir de cultivo celular, se recogió con *scrapel* un *pellet* de entre  $10^2$  y  $10^7$  células, lavadas previamente con 1mL de PBS, y se pasaron a un tubo preparado para su congelación a  $-80^{\circ}\text{C}$  hasta el momento de la extracción. Se empleó el kit comercial miRVana™ miRNA Isolation Kit de ambion® (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).

Resumiendo, a cada pellet se le añade 600  $\mu\text{l}$  de solución de lisis y se vortea vigorosamente para provocar la disrupción celular, en caso necesario se ayudó con disrupción mecánica mediante un *TissueLyser* LT (QIAGEN, Hilden, Germany), en cuyo caso se homogeneizan durante 30 segundos a 50 Hz. Tras esto, se añadieron 60  $\mu\text{l}$  de *Homogenate Additive* incluido en el kit que se mezclaron concienzudamente, dejando la mezcla en hielo 10 minutos. Tras este tiempo se procede a la extracción orgánica, para lo cual se añadieron 600  $\mu\text{l}$  de mezcla fenol ácido:cloroformo a pH 4,5 y se centrifugó a  $10.000 \times g$  a temperatura ambiente durante 5 minutos, tras lo cual se transfirió la fase acuosa a un nuevo tubo (aproximadamente 600  $\mu\text{l}$ ). En este momento se añadieron 750  $\mu\text{l}$  de etanol al 100% y se pasó la mezcla por columna mediante centrifugación durante 15 segundos a máxima velocidad, descartando el líquido sobrante, hasta haber pasado todo el volumen. Seguidamente se realizó un lavado con 700  $\mu\text{l}$  de la solución de lavado 1 y otro lavado de 500  $\mu\text{l}$  con la solución de lavado 2/3, ambas suministradas en el kit. Por último, se eluyó el RNA retenido en el filtro con 35  $\mu\text{l}$  del *Elution Buffer* proporcionado (o agua libre de RNasas en su defecto) a  $95^{\circ}\text{C}$  centrifugando a máxima velocidad durante 30 segundos, tras lo cual se repiten estos pasos con otros 35  $\mu\text{l}$  para obtener un volumen total de 70  $\mu\text{l}$ , de los que se alicuotaron 2  $\mu\text{l}$  para medir su concentración y, el resto, se congeló a  $-80^{\circ}\text{C}$ .

## 4. TÉCNICAS DE CUANTIFICACIÓN Y MEDIDA DE LA CALIDAD DEL MATERIAL EXTRAÍDO

---

### 4.1. Cuantificación del RNA total por espectrofotometría de absorción

El RNA extraído se cuantificó por espectrofotometría de absorción UV-Vis empleando el equipo Nanodrop 2000 (Thermo Fisher Scientific Inc., Wilmington DE, USA). Con tan solo 1  $\mu\text{l}$  de RNA, este equipo determina la concentración y la pureza de la muestra.

La concentración se determina a partir de la absorbancia a 260 nm, de acuerdo a la Ley de Lambert-Beer:

$$A = \varepsilon \cdot I \cdot c$$

siendo:  $A$  = absorbancia

$\varepsilon$  = coeficiente de extinción molar [para ssRNA a 260 nm,  $\varepsilon = 0,025(\mu\text{g}/\text{ml})\cdot\text{1cm}^{-1}$ ]

$I$  = distancia en cm

$c$  = concentración molar

La pureza de la muestra se determina en base a la relación de absorbancias 260/280 y 260/230, que deben ser ambos próximos a 2 para RNA puro y para asegurar que no quedan restos de reactivos empleados en la extracción, respectivamente.

#### 4.2. Análisis de la integridad del RNA total

La calidad y concentración del RNA total extraído se determinó con el RNA 6000 *LabChip*® junto con el 2100 *Bioanalyzer* (Agilent Technologies).

La integridad del RNA total se determinó empleando el equipo Agilent 2100 *Bioanalyzer* y el *LabChip*® Agilent RNA 6000 *Nano Assay* (Agilent Technologies Inc., Santa Clara, CA, USA). Esta tecnología se basa en la electroforesis capilar y este ensayo permite analizar muestras de RNA total de concentración entre 25 y 500 ng/μl, empleando 1 μl de muestra. El resultado del análisis se puede visualizar como una electroforesis convencional, en la que se observan los fragmentos de RNA separados por tamaño, además de proporcionar perfiles o electroferogramas de cada una de las muestras. Esta tecnología aplica un algoritmo complejo que determina automáticamente un valor de RIN (*RNA Integrity Number*), con valores entre 1 y 10, para cada una de las muestras analizadas. El RIN considera todo el perfil electroforético, incluyendo presencia y ausencia de productos de degradación, resultando una medida de la integridad del RNA fiable. El valor de RIN es muy reproducible y nos permite hacer una valoración numérica de la integridad del RNA y establecer comparaciones entre muestras.

Los *LabChip*® se prepararon siguiendo las instrucciones del fabricante. Los reactivos se dejaron 30 minutos a temperatura ambiente antes de su utilización. El gel se preparó utilizando una columna de 550 μl que se centrifugó a 1500 x g durante 10 min para el filtrado del gel.

Se desnaturalizó 1 μl del marcador de peso molecular (RNA 6000 *ladder*® Ambion) y 1 μl RNA de cada una de las muestras a analizar durante 2 min a 70°C y se introdujeron en los pocillos correspondientes. Finalmente el *LabChip*® se agitó en vórtex durante 1 min a máxima potencia y se introdujo en el *Agilent 2100 Bioanalyzer*.

El Agilent 2100 *Bioanalyzer* permitió visualizar los resultados en un ordenador mediante *Expert Software* v. B.01.02.SI136. El bioanalizador usa fluorescencia entre 670 nm y 700 nm para determinar la calidad e integridad del RNA total. Para cada muestra, el software calculó el cociente del área de los picos de RNA ribosomales (28S/18S), y el resultado se visualizó mediante un electroferograma. El cociente debe encontrarse alrededor de 2 ya que el peso molecular del 28S es el doble al del 18S.

#### 4.3. Análisis de la integridad del RNA de pequeño tamaño

La integridad del RNA de pequeño tamaño se determinó empleando el equipo *Agilent 2100 Bioanalyzer* y el *Agilent Small RNA Assay* (Agilent Technologies Inc., Santa Clara, CA, USA). Esta técnica permitió evaluar el contenido de especies de RNA de pequeño tamaño, en tan solo 10 ng de RNA total, para cada una de las muestras. Se trata de una alternativa más rápida y mucho más sensible que los geles de agarosa o poliacrilamida convencionales para el análisis y la detección de pequeños fragmentos de RNA en el rango entre 6 y 150 nucleótidos. El perfil o electroferograma de cada muestra permite visualizar fácilmente la región correspondiente a los miRNAs y, además, este ensayo determina el porcentaje y la concentración de miRNAs presentes en cada muestra. La metodología seguida es la misma que la empleada en el análisis de la integridad del RNA total arriba detallada.

## 5. TÉCNICAS DE ANÁLISIS DE LA EXPRESIÓN DE MATERIAL GENÉTICO (RNA Y MIRNA) MEDIANTE PCR CUANTITATIVA

---

La evaluación de la expresión de los genes se llevó a cabo mediante la cuantificación relativa de los mRNAs de dichos genes por PCR cuantitativa en tiempo real (qRT-PCR). Se trata de una reacción en cadena de la polimerasa (PCR) convencional, en la que se mide, en tiempo real, la cantidad de producto amplificado por cada ciclo de la PCR, obteniendo así una curva de amplificación de la que nos valemos para cuantificar la cantidad inicial de producto.

La técnica de la PCR se desarrolló por primera vez por Mullis et al, 1986, y desde entonces ha sido una de las técnicas básicas en laboratorios de todo el mundo (178). El objetivo básico de la misma consiste en la amplificación específica de nucleótidos a partir de un molde de DNA. Esta técnica se basa en dos conceptos: la desnaturalización y renaturalización del DNA dependiente de la temperatura, y la resistencia a altas temperaturas de la polimerasa del organismo *Thermus aquaticus* (Taq). La Taq polimerasa (como el resto de polimerasas) se encarga de sintetizar una secuencia de nucleótidos a partir de una hebra molde, actualmente se emplean diversas polimerasas resistentes a altas temperaturas. La PCR consiste en aumentar y disminuir la temperatura alternamente en ciclos, en los siguientes pasos:

- \* **Desnaturalización.** La temperatura sube por encima del nivel de desnaturalización del DNA. Ambas hebras del DNA se separan.
- \* **Hibridación.** La temperatura disminuye a un nivel en el que permite la unión de la Taq polimerasa y de los cebadores a las hebras del DNA desnaturalizadas, que sirven de molde.
- \* **Elongación:** La temperatura desciende permitiendo la síntesis de DNA de doble cadena por parte de la Taq polimerasa. Una vez sintetizado la Taq polimerasa se suelta de la hebra.

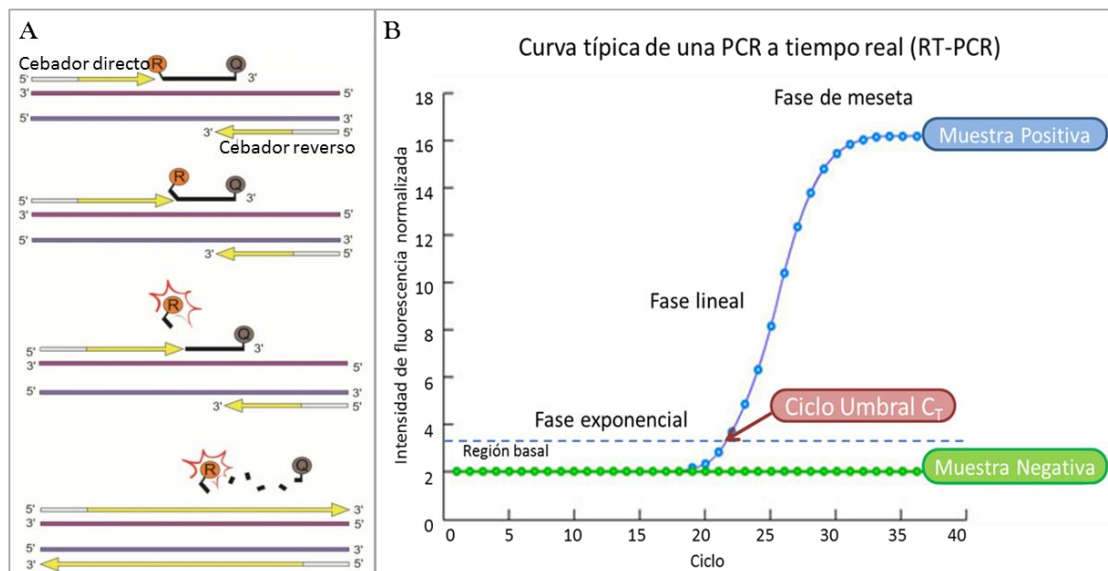
Tras cerrarse un ciclo de temperatura se inicia otro con los mismos pasos, esta vez, además de con la hebra completa de DNA como molde, cuenta con los fragmentos cortos de la secuencia específica de nucleótidos como sustrato, cuya cantidad aumentará de manera exponencial.

La especificidad de secuencia se consigue empleando unos cebadores complementarios a la secuencia nucleotídica problema, que se unen en el paso 2 del ciclo de temperatura y permiten el inicio de la polimerización del DNA por parte de la Taq. Obtenemos como producto final el fragmento de DNA de secuencia específica amplificado varios órdenes de magnitud, permitiéndonos detectar la presencia de pequeñas cantidades de dicha secuencia.

La qRT-PCR (PCR cuantitativa a tiempo real) es una variación de esta técnica, descrita en primer lugar por Higuchi et al. 1993, en la cual se detecta la cantidad de producto amplificado (amplicón) a tiempo real, es decir, en cada ciclo (179). Esto se consiguió mediante la incorporación en la reacción de un agente intercalante fluorescente (BrEt), que permitía detectar la cantidad de producto sintetizado a tiempo T, mediante la acumulación de fluorescencia. La cinética de la acumulación de la fluorescencia con cada ciclo depende directamente de la cantidad inicial de DNA, pudiendo extrapolar así dicha cantidad.

En la actualidad se emplean tanto agentes de unión al DNA de doble cadena (como el SYBR Green), como sondas fluorescentes unidas a los cebadores, que sólo emiten fluorescencia al tener lugar la elongación en cada ciclo de la PCR, aumentando así intensidad de fluorescencia al aumentar la cantidad de producto amplificado que es sintetizado. Al detectar únicamente la secuencia amplificada es un método mucho más específico y preciso.

En el análisis de expresión tanto de mRNAs como de miRNAs en este trabajo se han empleado sondas TaqMan®. Este método se caracteriza por un fluorocromo unido a una sonda específica con secuencia complementaria al fragmento que queremos amplificar, al igual que un elemento extintor de la fluorescencia, al producirse la fase de elongación de la PCR, el fluorocromo es escindido de la sonda y separado del extintor, emitiendo así su fluorescencia (Figura 9A). Para llevar a cabo la PCR y detectar la fluorescencia se requiere de instrumental adaptado para tal fin. Para la cuantificación de RNA mediante esta técnica, se requiere de un paso previo adicional, la síntesis mediante la retrotranscripción del DNA complementario (cDNA) empleando el RNA como molde, por parte de una transcriptasa reversa.



**Figura 9. A. Método empleado por las sondas TaqMan para reportar la fluorescencia.** R: fluorocromo o *reporter*; Q: extinguidor de la fluorescencia o *quencher*. **B. Representación de la curva típica obtenida en una RT-PCR.** El eje X muestra el ciclo de la reacción de PCR en el que se encuentra y en el eje Y viene dada la intensidad de fluorescencia emitida por el fluorocromo unido a la sonda, directamente relacionada con la cantidad de producto amplificado. La curva en verde muestra el patrón de una muestra negativa y correspondería con una expresión basal, mientras que la curva azul muestra el patrón típico de una muestra amplificada.

La cinética de la reacción sigue una curva sigmoidea típica de una función logística, por lo que obtendremos por cada muestra una curva similar a la representada en la Figura 9B. La curva consta de 3 fases, una fase exponencial, en la que la cantidad de producto aumenta de manera exponencial, hasta que la enzima alcanza su rendimiento máximo y se pasa a la fase lineal, donde la cinética de la enzima sigue un comportamiento lineal, de modo que se puede calcular su pendiente. La fase de meseta o estacionaria es la última fase, en la que la cantidad de producto llega al nivel de saturación de la enzima y se caracteriza por un aumento lento. El ciclo umbral o  $C_T$  es aquel ciclo en el que la curva corta con un nivel de intensidad llamado umbral, que marca un límite por encima del cual una muestra se considera significativamente positiva.

## 5.1. *Transcripción reversa o síntesis de cDNA.*

La síntesis del DNA complementario (cDNA) se llevó a cabo a partir de cantidades idénticas de RNA total (200ng), partiendo de una dilución de RNA total de 50 ng/μl.

### **Transcripción reversa de mRNA**

Para el análisis de expresión de genes se realizó la síntesis de cDNA mediante el kit *High-Capacity cDNA Reverse Transcription* (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA). En este caso la síntesis de cDNA se basó en el uso de cebadores aleatorios. Para cada reacción de 10 μl, 200 ng de RNA total se incubaron con 1 μl de tampón RT 10X, 0,4 μl de dNTPs 100mM, 1 μl de cebadores aleatorios 10X, 0,5 μl de MultiScribe™ Reverse Transcriptase 50 U/μl, 0,5 μl de inhibidor de RNasa y agua libre de nucleasas durante 10 minutos a 25°C, seguidos de 120 minutos a 37°C y 5 minutos a 85°C en el termociclador vapo.protect™ (Eppendorf, Hamburg, Germany) en placa de 96 pocillos.

### **Transcripción reversa de microRNA**

Para el análisis de RNAs de pequeño tamaño, se empleó el kit comercial TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA), para el análisis de los sRNAs. En este caso, la síntesis de cDNA es específica para cada miRNA o snoRNA maduro, por ello se empleó una mezcla de cebadores para realizar una única síntesis de cDNA por muestra, de acuerdo al protocolo *Protocol for Creating Custom RT and Preamplification Pools using TaqMan® MicroRNA Assays* (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA).

Se preparó una mezcla de los cebadores específicos de miRNA/snoRNA a estudiar a una concentración de 0,5X cada. Para cada reacción de 15 μl, 200 ng de RNA total se incubaron con 6 μl de la mezcla de cebadores, 0,3 μl de dNTPs 100 mM, 3 μl de MultiScribe™ Reverse Transcriptase 50 U/μl, 1,5 μl de tampón RT 10X y 0,2 μl de inhibidor de RNasa 20 U/μl durante 30 minutos a 16°C, seguidos de 30 minutos a 42°C y 5 minutos a 85°C en el termociclador vapo.protect™ (Eppendorf, Hamburg, Germany) en placa de 96 pocillos.

## 5.2. *Preamplificación*

La preamplificación consiste en realizar una primera fase de amplificación antes de llevar a cabo la qRT-PCR, de éste modo se aumenta la cantidad de base del sustrato y se facilita su posterior detección en los casos en que éste sea limitado.

Se llevó a cabo en los ensayos en los que se obtuvo sistemáticamente una baja detección de las mismas.

Para ello se preparó una mezcla a concentración 0,01 de las sondas que iban a ser empleadas en la qRT-PCR, 3 ul de esa mezcla y 6 ul de TaqMan® PreAmp Master Mix se añadieron a 2,5 ul de la muestra de cDNA obtenida de la reacción de transcripción reversa en placa de 96 pocillos. Las preparaciones se incubaron en termociclador 2 minutos a 50°C, 10 minutos a 95°C y 50 ciclos de desnaturalización a 95°C 15 segundos y de amplificación a 60°C 1 minuto, y se repitió 14 ciclos. Posteriormente se diluyó el producto de la preamplificación en una relación 1:10 previamente a la realización de la qRT-PCR para no saturar la reacción de amplificación.

### 5.3. PCR cuantitativa en tiempo real (qRT-PCR)

Los análisis de recuperación de RNAs de pequeño tamaño (miRNA y snoRNA) y mRNAs se llevaron a cabo mediante PCR cuantitativa en tiempo real (qRT-PCR) empleando el equipo Applied Biosystems™ 7900HT Fast Real-Time PCR system con el programa de análisis 7900HT Fast system SDS v2.4, y los kits TaqMan® MicroRNA Assays y TaqMan® Gene Expression Assays (Applied Biosystems™ by Life Technologies™, Carlsbad, California, USA), respectivamente.

Para cada reacción de 10 µl, se mezcló 1 µl del cDNA sintetizado con 0,5 µl del correspondiente TaqMan® Assay 20X, 5 µl de la correspondiente TaqMan® Universal Master Mix 2X y 3,5 µl de agua libre de nucleasas. Las condiciones de amplificación fueron las siguientes: 2 minutos a 50 °C, 10 minutos a 95 °C y 50 ciclos de desnaturalización a 95 °C 15 segundos y de amplificación a 60 °C 1 minuto. Se realizó triplicado de PCR para cada una de las muestras en placa de 384 pocillos. Cada placa se diseñó de modo que incluía el mismo número de ensayos junto con el control endógeno adecuado en cada caso, así como una representación equitativa de muestras de casos y controles para evitar posibles sesgos de metodología.

### 5.4. Método $2^{-\Delta\Delta C_t}$ para obtener el valor de expresión relativa

Este método derivativo asume ciertas suposiciones además de un diseño experimental concreto, previamente descrito en Livak et al (180). La cantidad inicial de moléculas normalizada mediante un gen de referencia y en relación a una muestra control viene dada por:

$$2^{-\Delta\Delta C_t}$$

Como se ha explicado anteriormente, la reacción de amplificación sigue una cinética sigmoideal, la ecuación que describe la amplificación en su base exponencial por PCR viene dada por:

$$X_n = X_0 (1 + E_X)^n$$

Donde:

$X_n$  = Número de moléculas del amplicón a ciclo n de la reacción.

$X_0$  = Número inicial de moléculas en la muestra de interés.

$E_X$  = Eficiencia de la amplificación en el segmento de la curva de interés.

$n$  = Número de ciclos.

El ciclo umbral ( $C_T$ ) indica el número de ciclos requerido para que la cantidad de producto amplificado alcance un umbral fijo siguiendo esta ecuación:

$$X_T = X_0 (1 + E_X)^{C_{T,X}} = K_X$$

Donde:

$X_T$  = Número de moléculas en el ciclo umbral.

$C_{T,X}$  = Ciclo umbral para una amplificación dada.

$K_X$  = Constante dada en la fase exponencial de la amplificación.

Tendríamos una ecuación similar para la amplificación del mRNA del gen de referencia o normalización en cada una de las muestras:

$$R_T = R_0 (1 + E_R)^{C_{T,R}} = K_R$$

Siendo:

$R_T$  = Número de moléculas de mRNA del gen de referencia en el ciclo umbral.

$R_0$  = Número inicial de moléculas de mRNA del gen de referencia.

$E_R$  = Eficiencia de la amplificación del gen de referencia.

$C_{T,R}$  = Ciclo umbral para la amplificación del gen de referencia.

$K_R$  = Constante dada en la fase exponencial del gen de referencia.

Dividiendo  $X_T$  por  $R_T$ , obtenemos la expresión:

$$\frac{X_T}{R_T} = \frac{X_0 (1 + E_X)^{C_{T,X}}}{R_0 (1 + E_R)^{C_{T,R}}} = \frac{K_X}{K_R} = K$$

Al asumir que la eficacia de la amplificación problema y del gen de referencia es la misma, tenemos que:

$$E_X = E_R = E$$

$$X_N (1 + E)^{\Delta C_T} = K$$

Donde:

$X_N$  = cantidad de molécula problema normalizada por muestra  $X_0 / R_0$

$\Delta C_T$  = la diferencia entre el ciclo umbral del gen problema y el ciclo umbral para el gen de referencia ( $C_{T,X} - C_{T,R}$ ).

Modificando esto en la expresión, obtenemos:

$$X_N = K(1 + E)^{-\Delta C_T}$$

El paso final consiste en dividir  $X_N$  de una muestra ( $q$ ) por  $X_N$  de la muestra empleada para la calibración ( $cb$ , muestra control):

$$\frac{X_{N,q}}{X_{N,cb}} = \frac{K(1 + E)^{-\Delta C_{T,q}}}{K(1 + E)^{-\Delta C_{T,cb}}} = (1 + E)^{-\Delta \Delta C_T}$$

Siendo:

$$\Delta \Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$$

Calculamos la expresión relativa, reemplazando en la fórmula:

$$2^{-\Delta \Delta C_T}$$

En la práctica, en primer lugar calculamos la media aritmética de los  $C_T$ :

$$\bar{M}_{C_T} = \frac{C_{T,1} + C_{T,2} + \dots + C_{T,n}}{n} = \frac{\sum_{i=1}^n C_{Ti}}{n}$$

Donde  $n$  es el número de réplicas y  $C_{T,i}$  cada uno de los valores de  $C_T$  replicados por muestra. Es útil calcular también la desviación estándar de los ciclos umbral ( $SD_C$ ), para garantizar una media confiable y una variación entre réplicas mínima, la  $SD_C$  deberá ser menor a 0,1, correspondiendo este valor a un coeficiente de variación ( $C_v$ ) del porcentaje. Siendo:



$$SD_{Ct} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (C_{T,i} - \bar{M}_{C_T})^2}$$

Dónde:

$n$  = número de réplicas de una muestra.

$$C_V = \frac{SD_{Ct}}{\bar{M}_{C_T}} 100$$

Una vez obtenidos los valores medios de  $C_T$ , para normalizar los valores, calculamos la diferencia por muestra entre el  $C_T$  propio y el  $C_T$  del gen de referencia correspondiente a la misma muestra problema.

$$\Delta C_T = C_{T,X} - C_{T,R}$$

Una vez obtenidos los valores de  $C_T$  normalizados por muestra ( $q$ ) ( $\Delta C_{T,q}$ ), se obtiene la diferencia de cada muestra con el valor de  $C_T$  normalizado del calibrador ( $\Delta C_{T,cb}$ ) (el calibrador consiste en una muestra con condición nula, es decir control, de la que nos servimos para establecer un límite basal de referencia).

$$\Delta \Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$$

Este valor será el que deberemos sustituir en la ecuación derivada de la expresión relativa, obteniendo así los valores de expresión del gen problema:

$$2^{-\Delta \Delta C_T}$$

Mediante este método se estimó la expresión relativa de todas las muestras en todos los ensayos, teniendo en cuenta los casos en los que la SD de valores triplicados excedía de 0,1 y, además, uno de los valores triplicados estaba desviado claramente de los otros dos (atribuyéndose a un error en la qRT-PCR), éste se eliminaba dejando los dos restantes como válidos. Con estos valores se realizaron los análisis estadísticos pertinentes. Para el tratamiento de los datos se emplearon los programas SDS 2.4 y RQ manager (Life Technologies), así como el procesador de hojas de cálculo Microsoft Office Excel 2010.

## 6. ANÁLISIS DE LA EXPRESIÓN GLOBAL DE MIRNAS MEDIANTE LA PLATAFORMA AFFYMETRIX

Un chip de RNA (del inglés *RNA microarray*) es una superficie sólida a la cual se une una colección de fragmentos de RNA. Las superficies empleadas para fijar el RNA son muy variables y pueden ser de vidrio, plástico e incluso de silicio. Los chips de RNA se usan para analizar la expresión diferencial de genes, y se monitorean de manera simultánea los niveles de miles de ellos. Su funcionamiento consiste, básicamente, en medir el nivel de hibridación entre la sonda específica y la molécula diana, que se indica generalmente mediante fluorescencia y a través de un análisis de imagen, lo cual indica el nivel de expresión del gen mediante su RNA mensajero. Del mismo modo y bajo el mismo principio, se puede analizar la cantidad de expresión de miRNAs mediante chips con sondas específicas para los mismos (181).

### 6.1. *Hibridación del RNA total de las muestras en chip de Affymetrix Genechip® miRNA 2.0*

En este estudio se empleó el chip de Affymetrix Genechip® miRNA 2.0 (Affymetrix Ltd., Santa Clara, CA) para medir la expresión de RNA no codificante. Con este fin, se diluyeron todas las muestras de RNA total a una concentración final homogénea, 40ng/μl en un volumen de 10μl. En el ensayo de hibridación se emplearon 8μl, es decir un total de 320ng. Las 37 muestras se marcaron con biotina y se hibridaron durante 18 horas (según el protocolo estándar establecido por el fabricante), cada muestra en un chip GeneChip® miRNA 2.0 Array (Affymetrix, Santa Clara, CA, USA), que comprende un total de 4553 secuencias de RNA de pequeño tamaño, incluyendo 1105 secuencias de microRNAs.

### 6.2. *Marcaje del RNA con Biotina*

Para el marcaje de las muestras de RNA se empleó el kit comercial FlashTag™ Biotin HSR RNA Labeling Kit (Affymetrix Ltd., Santa Clara, CA). Brevemente, a 8 μl de RNA total se le añadieron 2μl de *RNA Spike Control Oligos* proporcionado en el kit, a esta mezcla se le añadieron 5μl del Poli(A) Tailing Master Mix (que contiene Tampón de reacción 10X, Cloruro de Magnesio 25mM y 1μl de ATP Mix diluido 1:500 en tampón Tris 1mM) se mezcló suavemente sin vortear, y se incubó a 37°C durante 15 minutos para permitir la unión de las colas poli(A).

Para la reacción de ligamiento, las muestras fueron centrifugadas y se les añadió a cada muestra 4μl de FlashTag Biotin HSR Ligation Mix 5X y 2μl de ligasa T4 DNA, por separado para evitar auto-ligamientos, se mezcló suavemente sin vortear y se centrifugó, tras lo cual se incubó durante 30 minutos a temperatura ambiente. La reacción se detuvo añadiendo 2,5μl de HSR Stop Solution, del mismo kit. Tras mezclar y centrifugar los 23,5μl de muestra ligada se alicuotaron 2,5μl que sirvieron para comprobar la eficacia de reacción mediante ELOSA QC Assay. La muestra restante se almacenó a -20°C hasta el momento de hibridación en el chip, sin superar nunca las dos semanas.

### 6.3. *Hibridación del RNA con las sondas de RNA de pequeño tamaño incluidas en el chip*

Para la reacción de hibridación se preparó previamente el horno de hibridación de Affymetrix a 48°C, ajustando la agitación a 60rpm. Los chips se dejaron a temperatura ambiente durante 15 minutos y fueron marcados adecuadamente. Se colocó una punta de pipeta sin filtro en la sección superior derecha del chip para permitir una correcta inyección de la mezcla de hibridación y evitar burbujas.

Se preparó la mezcla de hibridación con los siguientes reactivos proporcionados, en este orde: 50 μl de Hybridization Mix 2X, 15μl de Formamida al 27,5%, 10μl de Dimetilsulfóxido (DMSO), 5 μl de la mezcla *Eukaryotic Hybridization Controls 20X*, proporcionada en GeneChip® Eukaryotic Hybridization Control Kit (Affymetrix Ltd., Santa Clara, CA), que previamente se incubó a 65°C durante 5 minutos, 1,7μl de oligos control B2 a una concentración 3nM (*Control Oligo B2*). Las cantidades mencionadas son por muestra, para la preparación de la mezcla se adaptaron a la cantidad de muestras a preparar, en nuestro caso 37, calculando un margen de error del 10%. Se añadieron 81,7μl de la mezcla de hibridación a 21,5 μl de la muestra de RNA marcada con biotina obteniendo un total de 103,2μl, que se incubaron a 99°C durante 5 minutos

seguidos de otros 5 minutos a 45°C. Se inyectaron 100µl de la mezcla preparada a cada chip, uno por cada muestra, tras lo cual se quitó la pipeta previamente introducida para facilitar la salida del aire y se cubrió con Tough-Spots® para minimizar la evaporación. Los chips se colocaron entonces en el horno de hibridación, previamente calentado a 48°C, donde se incubaron en agitación a 60 rpm durante 18 horas.

#### 6.4. Lavados y tinción del chip

Una vez pasado el tiempo de hibridación se pasó al lavado y revelado de los chips. Para ello se extrajo la mezcla de hibridación restante en el chip y se sustituyó por el tampón *Array Holding* dejándolos enfriar a temperatura ambiente. Tras lo cual se preparó la estación de lavado *Fluidics Station 450*, que se empleó para los lavados y revelado de los chips siguiendo el protocolo establecido para chips de formato 100. Brevemente consiste en 10 ciclos de lavado con tampón de lavado A (*Wash Buffer A*) a 30°C, seguido de 6 ciclos con tampón de lavado B (*Wash Buffer B*) a 50°C. Para el revelado se tiñó la sonda con *Stain Cocktail 1* durante 5 minutos a 35°C. Tras la tinción se lavó la muestra en 10 ciclos con tampón de lavado A nuevamente (*Wash Buffer A*) a 30°C, y se vuelve a incubar la muestra con *Stain Cocktail 2* a 35°C durante 5 minutos, seguida de otros 5 minutos en la *Stain Cocktail 1* a 35°C. Por último se lavó la muestra en 15 ciclos con tampón de lavado A (*Wash Buffer A*) a 35°C y se llenó el chip con *Array Holding Buffer*.

#### 6.5. Escaneado del chip y detección de la intensidad de fluorescencia

Tras los lavados y la tinción de las sondas se procedió al escaneado del chip mediante el GeneChip Scanner GCS3000 con autocargado (Affymetrix Ltd., Santa Clara, CA) que registró las imágenes con ayuda del programa proporcionado Affymetrix® GeneChip® Command Console® (AGCC) 4.0 (Affymetrix Ltd., Santa Clara, CA), que computó dichas imágenes, guardadas en un archivo con extensión tipo .DAT, y las transformó a valores de intensidad numéricos y analizables, almacenándolos en archivos tipo .CEL. Los archivos fueron depositados en la base de datos GEO (*Gene Expression Omnibus*) con código de acceso GSE48088 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48088>).

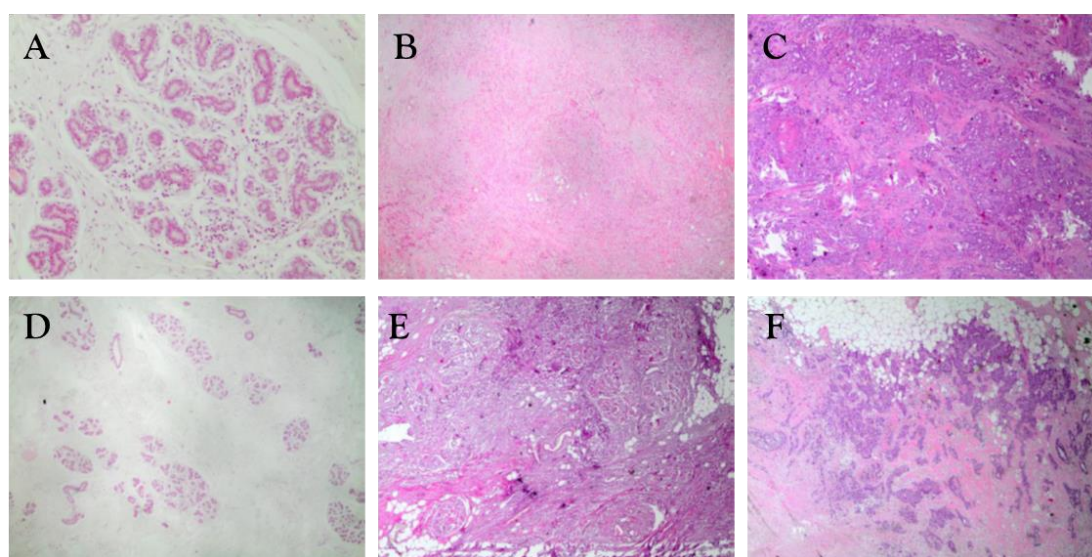
## 7. TÉCNICAS DE ANÁLISIS DE LA EXPRESIÓN DE PROTEÍNA IN SITU

---

### 7.1. Construcción de los Tissue-Microarrays

La construcción de *Tissue-MicroArray* (TMA) permite incluir en único un bloque de parafina (bloque receptor) cilindros de tejido parafinado de múltiples pacientes (bloques donadores), pudiendo así realizar análisis inmunohistoquímico de un gran número de muestras a la vez. Para este trabajo se construyeron 5 TMAs en los que se incluyeron dos cilindros de tejido tumoral duplicado de 90 pacientes, además de varias muestras de tejido normal de las pacientes empleado como control, con la finalidad de comprobar el correcto funcionamiento de la técnica. Del mismo modo se incluyeron en los TMA muestra de tejido normal procedente de mujeres libres de cáncer, para descartar tinción inespecífica. Cada uno de los TMA fue posteriormente cortado en secciones de 3µm de grosor, cada corte fue montado sobre un portaobjetos y embebido completamente en parafina para su almacenamiento.

De cada una de las muestras que se incluyó en este estudio se obtuvo un bloque de tejido parafinado. Para seleccionar la región del bloque más adecuada en cada uno de ellos se realizó, al microtomo, un corte de unos 4µm de grosor que se tiñó con hematoxilina-eosina y fue examinado y marcado por un patólogo (zona tumoral para los bloques a estudio y zona normal para los bloques control) (Figura 10). La misma región seleccionada en la tinción fue seleccionada en el bloque de parafina. Los tejidos se incluyeron en el bloque receptor mediante una aguja, se perforó parcialmente el bloque de parafina virgen (receptor) dejando un orificio en el que se incluyó un cilindro de tejido que previamente se extrajo, también mediante aguja, de uno de los bloques donadores. Este paso se repitió para cada una de las muestras que se incluyeron. Este proceso se realizó utilizando un Tissue-Arrayer (Manual Tissue Microarrayer, MTA1675. Beecher Instrument Inc, Wisconsin, USA).



**Figura 10. Cortes teñidos con Hematoxilina-eosina para el control histológico del porcentaje de celularidad tumoral. A y D, tejido mamario sano; B y E, tejido mamario tumoral procedente de mujer joven; C y F, tejido mamario tumoral procedente de mujer mayor.**

## 7.2. Estudio inmunohistoquímico de las muestras

Tanto la técnica inmunohistoquímica (IHC) como la de hibridación *in situ* fluorescente (FISH) se realizaron sobre cortes de 3µm de grosor, obtenidos de los TMA previamente realizados, recogidos sobre portaobjetos recubiertos de poli-L-lisina.

Para la tinción IHC, el tejido fue desparafinado mediante lavados sucesivos con xilol y posteriormente rehidratado mediante lavados consecutivos en un gradiente de concentraciones decrecientes de etanol: etanol absoluto, etanol al 95%, 80% y por último al 70%, se finalizó con un lavado con agua destilada.

Para el desenmascaramiento antigénico las muestras fueron sumergidas en tampón de desenmascaramiento y sometidas a presión mediante autoclave o incubadas en un baño a altas temperaturas, según el caso. Tras esto, las muestras fueron enfriadas a temperatura ambiente y se bloqueó la actividad de la peroxidasa endógena mediante incubación con peróxido de hidrógeno al 3% con azida sódica (NaN<sub>3</sub>) a 15mM durante 5 minutos a temperatura ambiente. Tras un lavado concienzudo con tampón salino Tris 1X (TBS), las secciones de tejido fueron incubadas con el anticuerpo primario durante 1 hora o 30 minutos, según el caso, tras lo cual, se volvieron a lavar con TBS 1X y se incubaron con el anticuerpo secundario por 30 minutos (EnVision™ FLEX anti rabbit, Dako, Glostrup, Denmark). Para la tinción se empleó el cromógeno

Diaminobenzidina (DAB) al 5% (Dako, Glostrup, Denmark) además de hematoxilina ácida para contrastar. Una vez revelada la tinción, se deshidrató la muestra mediante lavados sucesivos en xilol tras etanol en concentración creciente (70%, 80% y 95%) hasta etanol absoluto, una vez deshidratada la muestra, se añadió al portaobjetos medio de montaje Entellan® y se aisló con un cubreobjetos.

Las secciones fueron examinadas mediante microscopía y la intensidad de la tinción fue calificada mediante un patólogo en una escala relativa (0 = tinción negativa, 1+ = baja tinción, menor al 25% de las células tumorales, 2+ = tinción media, aproximadamente el 50%, 3+ = tinción elevada, el 100% de las células tumorales se encuentran teñidas), incluyéndose la localización de la tinción (membrana, citoplasmática o nuclear). En el caso de la cuantificación del Ki67 se dio como un porcentaje de células teñidas respecto del total de células tumorales.

La determinación inmunohistoquímica de HER2 se realizó mediante HercepTest™ (Dako, Glostrup, Denmark) que incluye el anticuerpo *Rabbit Anti-Human HER2* y el anticuerpo secundario *Visualization Reagent*, que contiene inmunoglobulinas de cabra y anti-inmunoglobulinas de conejo aisladas por afinidad. Se siguieron las recomendaciones del fabricante y se empleó un *Autostainer Immunostaining System*. El desenmascaramiento antigénico se hizo por calor en baño a 95°C en tampón citrato 40 minutos. El anticuerpo se incubó sobre el corte por 30 minutos.

Para la tinción de HER4 se empleó el anticuerpo monoclonal Anti-ERBB4/HER4 (phospho-Tyr1162, clone EP2270Y) (LifeSpan Bioscience, Inc, Seattle, WA) a una concentración 1:100 respecto a la proporcionada por el fabricante y se incubó con el anticuerpo secundario EnVision™ FLEX anti rabbit, (Dako, Glostrup, Denmark). Se incubó con tampón citrato a pH6 en autoclave a 1,5 atm durante 3 minutos para el desenmascaramiento antigénico. Posteriormente se incubó con el anticuerpo primario durante 60 minutos y 30 minutos con el anticuerpo secundario.

### 7.3. Hibridación de fluorescencia in situ para el estudio de la amplificación de HER2

La técnica de hibridación in situ fluorescente se emplea para comprobar la amplificación génica de *HER2*, de manera que obtenemos la expresión de HER2 de manera más eficiente que mediante IHC, ambas técnicas son complementarias, realizándose FISH en los casos en los que el resultado de la tinción IHC no sea concluyente.

La determinación de la amplificación de *HER2* se realizó de manera rutinaria en el departamento de Anatomía Patológica del Hospital Clínico Universitario de Valencia, mediante HER2 IQFISH pharmaDx™ (Dako, Glostrup, Denmark), siguiendo las instrucciones del fabricante.

Primeramente se realizó el desenmascaramiento antigénico del tejido empleando tampón citrato sódico a pH6 (*Dako REAL Target Retrieval Solution 1X*. Dako group. Glostrup, Denmark) en autoclave durante 4 minutos, tras lo cual se realizó un lavado de 5 minutos con tampón salino citrato sódico 2X.

Posteriormente se incubó con proteinasa K (Dako group, Glostrup, Denmark) durante 10 minutos a 37°C para la digestión y rotura de membranas celulares y se lavó con tampón citrato sódico 2X durante 5 minutos.

Tras la fijación del tejido en formol tamponado, se deshidrataron las muestras mediante lavado en etanol en concentración creciente (70%, 96% y 100%), y se hibridaron con 6µl de una mezcla de la sonda para el gen *HER2* y del centrómero del cromosoma 17 (HER2/CEN-17 IQFISH Probe Mix). Tras una breve incubación a 90°C 5 minutos para desnaturalizar el DNA, se mantuvo la hibridación a 37°C durante una noche en cámara húmeda. Se eliminaron las hibridaciones inespecíficas mediante lavado astringente por inmersión en solución de NP40 al 0,3% y tampón citrato sódico 2X a temperatura ambiente seguido de otro lavado a 72°C durante 5 minutos.

Finalmente se tiñeron los núcleos celulares mediante incubación con 4',6-diamino-2-fenilindol (DAPI) (Abbot laboratories. Illinois, EEUU) y se observaron las muestras mediante microscopía de fluorescencia. Se contaron entonces el número de genes, en color rojo, y de centrómeros, en verde, considerándose amplificación del gen *HER2* cuando el ratio de copias entre HER2 y el centrómero del cromosoma 17 fue superior o igual a 2,2.

#### 7.4. Determinación inmunohistoquímica de las muestras de cáncer de mama

La determinación inmunohistoquímica del tejido se llevó a cabo mediante la determinación del estatus hormonal, midiendo la expresión de receptores de estrógenos (ER) y progesterona (PR), la expresión del receptor para el factor de crecimiento epidérmico ERBB2 (*HER2*) o la amplificación génica del gen que codifica para el mismo (*HER2*), así como la medida de la proliferación celular, que vino dada por el porcentaje de células que expresan la proteína Ki67. Para la determinación se han seguido los criterios generales aprobados por el *12th St Gallen International Breast Cancer Conference 2011*.

**Tabla 8.** Interpretación de los resultados de la determinación del estatus hormonal y de HER2 en muestras de tejido tumoral mamario

Proteína	Positivo	Negativo	No concluyente
ER/PR-IHC	Tinción completa e intensa de membrana en > 10% de células infiltrantes	Tinción completa de membrana leve-moderada en < 10% de células infiltrantes	—
HER2-IHC	Tinción completa e intensa de membrana en > 30% de células infiltrantes	Ausencia de tinción o tinción débil e incompleta de membrana	Tinción completa de membrana leve-moderada en > 10% de células infiltrantes
<i>HER2</i> -FISH	Número de copias del gen por núcleo > 6 Ratio copias/centrómero > 2,2	Número de copias del gen por núcleo < 4 Ratio copias/centrómero < 1,8	Número de copias del gen por núcleo 4-6 Ratio copias/centrómero 1,8-2,2

ER: Receptor de estrógenos, PR, receptor de progesterona; HER2, receptor del factor de crecimiento epidérmico ERBB2, IHC, tinción inmunohistoquímica; FISH, Hibridación in situ fluorescente.

De manera complementaria, se evaluó el subtipo molecular de las muestras de tumor mamario mediante la realización del PAM50, empleando un mínimo de 100ng del RNA total extraído, con un ratio 260/280 superior a 1,75. El fundamento y metodología empleados se encuentra descrito en Bastien 2012 y Cheang 2015 (182, 183).

---

---

## 1. ANÁLISIS ESTADÍSTICOS

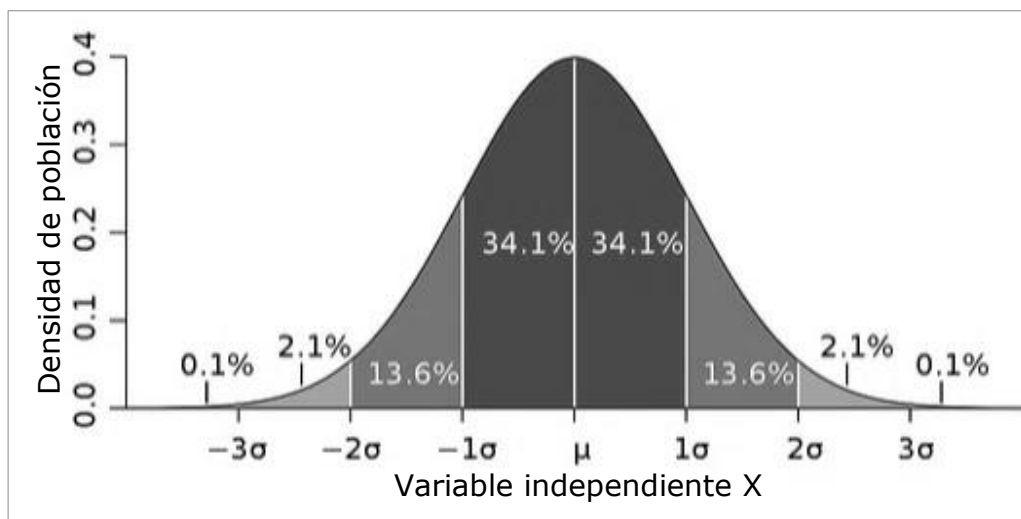
---

---

Los tests estadísticos que han sido implementados en este trabajo se han realizado bien mediante los programas informáticos SPSS y Graphpad Prism 6; mediante R (empleando los paquetes incluidos en el proyecto R/Bioconductor); y mediante las herramientas online POMELO y Statpages.

### 1.1. *Test de normalidad y de homogeneidad de varianzas*

Las distribuciones aproximadamente normales son muy comunes, como queda explicado por el teorema central del límite (Figura 11). Cuando en un fenómeno se sospecha la presencia de un gran número de pequeñas causas actuando de forma aditiva e independiente es razonable pensar que las observaciones serán “normales”. Hay métodos estadísticos para probar empíricamente esta asunción, por ejemplo, el test de D'Agostino-Pearson (184-186). El test de Kolmogorov-Smirnov, antes ampliamente usado, ahora parece estar desaconsejado.



**Figura 11. Representación de la curva de Gauss característica de una distribución normal.** Se representa la densidad de población en base a una variable independiente.  $\mu$  se refiere a la media poblacional y  $\sigma$  a la desviación estándar. Se muestran los porcentajes de la población contenidos por la curva, que corresponden con la probabilidad de seleccionar a un individuo aleatoriamente comprendido entre dos valores de la variable independiente dados.

En estadística, el Test de Shapiro-Wilk se usa para contrastar la normalidad de un conjunto de datos. Se plantea como hipótesis nula que una muestra  $x_1, \dots, x_n$  proviene de una población

normalmente distribuida. Se considera uno de los test más potentes para el contraste de normalidad, sobre todo para muestras pequeñas ( $n < 30$ ) (187).

El estadístico del test es: 
$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Donde  $x_{(i)}$  es el número que ocupa la  $i$ -ésima posición en la muestra;

$\bar{x}$  es la media de la muestra.

Las variables  $a_i$  se calculan  $(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$ ;

donde:  $m = (m_1, \dots, m_n)^T$ ;

siendo  $(m_1, \dots, m_n)$  los valores medios del estadístico ordenado (de variables aleatorias independientes e idénticamente distribuidas, muestreadas de distribuciones normales);  $V$  es la matriz de covarianzas de ese estadístico.

Siendo la hipótesis nula que la población está distribuida normalmente, si el p-valor es menor a  $\alpha$  (nivel de confianza) entonces la hipótesis nula es rechazada (se concluye que los datos no vienen de una distribución normal). Si el p-valor es mayor a  $\alpha$ , no se rechaza la hipótesis y se concluye que los datos siguen una distribución normal.

En este trabajo se han realizado tests de normalidad en aquellos conjuntos de datos a los que se les iba a someter a un test de comparación entre grupos; es decir, a los valores de expresión obtenidos para los diferentes miRNAs y mRNAs validados mediante qRT-PCR. Previamente, se identificaron los valores extremos mediante el método de ROUT con un valor de Q del 1%.

## 1.2. Comparaciones entre grupos

Según si asumimos que nuestra muestra sigue una distribución normal o no, los métodos empleados para comparar los grupos serán paramétricos (si asumimos una distribución normal) o no paramétricos (en caso de que la distribución de la muestra no se ajuste a la normalidad).

Si enumeramos los genes con  $i = 1; \dots; N$  entonces para el  $i$ -ésimo marcador molecular estamos considerando el contraste de hipótesis siguiente:

**$H_i$ :** El marcador  $i$  no tiene una expresión diferencial entre las condiciones consideradas.

**$K_i$ :** El marcador  $i$  tiene una expresión diferencial entre las condiciones consideradas.

Estos contrastes se plantean para cada uno de los  $N$  marcadores evaluados. Denotaremos  $G = \{1; \dots; M\}$  el conjunto de hipótesis nulas que estamos evaluando. El número de hipótesis que vamos a contrastar es conocido a priori ya que corresponde con el número de marcadores que estamos evaluando. Denotamos por  $G_0$  (con  $G_0 \subset G$ ) las hipótesis nulas que son ciertas. Denotamos por  $N_0 = |G_0|$  el cardinal del conjunto  $G_0$ . Debemos notar que el conjunto  $G_0$  no es desconocido. No se sabe ni cuántas ni cuáles son las hipótesis nulas ciertas. Esto supondría que se conocen los genes que se expresan diferencialmente y esto es precisamente el objetivo que perseguimos (188).



## Pruebas paramétricas de comparación de medias entre grupos: T-student y ANOVA

En estadística, una prueba  $t$  de Student, prueba T-Student, o Test-T es cualquier prueba en la que el estadístico utilizado tiene una distribución  $t$  de Student si la hipótesis nula es cierta. Se aplica cuando la población estudiada sigue una distribución normal pero el tamaño muestral es demasiado pequeño como para que el estadístico en el que está basada la inferencia esté normalmente distribuido, utilizándose una estimación de la desviación típica en lugar del valor real.

Esta prueba se puede utilizar únicamente si se puede asumir que las dos distribuciones poseen la misma varianza. El estadístico  $t$  para probar si las medias son diferentes puede ser calculado como sigue:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{x_1x_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde

$$S_{x_1x_2} = \sqrt{\frac{(n_1 - 1)S_{x_1}^2 + (n_2 - 1)S_{x_2}^2}{n_1 + n_2 - 2}}$$

Nótese que las fórmulas de arriba, son generalizaciones del caso que se da cuando ambas muestras poseen igual tamaño (sustituyendo  $n$  por  $n_1$  y  $n_2$ ).

Siendo:

$S_{x_1x_2}$  un estimador de la desviación estándar común de ambas muestras (esto se define así para que su cuadrado sea un estimador sin sesgo de la varianza común sea o no la media iguales),

$n$  = número de participantes de los grupos 1 y 2.

Con  $n - 2$  grados de libertad para cada grupo, y el tamaño muestral total menos dos (esto es,  $n_1 + n_2 - 2$ ) es el número de grados de libertad utilizados para la prueba de significancia.

En el caso de que ambos grupos posean diferentes varianzas se emplea una variante conocida como prueba  $t$  de Welch y es utilizada únicamente cuando se puede asumir que las dos varianzas poblacionales son diferentes (los tamaños muestrales pueden o no ser iguales) y por lo tanto deben ser estimadas por separado (189). El estadístico  $t$  se calcula como sigue:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

donde

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$s^2$  es el estimador sin sesgo de la varianza de las dos muestras,

$n$  = número de participantes de los grupos 1 y 2.

Nótese que la verdadera distribución de este estadístico de hecho depende (ligeramente) de dos varianzas desconocidas.

En estadística, el análisis de la varianza (ANOVA, *ANalysis Of VAriance*, en inglés) es una colección de modelos estadísticos y sus procedimientos asociados, en el cual la varianza está particionada en ciertos componentes debidos a diferentes variables explicativas. El análisis de la varianza parte de los conceptos de regresión lineal. Un análisis de la varianza permite determinar si diferentes tratamientos muestran diferencias significativas o por el contrario puede suponerse que sus medias poblacionales no difieren (190). El análisis de la varianza permite superar las limitaciones de hacer contrastes bilaterales por parejas (que son un mal método para determinar si un conjunto de variables con  $n > 2$  difieren entre sí).

La prueba ANOVA parte de algunos supuestos: la variable dependiente debe medirse al menos a nivel de intervalo, debe existir independencia de las observaciones, la distribución de los residuales debe ser normal, debe cumplirse homocedasticidad.

La técnica fundamental consiste en la separación de la suma de cuadrados (SS, del inglés *sum of squares*) en componentes relativos a los factores contemplados en el modelo. Como ejemplo, mostramos el modelo para un ANOVA simplificado con un tipo de factores en diferentes niveles:

$$SS_{Total} = SS_{Error} + SS_{Factores}$$

El número de grados de libertad (gl) puede separarse de forma similar:

$$gl_{Total} = gl_{Error} + gl_{Factores}$$

Para evaluar la asociación de los miRNAs procedentes del chip de Affymetrix, tanto el t-test como la ANOVA se realizaron por permutaciones (20000) mediante la herramienta online POMOLOII (<http://pomelo2.iib.uam.es/>) como se describe en Morrissey y Diaz-Uriarte 2009. (191). Para evaluar las diferencias en las medias de expresión de los miRNAs validados entre los grupos, se empleó el programa Graphpad Prism 6.

### Test de Tukey

Se trata de un método empleado para la corrección de múltiples comparaciones tras aplicar un test ANOVA de  $k$  grupos (*post-hoc*) para encontrar medias significativamente diferentes. Compara todos los posibles pares de medias entre sí e identifica aquellas diferencias entre medias que son mayores que el error estándar esperado. Se trata de un método conservador cuando existe diferente tamaño muestral entre grupos.

El estadístico del test de Tukey se define como:

$$q_s = \frac{Y_A - Y_B}{SE}$$

Siendo  $Y_A$  la mayor de las medias y  $Y_B$  la menor, y  $SE$  la desviación típica de los datos.

El valor  $q_s$  puede ser entonces comparado con el valor  $q$  obtenido tras un test de student, si en valor de  $q_s$  es mayor al valor  $q$  crítico obtenido, ambas medias son significativamente diferentes

El test de Tukey assume que las observaciones son independientes entre sí y entre grupos, que las medias se encuentran dentro de la normalidad y que existe homogeneidad de varianzas. Por ello esta corrección se ha llevado a cabo en aquellos casos en los que se ha realizado un test ANOVA mediante el programa Graphpad Prism 6.

## Pruebas no paramétricas de comparación de grupos: U de Mann Whitney y Kruskal-Wallis

El test de la U de Mann-Whitney (o prueba de rangos de Wilcoxon) es la versión no paramétrica de la t de Student, evalúa la diferencia entre dos grupos cuando estos no se ajustan a una distribución normal (185).

La prueba de Mann-Whitney asume que las observaciones de ambos grupos son independientes, las observaciones son variables ordinales o continuas. Que bajo la hipótesis nula, la distribución de partida de ambos grupos es la misma, y que bajo la hipótesis alternativa, los valores de una de las muestras tienden a exceder a los de la otra (192).

Para calcular el estadístico  $U$  se asigna a cada uno de los valores de las dos muestras su rango para construir:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

donde  $n_1$  y  $n_2$  son los tamaños respectivos de cada muestra;  $R_1$  y  $R_2$  es la suma de los rangos de las observaciones de las muestras 1 y 2 respectivamente.

El estadístico  $U$  se define como el mínimo de  $U_1$  y  $U_2$ .

Los cálculos tienen que tener en cuenta la presencia de observaciones idénticas a la hora de ordenarlas. No obstante, si su número es pequeño, se puede ignorar esa circunstancia.

En estadística, la prueba de Kruskal-Wallis (de William Kruskal y W. Allen Wallis) es un método no paramétrico para probar si un grupo de datos proviene de la misma población. Intuitivamente, es idéntico al ANOVA con los datos reemplazados por categorías. Es una extensión de la prueba de la U de Mann-Whitney para 3 o más grupos.

Ya que es una prueba no paramétrica, la prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA. Sí asume, bajo la hipótesis nula, que los datos vienen de la misma distribución.

El estadístico viene dado por:

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

donde:

$n_i$  es el número de observaciones del grupo  $i$ ,

$r_{ij}$  es el rango (entre las observaciones) de la observación  $j$  en el grupo  $i$ ,

$N$  es el número total de observaciones,

$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ ,  $\bar{r} = (N + 1)/2$  es el promedio de  $r_{ij}$ .

Finalmente, el p-valor es aproximado por  $Pr(\chi_{g-1}^2 \geq K)$ .

Estas pruebas se aplicaron a los análisis de validación de la expresión de miRNAs y mRNAs mediante Graphpad Prism 6 en el caso de que las distribuciones no cumplieren con la normalidad.

## Test de Dunn

El test de Dunn es un método de corrección *post-hoc* que se aplica tras un resultado significativo en el test de Kruskal-Wallis, ya que al realizar contrastes de múltiples muestras la tasa de error tipo I aumenta. Este test emplea correctamente los mismos rangos que la prueba de Kruskal-Wallis y mantiene la varianza propia de la hipótesis nula del test de Kruskal-Wallis para determinar cual de los pares de muestras evaluados son significativamente diferentes (193).

## Prueba de la Chi-cuadrado de Pearson ( $\chi^2$ )

La prueba  $\chi^2$  de Pearson es considerada como una prueba no paramétrica que mide la discrepancia entre una distribución observada y otra teórica (bondad de ajuste), indicando en qué medida las diferencias existentes entre ambas, de haberlas, se deben al azar en el contraste de hipótesis. También se utiliza para probar la independencia de dos variables entre sí, mediante la presentación de los datos en tablas de contingencia (194).

La fórmula que da el estadístico es la siguiente:

$$\chi^2 = \sum_i \frac{(\text{observada}_i - \text{esperada}_i)^2}{\text{esperada}_i}$$

Los grados de libertad (*gl*) vienen definidos por:

$$gl = (r - 1)(k - 1)$$

Siendo *r* el número de filas y *k* el de columnas de la tabla de contingencia.

Cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones. Por tanto, no se rechaza  $H_0$  cuando:

$$\chi^2 < \chi_t^2 (r - 1)(k - 1)$$

Donde *t* representa el valor proporcionado por las tablas de significación estadística.

Se empleó la prueba  $\chi^2$  para establecer la diferencia en la distribución de variables categóricas no aproximables a continuas. El número de muestras con tinción positiva (presencia) en mujeres jóvenes frente a mujeres mayores. Se calculó la tabla de contingencia de 2 x 2 de los casos observados en ambos grupos y se calculó el estadístico mediante la herramienta online stat-pages (<http://statpages.org/ctab2x2.html>).

## Coefficiente de correlación de Pearson

En estadística, el coeficiente de correlación de Pearson es una medida de la relación lineal entre dos variables aleatorias cuantitativas. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables (194).

En el caso de que se esté estudiando dos variables aleatorias *x* e *y* sobre una población; el coeficiente de correlación de Pearson se simboliza como  $\rho_{x,y}$ , siendo la expresión que nos permite calcularlo:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Donde  $\sigma_{XY}$  es la covarianza de  $(X, Y)$ ;  $\sigma_X$  es la desviación típica de  $X$  y  $\sigma_Y$  es la desviación típica de  $Y$ .

El valor del índice de correlación varía en el intervalo  $[-1, 1]$ . Una correlación de 0 indica que no existe correlación alguna entre las variables. Valores por encima de 0 indican que la correlación es positiva y por debajo, negativa.

En el análisis de la viabilidad del material parafinado, para evaluar la correlación entre los Ct medios de los mRNAs, snoRNAs y miRNAs evaluados entre los tipos de muestra, se empleó el programa Graphpad Prism 6 para realizar una correlación de Pearson con un intervalo de confianza del 95% y un p valor  $< 10^{-4}$ .

### Regresión logística

En estadística, la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictores. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados (GLM, del inglés *generalized linear model*) que usa como función de enlace la función logit (195). Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística.

La regresión logística analiza datos distribuidos de manera binomial de la forma  $Y_i \sim B(p_i, n_i)$ , para  $i = 1, \dots, m$ ; dónde los números de ensayos Bernoulli  $n_i$  son conocidos y las probabilidades de éxito  $p_i$  son desconocidas.

El modelo es entonces obtenido a base de lo que cada ensayo (valor de  $\lambda$ ) y el conjunto de variables independientes puedan informar acerca de la probabilidad final. Estas variables explicativas pueden pensarse como un vector  $X_i$   $k$ -dimensional y el modelo toma entonces la forma:

$$p_i = E\left(\frac{Y_i}{n_i} \mid X_i\right).$$

Los logits de las probabilidades binomiales desconocidas son modeladas como una función lineal de los  $X_i$ .

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}.$$

Nótese que un elemento particular de  $X_i$  puede ser ajustado a 1 para todo  $i$ , obteniéndose una constante independiente en el modelo. Los parámetros desconocidos  $\beta_i$  son usualmente estimados a través de máxima verosimilitud.

La interpretación de los estimados del parámetro  $\beta_i$  es como los efectos aditivos en el logaritmo de la razón de momios para una unidad de cambio en la  $j$ -ésima variable explicativa. En el caso de una variable explicativa dicotómica, por ejemplo género,  $e^{\beta}$  es la estimación de la razón de momion (en inglés *odds ratio*) de tener el resultado para dos grupos dados.

En este trabajo, se empleó regresión logística para estimar la influencia de las variables tamaño tumoral y porcentaje de Ki67 sobre la cantidad de miRNA expresada en las muestras procedentes de pacientes, analizadas mediante chip de Affymetrix. Se realizó mediante la herramienta online

POMELOII (<http://pomelo2.iib.uam.es/>) como se describe en Morrissey y Diaz-Uriarte 2009 (191).

### 1.3. Corrección por múltiples comparaciones

Conocidos el número de contrastes de hipótesis ( $M$ ), una vez hemos realizado todos los contrastes y tomado una decisión sobre si rechazamos o no cada hipótesis nula, nos indica cuántas hipótesis nulas hemos rechazado ( $R$ ) mediante una variable aleatoria. La variable aleatoria  $V$  nos está dando el número de falsos positivos o errores tipo I (no hay expresión diferencial pero decidimos que la hay de un modo erróneo) mientras que  $T$  nos da el número de falsos negativos o error tipo II (se acepta la hipótesis nula cuando ésta es falsa). Ambas variables indican error y son importantes, pero no son observables (196, 197).

#### Benjamini & Hochberg

Uno de los problemas que se presentan a la hora de analizar datos masivos, como son los microchips de expresión, es la aparición de errores de tipo I, ya que permiten evaluar miles de genes simultáneamente. La tasa de falsos positivos FDR (del inglés *false discovery rate*) es una manera de conceptualizar la tasa de dichos errores tipo I al evaluar la hipótesis nula cuando se realizan múltiples comparaciones (188).

La corrección por múltiples comparaciones FDR, proporciona un control de falsos positivos menos estricto que los métodos FWER (del inglés *familywise error rate*), proporcionando más poder, pero también una mayor tasa de errores tipo I, por lo que los resultados deben ser tomados con precaución. Aún así, es el método recomendado a la hora de realizar análisis de datos derivados de tecnologías de alto rendimiento como los microchips de expresión.

Se define  $Q$  como la proporción de falsos positivos  $V$  dado un número de positivos  $R$ .

$$Q = V/R$$

La tasa de falsos positivos (FDR), viene dada por la expresión:

$$FDR = Q_e = E[Q] = E\left[\frac{V}{V+S}\right] = E\left[\frac{V}{R}\right]$$

Siendo  $S$  el número de positivos ciertos.

La aproximación de Benjamini & Hochberg controla la tasa de falsos positivos bajo un umbral  $\alpha$  del siguiente modo:

Tenemos  $H_1, \dots, H_m$  hipótesis nulas para cada cual obtenemos un p-valor  $P_1, \dots, P_m$ , siendo  $m$  el número de tests realizado, para un  $\alpha$  dado, se busca la mayor  $k$ -ésima hipótesis para la cual se cumple que:

$$P_{(k)} \leq \frac{k}{m} \alpha.$$

Se rechazan, pues, todas las  $H_{(i)}$  cuando  $i=1, \dots, k$ .

Este procedimiento es válido mientras los  $m$  tests sean independientes.

Los recursos empleados en este trabajo de tesis para realizar los métodos estadísticos incluyen una corrección FDR por Benjamini & Hochberg, que fue realizada automáticamente.

## Test de Bonferroni

La probabilidad de rechazar la hipótesis nula cuando es cierta es baja, sin embargo esta se multiplica al aumentar el número de hipótesis evaluadas, aumentando la probabilidad de un evento que, en principio era poco probable y, por tanto, la probabilidad de cometer un error de tipo I.

La corrección por múltiples comparaciones de Bonferroni se basa en la idea de que, si un sujeto está evaluando  $m$  hipótesis, una manera de controlar la tasa de error FWER es evaluar cada una bajo un nivel de significación de  $\alpha/m$ , siendo  $\alpha$  el valor que se habría fijado para una única hipótesis (198).

Este método de corrección es muy conservativo si hay un gran número de pruebas simultáneas, aumentando considerablemente la probabilidad de cometer un error de tipo II (falsos negativos) (199).

### 1.4. Curvas de supervivencia

Para el estudio de asociación de la expresión diferencial de miRNAs con la supervivencia global se emplearon los datos de supervivencia, en días, obtenidos de los datos del estudio METABRIC. Los casos son censurados ya que no se conoce el destino final de todas las muestras. Se clasificaron como 0 a aquellas pacientes vivas y aquellas que murieron por causa no relacionada con la enfermedad o de las que se dejó el seguimiento por otros motivos. Se clasificaron como 1 a aquellas pacientes con exitus derivado de la enfermedad.

Mediante Graphpad Prism 6 se construyeron las curvas de supervivencia Kaplan-Meier y se calcularon los estadísticos Mantel-Haenszel y Gehan-Wilcoxon con su riesgo asociado (HR, del inglés *hazard's ratio*).

#### Estimador de Kaplan-Meier

En el análisis de la supervivencia, el estimador de Kaplan-Meier es un estimador no paramétrico de la función de supervivencia. Este estimador tiene en cuenta la censura. El estimador admite una representación gráfica por medio de una función escalonada (200).

Sea  $S(t)$  la función de supervivencia de una determinada población, es decir, la probabilidad de que uno de sus integrantes viva más allá de un tiempo  $t$ . Para una muestra de esta población de tamaño  $N$ , sean  $t_1 \leq t_2 \leq \dots \leq Tt_N$  los tiempos que discurren hasta la muerte todos ellos.

Entonces, para cada  $t_i$ , se define:

$d_i$  el número de muertes en el momento  $t_i$  y

$n_i$ , el número de sujetos en riesgo justo antes de  $t_i$ . De no haber censura,  $n_i$  es el número de supervivientes inmediatamente antes del momento  $t_i$ . Con censura es el número de supervivientes menos el número de casos censurados: sólo se observan los sujetos vivos que no se han caído del estudio en el momento en que ocurre una muerte.

El estimador de Kaplan-Meier de  $S(t)$  pues

$$\widehat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

El estimador de Kaplan–Meier tiene diversas aplicaciones. En medicina se usa para medir la fracción de pacientes todavía vivos tiempo después de un tratamiento. Uno de los problemas típicos en el análisis de la supervivencia consiste en determinar si dos curvas de Kaplan-Meier (por ejemplo, las correspondientes a dos cohortes de pacientes que han sido sometidos a tratamientos distintos) son o no significativamente distintas (200).

### **Test de Mantel-Haenszel-Cox**

La prueba Mantel-Cox, también conocida a veces como la prueba log-rank, es una técnica estadística para contrastar las funciones de supervivencia de dos poblaciones. Es una prueba no paramétrica que puede usarse en presencia de datos censurados (201).

La prueba de Mantel-Cox compara las estimaciones de la función de riesgo de dos grupos en cada unidad de tiempo en que ocurre un evento. Se construye comparando el número observado y esperado de sucesos en uno de los grupos en cada uno de esos momentos y agrupando los resultados para obtener un valor global (202).

### **Test de Gehan-Wilcoxon**

En ausencia de datos censurados, el test se reduce a un test de Mann-Whitney (o Wilcoxon de dos muestras). Aunque el test de log-rank suele ser más apropiado, el test de Wilcoxon modificado da mayor peso a las muertes que suceden antes en el tiempo, por lo que a veces aporta mayor información desde el punto de vista biológico (192).

## *1.5. Valor de significatividad y poder estadístico*

El valor de significatividad, llamado p-valor, se basa en asumir que la hipótesis nula es verdadera y calcula la probabilidad de que, dada la muestra, bajo la hipótesis nula puedan observarse valores tan extremos como los que presenta la muestra. Por tanto, cuanto menor sea el p-valor, menor la probabilidad de que encontremos esos valores extremos observados bajo la hipótesis nula, lo que se traduce en una mayor probabilidad de que la hipótesis nula sea falsa.

Típicamente se rechaza la hipótesis nula cuando el p-valor es menor que un  $\alpha = 0,05$ . Cuando el test es de una cola (es decir, si en nuestro caso solo evaluamos si la media es superior o inferior, no ambas condiciones),  $\alpha$  pasa a ser el doble (o lo que es lo mismo, el p-valor pasaría a ser su mitad), rechazando la hipótesis nula cuando p-valor es menor a un  $\alpha = 0,1$ .

El poder se define como la probabilidad de rechazar la hipótesis nula cuando, en efecto, ésta es falsa. Ergo, mayor poder estadístico nos garantiza que no se ha cometido un error de tipo II (que consiste en aceptar la hipótesis nula siendo falsa).

Sea la probabilidad de un error tipo II  $\beta$ , el poder se definirá como  $1 - \beta$ .

En última instancia el poder depende del tamaño efectivo de la muestra, de la significación y del tamaño muestral. Para calcular el poder para un t-test empleamos la función de R 'power.t.test()', en el caso de un test ANOVA 'power.anova.test()'. Si conocemos el valor de poder que se pretende alcanzar, se puede calcular del mismo modo el tamaño muestral necesario para alcanzar ese poder. Típicamente se acepta un poder que sea mayor que  $\Pi = 0,8$ , que es el valor límite que empleamos en este trabajo.



## 1.6. Combinación de $p$ -valores mediante el método de Fisher

Consideremos  $k$  estudios independientes en los se realiza un procedimiento de pruebas de hipótesis donde

$$H_{i,0} : \theta_i = \theta_{i,0} \quad \text{vs.} \quad H_{i,1} : \theta_i > \theta_{i,0}$$

Siendo  $i = 1, 2, \dots, k$ .

Si  $T_i$  y  $p_i$  corresponden al estadístico de prueba y al  $p$ -valor para el  $i$ -ésimo estudio, respectivamente,  $H_{0,i}$  será rechazada si  $p_i \leq \alpha$ , siendo  $\alpha$  una probabilidad de error tipo I definida de antemano. Ahora, es posible obtener una medida de resumen que tenga en cuenta los resultados obtenidos para cada uno de los  $k$  estudios.

Meta-análisis se puede definir como la identificación sistemática, valoración y síntesis de información proveniente de  $k$  estudios independientes que intentan responder la misma pregunta de investigación. Esta síntesis, que implica la combinación de la evidencia proporcionada por cada estudio, tiene como objetivo obtener un estadístico que resuma toda la información obtenida en ellos a través de la verificación de una hipótesis nula común, analizando simultáneamente un mismo parámetro  $\theta$  de interés. Esto, precisamente, responde la pregunta en el párrafo anterior.

En investigaciones biomédicas, uno de los métodos más utilizados para obtener este estadístico de resumen es la combinación de  $p$ -valores. Para su utilización deben darse los siguientes supuestos: los  $k$  estudios deben ser independientes, deben rechazar una hipótesis nula  $H_0$  común, y las pruebas de hipótesis realizadas deben ser unilaterales y en la misma dirección para los  $k$  estudios (203). El objetivo de esta metodología es determinar el efecto unidireccional cuando se combina la información de los diferentes estudios utilizando los  $p$ -valores. Obsérvese que bajo el procedimiento de hipótesis (1), las condiciones (i)-(iii) se satisfacen.

Dentro de los métodos de combinación de  $p$ -valores disponibles en la literatura, el método de Fisher es uno de los más empleados en el campo biológico, y está fundamentado en la probabilidad de la transformación integral (204).

Si  $F_{i,0}$  es la función de distribución acumulada de  $T_i$  bajo  $H_{i,0}$ , entonces  $F_i(T_i) \sim U(0, 1)$ ; por lo que  $p_i \sim U(0, 1)$ . Por tanto,  $C_F = -\log(p_i) \sim \text{Exp}(1)$ ; donde:  $i = 1, 2, \dots, k$ .

Siendo el estadístico de prueba para el método de Fisher, pues:

$$\chi_F^2 = -2 \sum_{i=1}^k \ln(p_i) \sim \chi_{2k}^2$$

La prueba de Fisher puede ser interpretada como la probabilidad de rechazar la hipótesis nula en al menos uno de los estudios de los  $k$  que hacen parte de la combinación. Sin embargo, la interpretación más adecuada consiste en indagar si la acumulación de información entre las pruebas sobre las hipótesis nulas similares puede rechazar una hipótesis nula compartida (205). Este método rechaza la hipótesis  $H_0$  común si  $P(\chi_F^2 > \chi_{2k}^2) < \alpha$ .

El principal inconveniente del método de Fisher es su sensibilidad a  $p$ -valores pequeños, permitiendo el rechazo de la hipótesis nula en favor de la hipótesis alternativa unilateral (205).

En este trabajo se ha realizado el método de Fisher para combinar los p-valores de los tres grupos de muestras estudiadas independientemente. Para ello se combinaron los p-valores en un estadístico que sigue una distribución  $\chi^2$  con  $2k$  grados de libertad, siendo  $k$  el número de p-valores que se combinan. Se empleó el paquete `CombinePValue v1.0` de R en el entorno R Studio, empleando la función `'competitive.test()'`.

## 2. ANÁLISIS BIOINFORMÁTICOS

---

### 2.1. Manejo, tratamiento y filtrado de datos

#### Muestras de pacientes y líneas

A partir de la base de datos del Hospital clínico de Valencia se recogieron los datos de las pacientes seleccionadas para el estudio en los siguientes campos: ID muestra, edad al diagnóstico, código de biopsia, subtipo, tipo histológico, %RE, %RP, %Ki67, HER2, grado histológico, tamaño tumoral, nódulos afectos, neoadyuvancia, tratamiento, metastasica de inicio, radioterapia, nacionalidad, consumo de tabaco, consumo de alcohol, dieta especial, IMC, HRT, edad de menarquia, edad de menopausia, número de hijos, edad del primer embarazo, lactancia, uso de anticonceptivos orales, estado de mutaciones en BRCA, antecedentes familiares, recaídas, otras comorbilidades.

Una vez seleccionadas las pacientes y extraído el material, se filtraron según la concentración de RNA obtenida y su calidad. En el caso de las líneas celulares se creó un archivo indicando la procedencia de la línea, la edad de la paciente, el subtipo, los receptores que expresa, las mutaciones que presenta y el medio de cultivo empleado.

Una vez obtenidos los archivos de expresión de miRNAs el filtrado de los datos se realizó de igual manera en los datos procedentes de pacientes o de líneas celulares de cáncer de mama. Una vez obtenido el archivo `"*.CEL"` tras el escaneado del chip de hibridación, se volcaron en el programa `Expresion Console (Affymetrix, Carlsband, USA)`, dónde se realizó la normalización de los valores obtenidos en cada chip.

#### Datos a partir de bases de datos online

Se obtuvieron los datos de expresión de miRNAs ya normalizados, así como 16 parámetros fenotípicos, demográficos y patológicos de las 1302 muestras estudiadas: edad de diagnóstico, edad en la última revisión, etnia, tamaño tumoral en milímetros (mm), numero de nodos linfáticos afectados, localización de la metástasis, grado histológico, estadio tumoral, resultado de la enfermedad, tipo histológico, estado de los receptores hormonales (de estrógenos y progesterona), estado receptor HER2 y subtipos intrínsecos PAM50.

Se construyó una matriz de expresión con los datos normalizados obtenidos y otra con los datos clínicos con identificador común.

#### Normalización mediante el método RMA

Al comparar chips diferentes procedentes de muestras diferentes, generalmente implica la corrección de errores sistemáticos debidos a variaciones en el protocolo o en la intensidad de tinción, al tratarse de sucesos independientes. El método de normalización RMA, del inglés

*Robust Multi-array Average* (media robusta multi-chip), consta de tres pasos: un primer paso de ajuste de fondo, normalización de cuartiles y por último una sumarización mediante un modelo de estimación que emplea el test de la mediana *polish* de Tukey (2006).

En este trabajo los valores de intensidad de los chips de hibridación se normalizaron empleando el programa proporcionado por Affymetrix Expression Console, empleando el método “RMA + DABG”. Los valores obtenidos fueron expresados en una escala  $\log_2$ .

### Filtrado de los datos

De los 20248 valores de intensidad por muestra disponibles se seleccionaron aquellos correspondientes a los 1100 miRNAs humanos, que nombrados con el prefijo hsa- (*homo sapiens*). Una vez seleccionados se calculó la expresión media umbral, por encima de la cual se considera expresión por encima de la basal. Para ello se seleccionaron miRNAs de plantas, concretamente *lotus japonica* (lja-), *pinus taeda* (pta-) y *gossypium raimondii* (gra-), y se calculó una media de los valores de expresión obtenidos. Ya que esos miRNAs no se expresan en humanos, todo valor por debajo del medio obtenido se considera no expresión. Se seleccionaron, pues, aquellos miRNAs que mostraban expresión por encima de este umbral en al menos 3 muestras.

Con los valores de expresión de los miRNAs seleccionados se construyó una matriz de datos de expresión normalizados, con los miRNAs en filas y nombre de muestra codificado, en columnas, a partir de la cual se realizaron los análisis.

## 2.2. Agrupación jerárquica

En minería de datos, el agrupamiento jerárquico es un método de análisis de grupos empleado para construir una jerarquía de grupos.

En este trabajo, se ha empleado el paquete de EisenSoftware que incluye el programa Cluster para la agrupación y Treeview, para la visualización de los datos. Constituyen un sistema de análisis de agrupamiento para datos masivos de expresión de genes obtenidos mediante chip de hibridación del DNA, aunque nuestros datos contengan la expresión de miRNAs, la técnica y métodos de análisis son los mismos, por lo que se decidió que este sistema era adecuado. La métrica de similitud de genes que se emplea es una variante del coeficiente de correlación y se encuentra descrita en Eisen y colaboradores, 1999 (207).

Brevemente, supongamos que  $G_i$  equivale a los datos de expresión transformados para el gen  $G$  dada una condición  $i$ . Para dos genes cualesquiera  $X$  e  $Y$  observados en una serie de condiciones  $N$ , un valor de similitud puede ser calculado como sigue:

$$S(X, Y) = \frac{1}{N} \sum_{i=1, N} \left( \frac{X_i - X_o}{\Phi_X} \right) \left( \frac{Y_i - Y_o}{\Phi_Y} \right)$$

Dónde

$$\Phi_G = \sqrt{\sum_{i=1, N} \frac{(G_i - G_o)^2}{N}}$$

Siendo  $G_o$  un parámetro de ajuste (del inglés *offset*) determinado a la media de las observaciones en  $G$ , así será la desviación estándar de  $G$  y  $S(X, Y)$  es equivalente al coeficiente de correlación de Pearson para las observaciones  $X$  e  $Y$ .

El criterio de enlace de los grupos elegido es una variante del agrupamiento de enlace promedio o UPGMA (del inglés *Unweighted Pair Group Method with Arithmetic Mean*) (208). Se trata de una estrategia de agrupamiento aglomerativa, cuyo algoritmo construye un dendograma que refleja la estructura presente en una matriz de distancias por pares.

En este trabajo se realizó el test de agrupación jerárquica sobre los miRNAs cuya expresión obtenida a partir del chip de hibridación resultó ser significativamente diferente según el grupo de edad. Esto se evaluó mediante un t-test como se describe antes en *Métodos estadísticos*. Se realizó sobre los datos de las 22 muestras procedentes de mujeres menores de 35 años, el grupo de 17 muestras mayores de 35 años y las 3 mujeres sanas, que se emplearon como control de edad. Se realizó también sobre las muestras de líneas celulares, donde sirvió para agrupar tanto las líneas celulares empleadas como los miRNAs expresados, por su similitud. Por último se empleó con los datos obtenidos a partir del estudio METABRIC, en este caso se realizó sobre un total de 83 muestras (33 procedentes de pacientes jóvenes y 50 muestras de mujeres mayores de 65 años seleccionadas al azar).

### 2.3. Estimación del efecto acumulativo de los miRNAs sobre los genes diana

Se parte de la expresión de los 91 miRNAs significativamente diferentes en mujeres jóvenes. Empleando a las mujeres mayores como referencia se ha obtenido un valor de *Fold Change*, que indica cuánto más se expresa un miRNA dado en función de mujeres mayores.

Un miRNA tiene varios genes diana a los que silencia, por lo tanto, si se asume un efecto aditivo de los mismos, el efecto de un miRNA siempre será el mismo y proporcional al número de dianas, suponiendo que la cantidad del gen es suficiente para que este efecto se acumule y nunca llegue a la saturación. En este caso, puede permitirse esta asunción ya que, lo que se pretende es estimar el efecto global que ejerce la desregulación de los miRNAs en los genes y, por ende, en las rutas metabólicas.

A su vez un gen puede estar reprimido por varios miRNAs, y el efecto acumulativo que estos ejerzan dependerá, únicamente, de la expresión de dichos miRNAs. Podemos, por tanto, asignar un valor de expresión a los miRNAs, que será en cierto modo equivalente a su acción represora, tomemos el *Fold Change* obtenido al realizar el t-test.

La acción represora que un miRNA ejerza se verá repartida entre la cantidad de genes diana que posea. Se puede deducir que, si se divide el valor de expresión de un miRNA (*Fold Change*) por el número de genes diana se obtendrá un valor de represión por cada gen, es decir, la represión que está sufriendo ese gen debida a ese miRNA.

Se ha de tener en cuenta que el valor obtenido por gen es su represión debida a los miRNAs, por tanto valores negativos indicarán sobreexpresión del gen.

Se descargaron tanto las dianas teóricas con un valor superior a 0,9 (microT-CDS) como las validadas (miRTarBase) de los 88 miRNAs desregulados en mujeres jóvenes (tres de los 91 miRNAs identificados en un primer análisis se excluyeron de la base de datos miRBase v21 por tratarse de fragmentos de tRNA). Por cada miRNA se obtuvo una lista de sus genes diana, a los que se le asignó un valor de represión individual, el *Fold Change* dividido entre el número de genes diana de dicho miRNA. A continuación, se combinaron los valores obtenidos para cada gen mediante un modelo aditivo, obteniendo una lista de genes con su valor de represión final.

Se generó un archivo de texto con el nombre de todos los genes diana de los miRNAs validados y su valor de represión global para cada uno de ellos. El procedimiento completo se explica en el Anexo III.

## 2.4. *Análisis de enriquecimiento de rutas*

El análisis de enriquecimiento de rutas pretende elucidar qué rutas se encuentran reguladas dada una serie de miRNAs, es decir, qué rutas están enriquecidas en genes que son, de hecho, dianas de dichos miRNAs y, por tanto, podrían estar influenciadas por esos miRNAs.

El análisis de enriquecimiento de rutas en este trabajo se llevó a cabo en R empleando las herramientas online enrichR, REACTOME y DIANA-miRPath 2.0, de DianaTOOLS, descritas en Chen 2013, Joshi-Tope 2005 y Vlachos 2012, respectivamente (209-211).

### **Análisis de enriquecimiento mediante Diana miRPath**

Para el análisis de enriquecimiento mediante Diana miRPath, se tomaron únicamente los miRNAs significativamente diferentes, se separaron por grupos según los nodos seleccionados a partir del análisis de agrupamiento jerárquico llevado a cabo sobre los datos de expresión de las muestras procedentes de pacientes de cáncer de mama. Se generó una lista de los miRNAs de cada nodo en base a su nomenclatura miRBase v.19. Los miRNAs seleccionados se cotejaron en la base de datos microT-CDS, en la que se encuentran todas las dianas teóricas de los miRNAs, se empleó un valor límite de 0.8, con el que se obtiene una media de 350 genes diana. Tras esto se llevó a cabo el análisis de enriquecimiento de los genes diana en la base de datos de rutas KEGG (versión 67).

El algoritmo empleado es el test exacto de Fisher de una cola, debido a que la hipótesis nula es detectar rutas únicamente enriquecidas en dianas específicas. Se obtuvo, además el valor EASE, un método de ajuste conservador para el estadístico de Fisher (212), a ambos p-valores se les aplicó el método FDR de Benjamini y Hochberg como corrección para múltiples comparaciones. Se abordaron dos estrategias para el análisis del enriquecimiento, en primer lugar el servidor combinó los genes diana de los miRNAs seleccionados en un supergrupo común (unión de genes), dicho supergrupo fue incorporado en el análisis de enriquecimiento; por otro lado, se calcula inicialmente la significación de los miRNAs individualmente entre todos los pares posibles de miRNA-ruta, realizando el análisis de enriquecimiento repetidamente; una vez hecho esto, el servidor combina los p-valores individuales calculados y proporciona un p-valor combinado mediante el método de combinación de probabilidad de Fisher, empleado como algoritmo de meta-análisis para unir los p-valores de dos tests independientes que se basen en la misma hipótesis.

### **Análisis de enriquecimiento global**

Para el análisis de enriquecimiento se descargaron los genes diana en la base de datos de dianas de miRNAs validadas miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>) y teóricas, microT-CDS (<http://www.microrna.gr/microT-CDS>), filtrando aquellas con un score de 0.9, posteriormente se cotejaron con las bases de datos de rutas KEGG (versión 74) y Reactome por separado. KEGG y Reactome son bases de datos online de libre acceso, en las que se recogen rutas metabólicas de organismos específicos, incluyendo el reactoma humano (213, 214).

Se realizó un análisis de enriquecimiento en ambas bases de datos empleando erichR (<http://amp.pharm.mssm.edu/Enrichr/enrich>) teniendo en cuenta las dianas teóricas y validadas, en su totalidad y tras haber seleccionado solo aquellas dianas cuyo valor de represión absoluto se encontraba por encima del tercer cuartil (75%), para dotas de sentido biológico al análisis y tener en cuenta únicamente aquellos genes cuya desregulación global debida a los miRNAs tenía como resultado una mayor o menor represión de los genes. El procedimiento seguido completo se encuentra detallado en el Anexo III.

El algoritmo implementado en EnrichR para la obtención de rutas enriquecidas calcula el p-valor a partir de la prueba exacta de Fisher, que prueba la proporción suponiendo una distribución binomial, así como la independencia de la probabilidad de cualquier gen perteneciente a cualquier grupo. Se emplea el método de corrección por múltiples comparaciones FDR de Benjamini & Hochberg.

Se calcula, además, una clasificación basada en el rango, derivada de comenzar el test exacto de Fisher para muchos conjuntos de genes al azar con el fin de calcular un rango medio y su desviación estándar para cada uno de los rangos esperados para cada término en la librería de genes, finalmente se calcula un z-score para evaluar la desviación del rango esperado. El valor combinado se calcula tomando el logaritmo del p-valor de la prueba exacta de Fisher y multiplicando ese valor por el Z-score de la desviación del rango esperado.

## 2.5. Construcción de la red de interacciones

Para construir la red de interacciones entre genes y miRNAs de manera integrada en una ruta metabólica, se empleó el programa Cytoscape v.3.2.1.

Empleando el lenguaje R, y tras instalar las librerías KEGGgraph, KEGG.db y org.Hs.eg.db, se descargaron de la base de datos KEGG, para cada una de las rutas seleccionadas, un archivo con las interacciones por su código KEGG (hsa\*\*\*\*), los datos se procesaron de modo que se construyó un archivo de texto con las parejas de interacciones de genes de cada una de las rutas, modificándose el código del gen al nombre del gen reconocible por Cytoscape.

El archivo de texto generado se cargó entonces en Cytoscape y se generó una red de interacciones para cada ruta mediante el diseño *Perfused force directed*. Se cargó también el archivo de atributos (compuesto por el nombre del gen y un valor de represión estimado calculado con anterioridad) que, según el valor de represión estimado, coloreó los nodos de la red de interacciones en función de la represión que ejercen el global de los miRNAs desregulados en mujeres menores de 35 años en referencia a mujeres mayores de 65 años.

En el caso de la ruta *miRNAs en cáncer* (hsa05206), se seleccionaron los genes (nodos) más relevantes para representar la ruta metabólica y aquellos que mostraban desregulación debida a los miRNAs con expresión diferente en mujeres muy jóvenes. Se le añadió una tabla de atributos representativo de la expresión de los miRNAs en mujeres jóvenes, que se empleó para colorear el borde de los nodos correspondientes a los miRNAs en función a la expresión obtenida.

## 2.6. Búsqueda de sitios de unión de miRNAs conservados en la zona 3'UTR de genes diana mediante TargetScan

Empleando la base de datos online TargetScan 6.2 ([http://www.targetscan.org/vert\\_61/](http://www.targetscan.org/vert_61/)), se buscaron sitios de unión a miRNAs conservados de los genes diana seleccionados previamente.

Para ello se calculan dos valores, el valor contexto y la probabilidad de unión conservada,  $P_{CT}$  (del inglés *conserved targeting probability*), como se describe en García y col 2011 y en Friedman y col 2009, respectivamente (215, 216).

Brevemente, el valor contexto tiene en cuenta los siguientes factores: contribución del tipo al que pertenece el sitio (octámero, heptámero,...), contribución de la posición en 3', contribución de la cantidad de parejas Adenina-Uracilo y su localización en el sitio (AU), posición del sitio, cantidad de sitios diana (TA) y estabilidad de la semilla del sitio de unión (SPS).

Para cada diana predicha de cada miRNA se calcula la suma de cada valor de contexto de cada sitio para un miRNA dado.

Para controlar por tipo de sitio y secuencia de unión, se calcula una razón señal-fondo (S/B) para cada sitio, esta razón se convierte a la probabilidad de unión conservada preferente ( $P_{CT}$ ).

$$P_{CT} = (S/(B-1))/(S/B)$$

Este valor refleja la estimación Bayesiana de la probabilidad de que un sitio se encuentre conservado debido a la selección positiva de la unión a miRNAs en lugar de mantenido por azar u otra cuestión no relacionada con los miRNAs. Los sitios predichos de una familia de miRNAs se pueden agrupar por su  $P_{CT}$  agregado, calculado como:

$$1 - ((1 - P_{CT})_{sitio 1} \cdot (1 - P_{CT})_{sitio 2} \cdot \dots \cdot (1 - P_{CT})_{sitio n})$$

Ya que sitios superpuestos no pueden estar ocupados por diferentes miRNAs al mismo tiempo, estos son eliminados de la tabla de dianas predictivas para maximizar el valor contexto. Estos sitios se emplean para calcular el  $P_{CT}$  agregado.

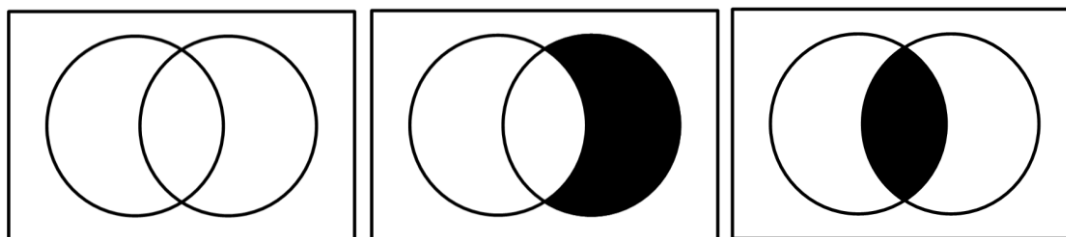
Cuando un sitio se encuentra en una región UTR que se superpone a una región de pauta de lectura abierta, el  $P_{CT}$  no se calcula, ya que la conservación puede estar debida a la traducción del sitio.

## 2.7. Análisis comparativo de los estudios

Para obtener resultados más robustos se unificaron y compararon los resultados de los tres estudios realizados: muestras procedentes de pacientes, muestras de líneas celulares y datos obtenidos a partir de bases de datos online del estudio METABRIC.

### Diagrama de Venn

Los diagramas de Venn son esquemas usados en la teoría de conjuntos, mediante los cuales es posible representar las relaciones de intersección, inclusión y disyunción dentro de un universo de elementos sin cambiar la posición relativa de los conjuntos (ver Figura 12).



**Figura 12. Representación de un diagrama de Venn de dos grupos.** Los grupos se encuentran representados por círculos, el espacio que comparten es su intersección. Se muestra el universo de exclusión de ambos grupos (izquierda), inclusión en uno de los grupos (centro) e inclusión en ambos grupos (derecha).

Se representó en un diagrama de Venn el universo de los miRNAs analizados, tomando como criterio de inclusión en un grupo el ser significativo para dicho grupo. Se hizo, en primer lugar, para el total de los miRNAs analizados en todos los grupos y, después, para los miRNAs sobreexpresados y reprimidos por separado. Los grupos establecidos fueron:

**BCA:** muestras de tumor procedente de pacientes con cáncer de mama analizadas mediante chip de Affymetrix.

**CLA:** muestra de células procedente de cultivo celular de líneas celulares procedentes de tumor mamario analizadas mediante chip de Affymetrix.

**MBD:** datos procedentes de muestras de pacientes obtenidos a partir de bases de datos online y depositados a partir del estudio realizado por el consorcio METABRIC.

Se construyó una matriz de 1 y 0, en la que 1 indica inclusión en el grupo y 0 exclusión, por cada miRNA analizado en cada uno de los grupos de muestras estudiadas. La inclusión en el grupo indica que dicho miRNA es significativo para ese grupo.

Una vez construida la matriz se representó el correspondiente diagrama de Venn mediante la función `VennCount` incluida en `R/Bioconductor`, que indica el número de miRNAs significativos para cada grupo, así como los coincidentes entre grupos.

#### **Meta-análisis mediante el método de Fisher**

Se combinaron los p-valores obtenidos en los tres estudios mediante el método de Fisher, descrito anteriormente en la sección *Métodos estadísticos*, empleando el paquete de R “CombinePValue v1.0” en el entorno R Studio, empleando la función ‘`competitive.test()`’. Al provenir de 3 estudios diferentes, los miRNAs de los que se disponía el p-valor no fueron los mismos en los tres casos, por ello, para el cálculo del estadístico F, se eliminaron aquellos miRNAs para los que no se contaba con p-valor de los tres estudios, quedando un total de 100 miRNAs. Los resultados se representaron mediante un gráfico de Manhattan en el que se representa el -logaritmo del p-valor obtenido.



# RESULTADOS



# ANÁLISIS DE LA VIABILIDAD DEL TEJIDO TUMORAL BIOPSIADO CONSERVADO EN PARAFINA PARA EL ESTUDIO MOLECULAR

## Parte 1

Para el análisis de la viabilidad del tejido biopsiado de tumor incluido en parafina como objeto de estudio en técnicas moleculares, se emplearon muestras de tejido de tumor de mama procedente de pacientes, de las que se disponía tanto de tejido almacenado en condiciones óptimas de congelación, como de tejido fijado en formalina e incluido en parafina (FFPE). Las muestras de tejido empleadas muestran un porcentaje de celularidad por encima del 50% en todos los casos.

### 1.1. *Análisis de la integridad del RNA total*

Se extrajo RNA total a partir de unas 10 criosecciones de 10  $\mu\text{m}$  de tejido congelado, y a partir de 4 secciones de 20  $\mu\text{m}$  de tejido parafinado. La cantidad de RNA total extraído fue suficiente en todos los casos para poder realizar este estudio (Tabla 9).

**Tabla 9.** Valores de concentración de RNA total y RIN de cada una de las muestras analizadas.

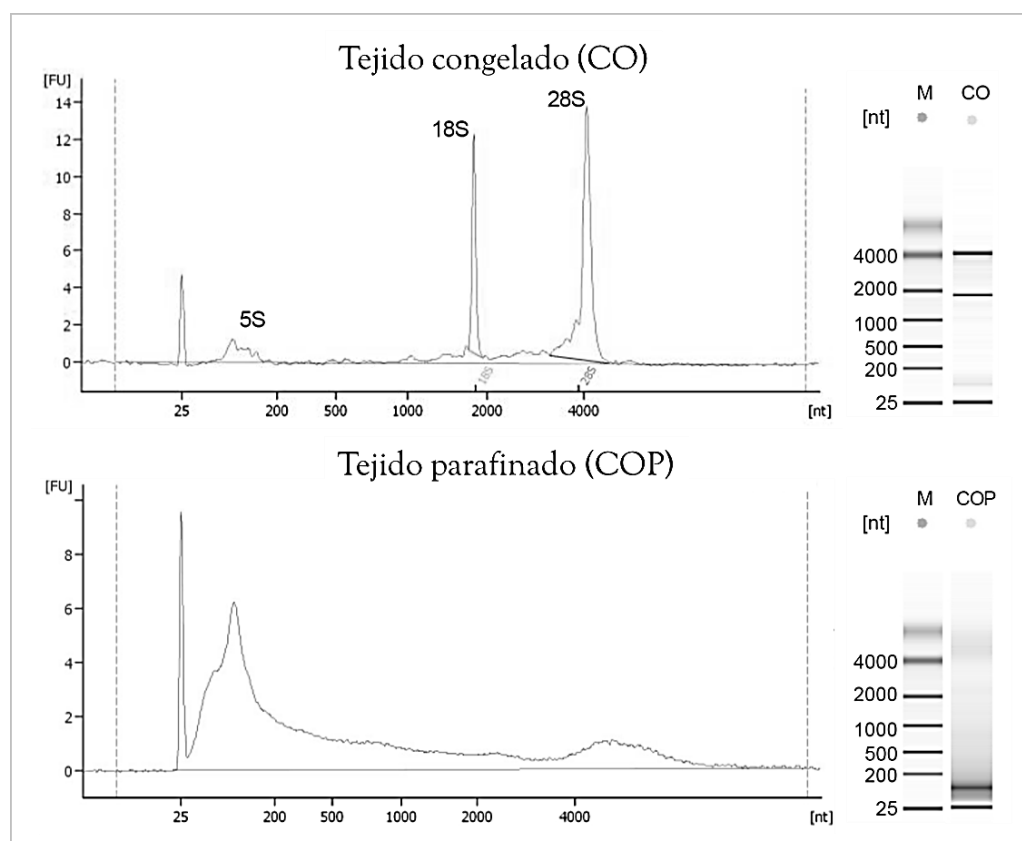
ID Muestra	[RNA] (ng/ $\mu\text{l}$ )	RIN	ID Muestra	[RNA] (ng/ $\mu\text{l}$ )	RIN
BC1CO	166,8	8,3	BC1COP	53,1	2,1
BC2CO	212,6	7,8	BC2COP	79,1	1,9
BC3CO	257,3	8,8	BC3COP	191,3	2,4
BC4CO	155,0	9,7	BC4COP	79,5	1,2
BC5CO	116,5	9,0	BC5COP	120,0	2,4
BC6CO	219,5	9,3	BC6COP	57,0	2,5
BC7CO	361,2	9,6	BC7COP	142,6	1,7
BC8CO	161,1	9,2	BC8COP	184,9	2,4
BC9CO	391,0	9,4	BC9COP	137,1	2,0
BC10CO	79,3	9,2	BC10COP	112,3	1,3
BC11CO	141,7	8,7	BC11COP	137,3	2,2
BC12CO	64,6	7,9	BC12COP	67,0	1,1
BC13CO	67,6	8,7	BC13COP	87,5	1,4
BC14CO	65,2	9,0	BC14COP	102,1	2,4

ID Muestra=identificación de la muestra; [RNA]=concentración de RNA expresada en ng/ $\mu\text{l}$ ; CO=congelado óptimo; COP=parafina del congelado óptimo.

Se analizó la integridad y calidad de este RNA total empleando el equipo Agilent 2100 Bioanalyzer, que permite hacer una estimación del nivel de integridad de una muestra de RNA total en base a un valor numérico, RIN (*RNA Integrity Number*), cuyo rango es de 1 a 10, indicando mayor integridad de la muestra cuanto más cercano a 10 es el valor. El RIN para cada una de las muestras, así como la concentración de RNA obtenida, se encuentran detallados en la Tabla 9.

El valor medio de RIN de las 14 muestras de tejido fresco congelado fue de 8,9 ( $\pm 0,6$ , rango entre 7,8 y 9,6), valor indicativo de muy alta calidad de las muestras. En cambio, el RIN medio de

sus correspondientes parafinas fue de 1,9 ( $\pm 0,6$ , rango entre 1,1 y 2,5), característico de muestras muy degradadas. Como cabía esperar, la calidad del RNA total aislado a partir de tejido fresco congelado y mantenido en condiciones óptimas fue muy superior a la calidad de las muestras descongeladas y parafinadas. En la Figura 13 se pueden ver los perfiles electroforéticos (electroferogramas) de un congelado óptimo junto al de su parafina asociada. En el Anexo III se muestran los perfiles electroforéticos de las muestras analizadas.



**Figura 13. Análisis de la integridad del RNA total.** En el electroferograma se muestran los perfiles electroforéticos e imágenes tipo electroforesis mostrando la integridad del RNA total extraído de una muestra congelada óptima (CO) y su correspondiente bloque de parafina (COP). M indica el patrón de tamaños empleado. En el eje de abscisas se representa la intensidad de fluorescencia emitida por la banda electroforética [FU]. En el eje de ordenadas se muestra el tamaño del fragmento de RNA recuperado en nucleótidos (nt). 5S, 18S y 28S hacen referencia a los fragmentos de RNA ribosomal del mismo nombre, cuya presencia da una medida de la integridad del RNA extraído.

En los perfiles electroforéticos de todas las muestras congeladas óptimas aparecen dos picos bien definidos correspondientes a los rRNAs 18S y 28S, e incluso se observa un pico adicional correspondiente al rRNA 5S, que indican que la integridad de estas muestras es muy buena. Por otro lado, en los perfiles de las muestras procedentes de bloques de parafina se observa la desaparición completa de estos picos ribosomales, que indica fuerte degradación. El perfil de muestras parafinadas muestra una acumulación de fragmentos de RNA de tamaño medio inferior a 200 nucleótidos (Ver Figura 13 y Anexo III).

## 1.2. Análisis de la integridad de los RNAs de pequeño tamaño

Para cada una de las muestras, se analizaron 10ng de RNA total con el ensayo *Agilent Small RNA Assay* (Agilent Technologies Inc., Santa Clara, CA, USA) empleando el equipo Agilent 2100 Bioanalyzer. Este ensayo permitió observar la distribución de las moléculas de RNA de

pequeño tamaño en el rango entre 6 y 150 nucleótidos, proporcionando además la concentración de moléculas de RNA de pequeño tamaño, la concentración de miRNAs y el porcentaje de microRNAs presentes en cada muestra, de acuerdo a la relación  $miRNA/small\ RNA$  [(moléculas de miRNAs) / (total de moléculas de RNA de pequeño tamaño)]. Todos estos valores se detallan en la Tabla 10.

**Tabla 10.** Valores de concentración de moléculas de RNA de pequeño tamaño.

ID Muestra	[pRNA] (pg/ $\mu$ l)	[miRNA] (pg/ $\mu$ l)	miRNA %	ID Muestra	[pRNA] (pg/ $\mu$ l)	[miRNA] (pg/ $\mu$ l)	miRNA %
1CO	666,6	132,4	20	1COP	5578,3	2276,6	41
2CO	1317,9	494,7	38	2COP	5527,9	2011,2	36
*3CO	4037,9	2393,1	59	3COP	3162,3	1100,5	35
4CO	1117,8	295,5	26	4COP	4398,9	1318,8	30
5CO	1024,3	280,5	27	5COP	2701,2	774,1	29
6CO	835,8	259,3	31	6COP	5012,8	1598,9	32
7CO	1104,1	225,7	20	*7COP	86643,8	32329,3	37
8CO	696,3	223,4	32	8COP	2311,7	717,5	31
9CO	827,1	204,7	25	9COP	1589,6	357,0	22
10CO	923,4	288,5	31	10COP	1482,1	341,3	23
11CO	610,2	180,2	30	11COP	1518,7	331,5	22
12CO	1234,4	562,2	46	12COP	1428,9	333,9	23
13CO	1222,7	489,4	40	13COP	1493,7	326,9	22
14CO	979,1	362,9	37	14COP	2421,7	606,8	25

ID Muestra=identificación de la muestra; pRNA=moléculas de RNA de pequeño tamaño; CO: congelado óptimo; COP: parafina del congelado óptimo.\*Valor eliminado en la determinación del valor medio por su elevada desviación. Concentración de miRNA y porcentaje de miRNAs determinados a partir de 10 ng de RNA total para cada una de las muestras analizadas.

Para una misma cantidad de RNA total de partida (10 ng), tanto la cantidad de pequeños RNAs como de miRNAs obtenida a partir de muestras parafinadas ([pRNA]<sub>media</sub> = 3837,4 $\pm$ 1856,4 pg y [miRNA]<sub>media</sub> = 1184,7 $\pm$ 685,5 pg) fue mayor que la obtenida a partir de muestras congeladas ([pRNA]<sub>media</sub> = 966,1 $\pm$ 229,9 pg y [miRNA]<sub>media</sub> = 307,6 $\pm$ 132,5 pg). En las muestras FFPE degradadas, las moléculas grandes se han perdido y aparecen fragmentos de RNA pequeños, productos de degradación, que pueden estar ocasionando una sobrestimación de la cantidad de moléculas de pequeño tamaño. Así que, para una misma cantidad de RNA de partida, siempre tendremos una mayor cantidad de fragmentos pequeños en muestras degradadas que en congeladas óptimas (56). En cambio, no se observan grandes diferencias de porcentaje de miRNAs recuperados (relación entre miRNAs/pequeños RNAs) entre los distintos tipos de muestras: 29 $\pm$ 5% para muestras parafinadas y 31  $\pm$ 8% para congeladas.

El análisis de RNAs de pequeño tamaño con el Agilent 2100 Bioanalyzer proporcionó el perfil electroforético para cada muestra, detectando distintas regiones: miRNAs (entre 10-40nt), tRNAs (entre 60-80 nt), rRNAs pequeños como el 5S y 5,8 S (entre 120-155 nt) y snoRNAs (entre 60-300 nt, aunque la escala llega hasta 150 nt). En el Anexo IV se pueden observar los perfiles electroforéticos de todas las muestra congeladas óptimas con sus correspondientes parafinas.

Las muestras congeladas óptimas presentan un perfil homogéneo entre ellas, se aprecian principalmente 3 picos con abundancia de moléculas de RNA, entre 60-80 nt, de alrededor de 90 nt y de cerca de 150 nt, correspondientes probablemente a tRNAs y rRNAs bien conservados, y a determinados snoRNAs.

En cambio, las parafinas muestran un perfil muy característico, con intensa señal de un único pico entre los 60-80 nt. La región comprendida entre los 10-40 nt es donde se encuentran los miRNA, el hecho de no ver un pico tan intenso en esta región indica que estos no son la población más abundante dentro de la región de pequeños RNAs, probablemente debido a que el análisis de pequeños RNAs se llevó a cabo con RNA total.

### 1.3. Análisis de la conservación de RNAs específicos

Se analizó mediante PCR cuantitativa en tiempo real (qRT-PCR) la recuperación de moléculas de RNA específicas en los distintos tipos de muestras (congelados y parafinas), partiendo de la misma cantidad de RNA total para cada una de las muestras. Los RNAs analizados fueron los miRNAs hsa-miR-21, hsa-miR-125b y hsa-miR191, el snoRNA RNU6B y los mRNAs GAPDH y HPRT1. Los valores de Ct medios y su desviación típica pueden encontrarse en el Anexo II.

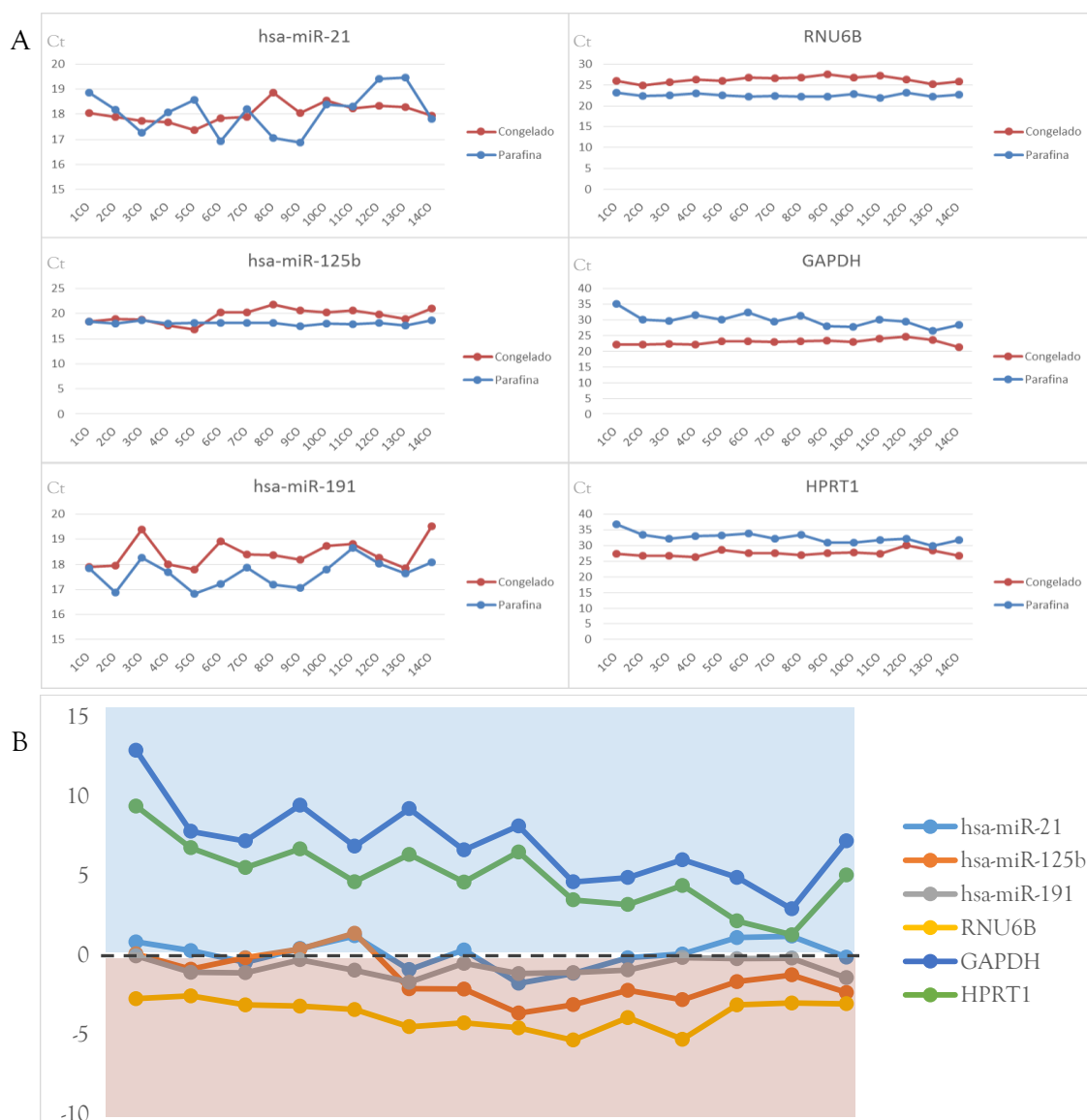
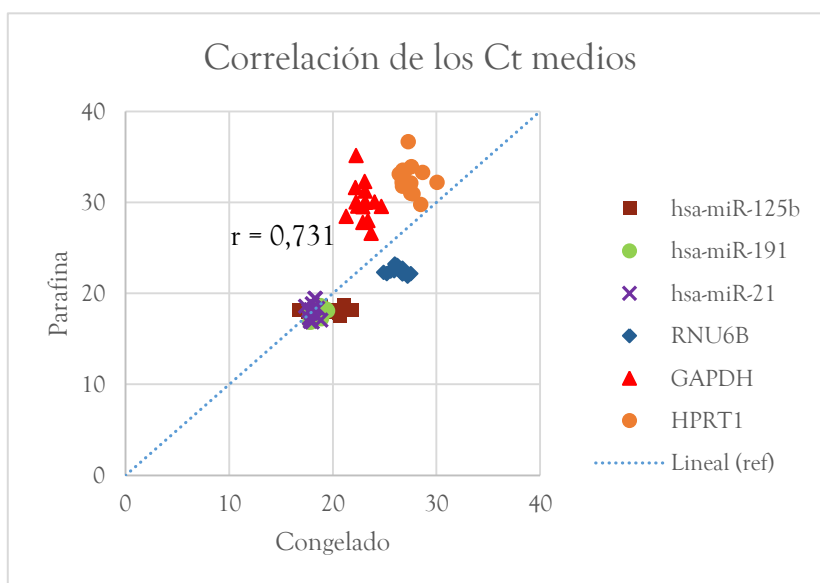


Figura 14. Valores de los Ct medios de las sondas estudiadas obtenidos para cada una de las muestras conservada en parafina y en congelación y la diferencia en Cts entre ambos formatos de congelación. **A**, Valores de los Ct obtenidos para cada una de las muestras. El eje de las Y representa los Ct, mientras que el X son las diferentes muestras. **B**, Representación de DcT (Ct medio muestra parafinada - Ct medio muestra congelada). Los valores en la zona azul indican diferencia positiva, es decir, mayores Ct en parafinas; en la zona roja se indican valores de diferencia de Cts negativos, mayores en muestras congeladas.

En la figura 14A se representan los valores de Ct medios en muestras congeladas y parafinadas, para las sondas empleadas. Puede observarse que, en general, los miRNAs se detectaron a Ct muy similares en todos los tipos de muestras. Por otro lado, los mRNAs se detectaron a Ct más bajos en muestras congeladas óptimas que en muestras sometidas a un proceso de degradación (parafinas). Y finalmente, el snoRNA se detectó un Ct menor en muestras parafinadas. Cabe destacar que un Ct menor implica mayor cantidad detectada y mejor extracción del RNA.

Con el fin de mostrar las diferencias entre muestras pareadas (congelado-parafina), se representó la diferencia del Ct medio de cada muestra congelada y su correspondiente muestra parafinada ( $DCt = Ct \text{ medio muestra parafinada} - Ct \text{ medio muestra congelada}$ ) (Figura 14B y Anexo IV).

Los resultados indican que para los mRNAs, GAPDH y HPRT1, el valor de Ct medio de cada muestra congelada óptima fue mucho menor, en general, que el de su correspondiente muestra parafinada, siendo el resultado de la diferencia positivo. En cambio, el comportamiento del snoRNA RNU6B fue completamente contrario: se detectó a un valor de Ct mucho más bajo en cada muestra procedente de parafina que en su correspondiente muestra congelada, siendo la diferencia un valor negativo. Los miRNAs se detectaron, en general, a valores de Ct muy similares entre las muestras congeladas y sus correspondientes muestras parafinadas, y por ello se encuentran alrededor del valor 0, aunque en general encontramos valores ligeramente más bajos en la muestra congelada.



**Figura 15. Correlación de los Ct entre las muestras de tejido conservado en parafina y su correspondiente par congelado.** Ct: ciclo umbral; r: coeficiente de correlación de Pearson. RNU6B: RNA nuclear de pequeño tamaño. GAPDH: Gliceraldehído-3-fosfato deshidrogenasa. HPRT1: hipoxantina-guanina fosforribosil-transferasa. Lineal(ref): referencia de correlación lineal positiva perfecta ( $r=1$ ).

Se analizó la correlación entre cada par de muestras, para lo que se representó el Ct medio de cada congelado óptimo con respecto al de su correspondiente muestra parafinada (Figura 15) para el total de moléculas de RNA analizadas. Se calculó el coeficiente de correlación de *Pearson* ( $r$ ) para un intervalo de confianza del 95% dando un p valor  $< 0,0001$  (test de dos colas) y una correlación estadísticamente significativa con  $r = 0,7312$  entre las muestras congeladas y sus correspondientes parafinas, siendo los valores de Ct correspondientes a los mRNAs *GAPDH* y *HPRT1* los que mostraron una peor correlación entre los pares congelado-parafina. En la Figura 18, la línea representa la correlación perfecta ( $r = 1$ ), los puntos cercanos a la línea de referencia

indican una buena correlación entre muestras pareadas, mientras que los puntos más alejados de la línea representan una correlación más débil entre las muestras. Considerando sólo los pequeños RNAs (los tres miRNAs y el snoRNA), la correlación entre muestras congeladas y sus parafinas mejoró, pasando de  $r = 0,7312$  a  $r = 0,9248$ .

Por otro lado, también se evaluó la variación de Ct medio de cada población muestral (congelados y parafinas). Para ello, se representó el Ct medio de los triplicados de cada una de las muestras agrupadas por población muestral para cada uno de los RNAs estudiados (Figura 16). A continuación, se realizó un test de normalidad y, por último, se llevó a cabo una prueba paramétrica del análisis de medias para dos grupos (t-test de muestras pareadas), ya que todas las poblaciones pasaron el test de normalidad. Se aplicó corrección de comparaciones múltiples de Bonferroni, estableciendo una significación estadística de p-valor  $< 0,008$ .

El análisis de medias mostró diferencias significativas entre las parafinas y congelado para los RNAs de pequeño tamaño *hsa-miR-125b* (p-valor 0,002, -1,48 IC95%-2,32 - -0,64), *hsa-miR-191* (p-valor  $< 0,0001$ , -0,79 IC95%-1,09 - -0,48) y *RNU6B*. Cabe señalar que para los miRNAs *hsa-miR-125b* y *hsa-miR-191* la diferencia de Ct medio entre parafinas y los otros dos grupos fue entre 1 y 2 veces menor para las parafinas. En cambio, esta diferencia de Ct medio en el caso del snoRNA *RNU6B* (p-valor  $< 0,0001$ , -3,74 IC95% -4,28 - -3,21) fue mucho más marcada, obteniéndose un Ct medio unas 4 veces menor en parafinas que en los otros dos grupos. En el caso de *hsa-miR-21* (p-valor 0,84, 0,048 IC95%-0,46 - 0,57) no se detectaron diferencias entre ninguno de los grupos. Respecto a los mRNAs, el estudio de Ct para el *GAPDH* (p-valor  $< 0,0001$ , 7,04 IC95% 5,61 - 8,48) mostró diferencias muy significativas entre las muestras congeladas y parafinas, con un Ct medio unas 7 veces inferior en congelados óptimos. Finalmente, para el mRNA de *HPRT1* (p-valor  $< 0,0001$ , 4,99 IC05% 3,77 - 6,21) se observaron diferencias significativas entre los grupos: el Ct medio de los congelados óptimos fue unas 5 veces inferior respecto al grupo de las parafinas.

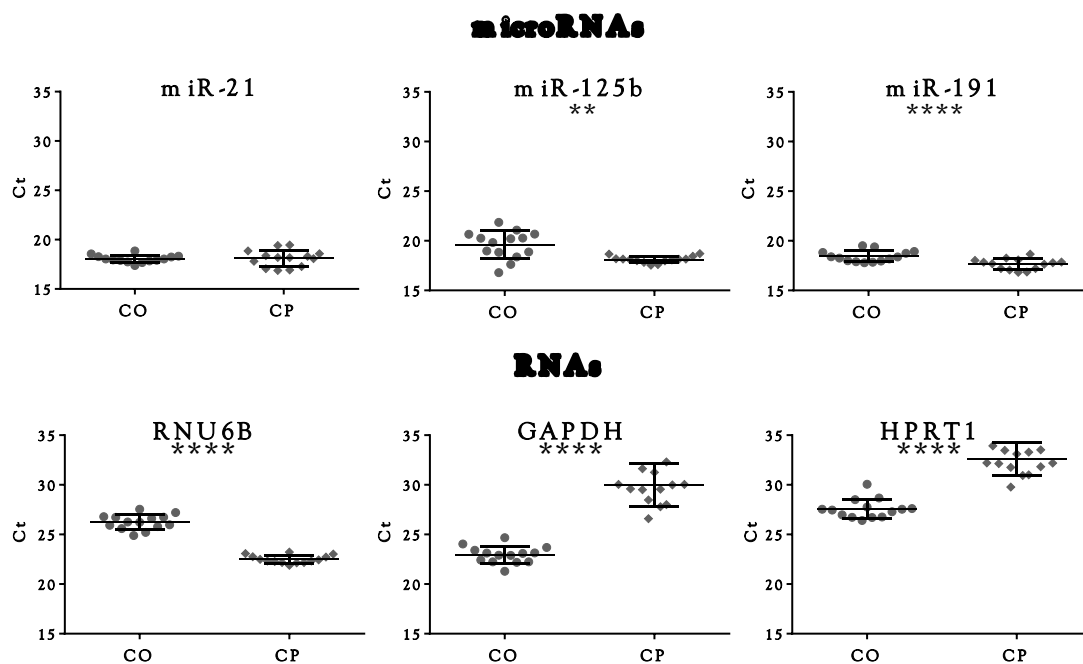


Figura 16. Diferencias en la recuperación de los RNA estudiados, medida por el Ct medio, en muestras congeladas y parafinadas. Las líneas marcan la media y la desviación estándar (SD). CO se refiere a las muestras congeladas y CP a las parafinadas. \*\*P-valor $<0,01$ . \*\*\*\*. P-valor  $< 0,0001$ .



# ANÁLISIS DEL PERFIL DE EXPRESIÓN DE MIRNAS EN CÁNCER DE MAMA DE MUJERES JÓVENES

## 1. ESTUDIO DE LA EXPRESIÓN DE LOS MIRNAS SOBRE MUESTRAS TUMORALES DE PACIENTES

### 1.1. Análisis de la intensidad de señal de los chips y control de calidad

Para llevar a cabo este estudio, se seleccionaron un total de 37 muestras, 22 mujeres menores de 35 años y 12 mayores de 65 años, empleamos 3 mujeres sanas como control para el análisis global de la expresión de miRNAs. Se procuró emparejar los grupos según su subtipo molecular, teniendo en cuenta también el grado de diferenciación, estadio, tamaño tumoral, afectación de los nodos linfáticos y sobreexpresión del factor de proliferación Ki67 (Tabla 7).

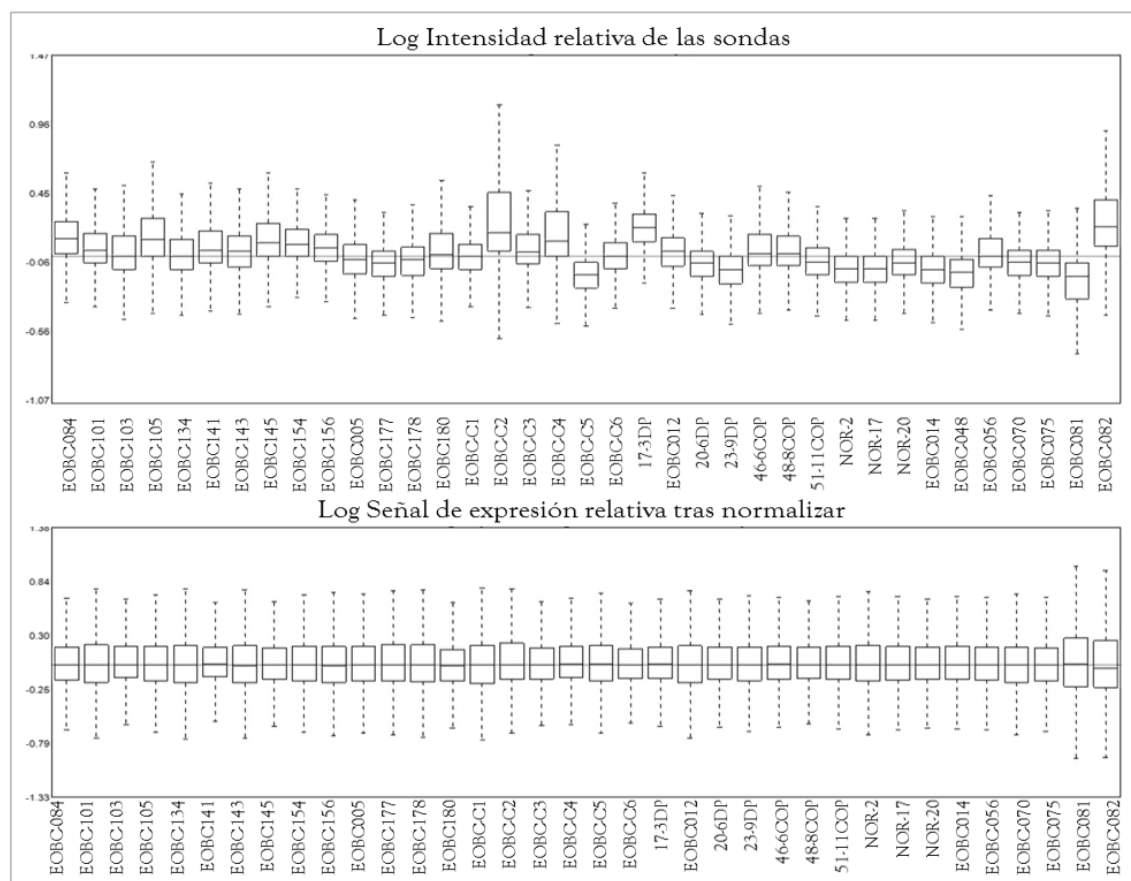
### 1.2. Análisis de la intensidad de señal de los chips y control de calidad

Los controles metodológicos y de calidad de la intensidad de señal se llevaron a cabo mediante el programa *Expression Console*, como se ha detallado en “Metodología”.



Figura 17. Histograma con los valores de intensidad obtenidos en cada uno de los chips de Affymetrix. A: sin normalizar. B: tras normalizar los valores mediante RMA-DABG. El eje de las X representa los rangos de intensidades del chip, el eje de las Y muestra el conteo de valores que se encuentran en dichos rangos. Obtenido mediante el programa *Expression Console* v.1.4.1 proporcionado por Affymetrix.

Los valores de intensidad obtenidos por cada chip se detallan en la Figura 17, donde se puede observar una distribución normal de las intensidades, con pico modal entre los valores 4,86 y 5,20, lo que nos indica que en este rango es dónde se sitúan la mayoría de valores del chip. El valor más bajo de intensidad es 4,392 y el máximo 15,998.



**Figura 18. Distribución de la intensidad de las sondas obtenida para cada chip antes y después de la normalización.** Diagrama de cajas que representa, para cada chip, el logaritmo de los valores de intensidad distribuidos en valor mínimo y máximo, primer y tercer cuartil y mediana. La línea roja marca la mediana de la intensidad entre todos los chips. Obtenido mediante el programa Expression Console v.1.4.1 proporcionado por Affymetrix.

Tras la normalización mediante RMA (*Robust Multi-array Average*) y DABG (*Detected Above Background*), el valor mínimo pasó a ser 0,519 y el máximo 14,056. Se tuvo que excluir del estudio el chip correspondiente a la muestra BCVY048, debido a que sus niveles de intensidad no se ajustaron a la distribución que el test RMA es capaz de procesar. Podemos observar en la Figura 18 que tras la normalización todos los chips tienen una distribución similar, garantizando que los resultados de intensidad obtenidos puedan ser comparables entre sí.

### 1.3. Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes

De los 1100 miRNAs humanos incluidos en el chip, tras el filtrado según controles de calidad, normalización y eliminación de aquellos con una expresión uniformemente baja o con una varianza menor a 1,5, nos quedamos con 251 microRNAs para su estudio.

De ellos, el test-t empleado para evaluar la diferencia de expresión de microRNAs de mujeres jóvenes frente a mujeres mayores, reveló 121 resultados significativos ( $p$  valor  $< 0,05$ ), de los que 91 aguantaron corrección por múltiples comparaciones de Benjamini & Hochberg (FDR). Los resultados están representados en la Figura 22 y los valores obtenidos para los microRNAs analizados se encuentran en el Anexo VI.



**Figura 19. Representación de la expresión de los microRNAs expresados en mujeres jóvenes respecto a mujeres mayores, ordenados por localización cromosómica.** El color rojo indica sobreexpresión en mujeres jóvenes, mientras que el verde indica represión en mujeres jóvenes. Los puntos grises corresponden a microRNAs cuya expresión no cambia.

Al representar los valores de expresión por orden cromosómico observamos que no existe agregación cromosómica y que la expresión de los miRNAs se distribuye homogéneamente por el espectro cromosómico. Aun así, existen zonas con tendencia a acumular miRNAs sobreexpresados o reprimidos. Los cromosomas 5, 10, 16, 18, 20 y 21 no contienen ningún miRNA reprimido y sí miRNAs sobreexpresados en mujeres jóvenes. Por otro lado, no se observan miRNAs sobreexpresados en los cromosomas 13, 17 y X. (Figura 19).

#### 1.4. *Análisis de las características tumorales como factor mediador en la significatividad obtenida.*

Se evaluó la distribución de las características tumorales en los diferentes grupos de estudio, para analizar la homogeneidad de la muestra y minimizar el efecto de un desequilibrio de una de las características en los resultados obtenidos (Figura 20).

Alternativamente, se realizaron test estadísticos de la misma manera que para la edad, teniendo en cuenta como variable dependiente las características tumorales por separado, de modo que, aplicando el concepto de mediación, si uno de los miRNAs, con resultados significativos al analizar los datos teniendo en cuenta la edad, exhibiese una expresión significativamente diferente para alguna de las características tumorales, esta significatividad atribuida a la edad podría ser resultado de un efecto mediador, en este caso la característica tumoral en cuestión.



Figura 20. Histograma de la distribución de las características tumorales de las pacientes estudiadas. Se representa el porcentaje de casos en cada una de las categorías por grupo de edad. ER+: casos positivos para los receptores de estrógenos; PR+: positivo para receptores de progesterona; HER2+: positivo para el receptor del factor de crecimiento epidérmico HER2. TN se refiere al subtipo triple negativo.

**Tabla 11.** Diferencias en la expresión de los miRNAs asociadas a las características tumorales estudiadas.

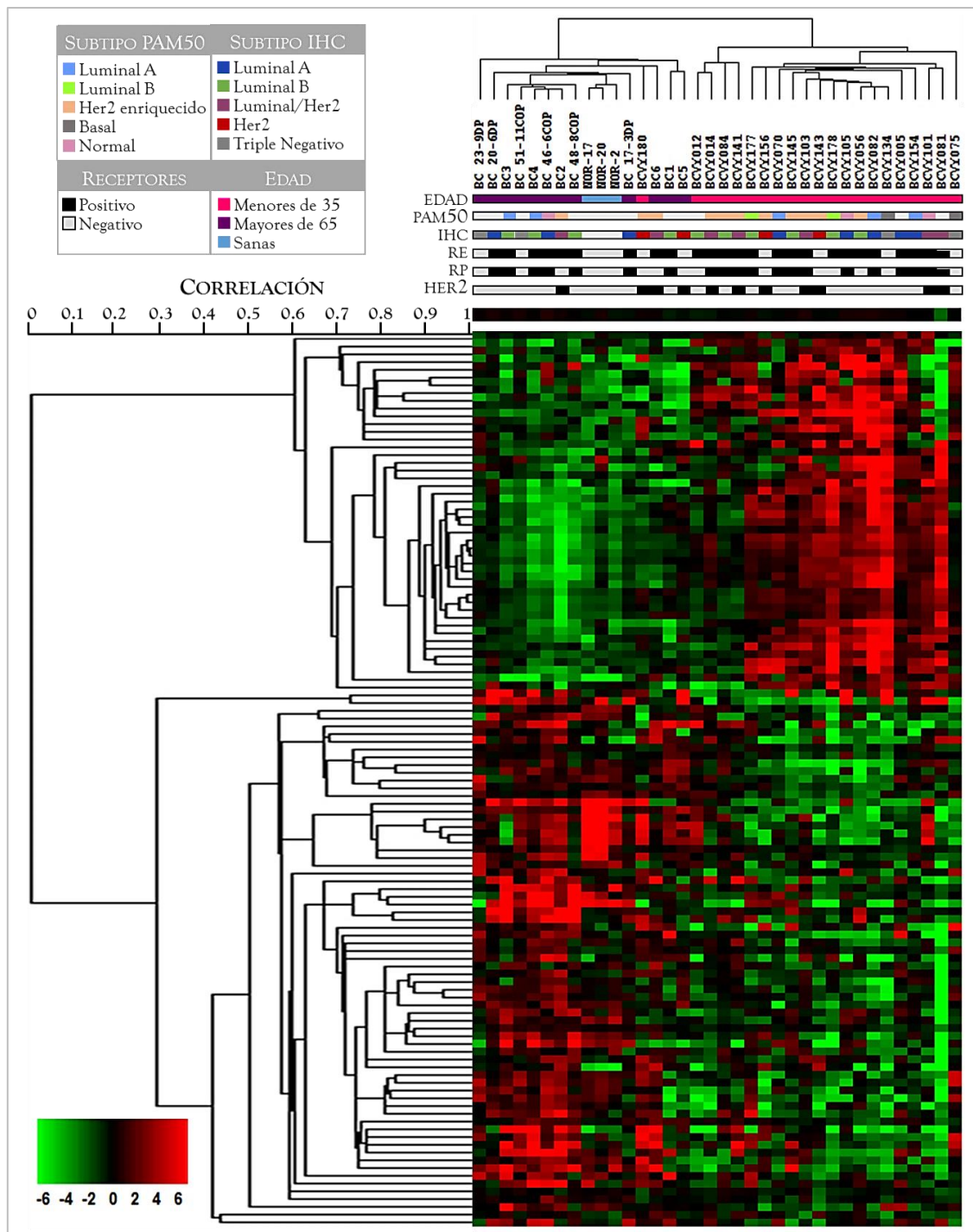
Tamaño tumoral				Grado histológico			
miRNA	p-valor	FDR	Expresión	miRNA	p-valor	FDR	Expresión
hsa-miR-145	9,48 x 10 <sup>-3</sup>	0,995	-2,731	hsa-let-7b	1,09 x 10 <sup>-3</sup>	0,274	8,061
hsa-miR-503	1,92 x 10 <sup>-2</sup>	0,995	2,443	hsa-let-7c	3,84 x 10 <sup>-3</sup>	0,461	7,270
hsa-miR-93	1,99 x 10 <sup>-2</sup>	0,995	2,437	hsa-miR-342-5p	5,50 x 10 <sup>-3</sup>	0,461	6,083
hsa-miR-30b*	3,64 x 10 <sup>-2</sup>	0,995	2,423	hsa-miR-342-3p	8,24 x 10 <sup>-3</sup>	0,517	5,613
hsa-miR-151-5p	2,56 x 10 <sup>-2</sup>	0,995	2,420	hsa-miR-98	1,58 x 10 <sup>-2</sup>	0,645	4,738
hsa-miR-342-3p	3,26 x 10 <sup>-2</sup>	0,995	-2,216	hsa-miR-99a	1,79 x 10 <sup>-2</sup>	0,645	4,596
Porcentaje de ki67				hsa-miR-193b*	1,80 x 10 <sup>-2</sup>	0,645	4,578
miRNA	p-valor	FDR	Expresión	Ganglios afectados			
hsa-miR-18a	8,60 x 10 <sup>-4</sup>	0,159	3,760	miRNA	p-valor	FDR	Expresión
hsa-miR-92a	1,27 x 10 <sup>-3</sup>	0,159	3,590	hsa-let-7b	1,09 x 10 <sup>-3</sup>	0,274	8,061
hsa-miR-505*	4,30 x 10 <sup>-3</sup>	0,360	3,063	hsa-let-7c	3,84 x 10 <sup>-3</sup>	0,461	7,270
hsa-miR-1301	1,10 x 10 <sup>-2</sup>	0,534	2,706	hsa-miR-342-5p	5,50 x 10 <sup>-3</sup>	0,461	6,083
hsa-miR-17	1,67 x 10 <sup>-2</sup>	0,534	2,525	hsa-miR-342-3p	8,24 x 10 <sup>-3</sup>	0,517	5,613
hsa-miR-98	2,08 x 10 <sup>-2</sup>	0,534	-2,435	hsa-miR-98	1,58 x 10 <sup>-2</sup>	0,645	4,738
hsa-miR-342-5p	2,16 x 10 <sup>-2</sup>	0,534	-2,419	hsa-miR-99a	1,79 x 10 <sup>-2</sup>	0,645	4,596
hsa-miR-664*	2,17 x 10 <sup>-2</sup>	0,534	-2,418	hsa-miR-193b*	1,80 x 10 <sup>-2</sup>	0,645	4,578
hsa-miR-503	2,34 x 10 <sup>-2</sup>	0,534	2,393	Tamaño tumoral			
hsa-miR-30a*	2,67 x 10 <sup>-2</sup>	0,558	-2,328	miRNA	p-valor	FDR	Expresión
hsa-miR-146b-5p	3,24 x 10 <sup>-2</sup>	0,608	2,239	hsa-miR-145	9,48 x 10 <sup>-3</sup>	0,995	-2,731
hsa-miR-29b-2*	3,70 x 10 <sup>-2</sup>	0,608	-2,179	hsa-miR-503	1,92 x 10 <sup>-2</sup>	0,995	2,443
hsa-let-7b	3,88 x 10 <sup>-2</sup>	0,608	-2,159	hsa-miR-93	1,99 x 10 <sup>-2</sup>	0,995	2,437
hsa-miR-200b	4,02 x 10 <sup>-2</sup>	0,608	-2,133	hsa-miR-30b*	3,64 x 10 <sup>-2</sup>	0,995	2,423
hsa-miR-1975	4,24 x 10 <sup>-2</sup>	0,608	-2,121	hsa-miR-151-5p	2,56 x 10 <sup>-2</sup>	0,995	2,420
hsa-miR-193b	4,95 x 10 <sup>-2</sup>	0,608	-2,041	hsa-miR-342-3p	3,26 x 10 <sup>-2</sup>	0,995	-2,216
hsa-miR-30a	4,98 x 10 <sup>-2</sup>	0,608	-2,045				

FDR se refiere al p-valor corregido para múltiples comparaciones de Benjamini & Hochberg.

Ninguno de los miRNAs significativamente diferentes en mujeres jóvenes produjo resultados significativos al realizar el correspondiente test estadístico con las características tumorales como variable, lo que indica que los resultados no son un artefacto debido a la distribución diferencial de las características tumorales en la población objeto de estudio. Los resultados obtenidos pueden observarse en la Tabla 11, donde se muestran los miRNAs para los que se obtuvieron valores de significación menores a 0,05; como puede observarse, ninguno de los miRNAs significativos asociados a características tumorales se mantuvieron significativos al aplicar la corrección por múltiples comparaciones de Benjamini & Hochberg (FDR).

### 1.5. Agrupación jerárquica de muestras y miRNAs obtenidos según el grupo de edad

Con los 121 microRNAs con expresión significativamente diferente en ambos grupos, se construyó el clúster no supervisado del perfil de expresión mediante agrupación jerárquica centrada en la mediana.



**Figura 21. Clustering jerárquico centrado en la mediana y visualización de la expresión de los miRNAs significativamente desregulados en mujeres jóvenes.** Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas por edad en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

En la figura 21 podemos ver como se separan claramente mujeres jóvenes de mayores en dos grupos, observamos también que las mujeres mayores son más similares al perfil de expresión del grupo de muestras de tejido normal de mujeres sanas, ya que se agrupan con las mayores. Se observa a primera vista dos grupos de expresión, el nodo 113 y el nodo 103, que dividen los microRNAs en aquellos sobreexpresados en mujeres jóvenes en la parte de arriba (en color rojo) y los reprimidos en mujeres jóvenes (abajo en color verde) (Tabla 12).

**Tabla 12.** Nodos desregulados en mujeres jóvenes seleccionados.

Sobre-expresados en mujeres jóvenes								
Nodo	Correlación	ID miRBase v.15	ID miRBase v.19	p-value	FDR	Expresión <sup>1</sup>		
Nodo 18	0,907	hsa-miR-1275	hsa-miR-1275	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	5,887		
		hsa-miR-1207-5p	hsa-miR-1207-5p	2,00 x 10 <sup>-4</sup>	1,78 x 10 <sup>-3</sup>	4,861		
		hsa-miR-4281	hsa-miR-4281	2,49 x 10 <sup>-4</sup>	1,93 x 10 <sup>-3</sup>	4,608		
		hsa-miR-1268	hsa-miR-1268a	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	5,013		
		hsa-miR-663	hsa-miR-663a	9,99 x 10 <sup>-5</sup>	9,99 x 10 <sup>-4</sup>	4,036		
		hsa-miR-149*	hsa-miR-149-5p	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	6,423		
		hsa-miR-762	hsa-miR-762	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	6,646		
		hsa-miR-3196	hsa-miR-3196	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	6,745		
		hsa-miR-1908	hsa-miR-1908	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	5,411		
		hsa-miR-1228*	hsa-miR-1228-5p	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	6,016		
		hsa-miR-3141	hsa-miR-3141	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	5,295		
		hsa-miR-92b*	hsa-miR-92b-5p	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	4,981		
		hsa-miR-1915	hsa-miR-1915-3p	5,99 x 10 <sup>-4</sup>	3,69 x 10 <sup>-3</sup>	3,836		
		hsa-miR-2861	hsa-miR-2861	9,99 x 10 <sup>-5</sup>	9,99 x 10 <sup>-4</sup>	4,355		
		hsa-miR-638	hsa-miR-638	4,99 x 10 <sup>-5</sup>	6,31 x 10 <sup>-4</sup>	4,691		
		hsa-miR-1469	hsa-miR-1469	7,83 x 10 <sup>-3</sup>	2,85 x 10 <sup>-2</sup>	2,771		
Nodo 60	0,723	hsa-miR-3172	hsa-miR-3172	2,81 x 10 <sup>-2</sup>	6,86 x 10 <sup>-2</sup>	2,322		
		hsa-miR-1274b	hsa-miR-1274b	1,72 x 10 <sup>-2</sup>	4,59 x 10 <sup>-2</sup>	2,567		
		hsa-miR-720	hsa-miR-720	9,08 x 10 <sup>-3</sup>	3,12 x 10 <sup>-2</sup>	2,897		
		hsa-miR-4284	hsa-miR-4284	2,04 x 10 <sup>-3</sup>	9,83 x 10 <sup>-3</sup>	3,143		
		hsa-miR-1973	hsa-miR-1973	1,09 x 10 <sup>-3</sup>	5,99 x 10 <sup>-3</sup>	3,948		
		hsa-miR-3175	hsa-miR-3175	9,99 x 10 <sup>-5</sup>	9,99 x 10 <sup>-4</sup>	4,867		
		hsa-miR-4299	hsa-miR-4299	9,99 x 10 <sup>-5</sup>	9,99 x 10 <sup>-4</sup>	5,776		
Reprimidos en mujeres jóvenes								
Nodo	Correlación	ID miRBase v.15	ID miRBase v.19	p-valor	FDR	Expresión <sup>1</sup>		
Nodo 48	0,754	hsa-miR-15b	hsa-miR-15b-5p	2,83 x 10 <sup>-2</sup>	6,86 x 10 <sup>-2</sup>	-2,003		
		hsa-miR-22	hsa-miR-22-3p	2,54 x 10 <sup>-2</sup>	1,11 x 10 <sup>-2</sup>	-2,446		
		hsa-miR-152	hsa-miR-152	7,49 x 10 <sup>-3</sup>	2,78 x 10 <sup>-2</sup>	-2,949		
		hsa-miR-28-5p	hsa-miR-28-5p	1,89 x 10 <sup>-3</sup>	9,49 x 10 <sup>-3</sup>	-3,137		
		hsa-miR-27b	hsa-miR-27b-3p	2,99 x 10 <sup>-3</sup>	2,18 x 10 <sup>-3</sup>	-3,637		
		hsa-miR-28-3p	hsa-miR-28-3p	4,99 x 10 <sup>-3</sup>	6,31 x 10 <sup>-4</sup>	-4,425		
		hsa-miR-29a	hsa-miR-29a-3p	2,31 x 10 <sup>-2</sup>	5,78 x 10 <sup>-2</sup>	-2,117		
		hsa-miR-195	hsa-miR-195-5p	4,84 x 10 <sup>-3</sup>	1,97 x 10 <sup>-2</sup>	-2,457		
		hsa-miR-26a	hsa-miR-26a-5p	5,49 x 10 <sup>-4</sup>	3,47 x 10 <sup>-3</sup>	-3,771		
		hsa-miR-30c	hsa-miR-30c-5p	3,49 x 10 <sup>-4</sup>	2,47 x 10 <sup>-3</sup>	-3,424		
		hsa-miR-30a	hsa-miR-30a-5p	3,42 x 10 <sup>-2</sup>	7,98 x 10 <sup>-2</sup>	-2,211		
		hsa-miR-26b	hsa-miR-26b-5p	1,51 x 10 <sup>-2</sup>	4,55 x 10 <sup>-2</sup>	-2,561		
		Nodo 51	0,751	hsa-miR-500	hsa-miR-500a-5p	1,34 x 10 <sup>-2</sup>	4,26 x 10 <sup>-2</sup>	-2,723
				hsa-miR-500*	hsa-miR-500a-3p	3,29 x 10 <sup>-3</sup>	1,41 x 10 <sup>-2</sup>	-3,287
hsa-miR-324-3p	hsa-miR-324-3p			1,71 x 10 <sup>-2</sup>	4,59 x 10 <sup>-2</sup>	-2,611		
hsa-miR-502-3p	hsa-miR-502-3p			1,44 x 10 <sup>-2</sup>	4,50 x 10 <sup>-2</sup>	-2,721		
hsa-miR-339-3p	hsa-miR-339-3p			2,15 x 10 <sup>-2</sup>	5,57 x 10 <sup>-2</sup>	-2,527		
hsa-miR-339-5p	hsa-miR-339-5p	1,31 x 10 <sup>-2</sup>	4,22 x 10 <sup>-2</sup>	-2,729				
Nodo 62	0,72	hsa-miR-139-5p	hsa-miR-139-5p	5,54 x 10 <sup>-3</sup>	2,22 x 10 <sup>-2</sup>	-3,131		
		hsa-miR-132	hsa-miR-132-3p	4,99 x 10 <sup>-3</sup>	6,31 x 10 <sup>-4</sup>	-5,572		
		hsa-miR-379	hsa-miR-379-5p	4,24 x 10 <sup>-2</sup>	9,36 x 10 <sup>-2</sup>	-2,140		
		hsa-miR-409-3p	hsa-miR-409-3p	1,64 x 10 <sup>-2</sup>	4,59 x 10 <sup>-2</sup>	-2,624		
		hsa-miR-134	hsa-miR-134	3,86 x 10 <sup>-2</sup>	8,83 x 10 <sup>-2</sup>	-2,199		
		hsa-miR-127-3p	hsa-miR-127-3p	2,01 x 10 <sup>-2</sup>	5,26 x 10 <sup>-2</sup>	-2,531		
		hsa-miR-433	hsa-miR-433	4,49 x 10 <sup>-2</sup>	9,73 x 10 <sup>-2</sup>	-2,149		
		hsa-miR-181a-2*	hsa-miR-181a-2-3p	9,28 x 10 <sup>-3</sup>	3,14 x 10 <sup>-2</sup>	-2,921		
Nodo 72	0,684	hsa-miR-23a	hsa-miR-23a-3p	2,49 x 10 <sup>-4</sup>	1,93 x 10 <sup>-3</sup>	-4,478		
		hsa-miR-183	hsa-miR-183-5p	1,09 x 10 <sup>-3</sup>	5,99 x 10 <sup>-3</sup>	-3,907		
		hsa-miR-421	hsa-miR-421	4,89 x 10 <sup>-2</sup>	1,04 x 10 <sup>-1</sup>	-2,107		
		hsa-miR-128	hsa-miR-128a-3p	3,25 x 10 <sup>-2</sup>	7,66 x 10 <sup>-2</sup>	-2,320		
hsa-miR-181d	hsa-miR-181d	2,29 x 10 <sup>-3</sup>	1,04 x 10 <sup>-2</sup>	-3,500				
Nodo 78	0,667	hsa-miR-30e*	hsa-miR-30e-3p	2,24 x 10 <sup>-3</sup>	1,04 x 10 <sup>-2</sup>	-3,490		
		hsa-miR-30c-2*	hsa-miR-30c-2-3p	1,71 x 10 <sup>-2</sup>	4,59 x 10 <sup>-2</sup>	-2,564		
		hsa-miR-30a*	hsa-miR-30a-3p	9,98 x 10 <sup>-4</sup>	5,85 x 10 <sup>-3</sup>	-3,776		
		hsa-miR-29c	hsa-miR-29c-3p	1,31 x 10 <sup>-2</sup>	4,22 x 10 <sup>-2</sup>	-2,740		
		hsa-miR-29b-2*	hsa-miR-29b-2-5p	7,54 x 10 <sup>-3</sup>	2,78 x 10 <sup>-2</sup>	-2,988		
Nodo 80	0,665	hsa-miR-361-5p	hsa-miR-361-5p	8,83 x 10 <sup>-3</sup>	3,12 x 10 <sup>-2</sup>	-2,862		
		hsa-miR-99b	hsa-miR-99b-5p	1,74 x 10 <sup>-2</sup>	4,60 x 10 <sup>-2</sup>	-2,594		
		hsa-miR-574-3p	hsa-miR-574-3p	3,99 x 10 <sup>-4</sup>	2,66 x 10 <sup>-3</sup>	-4,306		
		hsa-miR-125a-5p	hsa-miR-125a-5p	4,99 x 10 <sup>-3</sup>	6,31 x 10 <sup>-4</sup>	-4,764		
		hsa-miR-455-3p	hsa-miR-455-3p	3,64 x 10 <sup>-3</sup>	1,53 x 10 <sup>-2</sup>	-3,159		
		hsa-miR-23b	hsa-miR-23b-3p	2,49 x 10 <sup>-4</sup>	1,93 x 10 <sup>-3</sup>	-4,541		

FDR, p-valor tras la corrección por múltiples comparaciones de Benjamini & Hochberg, del inglés *false discovery rate*. 1. La expresión relativa del grupo de mujeres jóvenes respecto a la mediana global.

El nodo 103 contiene los microRNAs sobreexpresados en mujeres jóvenes. Presenta una correlación de 0,504, mostrándonos un grupo homogéneo y bien diferenciado, contiene los microRNAs con los resultados más significativos del test-t. Observamos que, a su vez, este nodo



consta de dos grupos diferenciados visiblemente diferentes, el nodo 60, con un buen valor de correlación de 0,723, en la parte superior del nodo; y el nodo 18, con un valor de correlación del 0,907, en el que observamos una gran homogeneidad de las muestras, así como de los miRNAs que conforman el nodo. La información relevante a estos nodos se encuentra en la Tabla 12.

Por otro lado, el nodo 113, con una correlación de 0,121, contiene los microRNAs menos expresados en mujeres jóvenes, en comparación con mujeres mayores y mujeres libres de cáncer. Se trata de un grupo menos homogéneo, lo que se refleja en su baja correlación, así como al visualizar la desviación de la mediana de los miRNAs. Se pueden identificar a su vez subgrupos dentro del mismo, detallados en las tablas 13 y 14. Seleccionamos los 8 nodos en función de varios criterios, como son la correlación obtenida, la homogeneidad como grupo y la diferenciación entre ambos grupos estudiados. Seleccionamos aquellos que mostraron un valor de correlación por encima de 0,65.

Los miRNAs que componen cada nodo se analizaron por separado para obtener una visión global de cada nodo, así como la posible función metabólica de cada uno de ellos. Con este fin se realizaron análisis de enriquecimiento de rutas con los miRNAs que componen cada nodo mediante DIANA miRPath.

### 1.6. *Análisis individual del enriquecimiento de rutas metabólicas en genes diana de los miRNAs que constituyen los nodos seleccionados*

A continuación se describen, en orden de aparición en el árbol de agrupación, cada uno de los nodos considerados, los miRNA que lo componen y las rutas enriquecidas obtenidas mediante los diferentes análisis con DIANA miRPath, considerando los posibles genes diana de los miRNAs incluidos y su papel en las rutas metabólicas (*gene union*).

#### Nodo 60

El nodo 60 es el primero que encontramos al analizar desde arriba el árbol, es uno de los dos sub-nodos que representan al grupo de los miRNAs sobreexpresados en mujeres jóvenes.

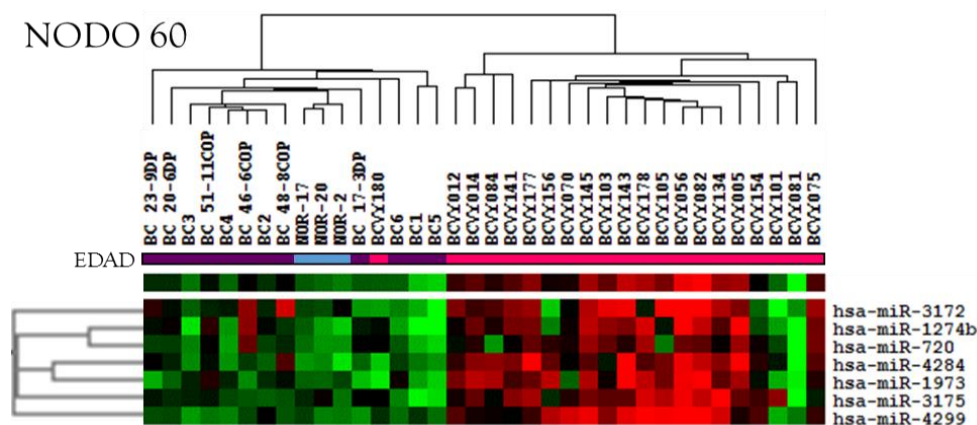


Figura 22. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 60. Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).



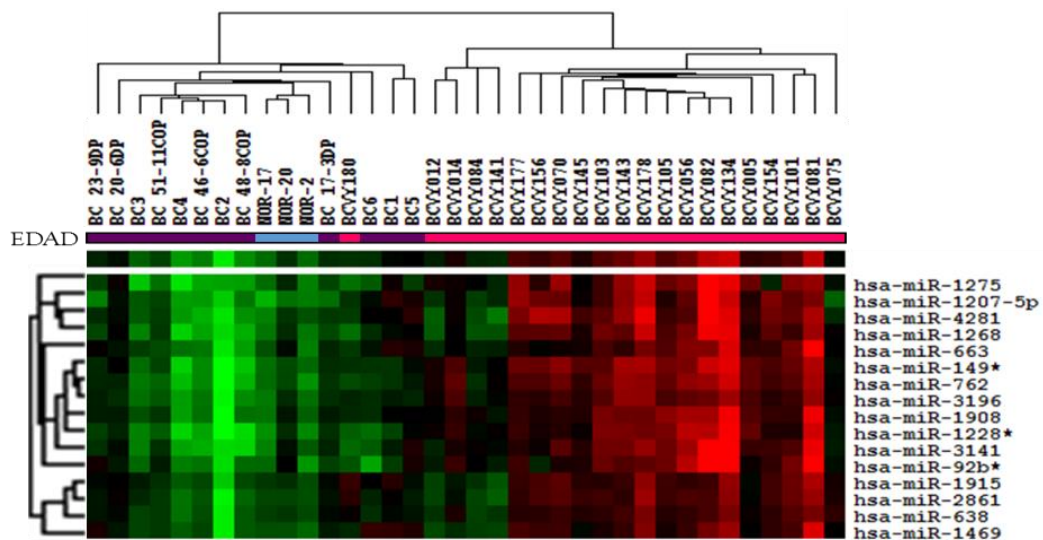
Este nodo muestra una correlación de 0,723, lo que indica que se trata de un nodo bastante homogéneo en cuanto a la expresión de los miRNAs. Los miRNAs que componen el nodo son hsa-miR-3172, hsa-miR-1274b, hsa-miR-720, hsa-miR-4284, hsa-miR-1973, hsa-miR-3175 y hsa-miR-4299.

Se observa una expresión generalizadamente mayor en mujeres jóvenes, a excepción de dos muestras jóvenes, BCVY101, y BCVY081, ambas muestras presentan un subtipo luminal/HER2, expresando tanto receptores hormonales como HER2, existen otras muestras jóvenes con este subtipo (BCVY014, BCVY141 y BCVY103) que se asemejan, sin embargo, al resto de las muestras del grupo (Figura 22). La expresión de la muestra joven BCVY180 se corresponde con las mujeres mayores, como ya se ha mencionado con anterioridad, por norma la expresión de esta muestra se parece más a la de mujeres mayores, agrupándose junto a éstas, por lo que se va a obviar en el resto de comentarios.

Al realizar el análisis de enriquecimiento de rutas de los genes diana se obtuvieron las rutas enriquecidas detalladas en el Anexo VII. El análisis se realizó con los miRNAs hsa-miR-720, hsa-miR-4284, hsa-miR-1973, hsa-miR-3175 y hsa-miR-4299, ya que los miRNAs hsa-miR-3172 y hsa-miR-1274b se han excluido de la base de datos de miRBase v. 19 al tratarse en realidad de tRNAs (RNA de transferencia). Entre las rutas enriquecidas se observan varias rutas sinápticas y neuronales, así como de transporte de RNA, rutas de señalización importantes como Wnt y ERBB, de síntesis de la matriz extracelular y una ruta directamente relacionadas con procesos cancerosos, como es la leucemia mieloide aguda. Cabe destacar que la mayoría de rutas están acopladas a la ruta de señalización del calcio.

## Nodo 18

Siguiendo en el grupo de miRNAs sobreexpresados en mujeres jóvenes, destaca el nodo 18 que, con una correlación de 0,924, es el grupo más homogéneo. En este nodo se encuentran los miRNAs con valores más elevados de significación del t-test entre mujeres jóvenes y mayores.



**Figura 23. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 18.** Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

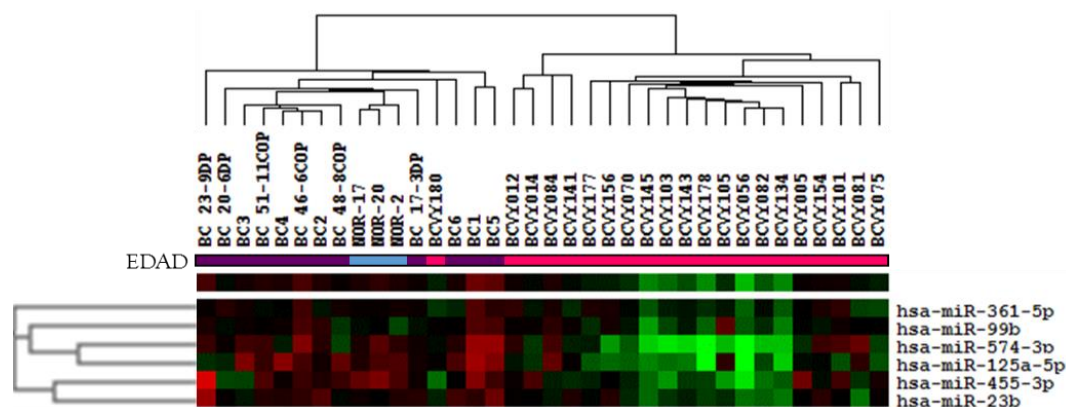
Se observa una expresión generalizadamente mayor en mujeres jóvenes en todas las muestras de los grupos a excepción de las muestras BCVY012, BCVY084, BCVY141 y BCVY075, que presentan una expresión menor a la mediana, más similar al grupo de mujeres mayores, las muestras BCVY012 y BCVY141 son luminal B, mientras que BCVY084 es luminal/HER2 y BCVY075 es triple negativa, por lo que no se observa ningún patrón diferente en cuanto a expresión de receptores. En cuanto al grupo de mujeres mayores, también observamos que existen dos muestras (BC1, luminal B, y BC5, HER2 positivo) que tienen la expresión en algunos de los miRNAs superior al resto de muestras de su grupo (Figura 23).

Se trata del grupo de miRNAs que más homogéneamente se expresa por muestra, aun así podemos diferenciar tres grupos, el primero formado por los miRNAs hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-4281, hsa-miR-1268a, el siguiente formado por hsa-miR-663a, hsa-miR-149-3p, hsa-miR-762, hsa-miR-3196, hsa-miR-1908, hsa-miR-1228-5p, hsa-miR-3141, hsa-miR-92b-5p; y por último el grupo formado por hsa-miR-1915-3p, hsa-miR-2861, hsa-miR-638 y hsa-miR-1469.

Al realizar el análisis de enriquecimiento de rutas de los genes diana se obtuvieron las rutas enriquecidas detalladas en la Tabla 7. Entre las rutas enriquecidas se observan varias rutas sinápticas y neuronales, al igual que en el nodo 60, además encontramos que la ruta *Guía de axones* tiene el p-valor más elevado. Encontramos también rutas de señalización canónicas importantes como MAPK, Hedgehog, PI3K/Akt, Wnt, HIF-1 y VEGF, alguna ruta relacionada con la movilidad celular, como son *Guía de axones* y *Regulación del citoesqueleto de actina*, y rutas directamente relacionadas con procesos cancerosos, como son Rutas en cáncer, *Cáncer de próstata*, *Carcinoma de célula basal* y *Desregulación transcripcional en cáncer*. Cabe destacar que, como en el nodo 60, muchas rutas están acopladas a la ruta de señalización del calcio, que, además, encontramos entre las rutas enriquecidas por este nodo con un p-valor de  $4,62 \times 10^3$ .

## Nodo 80

El nodo 80 es el primero de los nodos reprimidos en mujeres jóvenes seleccionados. Está compuesto por los miRNAs hsa-miR-361-5p, hsa-miR-99b, hsa-miR-574-3p, hsa-miR-125a-5p, hsa-miR-455-3p y hsa-miR-23b. Se puede observar en la Figura 24.



**Figura 24. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 80.** Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

Se trata de un nodo cuyos miRNAs poseen una elevada expresión en general, excepto en las muestras BCVY145, BCVY103, BCVY143, BCVY178, BCVY105, BCVY056, BCVY082 y BCVY134, en las que el total de los miRNAs que conforman el nodo se encuentran muy

reprimidos, agrupándose juntas. Estas muestras no pertenecen a tumores con el mismo subtipo, pero todas expresan receptor para estrógenos excepto BCVY143 y BCVY134.

Al realizar el análisis de enriquecimiento de rutas de los genes diana mediante estadísticos conservadores, se obtuvieron las rutas enriquecidas detalladas en el AnexoVI. Entre las rutas enriquecidas se observan principalmente rutas de señalización importantes como MAPK, Wnt, TGF- $\beta$ , mTOR, ERBB o por receptores tipo Toll; rutas relacionadas con síntesis de la matriz extracelular y de uniones celulares; también encontramos aquellas rutas directamente relacionadas con procesos cancerosos, como es rutas en cáncer, carcinoma de células renales o tipos de cánceres agresivos como son el de pulmón y páncreas. Se debe destacar la ruta del metabolismo del D-Glutamato y su implicación en la activación de la actividad celular y proliferación, por su asociación con la ruta de señalización del calcio, al ser un activador de los canales de calcio.

### Nodo 62

El nodo 62 se compone de los miRNAs hsa-miR-139-5p, hsa-miR-132, hsa-miR-379, hsa-miR-409-3p, hsa-miR-134, hsa-miR-127-3p, hsa-miR-433 y hsa-miR-151a-2\*. Con una correlación de 0,72, muestra una diferencia notable entre la expresión en las muestras de mujeres jóvenes y las de mujeres mayores y, especialmente, con aquellas libres de enfermedad.

Este nodo es particularmente interesante, debido a la elevada expresión de todos sus miRNAs en todas las muestras procedentes de tejido normal de pacientes sin cáncer (en especial las muestras NOR-17 y NOR-20), que, además, se corresponde con una expresión también elevada, aunque no tanto como en estas mujeres, de las muestras procedentes de mujeres mayores. En comparación, la expresión de estos miRNAs en mujeres jóvenes es notablemente menor, lo que identifica a este nodo como un candidato para establecer aquellos miRNAs y genes diana desregulados exclusivamente en mujeres jóvenes con cáncer de mama. Cabe destacar que la muestra BCVY180, que aun siendo joven es la única muestra que agrupa con las mujeres mayores, tiene unos niveles de sobre-expresión similares a los de las mujeres normales (Figura 25).

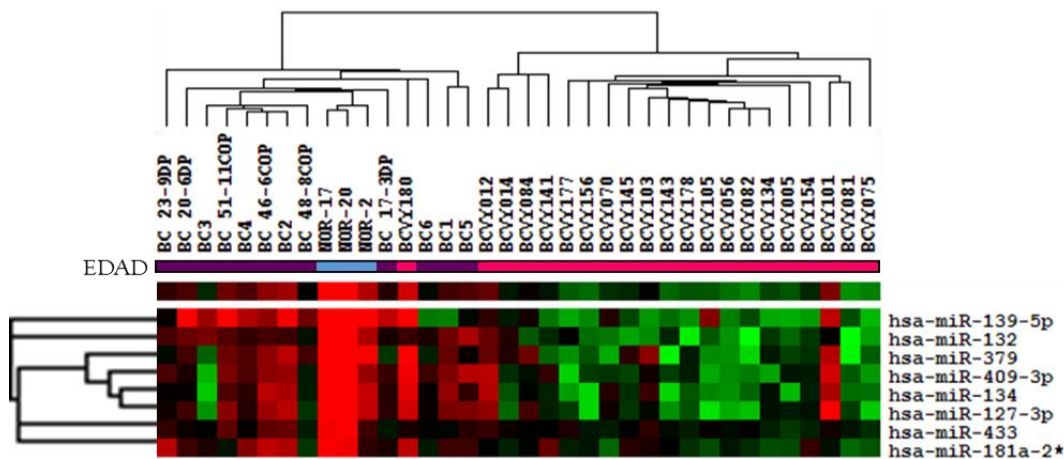


Figura 25. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 62. Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

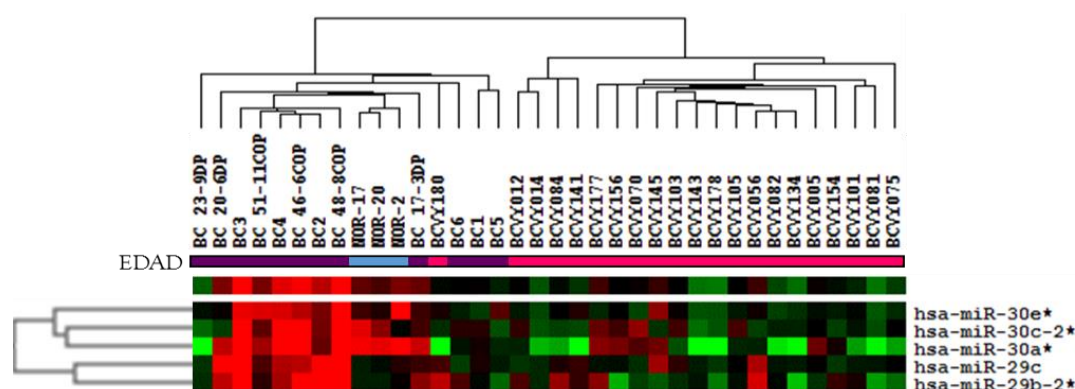
Tras el análisis de enriquecimiento de rutas por genes diana, se obtuvieron rutas de señalización como TGF- $\beta$ , PI3K/Akt, mTOR, ERBB, Wnt o P53, entre otras; rutas relacionadas con la adhesión y movilidad celular; rutas relacionadas con varios tipos de sinapsis y con procesos neuronales como guía de axones y potenciación y depresión a largo plazo; y rutas directamente

involucradas en procesos cancerosos como en leucemia mieloide aguda, Cáncer de pulmón, de páncreas, de próstata o de endometrio, por señalar algunos. La lista completa de rutas se puede observar en el Anexo VI.

## Nodo 78

Este nodo con correlación de 0,684 se encuentra enriquecido en la familia miR-30 (hsa-miR-30e\*, hsa-miR-30c-2a, hsa-miR-30a\*, hsa-miR-29c y hsa-miR-29b-2\*). Se encuentra más expresado en mujeres mayores, excepto en las muestras BC23-9DP, BC6, BC1 y BC5.

Se observa que la expresión de los miRNAs de la familia miR-30 en mujeres sin cáncer es similar a la de mujeres mayores, con una mayor expresión, sin embargo, de los miRNAs de la familia miR-29, las mujeres normales tienen una expresión menor, similar a la de mujeres jóvenes (Figura 26).



**Figura 26. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 78.** Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

Las rutas obtenidas mediante el análisis de enriquecimiento de genes muestran unos p-valores corregidos muy significativos. Se deben destacar las rutas relacionadas con adhesión celular, la interacción entre la matriz extracelular y el receptor ( $p$ -valor  $5,02 \times 10^{57}$ ) con 23 genes de la ruta enriquecidos y adhesiones focales ( $p$ -valor  $1,31 \times 10^{15}$ ), con 40 genes enriquecidos en los 5 miRNAs que componen el nodo.

Entre las rutas destacadas se encuentran otras relacionadas con señalización PI3K/Akt, MAPK, Wnt, con procesos cancerosos como cáncer de pulmón, de próstata o melanoma, entre otros y relacionadas con sinapsis y rutas neuronales como depresión y potenciación a largo plazo.

## Nodo 48

El nodo 48 tiene la mayor correlación dentro de los nodos reprimidos en mujeres jóvenes (0,754), en el cual se observa una menor expresión en mujeres jóvenes frente a mujeres mayores y normales (Figura 27).

Este nodo incluye dos sub-nodos diferenciados, el nodo 37, compuesto por hsa-miR-22, hsa-miR-152, hsa-miR-28-5p, hsa-miR-27b y hsa-miR-28-3p, con una correlación de 0,793; y el nodo 35 (correlación de 0,815) formado por hsa-miR-29a, hsa-miR-195, hsa-miR-26a, hsa-miR-30c y hsa-miR-30a. Quedarían excluidos de estos dos nodos los miRNAs hsa-miR-15b y hsa-miR-26b.

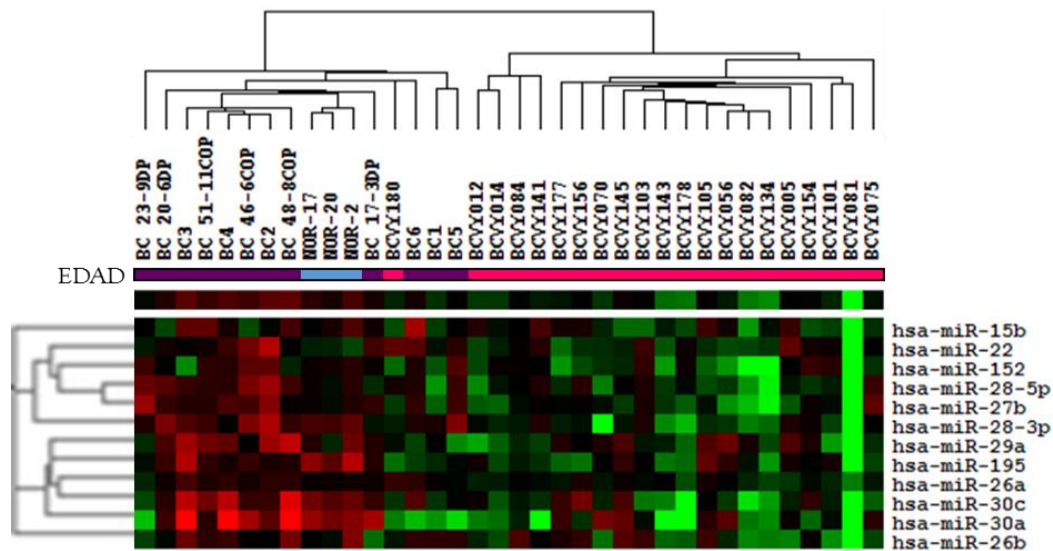


Figura 27. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 48. Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

Tras el análisis de enriquecimiento, se obtuvieron un gran número de rutas de señalización posiblemente desreguladas, en concreto *Adhesiones focales* obtuvo un p-valor corregido de  $2,90 \times 10^{-35}$ . Otras rutas también relacionadas con movilidad celular y adhesión mostraron p-valores corregidos muy significativos, como son *Interacción ECM-receptor* (p-valor  $3,60 \times 10^{-17}$ ) y *Regulación del citoesqueleto de actina* (p-valor  $1,37 \times 10^{-12}$ ).

Encontramos otras rutas relacionadas con señalización celular, como son PI3K/Akt, mTOR, p53, MAPK, ERBB, Wnt y TGF- $\beta$ , todas ellas obtuvieron valores de significación corregidos inferiores a  $1 \times 10^{-10}$ . También rutas relacionadas con varios tipos de sinapsis neuronales y con procesos como guía de axones, potenciación y depresión a largo plazo. También encontramos rutas directamente involucradas en procesos cancerosos como en cáncer de pulmón, de próstata, rutas en cáncer, melanoma, cáncer colo-rectal o glioma, entre otros. Las rutas obtenidas junto con sus valores de significación se encuentran listadas en el Anexo VI.

## Nodo 51

El nodo 51 contiene los miRNAs de la familia miR-500 y miR-339: hsa-miR-500, hsa-miR-500\*, hsa-miR-324-3p, hsa-miR-502-3p, hsa-miR-339-3p y hsa-miR-339-5p. Presenta un valor de correlación de 0,751 y muestra una clara diferencia de expresión entre ambos grupos. Todas las muestras del grupo de mujeres mayores, junto con las muestras normales, muestran una mayor expresión, excepto las muestras BC1 y BC5 y, ligeramente, la muestra BC17-3DP. Por su parte, las muestras procedentes de mujeres jóvenes muestran una expresión notablemente menor a la mediana, a excepción de las muestras BCVY177 y BCVY075, que muestran valores de expresión más similares a los encontrados en las muestras procedentes de mujeres mayores junto con las mujeres normales (Figura 28).



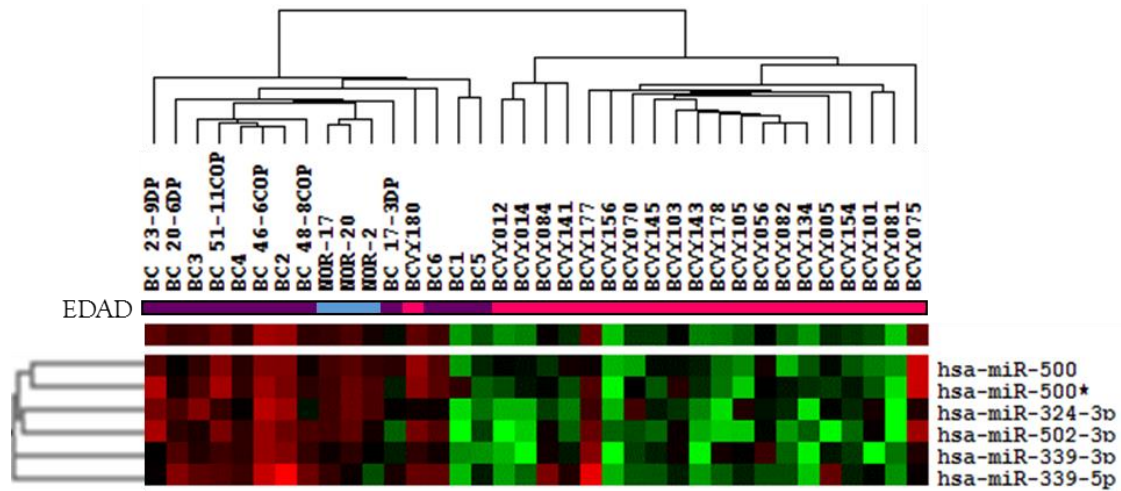


Figura 28. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 51. Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

El análisis de enriquecimiento de rutas de los genes diana de los miRNAs que componen el nodo 51, revelaron pocas rutas posiblemente relacionadas, de las cuales la más significativa fue infección por *Escherichia coli* ( $p$ -valor  $1,32 \times 10^5$ ). Cabe destacar la presencia de la ruta *Desregulación transcripcional en cáncer*, que obtuvo un  $p$ -valor corregido de  $3,20 \times 10^3$  y las rutas *Cáncer de pulmón* y *Leucemia mieloide crónica*, aunque el  $p$ -valor corregido por Benjamini & Hochberg alcanzado por estas rutas fue  $1,97 \times 10^2$  y  $2,28 \times 10^2$ , respectivamente. El resto de rutas no se consideraron relevantes desde un punto de vista traslacional.

### Nodo 72

El nodo 72 (correlación 0,684) lo componen los miRNAs hsa-miR-23a, hsa-miR-183, hsa-miR-421, hsa-miR-128 y hsa-miR-181d.

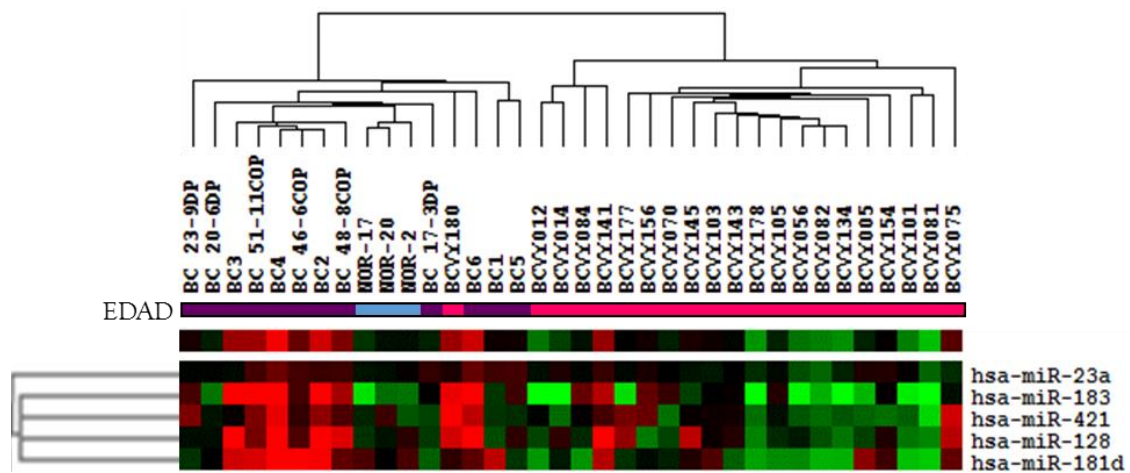


Figura 29. Expresión relativa centrada en la mediana de los miRNAs que componen el nodo 72. Se representa la expresión de los microRNAs, en vertical, por cada muestra analizada, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa), mayores de 65 años (morado) y libres de cáncer (azul).

Los valores de expresión en este nodo se encuentran menos diferenciados por edad, aunque sigue la tendencia de una menor expresión en mujeres jóvenes. En este caso, las sanas tienen una

expresión más similar a las mujeres jóvenes, al igual que las muestras BC20-6DP, BC17-3DP y BC1. Algunas muestras jóvenes tienen una expresión más alta que el resto, especialmente BCVY141 y BCVY075. La expresión más uniforme en ambos grupos y que muestra mayor diferenciación entre grupos, es la del miRNA hsa-miR-23a (Figura 29).

En cuanto a las posibles rutas desreguladas por este nodo, el análisis de enriquecimiento de los genes diana, la ruta con p-valor corregido más significativo fue potenciación a largo plazo (p-valor  $7,50 \times 10^{-13}$ ), encontramos otras rutas relacionadas con procesos neuronales, como varias sinapsis, señalización por neurotrofina y depresión a largo plazo. Entre las rutas relacionadas con señalización, destacamos la señalización mediante ERBB (p-valor  $7,33 \times 10^{-8}$ ), así como las rutas de señalización por mTOR, MAPK, insulina y p53, todas con p-valor  $< 1 \times 10^{-3}$  (Anexo VII).

### Clasificación de las rutas desreguladas según los genes diana

Las rutas obtenidas se clasificaron en categorías y se calculó el porcentaje de representación para cada una de esas categorías.

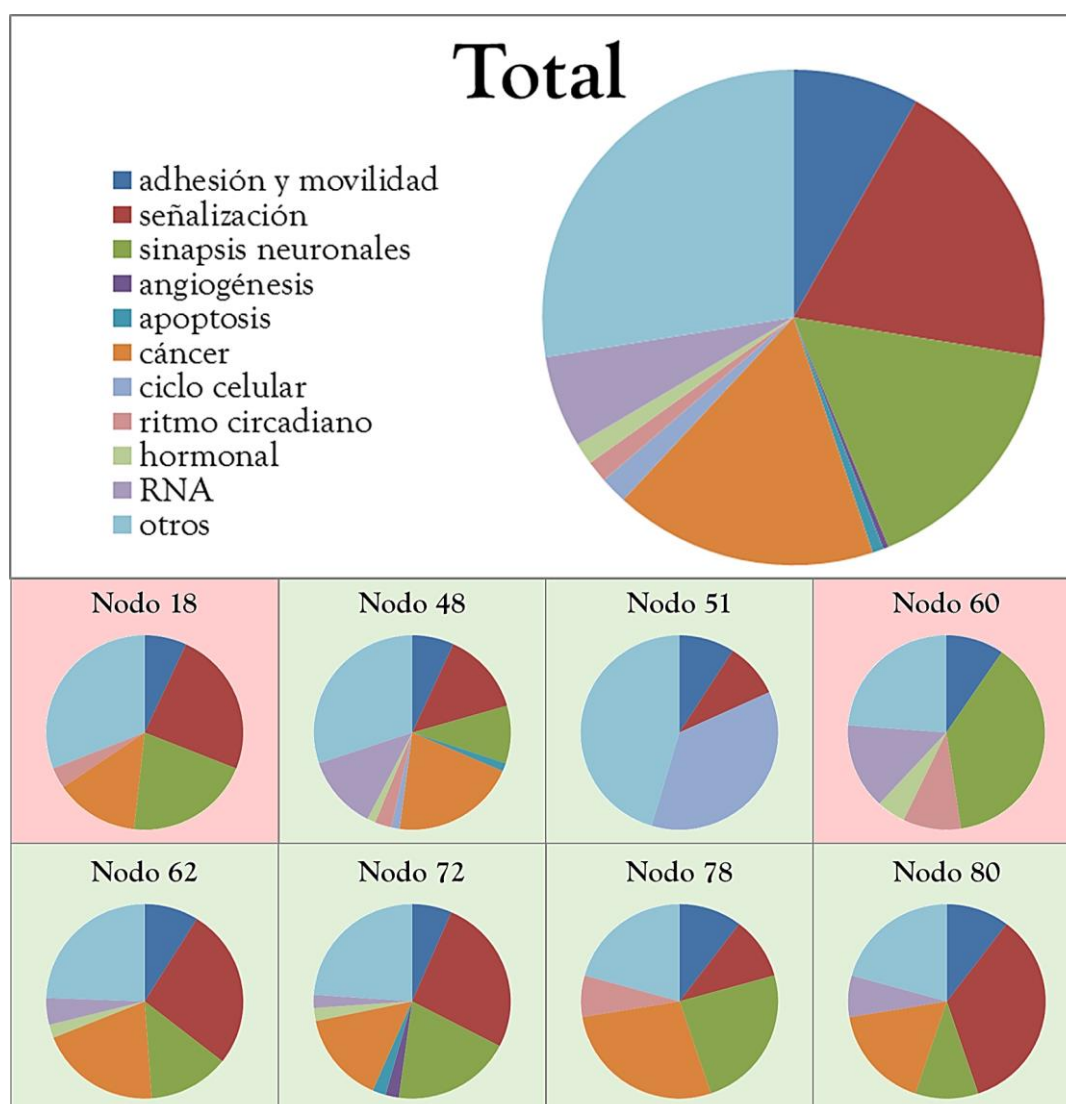


Figura 30. Distribución de las rutas asociadas a los genes diana predichos por DIANA miRPath de los miRNAs desregulados en total y por nodo seleccionado. Las gráficas correspondientes a nodos sobre-expresados y reprimidos en mujeres jóvenes se muestran con el fondo rojo y verde, respectivamente. Las rutas de KEGG obtenidas se categorizaron según su pertenencia a uno de los grupos representados en la leyenda. Se representa el número de rutas de una categoría respecto al total.

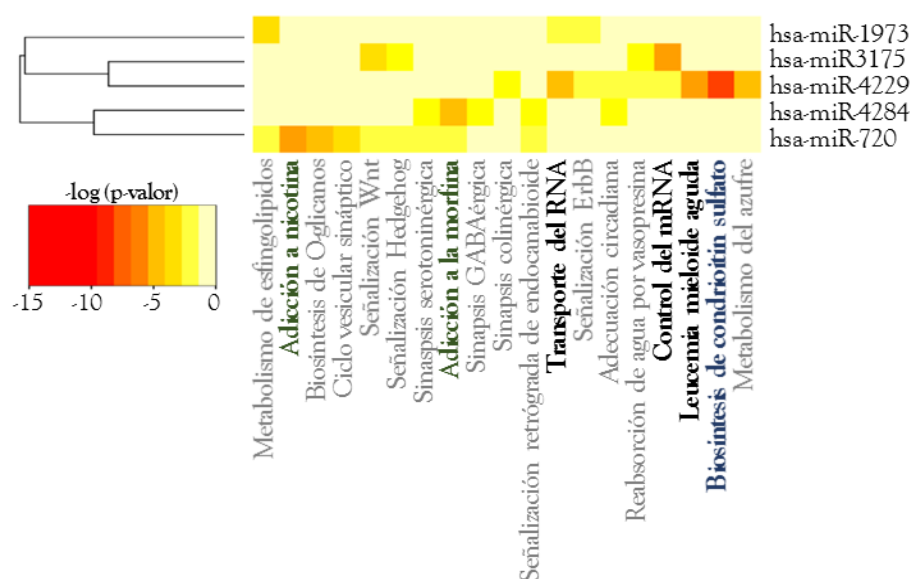
Encontramos en todos los análisis de enriquecimiento de rutas de los genes diana, rutas que pueden incluirse en cuatro categorías: relacionadas con rutas de señalización celular canónicas (19,43%), que se asocian a procesos de proliferación celular, apoptosis y diferenciación; relacionadas con procesos neuronales como sinapsis, guía de axones, potenciación y depresión a largo plazo y síntesis de neurotransmisores (16,25%); rutas relacionadas con adhesión, movilidad y migración celular (8,13%); y rutas relacionadas directamente con procesos cancerosos y tumorigénicos (16,96%), sobre todo encontramos cáncer de páncreas, pulmón, de células renales, melanoma y otros cánceres agresivos y metastásicos (Figura 30).

### 1.7. *Análisis de enriquecimiento de rutas en los miRNAs que componen cada nodo*

Para la selección de los miRNAs más relevantes, en primer lugar, se realizó el análisis de enriquecimiento de rutas por parte de los miRNAs directamente, obteniendo así qué miRNAs están más asociados a qué rutas de las obtenidas. La lista de las rutas y los p-valores asociados se encuentra en el Anexo VII.

#### Nodo 60

Respecto a los miRNAs del nodo 60, de las rutas obtenidas, sólo 6 se mantuvieron significativas tras aplicar estadísticos más conservadores, destacando la ruta de biosíntesis del condroitín sulfato y el papel de hsa-miR-4229, con un p-valor inferior a  $1 \times 10^{-16}$  (Figura 31).



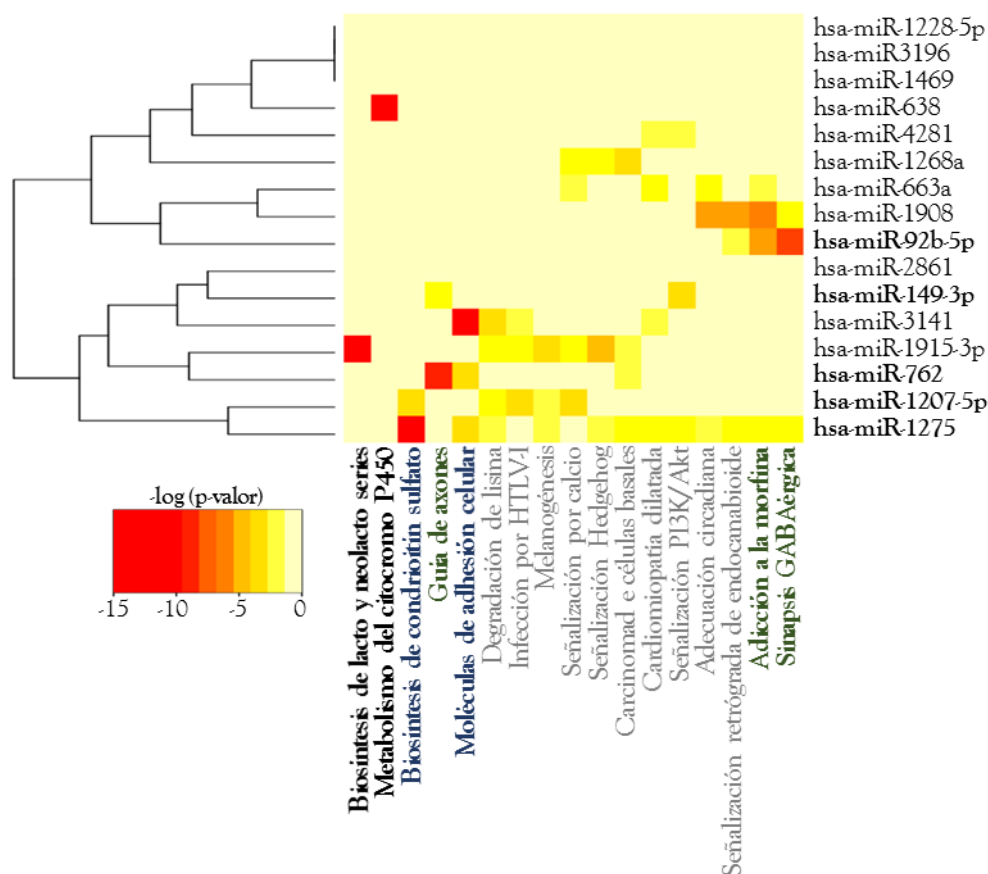
**Figura 31.** Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 60. El análisis se ha realizado mediante unión de rutas (pathway union) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul las relacionadas con matriz extracelular y uniones celulares.



## Nodo 18

Los miRNAs del nodo 18 se asociaron principalmente a rutas de síntesis de la matriz extracelular y adhesión celular, así como relacionadas con procesos neuronales. Se destaca el enriquecimiento en los miRNAs hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-762 y hsa-miR-92b-5p de dichas rutas, además, hsa-miR-149-3p es el miRNA con un p-valor más significativo asociado a la ruta de señalización PI3K/Akt (Figura 32).

## Nodo 18



**Figura 32 Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 18.** El análisis se ha realizado mediante unión de rutas (*pathway union*) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul las relacionadas con matriz extracelular y uniones celulares.

## Nodo 80

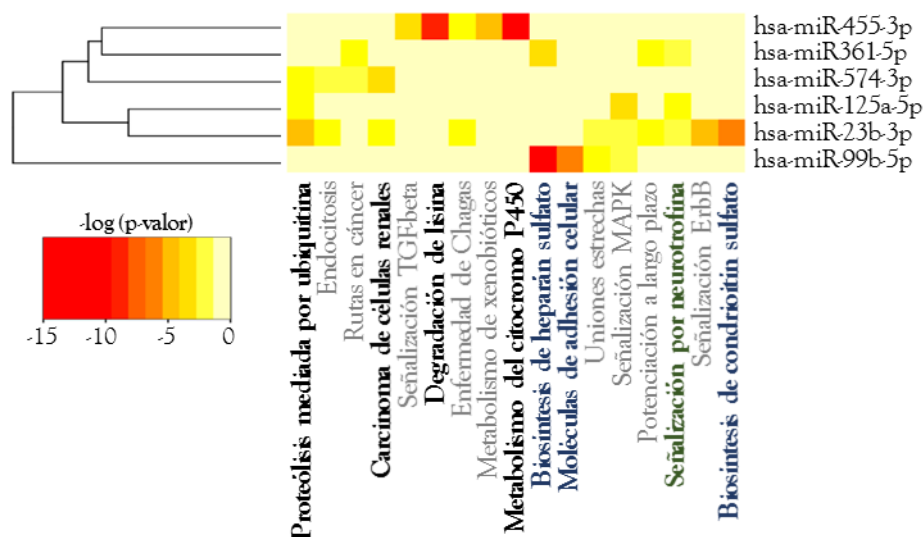
De las 17 rutas enriquecidas en los miRNAs del nodo 80, ocho mantuvieron resultados significativos mediante estadísticos conservadores. Se encontraron rutas relacionadas con la movilidad y adhesión y de señalización, además de, la ruta de señalización MAPK, asociada a hsa-miR-125a-5p. Destacar los miRNAs hsa-miR-23b-3p y hsa-miR-99b-5p, asociados principalmente a rutas de movilidad celular (Figura 33).

## Nodo 62

El análisis de enriquecimiento de rutas del nodo 62 apuntaron a las rutas de señalización de TGF- $\beta$  (p-valor de  $1,79 \times 10^{-4}$ ) y PI3K/Akt (p-valor de  $3,58 \times 10^{-2}$ ). Ocho de las 29 rutas aguantaron

estadísticos más conservadores. Destacamos entre las que no aguantan estos estadísticos las rutas de señalización de mTOR, Wnt y ERBB (p-valores de  $2,24 \times 10^{-2}$ ,  $2,82 \times 10^{-2}$  y  $3,46 \times 10^{-2}$ , respectivamente) el enriquecimiento en los miRNAs hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-762 y hsa-miR-92b-5p de dichas rutas, además, hsa-miR-149-3p es el miRNA con un p-valor más significativo asociado a la ruta de señalización PI3K/Akt (Figura 33).

### Nodo 80



### Nodo 62

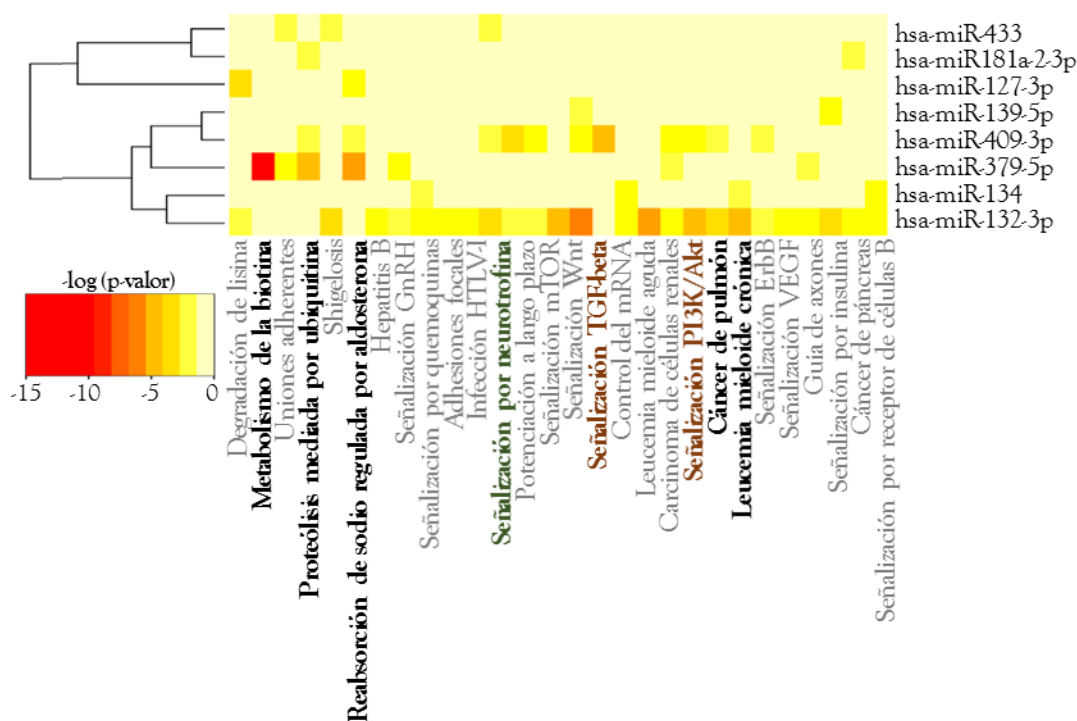


Figura 33. Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs de los nodos 80 y 62. El análisis se ha realizado mediante unión de rutas (*pathway union*) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul las relacionadas con matriz extracelular y uniones celulares.

## Nodo 78

De las 21 rutas de señalización obtenidas del nodo 78, 14 se mantuvieron significativas, entre las cuales se encontraron adhesiones focales e interacción receptor-ECM, asociada a hsa-miR-29; señalización PI3K/Akt asociada a los miRNAs hsa-miR-29c, hsa-miR-30a y hsa-miR-30e\* y sinapsis glutamatérgica asociada al hsa-miR-30c-2. Además se encontró a hsa-miR-30e\*, asociado a varias rutas directamente relacionadas con varios tipos de cáncer (Figura 34).

## Nodo 78

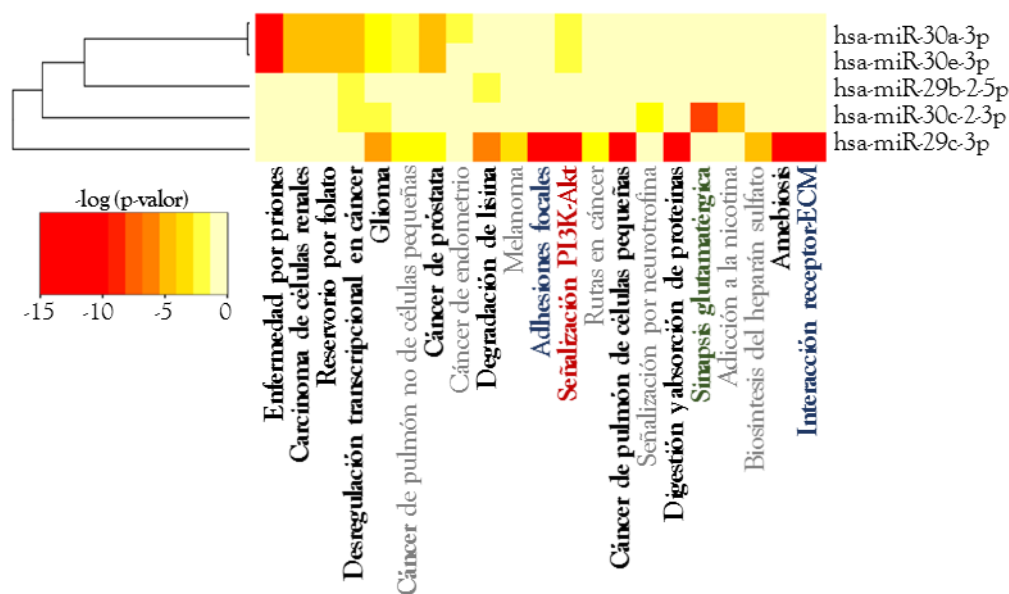


Figura 34. Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 78. El análisis se ha realizado mediante unión de rutas (*pathway union*) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul las relacionadas con matriz extracelular y uniones celulares y en marrón las relacionadas con señalización celular.

## Nodo 48

El análisis de enriquecimiento de rutas del nodo 48 dio como resultado 48 rutas significativas tras aplicar métodos conservadores, indicando que se trata de un nodo cuyos miRNAs cuentan con numerosas dianas y que su desregulación puede ser relevante para la célula. Entre las rutas enriquecidas destacan varias relacionadas con matriz extracelular y adhesión, destacando biosíntesis de condroitín sulfato (p-valor  $\times 10$ ), interacción receptor-matriz extracelular, biosíntesis de heparán sulfato y adhesiones focales, asociados principalmente a hsa-miR-29a, además de hsa-miR-195, hsa-miR-27b y hsa-miR-159.

Entre las rutas relacionadas con procesos neuronales destacan potenciación y depresión a largo plazo (con los miRNAs hsa-miR-30a y hsa-miR-30c, y hsa-miR-27), regulación por neurotrofina (hsa-miR-30a, hsa-miR-30c, hsa-miR-29, hsa-miR-27 y hsa-miR-195) sinapsis dopaminérgica (hsa-miR-30a, hsa-miR-30c y hsa-miR-27) y por último guía de axones con fuerte presencia de hsa-miR-30a y hsa-miR-30c. En cuanto a las relacionadas con rutas de señalización destaca la ruta de señalización PI3K/Akt, enriquecida en todos los miRNAs del nodo excepto hsa-miR-28-3p y hsa-miR-28-5p. Además encontramos la ruta de señalización de ERBB y otras como MAPK, mTOR, Wnt y TGF- $\beta$ .

Destaca el papel de hsa-miR-27b, hsa-miR-29a, hsa-miR-30a, hsa-miR-30c y hsa-miR-195. Se seleccionó también a hsa-miR-28-5p ya que los miR-28 fueron los menos correlacionados con el resto del nodo y, de ellos el hsa-miR-28-5p mantenía asociación con más rutas (Figura 35).

### Nodo 48

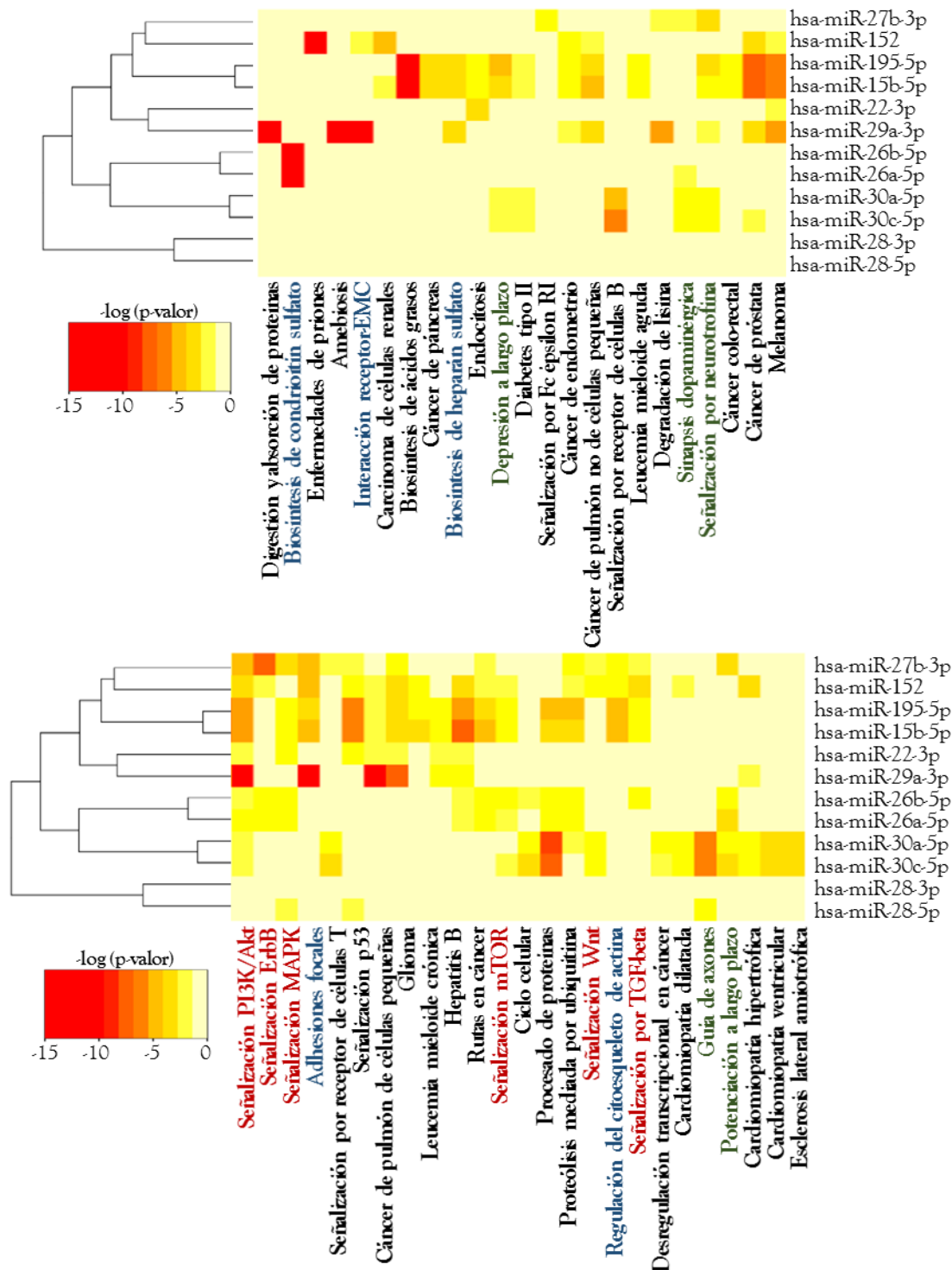
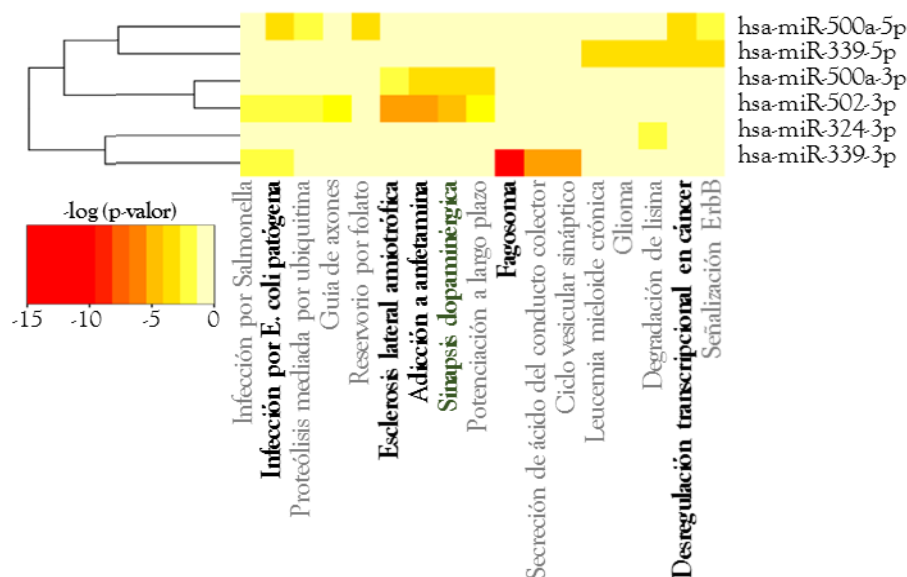


Figura 35. Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs del nodo 48. El análisis se ha realizado mediante unión de rutas (*pathway union*) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul con matriz extracelular y uniones celulares y en rojo con señalización celular.

## Nodo 51

El análisis de enriquecimiento de rutas del nodo 51 mostró 17 rutas enriquecidas, de las que 6 se mantuvieron significativas. De ellas destaca únicamente la sinapsis dopaminérgica. Aunque no se mantuvieron significativas tras analizar los resultados mediante estadísticos conservadores, destacan entre las rutas *Guía de axones* y *Señalización ERBB* (Figura 36).

### Nodo 51



### Nodo 72

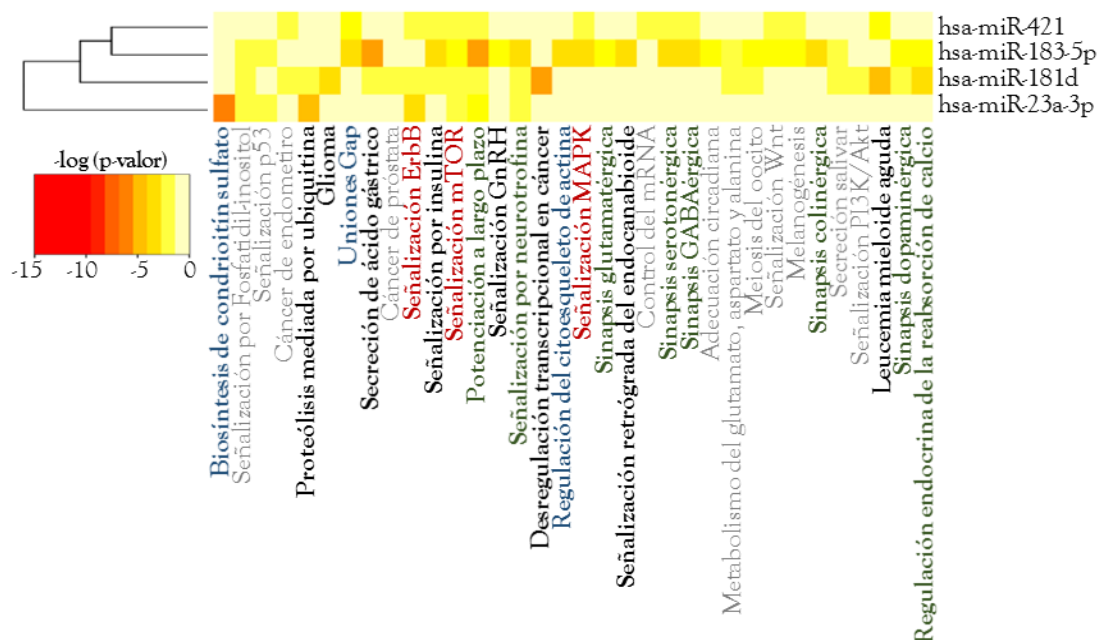


Figura 36. Heat-map del p-valor corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs de los nodos 51 y 72. El análisis se ha realizado mediante unión de rutas (*pathway union*) realizado mediante DIANA miRPath. Se representa el inverso del logaritmo en base 10 del p-valor en escala de colores desde el blanco al rojo. Las rutas remarcadas son aquellas que se mantienen significativas tras aplicar pruebas estadísticas más conservadoras. En verde oscuro se marcan aquellas relacionadas con procesos neuronales, en azul las relacionadas con matriz extracelular y uniones celulares y en marrón las relacionadas con señalización celular.

## Nodo 72

El análisis de enriquecimiento en los miRNAs del nodo 72 en las rutas metabólicas produjo 34 resultados significativos, de los cuales 22 se mantuvieron significativos mediante estadísticos más conservadores. La ruta con mayor valor de significación obtenido fue la biosíntesis de condroitín sulfato (p-valor corregido  $< 10^{-15}$ ), enriquecida en hsa-miR-23a-3p (Figura 36).

### Selección de los miRNAs más relevantes para su estudio como marcadores moleculares

La selección de los miRNAs se realizó teniendo en cuenta tres criterios de selección; en primer lugar se seleccionaron aquellos más relevantes según las rutas enriquecidas en dichos miRNAs de las seleccionadas como relevantes desde un punto de vista traslacional; se tuvo en cuenta el valor de significación obtenido en el test-t; por último, se llevó a cabo una búsqueda bibliográfica para hallar aquellos miRNAs cuyo papel en procesos cancerosos había sido confirmado en otros trabajos previamente y podían tener valor transcripcional. Los criterios de selección se encuentran en la Tabla 13.

**Tabla 13.** Selección de miRNAs para la validación y los criterios empleados.

Nodo	miRNA	Significación	Representación rutas	Relevancia traslacional	Otros
Nodo 18	hsa-miR-92b-5p	✓	✓	✓	
	hsa-miR-149-5p	✓	✓		
	hsa-miR-1207-5p	✓	✓		
	hsa-miR-1228-5p	✓		✓	
	hsa-miR-1275	✓	✓		
	hsa-miR-3196	✓			
Nodo 80	hsa-miR-125a-5p		✓	✓	
Nodo 62	hsa-miR-132-3p	✓	✓		Sobre-expresión en mujeres sanas
	hsa-miR-139-5p			✓	
	hsa-miR-379-5p			✓	
	hsa-miR-409-3p		✓		
	hsa-miR-433		✓		
Nodo 48	hsa-miR-27b-3p	✓	✓		
	hsa-miR-195-5p		✓	✓	
	hsa-miR-30c-5p	✓	✓		
	hsa-miR-28-5p	✓			Representación del nodo
Nodo 72	hsa-miR-23a-3p	✓	✓		
Nodo 78	hsa-miR-30e*	✓	✓		

Los cuadros en gris con el símbolo “✓” resaltan los criterios seguidos para la selección de cada miRNA.

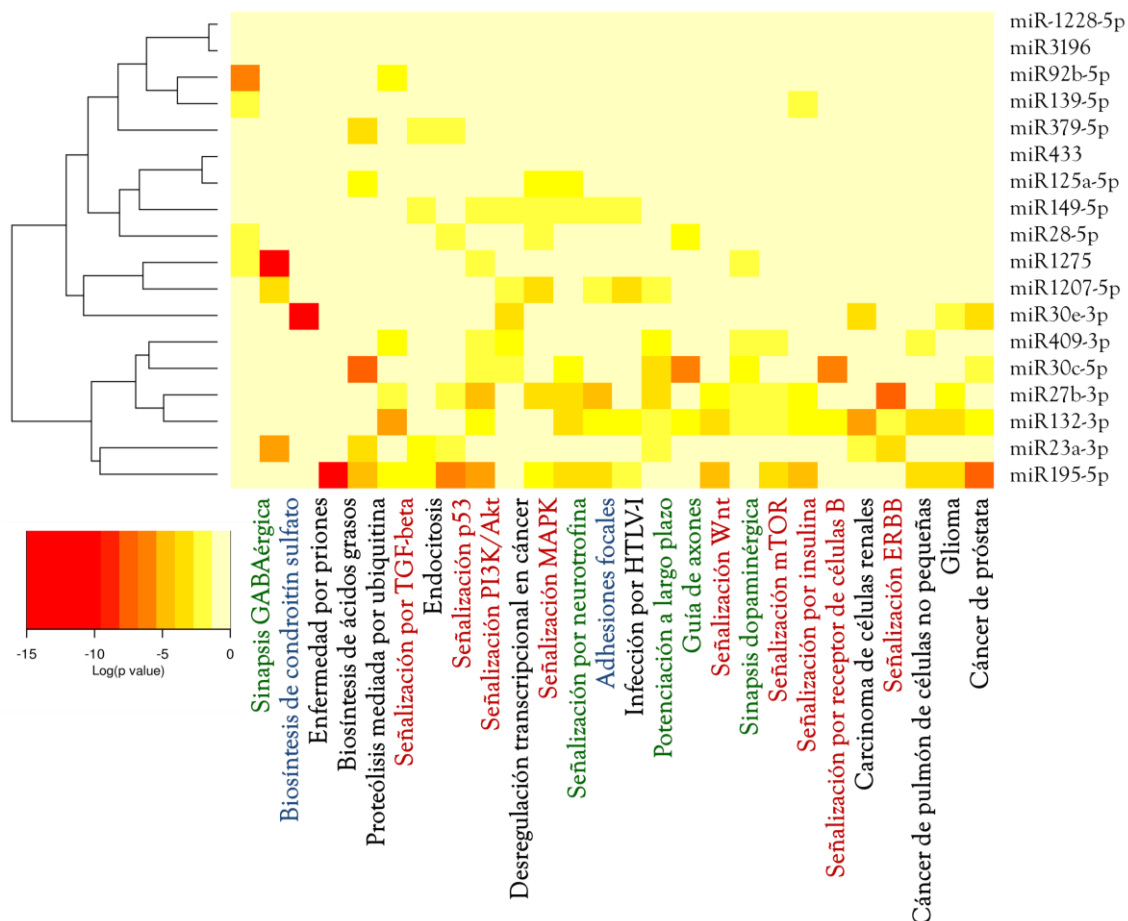
Los miRNAs seleccionados se encuentran enriquecidos en las mismas rutas que consideramos relevantes en el análisis de enriquecimiento de rutas de los miRNAs que componen los distintos nodos; esto quiere decir que con los miRNAs seleccionados reconstruimos las rutas enriquecidas en la totalidad los miRNAs desregulados en el perfil molecular diferencial de mujeres jóvenes con cáncer de mama (Tabla 14, Figura 37 y Anexo XVIII). A excepción de los miRNAs hsa-miR-433, hsa-miR-1228-5p y hsa-miR-3196, que fueron seleccionados por su elevado p-valor en el test de expresión diferencial en mujeres jóvenes, en el caso de hsa-miR3196 y hsa-miR-1228-5p, así como su implicación probada en cáncer de mama y en otros cánceres (217-219).

**Tabla 14.** Rutas significativamente enriquecidas en los genes diana de los miRNAs seleccionados.

KEGG ID	Ruta KEGG	p-valor ajustado	Nº genes	Nº miRNAs
hsa04722	Señalización por neurotrofina	3,08E-16	44	12
hsa04151	Señalización por PI3K/Akt	4,34E-14	87	14
hsa04510	Adhesiones focales	3,46E-12	58	14
hsa04150	Señalización por mTOR	4,22E-12	25	11
hsa04720	Potenciación a largo plazo	2,63E-11	26	11
hsa04910	Señalización por Insulina	4,55E-11	42	11
hsa05211	Carcinoma de células renales	1,70E-10	27	9
hsa04120	Proteólisis mediada por ubiquitina	4,19E-10	42	10
hsa05223	Cáncer de pulmón no de células pequeñas	4,59E-10	21	10
hsa04010	Señalización por MAPK	6,32E-10	68	13
hsa05215	Cáncer de próstata	1,03E-09	30	9
hsa05220	Leucemia mieloide crónica	2,92E-09	25	8
hsa05222	Cáncer de pulmón de células pequeñas	4,52E-09	28	12
hsa04012	Señalización por ERBB	7,90E-08	29	11
hsa05212	Cáncer de páncreas	1,03E-07	23	10
hsa05213	Cáncer de endometrio	1,87E-07	19	9
hsa04370	Señalización por VEGF	2,59E-07	22	11
hsa04360	Guía de axones	2,96E-07	40	14
hsa04960	Reabsorción de sodio regulada por aldosterona	3,37E-07	15	8
hsa05161	Hepatitis B	5,68E-07	43	11
hsa04914	Maduración del oocito mediada por progesterona	8,49E-07	26	11
hsa05221	Leucemia mieloide aguda	9,65E-07	19	9
hsa04930	Diabetes tipo II	1,39E-06	17	8
hsa05200	Rutas en cáncer	7,26E-06	77	13
hsa00310	Degradación de lisina	9,59E-06	16	10
hsa04728	Sinapsis dopaminérgica	9,59E-06	34	12
hsa00061	Biosíntesis de ácidos grasos	1,17E-05	1	1
hsa05210	Cáncer colorrectal	1,73E-05	20	9
hsa04115	Señalización por p53	2,71E-05	21	13
hsa04912	Señalización por GnRH	2,80E-05	25	11
hsa04664	Señalización por Fc epsilon RI	4,73E-05	21	9
hsa04380	Diferenciación del osteoclasto	5,52E-05	34	10
hsa04350	Señalización por TGF-β	7,30E-05	25	11

Se muestran las rutas con un p-valor < 10<sup>-4</sup>. Se empleó un valor umbral en microT-CDS de 0,9.





**Figura 37. Heat-map del p-valor aplicando estadísticos conservadores y corregido mediante FDR del análisis de enriquecimiento de rutas de los miRNAs seleccionados.** El análisis se ha realizado mediante unión de rutas con DIANA miRPath. Se representa el -logaritmo del p-valor en escala de colores del blanco al rojo. En verde oscuro se marcan rutas relacionadas con procesos neuronales, en azul con matriz extracelular y uniones celulares y en rojo con señalización celular.

### 1.8. *Análisis de enriquecimiento de rutas en bases de datos online de dianas del total de miRNAs desregulados*

Implementando la herramienta online EnrichR, se llevó a cabo el análisis de la desregulación de rutas global sobre dianas validadas (miRTarBase) y teóricas (microT-CDS), empleando en un primer lugar la totalidad de los genes diana y después tras seleccionar aquellos con sentido biológico y desregulación real, seleccionando aquellos cuyo valor absoluto de represión debida a los miRNAs se encontraba por encima del 75%, es decir, aquellos genes más afectados por la desregulación de los miRNAs (Anexo IX).

#### **Enriquecimiento KEGG y REACTOME a partir del global de los genes diana validados**

Se obtuvieron 64 de los 88 miRNAs con dianas validadas en miRTarBase, obtenemos 6955 dianas, con las que se realizó el análisis de enriquecimiento tanto en KEGG como en Reactome.

Empleando KEGG se obtuvieron únicamente tres rutas con p-valor ajustado menor a 0,05: “ribosoma”, “ciclo celular” y leucemia mieloide crónica”, sin embargo, en Reactome se obtuvieron 109 rutas con p-valor ajustado significativo (Anexo X). De ellas la gran mayoría están relacionadas con el control del ciclo celular (30,91%), con el procesado, regulación, transporte y degradación del RNA (25,45%) y un 16,36% se relacionan con rutas de señalización como Wnt, Hedgehog y



PI3K, entre otras, encontramos también rutas relacionadas con uniones y movilidad celular y con apoptosis, con una presencia del 7,27% y del 6,36%, respectivamente.

### Enriquecimiento KEGG y REACTOME a partir de los genes diana validados más afectados

Tras obtener las dianas validadas de los 64 miRNAs desregulados con dianas validadas, se seleccionaron aquellos cuyo valor de represión se encontraba por encima del tercer cuartil y, por tanto, estaban más afectados por la desregulación de miRNAs. Se puede observar en el histograma de frecuencias del Anexo II que el grueso de los valores está en torno a 0, por lo que, pese a que haya muchos miRNAs desregulados y muchos genes diana, realmente se ven afectados una parte mucho menor. Así, de las 6955 dianas que obtuvimos anteriormente, 1321 fueron seleccionadas según este criterio.

Empleando KEGG se obtuvieron 15 rutas enriquecidas, la mayoría relacionadas con procesos cancerosos, destacan señalización por ERBB, adhesiones focales y uniones adherentes (Tabla 15).

Empleando la base de datos Reactome se obtuvieron 121 rutas enriquecidas con p-valor ajustado significativo. Entre ellas destacan rutas asociadas al procesado, transporte y traducción del RNA (27,27%); relacionadas con receptores RTK (14,05%), de los cuales la familia de EGFR representó el 58,82%; las rutas relacionadas con el ciclo celular y su progresión representaron un 10,74% del total de rutas; un 9,9% de las rutas estaban asociadas a respuesta inmune, principalmente mediante receptores tipo Toll; en menor medida se encontraron rutas relacionadas con apoptosis, reparación del DNA, MAPquinasas y respuesta a estrés, cada una con menos de un 3% de representación del total de rutas. La lista completa de las rutas obtenidas, su categorización así como sus p-valores se encuentra detallada en el Anexo XII.

**Tabla 15.** Rutas KEGG enriquecidas en genes diana validados afectados por los miRNAs seleccionados por encima del 75%.

Ruta KEGG	Nº genes/ Total	p-valor	p-valor ajustado*	Z-score	Score combinado
Cáncer de próstata	28/86	1,40E-06	1,38E-04	-2,058	18,298
Ciclo celular	31/104	1,81E-06	1,38E-04	-1,824	16,218
Cáncer de páncreas	23/73	1,95E-05	9,86E-04	-1,903	13,175
Leucemia mieloide aguda	23/75	2,78E-05	1,06E-03	-1,880	12,882
Ribosoma	22/71	3,65E-05	1,11E-03	-1,587	10,799
Cáncer de pulmón de células no pequeñas	17/53	2,00E-04	5,08E-03	-1,726	9,121
Cáncer de pulmón de células pequeñas	22/85	3,32E-04	6,17E-03	-1,550	7,886
Glioma	18/62	3,65E-04	6,17E-03	-1,491	7,588
Cáncer de vejiga	14/40	3,54E-04	6,17E-03	-1,319	6,712
Señalización por ERBB	21/85	7,61E-04	1,16E-02	-1,464	6,528
Carcinoma de células renales	18/69	1,06E-03	1,34E-02	-1,236	5,331
Uniones adherentes	19/75	1,04E-03	1,34E-02	-1,187	5,117
Adhesiones focales	36/192	1,26E-03	1,47E-02	-1,403	5,922
Cáncer colorrectal	20/84	1,49E-03	1,61E-02	-1,172	4,837
Melanoma	17/71	3,15E-03	3,19E-02	-1,001	3,449

\*p-valor ajustado para múltiples comparaciones.

## Enriquecimiento KEGG y REACTOME a partir de los genes diana teóricos

Del mismo modo se llevó a cabo el análisis de enriquecimiento teniendo en cuenta las dianas teóricas, para tener una visión global y evitar introducir sesgos al dejar fuera posibles genes diana que aún no se han validado experimentalmente.

Se filtraron las dianas teóricas según el score dado por microT-CDS, quedando aquellas con un valor mayor a 0,9. Se obtuvieron 12558 genes diana, tras el enriquecimiento se obtuvieron rutas principalmente relacionadas con procesos neuronales (NGF es el factor de crecimiento nervioso), destaca las rutas guía de axones, presente en ambas bases (Tabla 16).

**Tabla 16.** Rutas enriquecidas en genes diana teóricos de los miRNAs desregulados.

Ruta	Nº genes/ Total	p-valor	p-valor ajustado*	Z-score	Score combinado
<b>KEGG</b>					
Guía de axones	82/126	6,73E-06	1,19E-03	-1,898	12,786
Señalización por MAPK	126/248	2,22E-04	1,95E-02	-2,105	8,289
Potenciación a largo plazo	45/67	5,00E-04	2,93E-02	-1,789	6,315
<b>Reactome</b>					
Guía de axones	187/327	6,64E-08	9,08E-05	-2,211	20,579
Señalización por NGF	154/287	1,45E-05	9,88E-03	-2,335	10,784

\*p-valor ajustado para múltiples comparaciones.

## Enriquecimiento KEGG y REACTOME a partir de los genes diana teóricos más afectados

A continuación se realizó el mismo análisis considerando sólo las dianas más afectadas por la desregulación de miRNAs, es decir, aquellas con un valor de represión por encima del tercer cuartil, una vez calculado de nuevo con todas las dianas teóricas. Tras filtrar quedaron 1660 genes diana teóricos, cuyo análisis destacó como enriquecidas las rutas detalladas en la tabla 17.

**Tabla 17.** Rutas enriquecidas en genes diana teóricos afectados por los miRNAs

Ruta	Nº genes/ Total	p-valor	p-valor ajustado*	Z-score	Score combinado
<b>KEGG</b>					
Leucemia mieloide crónica	18/75	1,24E-04	1,60E-02	-2,084	8,613
Señalización MAPK	37/248	4,36E-04	1,60E-02	-1,979	8,178
Carcinoma de células renales	16/69	4,10E-04	1,60E-02	-1,901	7,858
Potenciación a largo plazo	16/67	3,09E-04	1,60E-02	-1,863	7,702
<b>Reactome</b>					
Señalización por NGF	51/287	4,23E-06	4,65E-03	-2,345	12,593
Señalización NGF via TRKA	39/206	1,71E-05	7,68E-03	-2,297	11,187
Señalización por FGFR en enfermedad	35/177	2,10E-05	7,68E-03	-2,294	11,169
Señalización por FGFR	31/162	1,05E-04	2,89E-02	-2,251	7,976
Señalización por PDGF	33/184	1,87E-04	3,74E-02	-2,251	7,395
Transducción de señales	30/161	2,04E-04	3,74E-02	-2,196	7,215
Señalización de activación de FGFR	28/149	2,93E-04	4,02E-02	-2,165	6,960
Biogénesis de la mitocondria	13/42	2,73E-04	4,02E-02	-1,924	6,184
Señalización EGFRvIII en cáncer	31/181	5,84E-04	4,93E-02	-2,237	6,735
Señalización de sobre-expresión de	31/179	4,95E-04	4,93E-02	-2,221	6,685
Señalización EGFR	31/179	4,95E-04	4,93E-02	-2,221	6,685
Señalización por union al ligando de	31/181	5,84E-04	4,93E-02	-2,210	6,651
Señalización por EGFR en cáncer	31/181	5,84E-04	4,93E-02	-2,209	6,651

\*p-valor ajustado para múltiples comparaciones.

## 1.9. Red de interacciones de las rutas desreguladas seleccionadas

Se construyó una red de interacciones a partir de la base de datos KEGG por cada ruta seleccionada, sobre dicha red se visualizaron los genes diana validados de los miRNAs desregulados en mujeres jóvenes para estimar el efecto global del conjunto de los miRNAs sobre las rutas metabólicas más relevantes seleccionadas según el número total de miRNAs desregulados y su p-valor. Los valores de represión estimados para cada gen diana explicados por los miRNAs desregulados en mujeres jóvenes en relación a mujeres mayores indican la desregulación global de la ruta en mujeres jóvenes debido a la desregulación de miRNAs, por lo que a menor valor de represión, mayor expresión de los genes diana. Los resultados se muestran en el Anexo XII y se describen a continuación en el mismo orden en el que se encuentran en el aAnexo.

### Adhesiones focales

La ruta de *Adhesiones focales* mostró una desregulación en numerosos genes de la misma. Tras el análisis del efecto de los miRNAs desregulados sobre dianas validadas se obtuvo un valor de represión por parte de los miRNAs menor en mujeres jóvenes que en mujeres mayores, lo que indica que la ruta está principalmente sobre-expresada en mujeres jóvenes. Por otro lado, los genes *PTEN* y *AKT1* muestran valores de represión por parte de los miRNAs mayores, encontrándose éstos reprimidos. Destacamos la sobre-expresión del colágeno, integrinas, caveolinas y lamininas, por otro lado *AKT2* y *AKT3* están sobre-expresados, al igual que varios receptores de crecimiento. Encontramos también mayor expresión de *BCL2*, *IGFR*, *HRAS*, *MAPK* y *CCND1*, *CCND2* y *CCND3*, que controlan la proliferación. Al tener en cuenta las dianas teóricas, la represión de algunas dianas se ve afectada, apareciendo más dianas con mayor represión, como es el caso de algunos genes codificantes de proteínas de colágeno, las proteínas *AKT* se encuentran ahora principalmente reprimidas, al igual que *BRAF*, *IGFR* y *MAPK1*. *PTEN*, sin embargo, aparece menos reprimido.

### Biosíntesis de condroitín sulfato

La *Biosíntesis del condroitín sulfato* se encuentra ligeramente afectada, teniendo como dianas validadas de los miRNAs desregulados en mujeres jóvenes, a cuatro de los 11 genes responsables en el caso de dianas validadas y seis en el caso de las dianas teóricas, que muestran una mayor desregulación.

### Regulación del citoesqueleto de actina

La ruta *Regulación del citoesqueleto de actina* se encuentra sobre-expresada, a excepción de los genes *WASF2* y *ABI2*, que están muy reprimidos. De entre las dianas sobre-expresadas destacamos integrinas y actininas, oncogenes como *NRAS*, *HRAS*, *RAC1* y *RAC3*, *MAPK*, y receptores de factores de crecimiento como *PDGFRA* y *FGFR3*. Al considerar las dianas teóricas la expresión de los genes dianas pasa a estar más reprimida generalizadamente, aunque *HRAS* permanece menos reprimido.

### Potenciación a largo plazo

Todas las dianas validadas que pertenecen a la ruta *Potenciación a largo plazo* se encuentran sobre-expresadas. Destacando varios fosfodiesterasas del fosfatidil inositol (*PLCB*), así como fosfatasa, calmodulinas, los oncogenes *NRAS*, *HRAS* y *RAF1*, y varias MAPquinas. Cuando se consideran las dianas teóricas, aumenta la represión sobre la ruta, especialmente en *MAPK1*,

*BRAF* y calmodulinas, sin embargo, aparecen nuevas dianas con valores de represión menores, destacando *CAMK4*, *RAP1B* y subunidades de receptores de glutamato, *GRIN2A*, *GRIA2* y *GRM5*.

### Guía de axones

En cuanto a la ruta *Guía de axones*, a excepción del gen *SRGAPI*, cuya expresión se encuentra reprimida, está ligeramente sobre-expresada en pacientes jóvenes de cáncer de mama. Destacamos el papel de *RAC3*, *CDC42* y varias cinasas de la familia PAK (*PAK1*, *PAK2* y *PAK6*). *HRAS* y *MAPK1*, así como varios genes codificantes de efrina (*EFNA4*, *EFNA5*, *EFNB1*) y sus receptores (*EPHA2*, *EPHA4*, *EPHA7* y *EPHB2*). Al mirar las dianas teóricas, muchas de ellas pasan a estar muy reprimidas, especialmente algunas semaforinas, como *SEMA3F*, receptores de efrina y *MAPK1*.

### Sinapsis dopaminérgica y Sinapsis GABAérgica

Las rutas responsables de las *Sinapsis dopaminérgica* y *Sinapsis GABAérgica* se encuentran ligeramente sobre-expresadas, a excepción de los genes *AKT1* y *ARNTL*, en la sinapsis dopaminérgica. Destaca *FOS*, numerosas proteínas G (*GNAS*, *GNB2*, *GNB4*, *GNG5*, *GNG7*, etc), fosfatasa, receptores de GABA (*GABARAP*, *GABARAPL2*, *GABRB3*, entre otros), varias MAP cinasas, *CREB1*, *CREB5* y *CREB3L2*, *FOS* y *CLOCK*. Al tener en cuenta las dianas teóricas, la represión se ve fuertemente aumentada en ambas rutas, estando reprimidos los genes *AKT*, *CREB*, calmodulinas y proteínas G. *FOS*, sin embargo, mantiene su bajo valor de represión.

### Señalización por neurotrofina

La *Señalización por neurotrofina* está sobre-expresada, a excepción de *AKT1*, destacando, además de numerosos genes diana compartidos con las otras rutas comentadas, *PTPN11*, *FOXO3*, receptores de NFκB, MAP cinasas, *AKT2*, *NTRK3*, *SHC2* y *SHC3*, *JUN*, *TP53* y *TP73*. Si se consideran las dianas teóricas muchas dianas ven su represión aumentada, como *PTPN11*, *NGFR*, *BRAF*, calmodulinas y *RHOA*, entre otras.

### Señalización mediante mTOR

El gen de la proteína mTOR se encontraría sobre-expresado, al igual que la mayoría de genes de esta ruta, como *IRS1* y *RPTOR*, excepto, como se observa, *AKT1*. Al tener en cuenta las dianas teóricas la expresión de muchas dianas se encuentra más reprimida, sin embargo mTOR está ahora fuertemente sobre-expresado en mujeres jóvenes.

### Señalización de PI3K

La ruta de *Señalización de PI3K* está globalmente sobre-expresada, excepto los genes *AKT1* y *BCL2L11*. Destacando entre sus genes diana proteínas G, colágeno, lamininas, integrinas, interleucinas, receptores de factores de crecimiento, *NFκB*, *IGFR*, *IRS1*, *HRAS*, *YWHAE*, MAPcinasas, cinasas dependientes de ciclina, *BCL2* y *BCL2L1* (*BCL2L11* estaría reprimido), *TP53*, *FOXO3*, *RHEB*, y el resto de genes compartidos con la ruta de mTOR. Como sucede hasta ahora, al considerar las dianas teóricas, aparecen muchos más genes desregulados y aumenta la represión ejercida sobre la ruta por los miRNAs desregulados en mujeres jóvenes con cáncer de mama, encontrando los genes de las proteínas de la familia AKT y los, receptores tipo RTK, principalmente reprimidos.

## Señalización ERBB y Señalización TGF- $\beta$

Las rutas de *Señalización por ERBB* y *Señalización por TGF $\beta$*  comparten muchos de los genes ya comentados, destacan los genes de los receptores de membrana *ERBB2* y *ERBB3*, *TGFA*, *Myc* y *JUN*, en relación a *ERBB*; en la ruta de señalización de TGF $\beta$ , destacan *TGFBR2*, *BAMBI*, genes de la familia SMAD y de proteínas morfogénicas del hueso (BMPs). En el caso de considerar las dianas teóricas, destacamos que varía la represión sobre *ERBB2*, que se encontraría ahora más reprimido que en mujeres mayores, sin embargo aparece una nueva diana interesante, *ERBB4*, que, junto con *ERBB3*, se encontrarían menos reprimidos, es decir sobre-expresados, en tumor de mama de mujeres jóvenes.

## Señalización MAPK

La *Señalización por MAPK* está generalizadamente sobre-expresada con excepción de *AKT* y *JUND* y *CACNG8*, que se encuentran ligeramente reprimidos. Muchas de las proteínas en esta ruta ya han sido comentadas arriba, pudiendo añadir *NFATC1* y *NFATC3*, *HSPA8*, *DUSP*, *ELK4*, canales de calcio, caspasas y *RAP1B*. Se observa un aumento de las dianas menos reprimidas al tener en cuenta las dianas validadas, pero, como en el resto de rutas comentadas, aparecen ahora numerosas dianas muy reprimidas.

## MicroRNAs en cáncer

Por último se evaluó la expresión de los miRNAs y los genes diana en la ruta *miRNAs en cáncer* (hsa05206). En este caso se representó la expresión de los miRNAs y la represión de éstos sobre los genes (Figuras 38 y 39). Se observa que, en efecto, algunos de los miRNAs desregulados se encuentran descritos en la ruta, y que su represión (en la gran mayoría de ellos) coincide con una menor represión de los genes diana (por tanto estando estos sobre-expresados). Sin embargo, se aprecia que la expresión de muchos genes diana de los miRNAs desregulados en mujeres jóvenes, menos reprimidos en mujeres jóvenes, no se explica por los miRNAs incluidos en esta ruta, además los miRNAs responsables de su represión descritos en la ruta (no se representan en la Figuras 38 y 39) no se encuentran desregulados en mujeres jóvenes.

Muchos de los genes diana de la ruta miRNAs en cáncer, son también dianas validadas de otros miRNAs, no descritos en la ruta, pero desregulados en cáncer de mama de mujeres jóvenes.

Observamos que al analizar las dianas teóricas los genes diana varían, así como su valor de represión. Como hemos visto, en el caso de *PTEN* pasa de estar poco reprimidos a muy reprimido, lo que indica que existen más genes diana no validados que podrían cambiar radicalmente la acción sobre *PTEN*. Otros genes también ven afectada su represión, como *ZEB1*, *ZEB2*, *IGF2BP1* y *MAPK1*, que muestra valores de represión mucho menores al considerar las dianas teóricas.

Los datos obtenidos se deben tomar con precaución, ya que la desregulación de la ruta varía mucho según se consideren las dianas validadas o las teóricas.

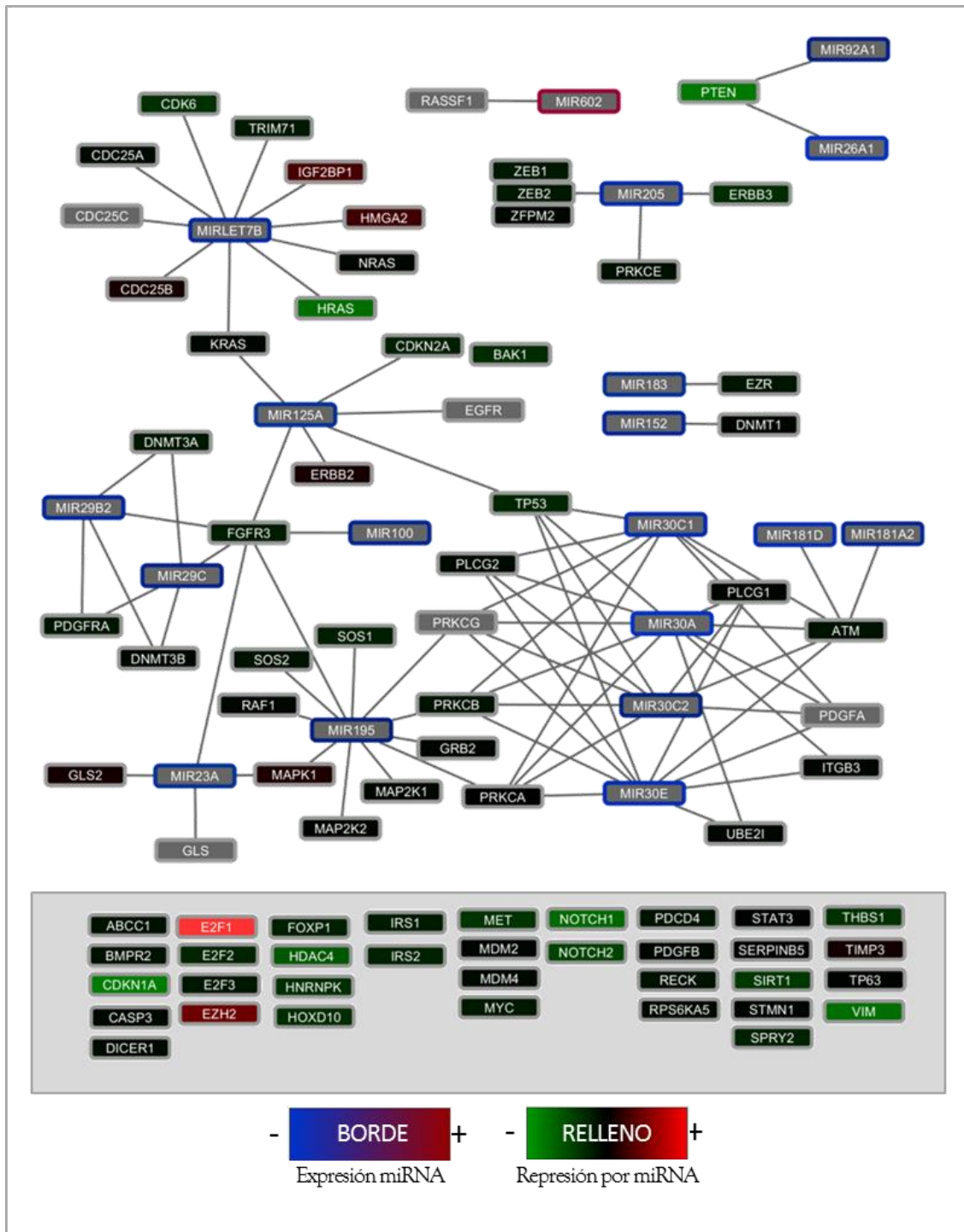


Figura 38. Red de interacciones de los miRNAs desregulados en cáncer de mama y sus dianas validadas en la ruta metabólica miRNAs en cáncer (hsa05206). Los genes diana de los miRNAs desregulados se muestran en escala de colores desde el verde (poca represión por parte de los miRNAs) hasta el rojo (muchas expresiones por parte de los miRNAs). El borde marcado muestra aquellos miRNAs de la ruta que se encuentran desregulados en mujeres jóvenes, en escala de colores desde el azul (reprimidos en mujeres jóvenes) hasta el rojo oscuro (sobre-expresado en mujeres jóvenes). Aquellos genes incluidos en el recuadro gris corresponden a aquellos genes de la ruta que son dianas de los miRNAs desregulados en mujeres jóvenes, pero que, sin embargo, estos miRNAs no aparecen en la ruta.

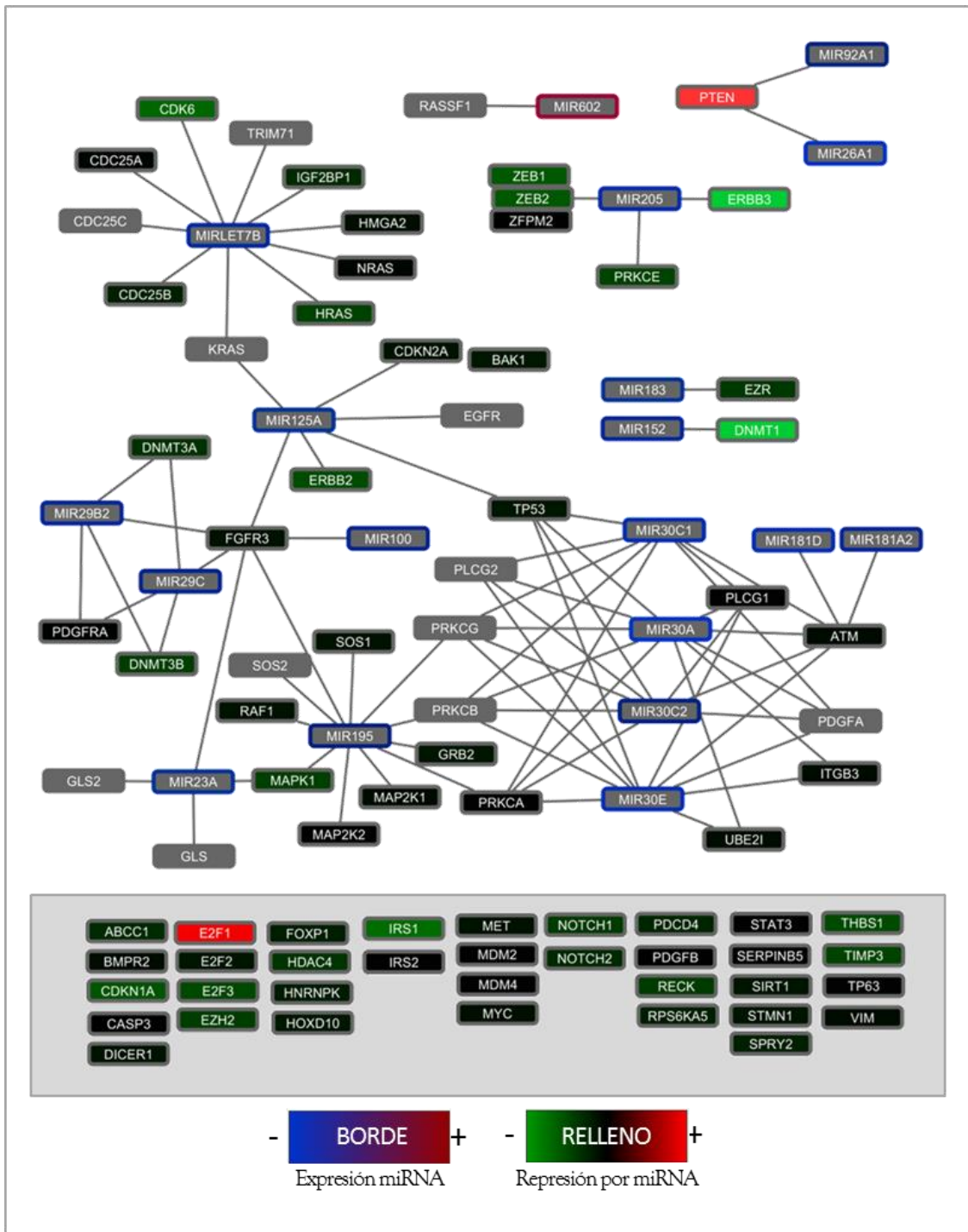


Figura 39. Red de interacciones de los miRNAs desregulados en cáncer de mama y sus dianas teóricas en la ruta metabólica miRNAs en cáncer (hsa05206). Los genes diana de los miRNAs desregulados se muestran en escala de colores desde el verde (poca represión por parte de los miRNAs) hasta el rojo (muchas expresión por parte de los miRNAs). El borde remarcado muestra aquellos miRNAs de la ruta que se encuentran desregulados en mujeres jóvenes, en escala de colores desde el azul (reprimidos en mujeres jóvenes) hasta el rojo oscuro (sobre-expresado en mujeres jóvenes). Aquellos genes incluidos en el recuadro gris corresponden a aquellos genes de la ruta que son dianas de los miRNAs desregulados en mujeres jóvenes, pero que, sin embargo, estos miRNAs no aparecen en la ruta.

## 2. ESTUDIO DEL PERFIL DE EXPRESIÓN DE MIRNA EN LÍNEAS CELULARES DE CÁNCER DE MAMA EN MUJERES JÓVENES

---

Para evaluar la reproducibilidad del perfil hallado en muestras de tumor de pacientes en líneas celulares y establecer así su adecuación como modelo para estudios funcionales, se llevó a cabo el análisis de la expresión de miRNAs en las líneas celulares de cáncer de mama procedentes de mujeres jóvenes, HBL100, HCC1500, HCC1937, Hs566(B).T y se comparó con la expresión de líneas procedentes de mujeres mayores (MDA-MB-231, MDA-MB-468 y MCF7). Las características de las líneas celulares y de las pacientes de origen han sido descritas en el apartado Material y Métodos, Tabla 7.

Las líneas celulares se incubaron por triplicado el día 0 con una población inicial de células viables. Los recuentos se hicieron cada 24-48 h durante 7-19 días en función de la línea celular. El tiempo de doblaje la línea celular HBL100 fue de 36,48 horas; de 66,92 horas para la línea HCC1500; y HCC1937 dobló la cantidad de sus células en 35,32 horas. El tiempo de doblaje para la línea celular Hs566.B(T) no pudo ser calculado por tener un crecimiento anormalmente lento, manteniéndose el número de células constante desde del día 0 hasta el día 14.

Para el análisis de expresión de miRNAs mediante chio de Affymetrix se obtuvieron tres réplicas procedentes del mismo cultivo celular de las que se recogió en el pase 4 el pellet de células y se extrajo el RNA como está descrito en material y métodos, se midió la cantidad y calidad del RNA extraído y se homogeneizaron las concentraciones de las muestras a 50ng/ $\mu$ l. Una de las líneas tuvo que ser excluida del estudio al no alcanzar los estándares de concentración necesarios debido a su lento e inusual crecimiento en cultivo.

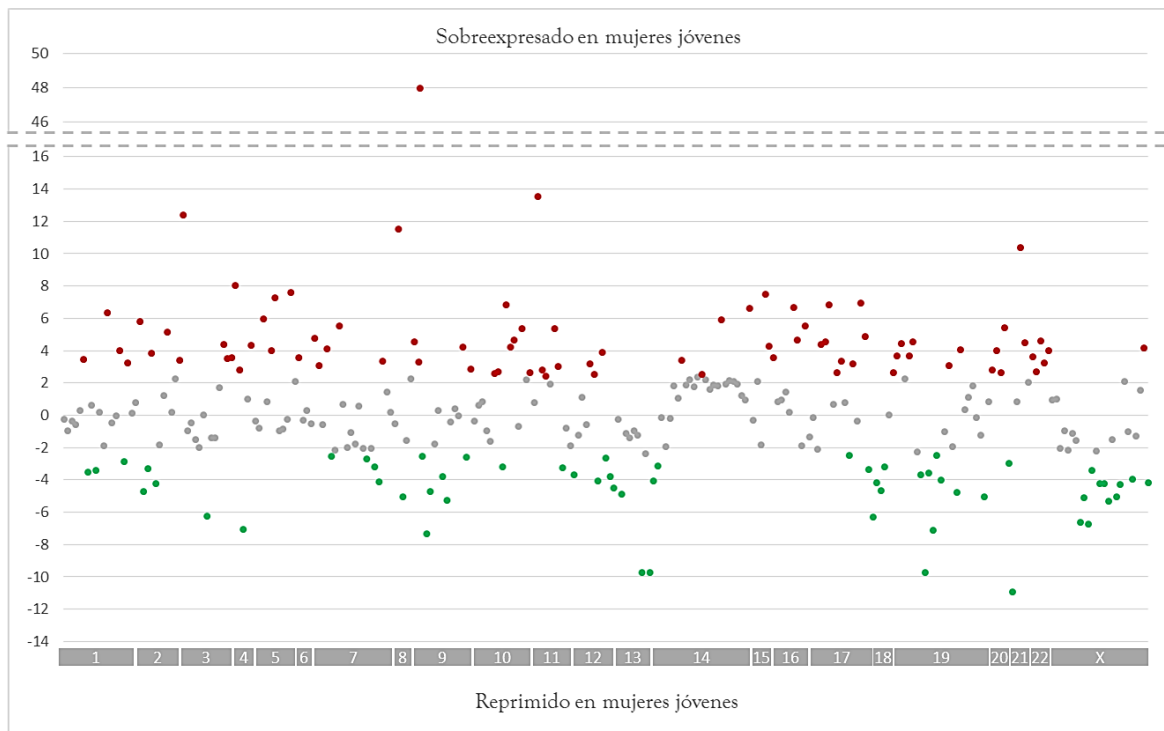
Para el análisis se llevaron a cabo los mismos controles de calidad y métodos de normalización y filtrado que los realizados con las muestras de tejido procedente de pacientes del Hospital Clínico Universitario de Valencia. Tras este procedimiento se obtuvieron un total de 279 miRNAs humanos maduros para el estudio de su expresión según la edad de las pacientes de procedencia de la línea celular.

### *2.1. Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes*

El t-test realizado sobre estos datos mostró 148 miRNAs con una expresión significativamente diferente en líneas de mujeres jóvenes, cuyo FDR era menor a 0,05 y, por tanto, aguantaban la corrección para múltiples comparaciones de Benamini & Hochberg. Los resultados obtenidos se pueden encontrar en el Anexo V y los valores de expresión se representan en la figura 40.

Al representar los valores de expresión obtenidos para cada uno de los miRNAs por orden cromosómico del gen codificante de los miRNAs estudiados, observamos que los miRNAs se expresan de manera homogénea a lo largo de los cromosomas. Sin embargo, se observa una agregación de miRNAs sobreexpresados en mujeres jóvenes en los cromosomas 5, 6, 11, 14, 15 y 16, en los que no se observa ningún o casi ningún cromosoma reprimido. Del mismo modo, en los cromosomas 13 y X se observa una tendencia a acumular miRNAs menos expresados en mujeres jóvenes (Figura 40).





**Figura 40. Representación de la expresión de los microRNAs en líneas celulares de cáncer de mama.** Estudio realizado sobre líneas procedentes de mujeres jóvenes (HBL100, HCC1500, HCC1937) respecto a procedentes de mujeres mayores (MDA-MB-231, MDA-MB-468 y MCF7). Los valores de expresión de los miRNAs se encuentran ordenados por localización cromosómica. El color rojo indica sobreexpresión en mujeres jóvenes, mientras que el verde indica represión en mujeres jóvenes. Los puntos grises corresponden a microRNAs cuya expresión no cambia.

## 2.2. Agrupación jerárquica por líneas celulares y miRNAs según la edad

Mediante los programas Clúster y Treeview se construyó un *heatmap* representativo de la expresión de los miRNAs significativos en cada muestra para visualizar el perfil de expresión obtenido, se encuentra representado en la Figura 41. En él se observa que la agrupación jerárquica centrada en la mediana llevada a cabo, agrupa las muestras en dos grupos principales, muestras de líneas celulares procedentes de mujeres jóvenes y muestras procedentes de mujeres mayores; además, las muestras se agrupan con aquellas procedentes de la misma línea celular, pero de diferente réplica de crecimiento del cultivo, lo que asegura la fiabilidad de la agrupación y de los datos de expresión obtenidos.

Estos resultados refuerzan la hipótesis inicial de que el cáncer de mama de mujeres jóvenes se diferencia del cáncer de mama general a nivel molecular, pudiéndose establecer un perfil único y distintivo de tumor procedente de mujeres jóvenes, tanto en muestras de pacientes como en líneas celulares.

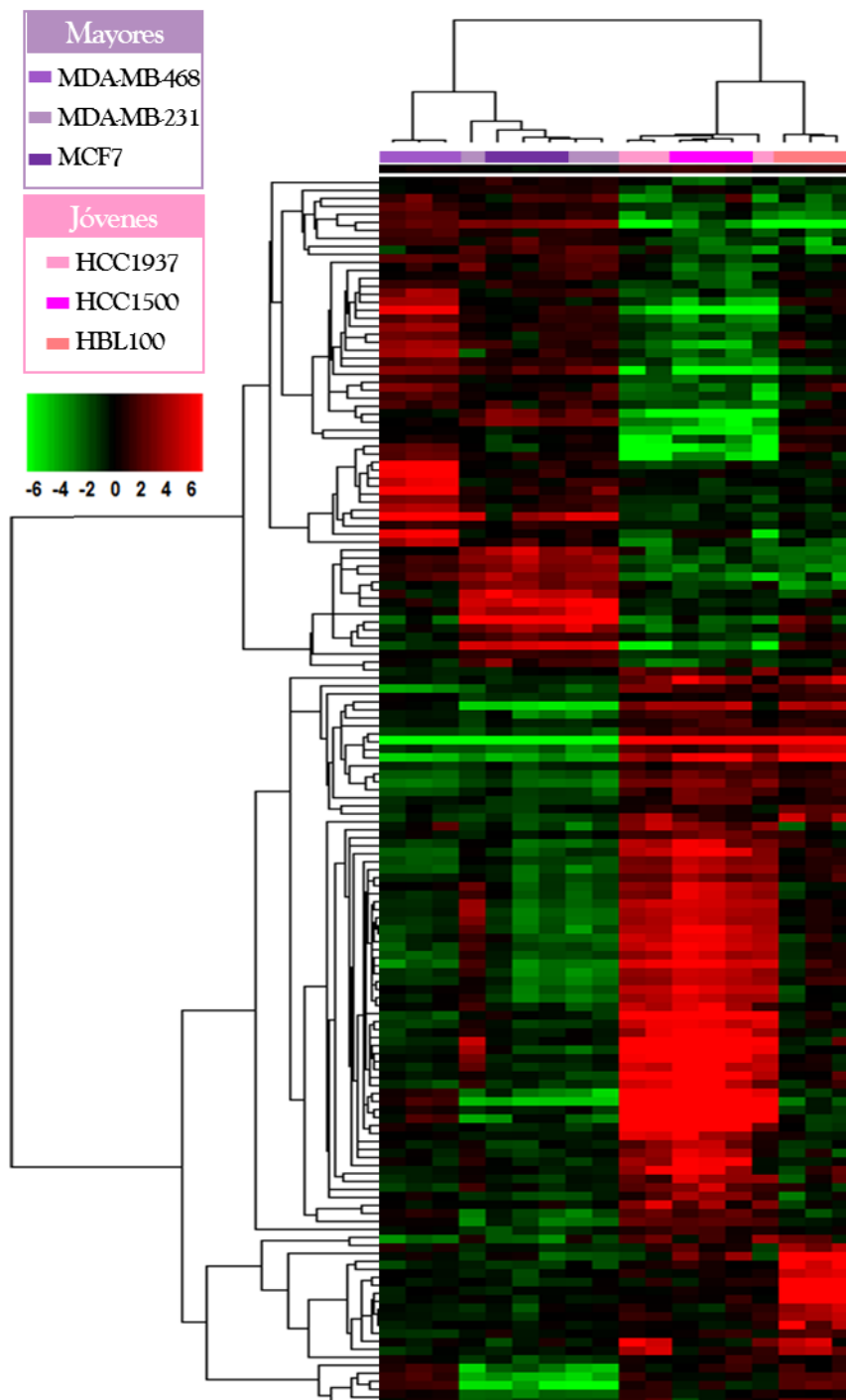


Figura 41. Representación del perfil de miRNAs desregulado en líneas celulares procedentes de mujeres jóvenes comparado con aquellas procedentes de mujeres mayores. El clúster se ha obtenido mediante agrupación jerárquica centrada en la mediana. Se representa la expresión de los microRNAs, en horizontal, por cada línea analizada, en vertical. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las líneas se encuentran categorizadas en procedentes de mujeres jóvenes (tonos de rosa) y mayores de 45 años (tonos de morado).

## Comparación de los miRNAs obtenidos en el estudio de las muestras procedentes de pacientes y de líneas celulares

De los miRNAs con expresión significativamente diferente tanto en el estudio de muestras de pacientes como procedentes de líneas celulares un total 50 miRNAs se encontraban compartidos entre ambos grupos; sin embargo, 7 de esos miRNAs dieron valores contradictorios de expresión, por lo que no se tuvieron en cuenta, quedándonos así 43 miRNAs significativos y con expresión coherente en ambos grupos. De ellos, 35 se encontraban sobre-expresados en mujeres jóvenes y 8 reprimidos (Anexo XIII).

Posteriormente, se reconstruyó el perfil de expresión de miRNAs obtenido en muestras a partir de tumor biopsiado de pacientes. Para ello, los chips de hibridación de muestras de pacientes y de líneas celulares se volvieron a analizar y normalizar todos juntos, tal y como se hizo para los análisis de los chips de hibridación de muestras procedentes de pacientes.

Se seleccionaron los miRNAs cuya expresión en todas las muestras mostró expresión diferente entre mujeres jóvenes y se construyó el perfil de expresión mediante agrupación jerárquica sobre las muestras para comprobar su similitud. En la Figura 42 se muestra la correlación de las réplicas de cada línea celular.

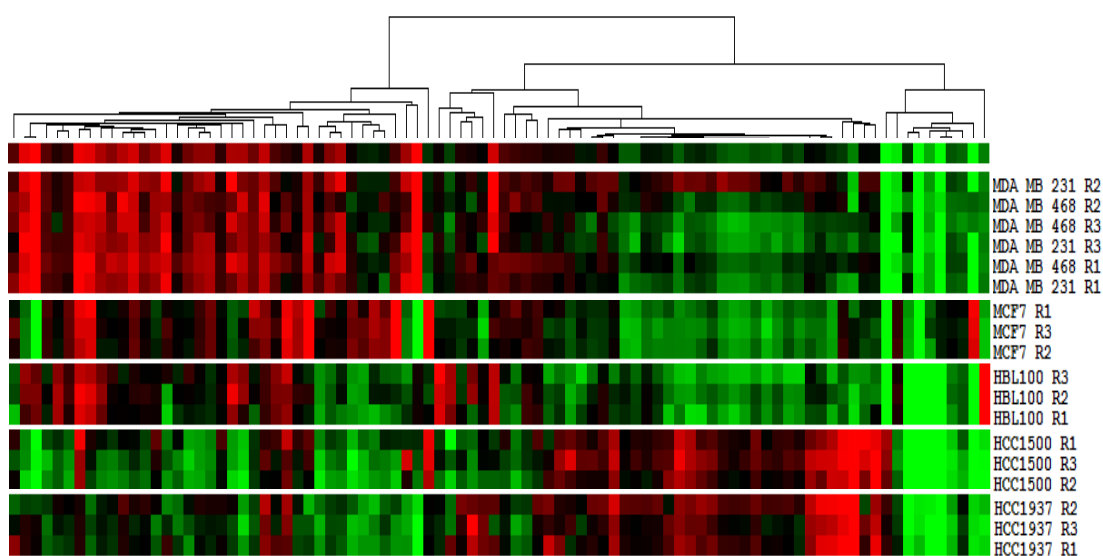


Figura 42. Agrupación jerárquica de la expresión de los miRNAs desregulados en pacientes muy jóvenes de cáncer de mama por réplica de las líneas celulares analizadas. Se representa la expresión de los microRNAs, en vertical, por cada línea celular, en horizontal. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana.

El clúster de las líneas celulares MDA\_MB231 y MDA\_MB468 obtuvo un valor de correlación de 0,844, el clúster de las réplicas de la línea MCF7 una correlación de 0,977. En cuanto a las líneas de mujeres jóvenes, las réplicas de HBL100 se correlacionaron con un valor de 0,941; sin embargo, su expresión no se asemejó al resto de líneas, HCC1500 y HCC1937 obtuvieron una correlación de 0,920 y 0,877, respectivamente, agrupándose todas ellas en un nodo con correlación 0,790.

Al realizar la agrupación jerárquica por miRNAs se comprobó que las muestras de líneas celulares procedentes de mujeres jóvenes HCC1500 y HCC1973 agruparon con las muestras de pacientes jóvenes, mientras que las líneas celulares de mujeres mayores agruparon con las muestras de pacientes mayores y las muestras de mujeres sanas. Dos de las tres réplicas de HBL100, sin embargo, agruparon junto con las mujeres mayores, confirmando lo que se obtuvo

en el perfil de líneas celulares, que HBL100 es la línea celular que menos reproduce el perfil de expresión de miRNAs en mujeres jóvenes (Figura 43).

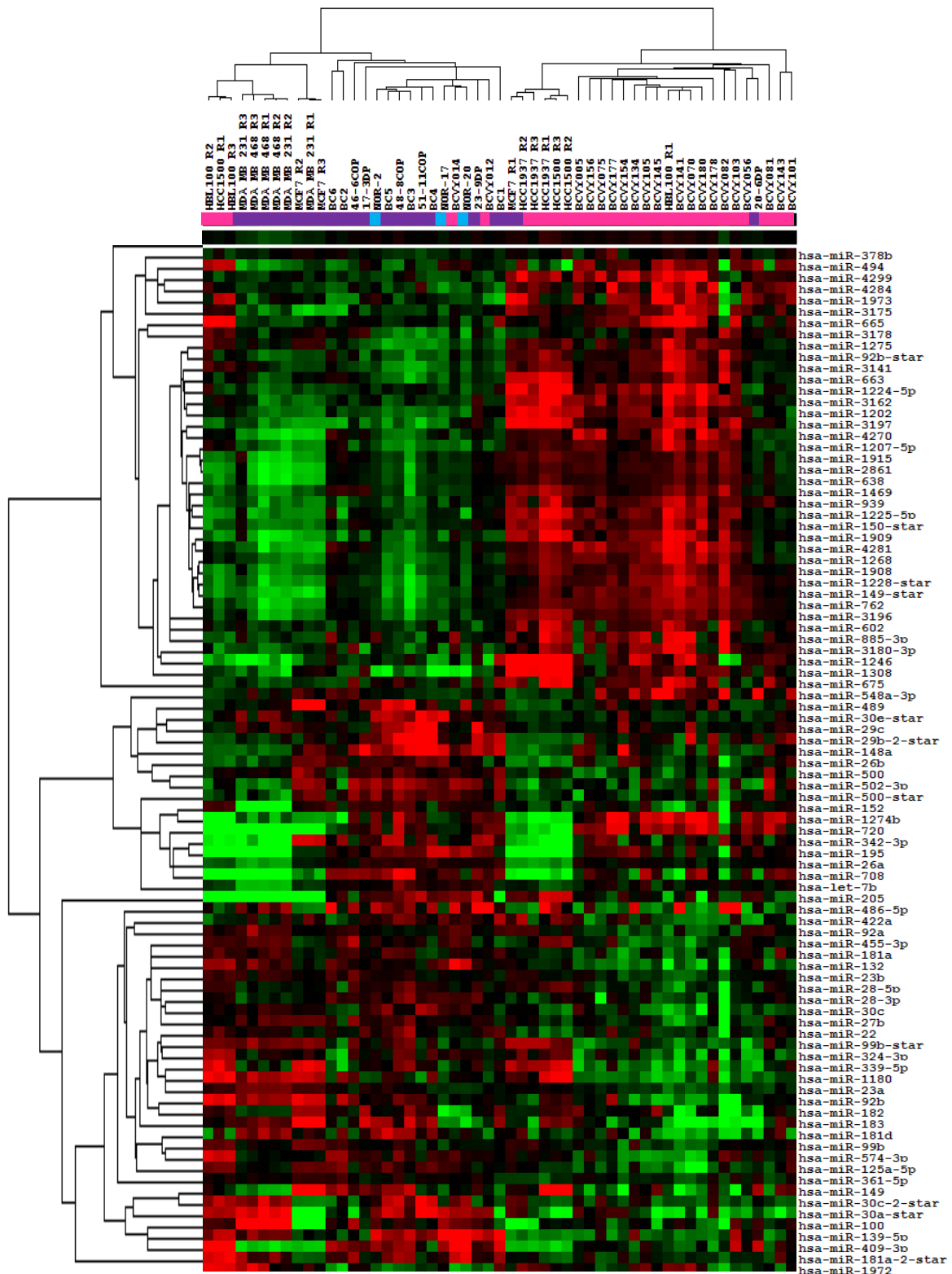


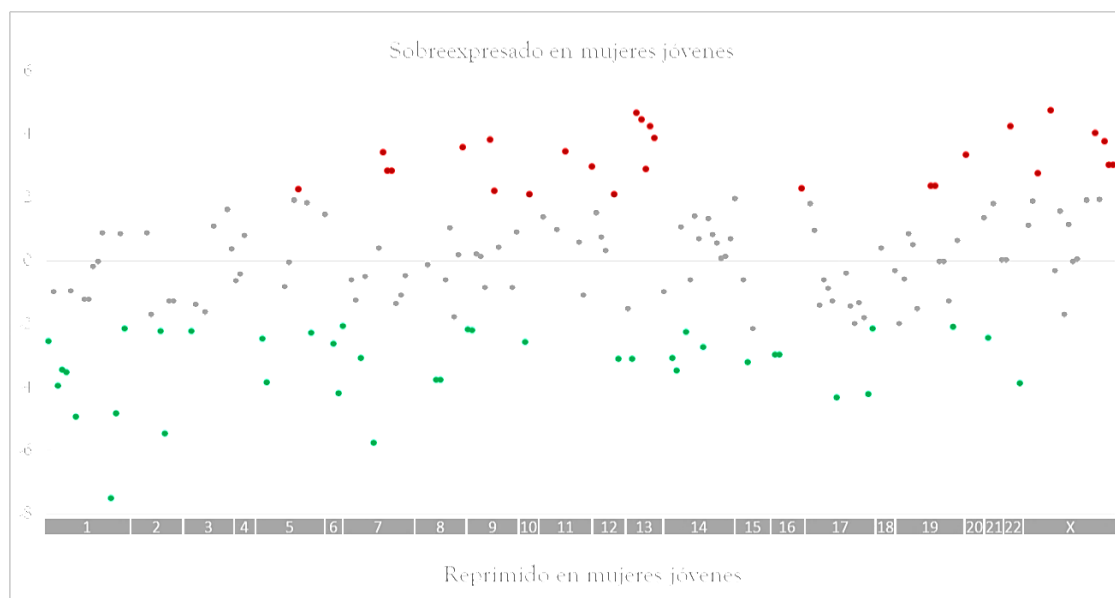
Figura 43. Representación del perfil de expresión de los miRNAs desregulados en muestras procedentes de mujeres jóvenes tras analizar las muestras de pacientes y líneas celulares conjuntamente. Se representa la expresión de los microRNAs, en horizontal, por cada muestra analizada, vertical. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en tumores procedentes de mujeres menores de 35 años (rosa), mayores de 65 años (morado) y tejido de mujeres libres de cáncer (azul).

### 3. ANÁLISIS DEL PERFIL DE EXPRESIÓN DE MIRNAS EN UNA SERIE CON UN TAMAÑO MUESTRAL MAYOR OBTENIDA A PARTIR DE BASES DE DATOS ONLINE (METABRIC)

Para validar la existencia de un perfil de miRNAs en cáncer de mama de mujeres jóvenes, se decidió emplear los datos disponibles en bases de datos online, debido a su accesibilidad y a la gran cantidad de muestras disponibles, que de otro modo son difícilmente accesibles en la práctica experimental. Para ello, se descargaron los datos de expresión de miRNAs en cáncer de mama depositados en la base de datos europea de datos genómicos (*European Genome Atlas*) con código de acceso EGAD00010000438. Tras normalizar los datos de expresión obtenidos, se construyó una matriz con los datos de expresión y las características clínicas, se filtraron los datos según su edad, obteniendo 33 mujeres jóvenes menores de 35 años y 558 mayores de 65 años cuyas características clínicas se muestran en el Anexo XIV.

#### 3.1. *Análisis de la expresión de miRNAs en tumor de mama de mujeres jóvenes*

Tras filtrar aquellos miRNAs con expresión uniformemente baja, nos quedaron 485 miRNAs que sometimos a un test t para evaluar las diferencias entre ambos grupos, mujeres jóvenes y mayores de 65 años. Se obtuvieron 56 miRNAs con resultados significativos, de los cuales 31 tenían una mayor expresión en tumor de mujeres jóvenes, mientras que 25 mostraban una expresión menor en este grupo de pacientes. Los resultados obtenidos se detallan en el Anexo V.



**Figura 44.** Representación de la expresión de los microRNAs expresados en mujeres jóvenes respecto a mujeres mayores obtenida a partir de los datos depositados en bases de datos online (METABRIC), ordenados por localización cromosómica. El color rojo indica sobreexpresión en mujeres jóvenes, mientras que el verde indica represión en mujeres jóvenes. Los puntos grises corresponden a microRNAs cuya expresión no cambia significativamente.

En cuanto a la distribución cromosómica de los miRNAs de los datos obtenidos a partir de los datos depositados en bases de datos online, se observa en la Figura 44 que los cromosomas con mayor concentración de miRNAs sobreexpresados en mujeres jóvenes son los cromosomas 7, 9,

10, 11, 12 y 13, 19, 20, 22 y el cromosoma sexual X, destacando especialmente la agregación en los cromosomas 13 y X. Por otro lado, la representación de miRNAs reprimidos se halla más generalizada, encontrándose miRNAs reprimidos en los cromosomas 1, 2, 3, 5, 6, 7, 8, 10, 12, 13, 13, 14, 16, 17, 20, 21 y 22. Sin embargo, se observa una mayor agregación de estos en los cromosomas 1, 2, 6, 8, 14, 15 y 17.

### 3.2. Perfil de expresión mediante agrupación jerárquica de los datos analizados

Para realizar el clustering jerárquico y construir el perfil de expresión de miRNAs centrado en la mediana, se realizó un re-muestreo aleatorio de 50 muestras de la población de mujeres mayores debido a su mayor tamaño comparado con la población de mujeres jóvenes (558 mujeres mayores frente a 33 jóvenes). Se realizó la agrupación tanto supervisada (Figura 45) como no supervisada, en ambos casos se obtuvo un perfil de expresión de miRNAs según el grupo de edad en el que se observa un nodo reprimido en mujeres jóvenes y otro sobre-expresado. Los subtipos se encuentran bastante equilibrados en ambos grupos de edad y no se observa agrupación por subtipos.

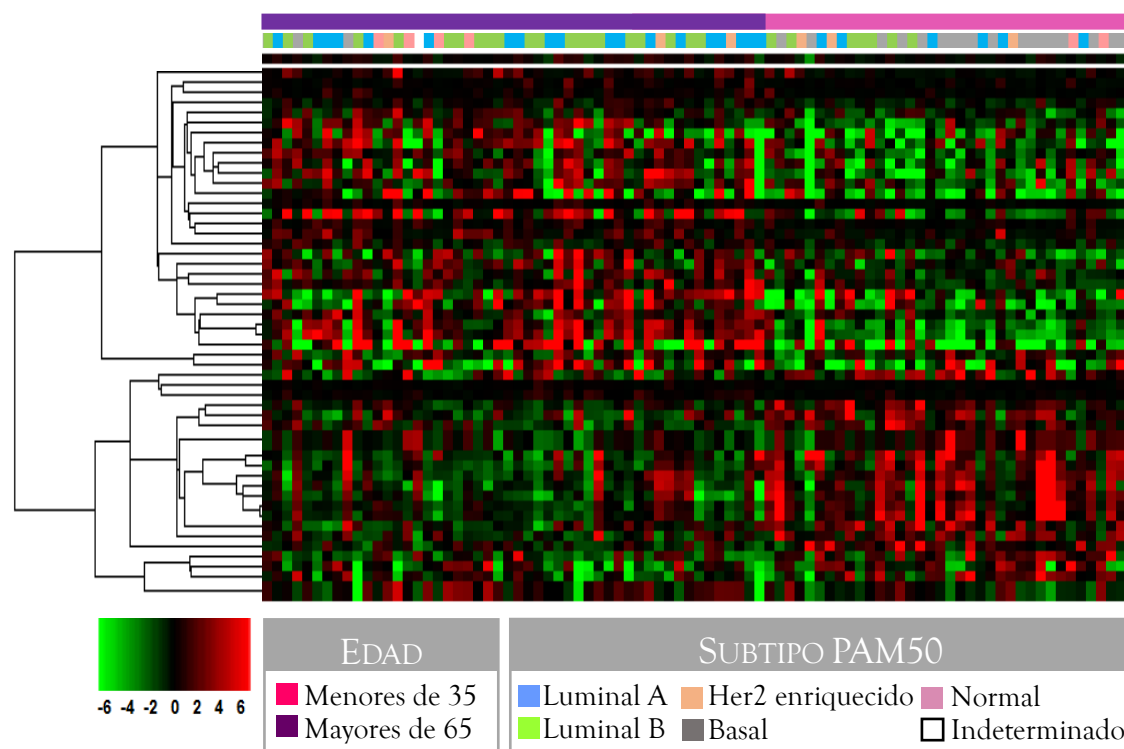


Figura 45. Clustering jerárquico supervisado centrado en la mediana y visualización de la expresión de los miRNAs significativamente desregulados en mujeres jóvenes en el estudio de los datos procedentes del estudio METABRIC. Se representa la expresión de los microRNAs, en horizontal, por cada muestra analizada, en vertical. La expresión se muestra en diferentes intensidades de verde cuando está por debajo de la mediana, rojo por encima y en negro sin cambios respecto a la mediana. Las muestras se encuentran categorizadas en mujeres menores de 35 años (rosa) y mayores de 65 años (morado).

### 3.3. Comparación y meta-análisis de los resultados obtenidos en los tres estudios

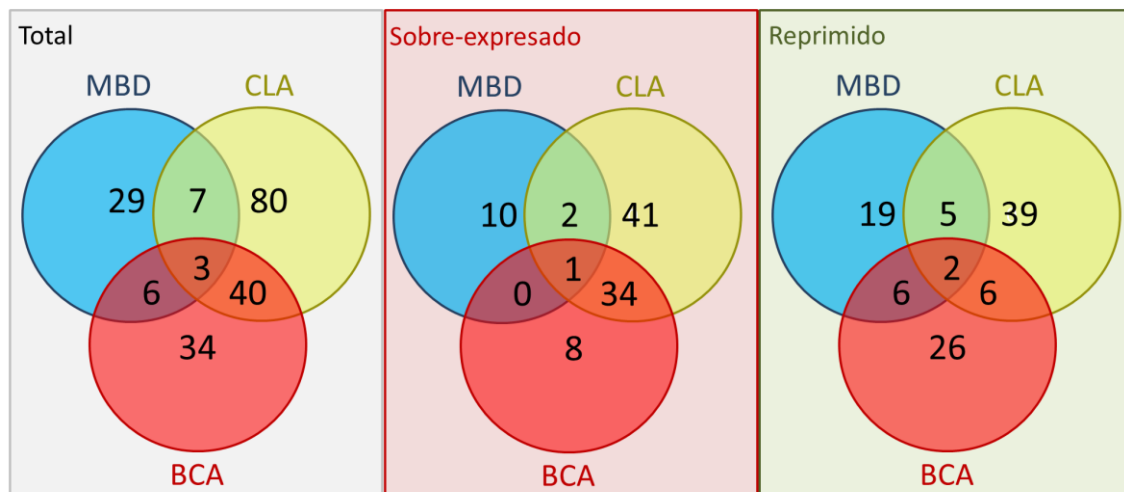
Se validó la existencia de un perfil de expresión característico de pacientes de cáncer de mama menores de 35 años en tres estudios realizados: a partir de muestras tumorales de pacientes de cáncer de mama del Hospital Clínico Universitario de Valencia (BCA); a partir de líneas celulares procedentes de mujeres jóvenes y de mujeres mayores (CLA) y a partir de datos depositados en bases de datos online obtenidos del estudio previo sobre cáncer de mama METABRIC (MBD).

#### MiRNAs coincidentes en los tres perfiles moleculares de mujeres jóvenes generados

Mediante un diagrama de Venn se representaron los miRNAs con expresión significativamente diferente en al menos uno de los estudios realizados, observándose los miRNAs que comparten los grupos de muestras (Figura 46).

De los miRNAs coincidentes, se eliminaron 14 miRNAs cuya expresión no mostró la misma tendencia en los resultados (es decir, sobre-expresado o reprimido). Los p-valores corregidos y la expresión para cada uno de los miRNAs significativos en, al menos, dos de los tres estudios se puede observar en la tabla del Anexo XIII.

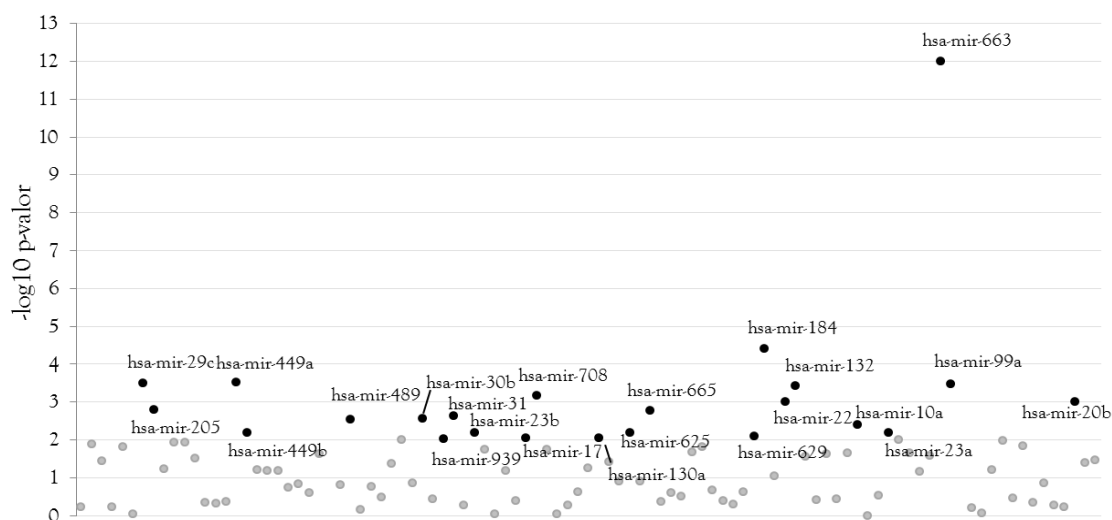
Únicamente se hallaron 3 miRNAs con expresión diferente y coherente en los tres estudios, los miRNAs hsa-miR-30c, hsa-miR-489, hsa-miR-663. Concretamente el miRNA hsa-miR-30c mostró buenos valores de significación corregidos por FDR en los tres estudios (p-valor BCA: 0,002, p-valor CLA: 0,008, p-valor MBD: 0,002. En el caso de los miRNAs hsa-miR-489 y hsa-miR-663 mostraron un valor cercano a 0,05 en uno de los tres estudios (0,046 en BCA y 0,031 en MBD, respectivamente). Se observa que los estudios BCA y CLA son los que comparten mayor número de miRNAs significativos.



**Figura 46. Comparación de miRNAs significativos entre los diferentes estudios.** Se muestran las coincidencias entre los microRNAs significativos en los tres grupos de muestras estudiados. BCA: muestras de tumor procedente de pacientes con cáncer de mama (círculo rojo); CLA: muestra de células procedente de cultivo celular de líneas celulares procedentes de tumor mamario (círculo amarillo); MBD: datos procedentes de muestras de pacientes obtenidos a partir de bases de datos online y depositados a partir del estudio realizado por el consorcio METABRIC (círculo azul). Se representa mediante diagrama de ven las coincidencias entre los miRNAs estadísticamente significativos totales (se representa en el cuadro color gris), los más expresados en tumores de mujeres jóvenes (en color rojo) y las coincidencias entre los miRNAs estadísticamente significativos menos expresados en mujeres jóvenes (color verde).

## Meta-análisis de los tres estudios realizados mediante el método Fisher

Se combinaron los p-valores obtenidos en los tres estudios mediante el método de Fisher de aquellos miRNAs analizados en los tres estudios. Se ha de tener en cuenta que los miRNAs analizados en los tres estudios no fueron los mismos, debido a que la plataforma empleada en BCA y CLA fue Affymetrix y en MBA fue EXIQON, tras la normalización y el filtrado de los miRNAs el número restante fue diferente en los tres casos. Los p-valores obtenidos de los 100 miRNAs combinados se representan en la Figura 47.



**Figura 47. Significación de la expresión de los miRNAs tras combinar los p-valores de los tres estudios realizados.** Se muestra el  $-\log$  de los 100 p-valores combinados mediante Fisher. Los valores representados en negro son menores a un  $\alpha = 0,1$  y se encuentran al lado del nombre del miRNA correspondiente, los menores de  $\alpha$  se representan en gris. Los miRNAs se encuentran ordenados según su localización cromosómica.

De los 100 miRNAs analizados en el meta-análisis, 23 mostraron resultados significativamente diferentes en cuanto a la expresión de miRNAs en mujeres jóvenes, con un p-valor  $< 0,1$ . El miRNA más significativo fue hsa-miR-663, con un p-valor  $< 10^{-12}$ . Encontramos también entre los miRNAs significativos los miRNAs hsa-miR-29c, hsa-miR-30b, hsa-miR-23a y hsa-miR-23b, hsa-miR-17, hsa-miR-132, hsa-miR-449a y hsa-miR-449b, hsa-miR-489 entre otros (Figura 46). La lista completa de los p-valores combinados se detalla en el Anexo XV.



---

## 1. VALIDACIÓN DE LA EXPRESIÓN DE MIRNAS EN TEJIDO TUMORAL MAMARIO DE MUJERES JÓVENES

---

Debido a la naturaleza de los datos y de los propios métodos estadísticos, los resultados pueden estar mostrando falsos positivos, debido a la técnica empleada o debido a procesos estocásticos dado el tamaño muestral. Se debe, por tanto, evaluar la robustez de los datos obtenidos y confirmar que, en efecto, los miRNAs seleccionados se encuentran desregulados en mujeres jóvenes. Con este fin, se llevó a cabo un estudio de la expresión de los miRNAs seleccionados según el proceso antes detallado, empleando una técnica diferente, más concreta/directa, como es el análisis de la expresión mediante PCR cuantitativa a tiempo real (qRT-PCR). En este estudio se analizó la expresión de 18 miRNAs en una serie independiente de 55 pacientes de cáncer de mama.

Para la selección del segundo grupo de pacientes, se siguió el mismo criterio que el empleado para la primera fase del estudio, y detallado en *Material y Métodos*. El estudio se llevó a cabo sobre 55 muestras procedentes de tejido de tumor mamario de pacientes de cáncer de mama. Las muestras se agruparon en tres categorías atendiendo a su edad: 21 de las mujeres eran menores de 35 años; 17 tenían edades comprendidas entre los 45 y los 65 años; por último, 17 pacientes eran mayores de 65 años.

Se añadió un grupo de edad adicional a los establecidos para el estudio de la primera fase. Debido a que el cáncer de mama es una enfermedad que, aunque principalmente postmenopáusica, presenta una alta frecuencia de aparición alrededor de los 50 años, era de interés establecer si las diferencias encontradas al comparar mujeres muy jóvenes con mujeres postmenopáusicas de avanzada edad, se mantendrían al compararlas con mujeres perimenopáusicas, lo que añadiría robustez a la hipótesis de que el perfil obtenido es exclusivo de mujeres muy jóvenes.

### 1.1. *Análisis de la expresión de miRNAs en tumor de mama de mujeres muy jóvenes*

Para validar la expresión de los miRNAs seleccionados en la fase I del estudio (hsa-miR125b, hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-149\*, hsa-miR-3196, hsa-miR-1228\*, hsa-miR-92b, hsa-miR-139-5p, hsa-miR-132, hsa-miR-379, hsa-miR-409-3p, y hsa-miR-433, hsa-miR-23a, hsa-miR-27b, hsa-miR-28-5p, hsa-miR-30c, hsa-miR-30e, hsa-miR-195-5p), se realizó una qRT-PCR en el RNA total extraído a partir de 55 muestras de tejido tumoral de mama.

## MicroRNAs y control endógeno seleccionados para la validación

Tras realizar la qRT-PCR sobre el cDNA obtenido de la retrotranscripción selectiva de los miRNAs objeto de estudio se obtuvieron las medidas de Ct, a partir de las cuales se pudo obtener la cantidad relativa de miRNA en cada muestra respecto a una muestra tomada como calibrador, mediante el método  $\Delta\Delta Ct$ .

En cuanto a la expresión de los controles endógenos empleados (RNU43 y hsa-let-7d) sigue unos patrones muy similares, sin embargo RNU43 se detecta mucho antes en la reacción de amplificación, del orden de 5 ciclos antes, por lo que se considera su detección más fiable. Cabe destacar que su expresión es similar excepto en la muestra BCVY40, en la que se observa un retraso notable en el Ct en cuanto al control RNU43, no así en el control hsa-let-7d. Sin embargo, el hecho de que varias sondas muestren una expresión anormal en esta muestra (mir-1275 y miR-409-3p), lleva a interpretar sus resultados con precaución.

## Expresión diferencial de los microRNAs en tejido tumoral de mujeres jóvenes

Tras la normalización de los valores mediante el control endógeno RNU43, observamos que todos los valores de expresión relativa (RQ) se encuentran dentro de un rango normal entre 0 y 30, a excepción de la muestra BCVY40, que ya hemos comentado antes, que obtiene unos valores anormales (del orden de miles), lo que nos lleva a descartar esta muestra debido a problemas técnicos. Por otro lado, la muestra BCVY4 presenta valores algo elevados, siendo de 97,78 para la sonda miR-3196 y de 44,78 para la sonda miR-1228, el resto de valores entran dentro de lo normal, si bien no son irracionalmente elevados, conviene tenerlos en cuenta para futuros análisis. Los valores de expresión relativa obtenidos se encuentran en el Anexo XVI.

**Tabla 18.** Diferencias en la expresión de los miRNAs seleccionados en los grupos de edad.

miRNA	Distribución normal	ANOVA p-valor <sup>1</sup>	Homogeneidad de varianzas	Media (ES)		
				<35	45-64	>65
hsa-miR-23a	NO	0,282	0,1269	3,26(0,40)	3,95(0,49)	4,20(0,89)
<b>hsa-miR-27b</b>	NO	<b>0,067</b>	0,463	1,88(0,26)	2,24(0,23)	2,66(0,58)
hsa-miR28-5p	NO	0,583	0,2152	1,42(0,16)	1,36(0,19)	1,58(0,31)
<b>hsa-miR30c</b>	NO	<b>0,0003*</b>	0,004	0,45(0,06)	0,96(0,13)	0,88(0,59)
hsa-miR30e-3p	SI	0,83	0,8695	2,48(0,29)	2,20(0,46)	2,63(0,58)
<b>hsa-miR-92b-3p</b>	NO	<b>0,0002*</b>	0,013	0,61(0,11)	1,74(0,27)	1,22(0,17)
<b>hsa-miR-125a</b>	SI	<b>0,082</b>	0,0022	1,42(0,14)	2,13(0,40)	2,08(0,42)
<b>hsa-miR-132</b>	SI	<b>0,023</b>	0,0054	0,81(0,08)	1,31(0,17)	1,56(10,28)
<b>hsa-miR-139</b>	SI	<b>0,001*</b>	0,0134	0,23(0,04)	0,92(0,14)	1,03(0,24)
hsa-miR-149-5p	NO	0,957	0,039	4,99(0,91)	4,90(1,03)	8,16(2,33)
hsa-miR-195	NO	0,324	0,361	1,56(0,16)	1,14(0,15)	1,50(0,23)
hsa-miR-379	NO	0,185	0,129	0,87(0,13)	1,23(0,17)	1,26(0,29)
<b>hsa-miR-409</b>	NO	<b>0,028</b>	0,179	0,55(0,08)	1,02(0,14)	0,79(0,19)
<b>hsa-miR-433</b>	NO	<b>0,0312</b>	0,145	0,72(0,11)	1,41(0,23)	1,15(0,29)
<b>hsa-miR-1207-5p</b>	NO	<b>0,094</b>	0,031	7,57(1,60)	5,86(2,47)	1,81(0,54)
<b>hsa-miR-1228-5p</b>	SI	<b>&lt;0,0001*</b>	<0,0001	7,89(1,46)	1,92(0,26)	1,42(0,22)
<b>hsa-miR-1275</b>	SI	<b>0,0006*</b>	0,031	7,33(0,92)	3,32(0,46)	4,16(0,61)
<b>hsa-miR-3196</b>	NO	<b>0,0011*</b>	<0,0001	10,87(2,29)	1,57(0,18)	1,54(0,31)

\*Mantiene la significación tras aplicar Bonferroni. <sup>1</sup>Cuando las distribuciones no cumplen con la normalidad, se realizó un test de Kruskal-Wallis en lugar de ANOVA. La homogeneidad de varianzas se ha evaluado mediante el test de Brown-Frosythe. Las medias se muestran por grupos de edad (<35: menores de 35 años; 45-65: entre 45 y 65 años; >65: mayores de 65 años). ES: error estándar.

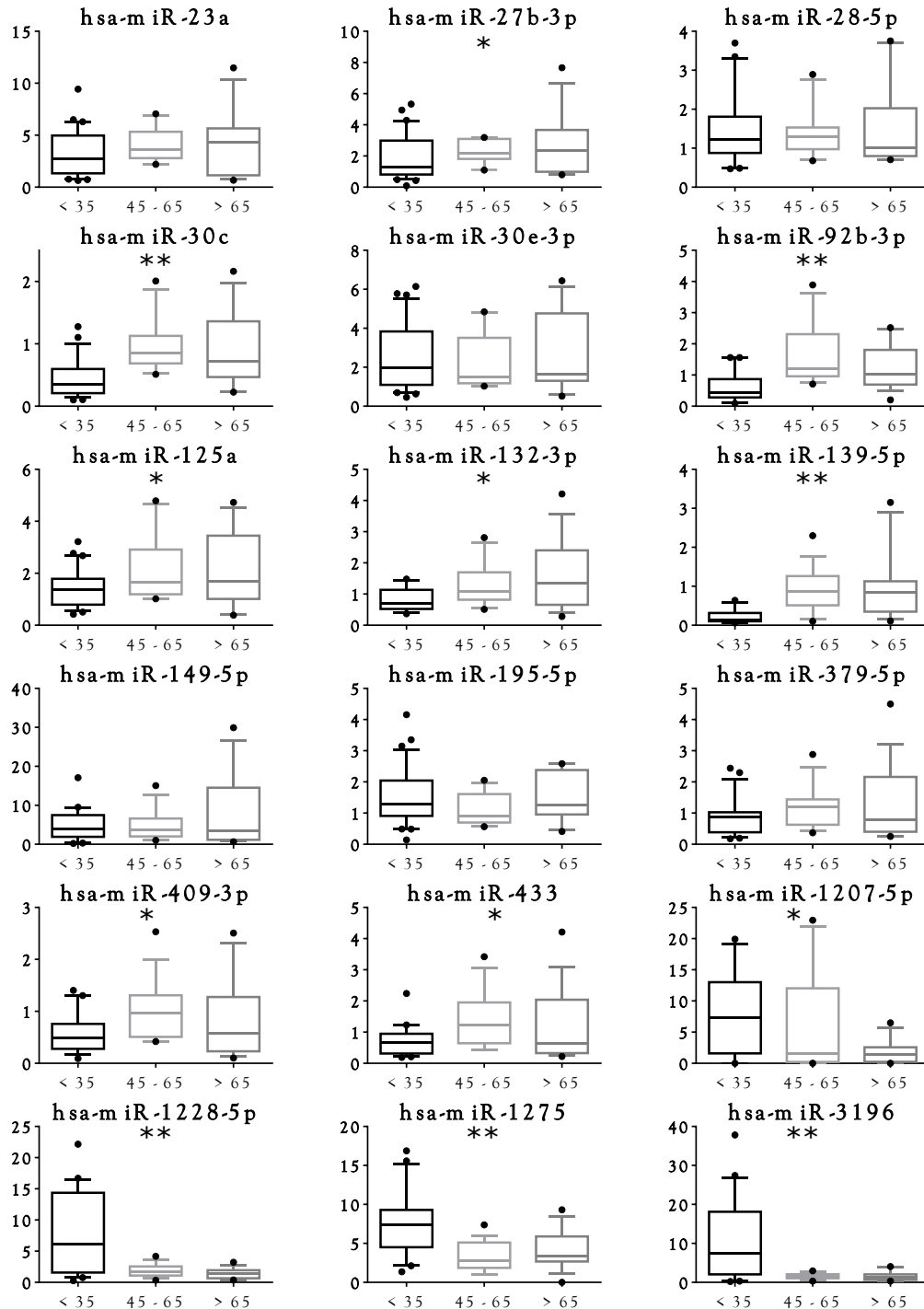


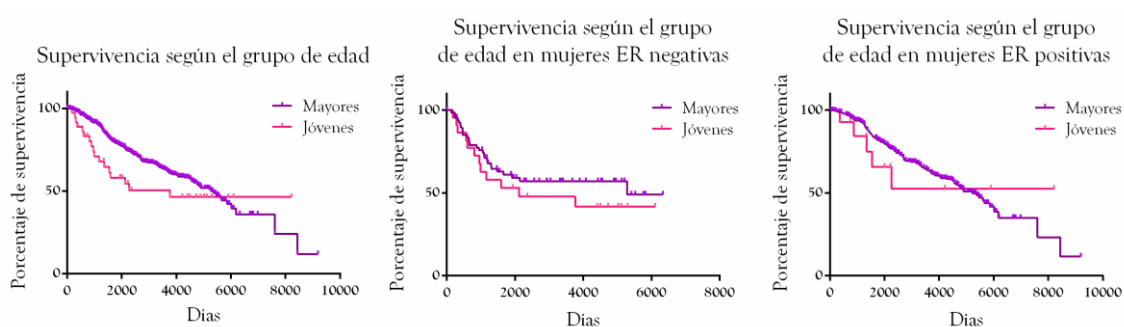
Figura 48. Expresión relativa de los miRNAs seleccionados de los nodos según el grupo de edad mediante qRT-PCR. Se muestran las diferencias en la expresión de los miRNAs seleccionados según el grupo de edad al que pertenecen de los grupos menores de 35 años, de entre 45 y 65 años y mayores de 65. En el eje de las Y se representa la expresión relativa. Se representa mediante diagrama de cajas la media, los cuartiles primero y tercero y los valores más allá del 10% y 90% de la muestra como puntos. \*P-valor < 0,1. \*\*P-valor < 0,0056.

En la Tabla 18 se observa que, de los 18 miRNAs seleccionados cuya expresión se evaluó mediante qRT-PCR, teniendo en cuenta un  $\alpha = 0,1$  (debido a que se trata de un test de una cola) 12 dieron resultados significativamente diferentes entre mujeres jóvenes y mayores: hsa-miR-27b, hsa-miR-30c, hsa-miR-92b, hsa-miR-125a, hsa-miR-132, hsa-miR-139, hsa-miR-409, hsa-miR-433,

hsa-miR-1207-5p, hsa-miR-1228, hsa-miR-1275 y hsa-miR-3196. Si aplicamos corrección por Bonferroni dividiendo el valor de  $\alpha$  entre el número de estudios simultáneos realizados, 18, el valor  $\alpha$  corregido pasó a ser 0,0056, por lo tanto, sólo aguantaron corrección por Bonferroni seis miRNAs: hsa-miR-30c, hsa-miR-92b, hsa-miR-139, hsa-miR-1228, hsa-miR-1275 y hsa-miR-3196. Se debe tener en cuenta que Bonferroni es un método muy estricto y que el resto de miRNAs significativos mantuvieron, en efecto, la misma tendencia que en los resultados de la primera fase. La distribución de los valores de cada grupo por miRNA puede observarse en la Figura 48.

## 1.2. Estudio de los miRNAs desregulados seleccionados como marcadores de la supervivencia

Se llevó a cabo el estudio de miRNAs desregulados seleccionados como marcadores moleculares asociados a la supervivencia, para ello se emplearon los datos de supervivencia de la base de datos de METABRIC. Tras construir las respectivas curvas de supervivencia según la edad, se comprobó que la supervivencia es diferente según el grupo de edad, con un p-valor de 0,11 en el test de Mantel-Cox y 0,0028 en el test Wilcoxon. Al separar las muestras en ER negativas y ER positivas, se observó que las ER negativas no mostraron diferencias estadísticas en cuanto a la supervivencia, siendo estas más acusadas en el grupo de ER positivas, estos datos deben tomarse con precaución, debido al reducido número de muestras procedentes de mujeres jóvenes tras fragmentar la muestra (Figura 49).

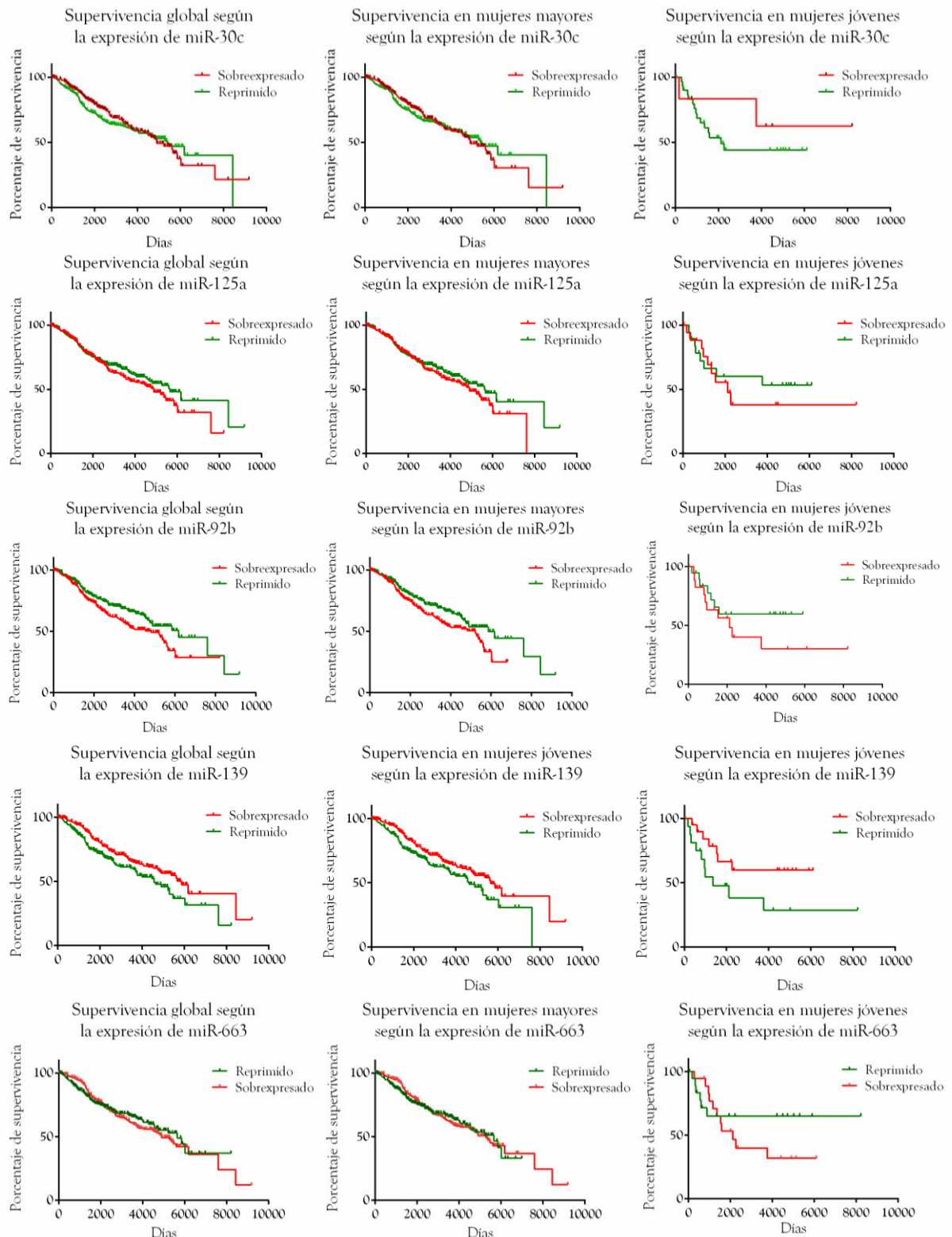


**Figura 49. Curvas de supervivencia asociada a los grupos de edad y comparación entre mujeres ER negativas y ER positivas.** En cada gráfica se representan las curvas de supervivencia según la pertenencia al grupo de mujeres mayores de 65 años (color morado) o menores de 35 años (jóvenes, en color rosa).

Se estudió también la supervivencia asociada a los miRNAs hsa-miR-30c, hsa-miR-125a, hsa-miR-17, hsa-miR-92b, hsa-miR-139 y hsa-miR-663, que fueron considerados como relevantes marcadores moleculares, bien por sus niveles de significatividad en los tres grupos de muestras analizados mediante microchip (pacientes, líneas celulares y datos procedentes de bases de datos online), por haber sido validados mediante qRT-PCR o bien por resultar funcionalmente relevantes en la bibliografía. Las curvas de supervivencia asociadas se representan en la Figura 50.

Se observaron diferencias significativas de supervivencia según la expresión de los miRNAs hsa-miR-92b y hsa-miR-139 al analizar la muestra global. En cuanto a hsa-miR-92b, mostró un p-valor de 0,018 en el test de Mantel-Cox y un 0,043 en el test Wilcoxon, mostrando una mejor supervivencia aquellas muestras con expresión del miRNA por debajo de la mediana (reprimido) con un riesgo de 0,72 (IC95%: 0,54 - 0,95), esta diferencia fue más evidente en mujeres mayores (p-valor 0,028) que en jóvenes (p-valor 0,22), probablemente debido al poco número de casos en el grupo de mujeres jóvenes. El miRNA hsa-miR-139 mostro un p-valor de 0,0091 en la prueba Mantel-Cox y 0,0055 en la de Wilcoxon con un HR de 1,29 (IC95% 0,97-0,78), presentando un mayor riesgo las muestras cuya expresión de hsa-miR-139 se encuentra reprimida. Al separar por

edad se obtuvo en mujeres mayores un p-valor de 0,019 en Mantel-Cox y 0,013 en Wilcoxon HR 1,22 (IC95% 0,91-1,63); en mujeres jóvenes el p-valor fue menos robusto, con un 0,074 en Mantel-Cox y un 0,070 en la prueba de Wilcoxon, además no se pudo calcular el riesgo por la poca cantidad de muestras reprimidas.



**Figura 50. Curvas de supervivencia asociadas a la expresión de miRNAs seleccionados como posibles marcadores.** En cada gráfica se representan las curvas de supervivencia según la expresión por encima de la mediana (sobrexpresado, en color rojo) o por debajo (reprimido, en color verde).

Los miRNAs siguientes mostraron p-valores cercanos a la significación en el test de Wilcoxon: hsa-miR-30c (p-valor 0,089), hsa-miR-27b (p-valor 0,072), aunque los resultados no fueron significativos. Se observó cierta correspondencia entre la supervivencia y la expresión del miRNA por encima de la mediana en el grupo de mujeres jóvenes, al separar las muestras según su grupo de edad, aunque, debido al poco número de muestras procedentes de mujeres jóvenes, no se alcanzó la significación estadística (ver Figura 50).

## 2. ANÁLISIS DE LA EXPRESIÓN DIFERENCIAL DE LOS GENES DIANA SELECCIONADOS Y SU PROTEÍNA EN MUESTRAS TUMORALES DE MUJERES JÓVENES

---

Con el fin de analizar la expresión de los receptores seleccionados, HER4 (gen *ERBB4*), NR2B (gen *GRIN2B*), NR3A (gen *GRIN3A*) y mGluR3 (gen *GRM3*), se midió la cantidad de RNA mensajero de los mismos mediante qRT-PCR en una serie de muestras de tejido tumoral biopsiado independiente, procedentes de pacientes de cáncer de mama. El análisis se llevó a cabo sobre un total de 50 muestras divididas en tres grupos: mujeres menores de 35 años, de entre 45 y 65 años y mayores de 65 años.

### 2.1. Estudio de la expresión en mujeres jóvenes del mRNA de los genes dianas seleccionados mediante qRT-PCR

Tras la pre-amplificación y la amplificación por PCR, los Ct medios obtenidos, su desviación típica y los valores RQ se encuentran en la Tabla 20. No se obtuvieron resultados en el caso de la sonda para el mRNA de *GRM3*, sí se obtuvo expresión para las mismas muestras del gen empleado como control, por lo que, tras repetir el experimento y obtener los mismos resultados nulos, se determinó que o bien las muestras seleccionadas no expresan *GRM3*, o bien la técnica empleada no permitió detectar su nivel de expresión.

#### Expresión anormal de los mRNAs de genes *GRIN2B* y *GRIN3A*

De modo similar, al observar la expresión de los genes *GRIN2B* y *GRIN3A* se observa expresión normal en algunas muestras, mientras que en otras se obtienen valores nulos de expresión. Los análisis de medias en estos casos se realizaron empleando las muestras que dieron valores de expresión positivos. Por otro lado, los resultados obtenidos se codificaron en una variable dicotómica de presencia/ausencia de receptor y se analizó su asociación con los grupos de edad mediante  $\chi^2$ . Ninguno de los grupos analizados dio resultados significativos, si bien, parece existir una tendencia a una menor presencia de ambos receptores en mujeres menores de 35 años (con un p-valor de 0,150). Particularmente se observó que (a excepción de un caso) cuando el mensajero de *GRIN2B* se encontraba ausente, también lo estaba *GRIN3A*, no ocurrió así cuando *GRIN2B* se encontraba presente, pudiendo en este caso *GRIN3A* encontrarse o no (Figura 51).

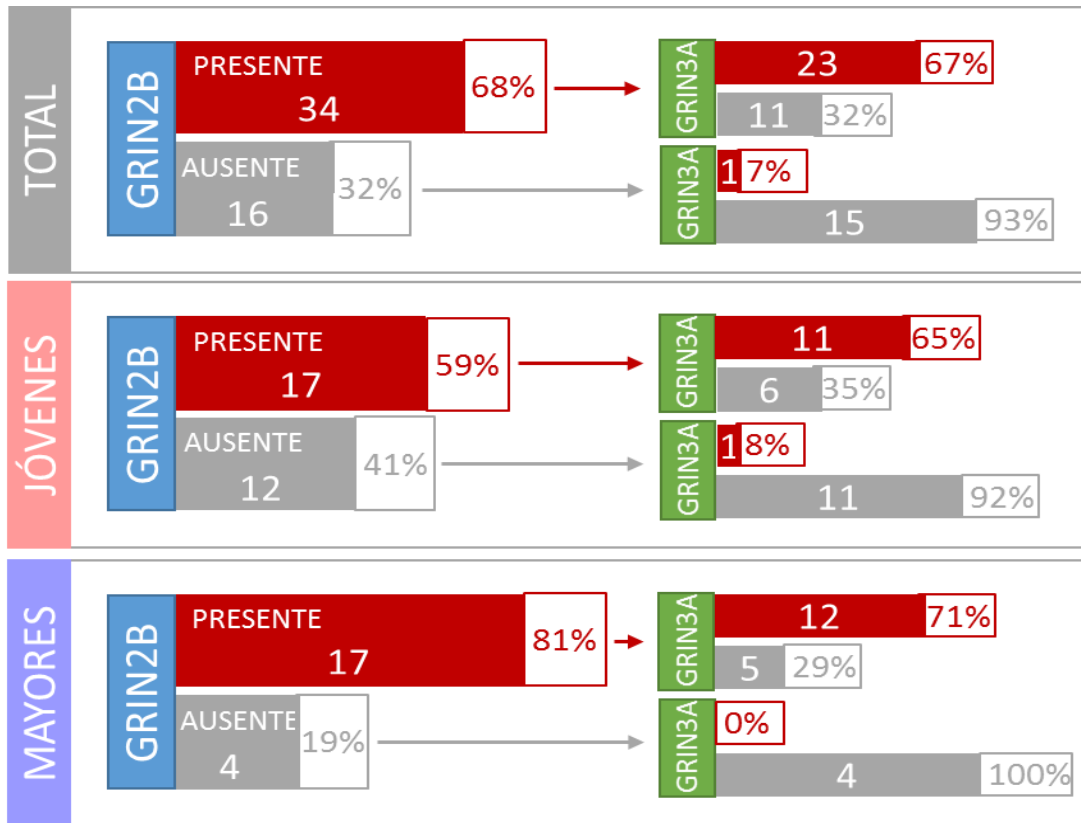


Figura 51. Esquemático de la presencia de expresión de las subunidades GRIN2B y GRIN3A del receptor de glutamato NMDAR. Se muestran los resultados en conjunto y separados por pacientes jóvenes (>35 años) y mayores (<45 años). Se indica el número total y el porcentaje calculado del total de su grupo por gen. La columna en rojo muestra el recuento de muestras con presencia del mRNA, las de color gris, aquellas con ausencia de mRNA.

### Análisis de la expresión diferencial en mujeres jóvenes de los receptores seleccionados

Tras obtener los valores de expresión relativa mediante el método  $\Delta\Delta C_t$ , descrito anteriormente, y eliminar los valores extremos, se realizó un análisis de medias entre los grupos. En primer lugar se realizó un test de normalidad, tras el cual se decidió emplear un test paramétrico (ANOVA) o no paramétrico (Kruskal-Wallis) entre los tres grupos de edad y, posteriormente, un t-test (o un test de Mann-Whitney en el caso de realizarse una prueba estadística no paramétrica) entre las muestras procedentes de mujeres menores de 35 y aquellas mayores de 45 años (este grupo es el resultante de unir el grupo de las mujeres de entre 45 y 65 años y las mayores de 65).

Los valores de expresión relativa empleados se detallan en la Tabla 19 y se representan en la Figura 52. En el caso de *GRIN2B* y *GRIN3A* se emplearon los valores de aquellas muestras que resultaron positivas en expresión para éstos genes.

**Tabla 19:** Valores de expresión en tumor mamario obtenidos para los receptores *ERBB4*, *GRIN2B* y *GRIN3A*

Muestra	Edad	<i>ERBB4</i>			<i>GRIN2B</i>			<i>GRIN3A</i>		
		MCt	SD	RQ	MCt	SD	RQ	MCt	SD	RQ
BCVY18579	< 35	26,38	0,19	2,35	-	-	-	-	-	-
BCVY12681	< 35	23,07	0,11	0,26	30,28	0,01	1,56	30,25	0,04	15,36
BCVY133	< 35	29,29	0,06	0,14	29,29	0,42	25,96	36,22	0,43	2,07
BCVY14310	< 35	22,71	0,19	6,93	-	-	-	35,16	0,18	10,64
BCVY15243	< 35	22,49	0,32	0,07	30,79	0,00	0,24	34,40	0,09	0,19
BCVY15244	< 35	22,69	0,07	0,21	28,70	0,08	0,68	32,00	0,11	0,66
BCVY162	< 35	24,75	0,29	1,58	29,36	0,04	37,94	33,87	0,11	16,17
BCVY176	< 35	25,06	0,36	0,41	-	-	-	-	-	-
BCVY18368	< 35	28,38	0,25	0,00	27,71	0,34	0,72	34,03	0,03	0,09
BCVY189	< 35	28,14	0,29	0,01	30,28	0,16	-	33,38	0,12	-
BCVY1920	< 35	25,09	0,15	0,01	27,03	0,09	3,00	32,87	0,05	0,51
BCVY1924	< 35	30,22	0,07	0,08	27,22	0,03	127,98	31,17	0,06	80,05
BCVY25503	< 35	24,22	0,19	1,21	30,92	0,01	9,62	31,94	0,06	46,02
BCVY2734	< 35	22,78	0,34	0,11	31,91	0,09	0,33	31,31	0,30	4,79
BCVY2906	< 35	31,06	0,20	0,21	-	-	-	-	-	-
BCVY4988	< 35	26,53	0,54	0,06	31,55	0,35	3,74	-	-	-
BCVY96	< 35	32,71	0,03	0,06	-	-	-	-	-	-
BCVY9649	< 35	26,33	0,25	0,97	32,39	0,05	18,13	-	-	-
BCVYV10	< 35	25,94	0,13	1,82	-	-	-	-	-	-
BCVYV12	< 35	22,62	0,33	0,97	29,41	0,03	15,48	-	-	-
BCVYV15	< 35	27,00	0,10	6,85	32,98	0,09	55,57	-	-	-
BCVYV2	< 35	23,85	0,17	7,83	28,27	0,04	1,01	-	-	-
BCVYV3	< 35	28,96	0,25	4,74	-	-	-	-	-	-
BCVYV4	< 35	-	-	-	-	-	-	-	-	-
BCVYV5	< 35	29,06	0,23	0,42	-	-	-	-	-	-
BCVYV6	< 35	23,81	0,09	0,46	28,40	0,09	3,89	-	-	-
BCVYV7	< 35	22,97	0,49	13,13	-	-	-	-	-	-
BCVYV8	< 35	26,16	0,02	2,30	-	-	-	-	-	-
BCVYV9	< 35	24,60	0,21	35,22	-	-	-	-	-	-
10COP	45 - 65	22,53	0,32	0,31	29,34	0,11	3,04	-	-	-
10DP	45 - 65	24,68	0,15	0,17	33,25	0,10	0,22	30,95	0,05	10,41
1DP	45 - 65	24,36	0,07	1,00	-	-	-	-	-	-
BC16	45 - 65	22,22	0,26	0,85	31,81	0,26	1,04	-	-	-
BC17	45 - 65	21,97	0,02	0,20	31,04	0,03	0,32	31,36	0,06	2,49
BC21	45 - 65	19,61	0,29	2,91	33,32	0,08	0,29	33,06	0,13	3,41
BC24	45 - 65	28,86	0,01	0,01	30,29	0,11	0,61	32,78	0,05	1,05
BC26	45 - 65	23,39	0,33	0,08	30,92	0,43	0,08	34,83	0,01	0,05
7DP	> 65	22,97	0,09	1,16	30,71	0,08	6,58	34,24	0,07	5,51
BC10	> 65	21,68	0,15	1,71	31,73	0,10	0,33	-	-	-
BC28	> 65	19,66	0,08	1,57	30,11	0,05	1,05	31,10	0,15	5,15
BC30	> 65	20,82	0,35	0,82	27,61	0,01	15,12	32,32	0,04	5,58
BC32	> 65	26,50	0,07	0,36	-	-	-	-	-	-
BC33	> 65	21,28	0,22	5,48	31,36	0,08	3,45	33,23	0,20	9,12
BC34	> 65	26,58	0,08	0,03	30,50	0,16	0,39	30,11	0,07	4,95
BC35	> 65	24,76	0,09	2,69	-	-	-	-	-	-
BC36	> 65	23,75	0,24	0,10	31,54	0,13	0,09	-	-	-
BC38	> 65	19,72	0,36	1,54	31,63	0,05	0,65	33,53	0,07	1,68
BC40	> 65	24,34	0,10	0,61	32,50	0,03	2,55	-	-	-
BC41	> 65	21,07	0,02	0,69	30,56	0,05	1,00	33,84	0,09	1,00
BC43	> 65	24,28	0,36	0,94	-	-	-	-	-	-

MCt: Ciclo umbral medio; SD: desviación típica de los Ct; RQ: valor de expresión relativa, del inglés *Relative Quantification*.



### ***ERBB4***

El test de normalidad D'Agostino realizado reveló que las poblaciones de mujeres mayores de 45 años y mayores de 65 se ajustan a la normalidad (p-valor de 0,053), sin embargo la población de mujeres jóvenes no lo hace, con un p-valor de 0,033. El test de normalidad Shapiro-Wilk rechazó la normalidad para ambas poblaciones, por lo que fueron tratadas como poblaciones que no siguen una distribución normal.

Se realizó un test de medias para  $k$  poblaciones no paramétrico (Kruskall-Wallis), que dio un p-valor de 0,244 al comparar los tres grupos de edad por separado. Al juntar las poblaciones de mujeres de mediana edad y mayores de 65 años y realizar un test U de Mann-Whitney, el p-valor bajó hasta 0,112. Aun así, no se puede rechazar la hipótesis nula, por lo que no existen evidencias estadísticas significativas de que la expresión del gen *ERBB4* sea diferente en mujeres jóvenes.

### ***GRIN2B***

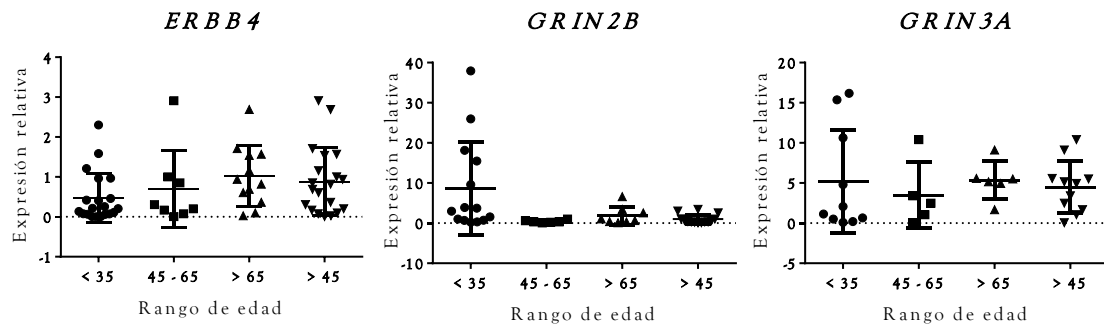
El test de normalidad D'Agostino & Pearson reveló que las poblaciones de muestras jóvenes y mayores no se ajustan a la normalidad, con un p-valor de 0,0129 para las mujeres jóvenes y de 0,043 para las mayores, por lo que se descarta la hipótesis nula de ajuste a la normalidad. Para el grupo de mujeres de mediana edad no se pudo calcular el test de normalidad debido al bajo número de muestras, por lo que se realizó el mismo test incluyéndolas en el grupo de mujeres mayores. De este modo el test dio un p-valor de 0,0129 para mujeres jóvenes y de 0,0598 para mayores. Para comprobar la desviación de la normalidad, debido a la discrepancia de ambos resultados, se realizó un test de Shapiro-Wilk que rechazó la hipótesis nula en ambas poblaciones (p-valor en mujeres jóvenes de 0,0018; p-valor en mujeres mayores de 0,0015), por lo que se determinó que ambas poblaciones no se ajustan a la normalidad.

El test no paramétrico para comparar medias entre más de dos grupos empleado fue el test de Kruskal-Wallis. Se obtuvo un p-valor de 0,0108, un valor significativo que confirma la diferencia en la expresión de *GRIN2B* entre los tres grupos. Tras unir los grupos de mujeres de mediana edad y mujeres mayores se realizó un test U de Mann-Whitney para comparar medias, obteniendo un valor de significación inferior a 0,1 (0,0029). Ambos valores aguantan una corrección por Bonferroni para 3 comparaciones (el valor  $\alpha$  resultante fue 0,016), confirmando que las mujeres jóvenes tienen una expresión de *GRIN2B* superior a mujeres de mediana edad y mayores de 65 años, cuya expresión es similar entre sí.

### ***GRIN3A***

Los test de normalidad empleados revelaron que la expresión de *GRIN3A* sigue una distribución normal en las poblaciones evaluadas, dando p-valores  $> 0,05$  por lo que se admite la hipótesis nula.

El test paramétrico ANOVA para evaluar las diferencias entre los tres grupos dio un p-valor de 0,860 y el t-test, realizado en los dos grupos resultantes tras unir las poblaciones de mujeres de entre 45 y 65 años y mayores de 65 años, un p-valor de 0,381. Ambos p-valores están por encima del umbral de significación, por lo que se admite que las poblaciones no difieren entre sí.



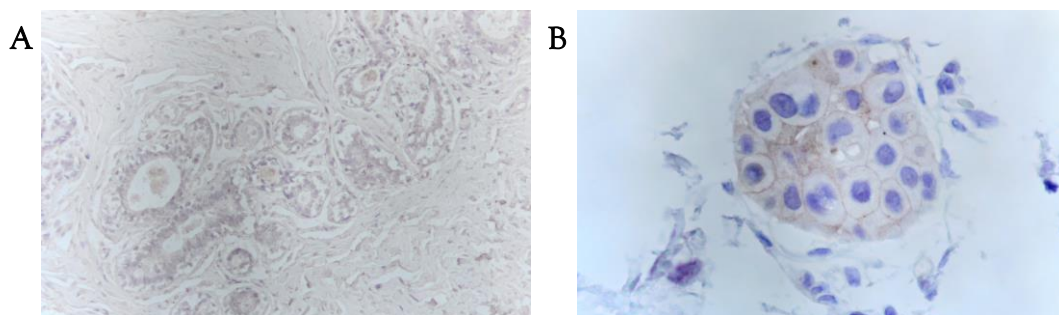
**Figura 52. Expresión relativa de los mRNAs de los genes *ERBB4*, *GRIN2B* y *GRIN3A*.** El diagrama de puntos de la izquierda representa la mediana, los cuartiles 1° y 3°, y los valores mínimo y máximo. El diagrama de la derecha representa los valores individualmente, junto con la media y la desviación típica. El eje de las X se representa el grupo de edad, mientras que en el eje de las Y se muestra la expresión relativa para cada mRNA, nótese que la escala de cada gráfica es diferente para comodidad del lector y por considerar los análisis independientes entre sí. El grupo denominado “>45” es el grupo resultante de juntar el grupo de mujeres de mediana edad (45-65) con el grupo de mujeres mayores (65).

## 2.2. Validación de la presencia diferencial de HER4 según el grupo de edad en células tumorales mediante inmunohistoquímica en tejido tumoral parafinado

Para evaluar la expresión diferencial de los receptores en tejido tumoral de mujeres jóvenes, se midió la cantidad de proteína expresada *in situ*, mediante tinción inmunohistoquímica empleando anticuerpos contra el receptor HER4. Se evaluaron un total de 60 muestras de pacientes de cáncer de mama por triplicado, junto con tejido mamario sano de 15 de ellas y 4 muestras de tejido procedente de mujeres sanas que se emplearon como control.

La tinción se evaluó en una escala subjetiva pero estandarizada según la evaluación de la tinción del receptor HER2, ampliamente evaluado en tumores mamarios.

Posteriormente se codificó de 0 a 5, según la intensidad de tinción. La tinción obtenida fue específica y limpia pudiéndose evaluar correctamente la tinción (Figuras 53 y 54).

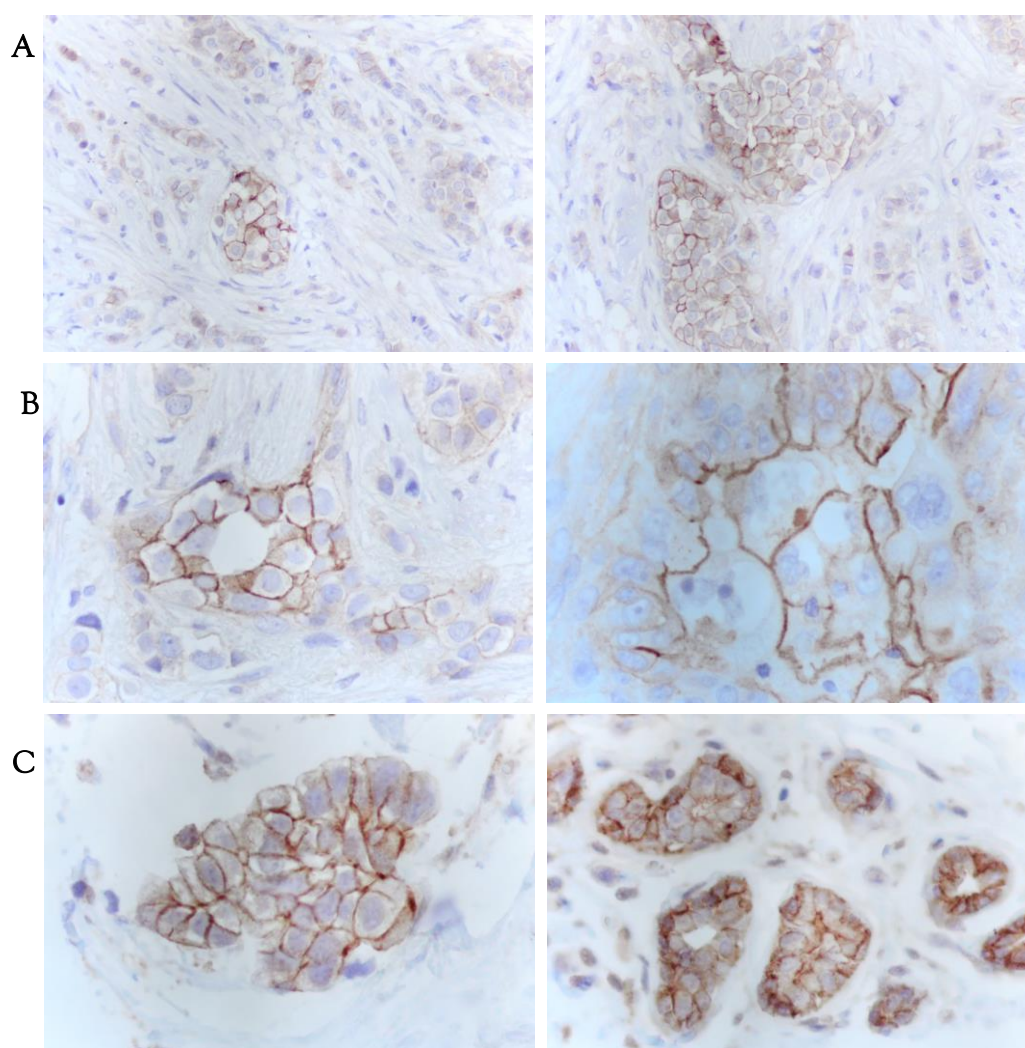


**Figura 53. Tinción negativa (A) y débil (B) mediante IHC anti-HER4 de tejido mamario FFPE.** A, tejido mamario procedente de una mujer joven libre de cáncer, no se observa tinción. B, tinción débil de membrana en tejido tumoral de mama. Las tinciones se realizaron empleando desenmascaramiento antigénico por presión en autoclave, seguida de incubación 60' del anticuerpo primario y 30' del anticuerpo secundario, la tinción se visualizó mediante el método EnVision, con tinción de DAB al 5%.

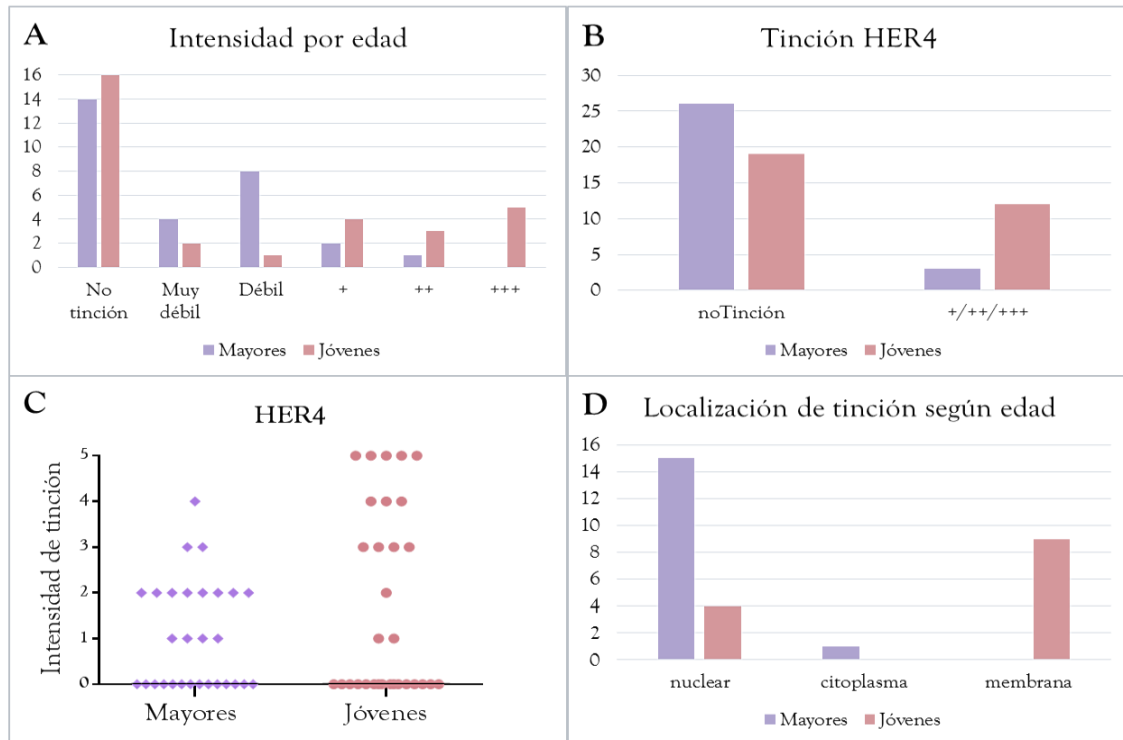
De las 60 muestras analizadas, 29 fueron mayores, de las cuales 26 no mostraron tinción o esta fue muy débil, mientras que sólo 3 tuvieron tinción positiva; en el caso de las 31 muestras

procedentes de mujeres jóvenes, 19 no mostraron tinción específica para HER4 o muy débil, sin embargo 12 muestras dieron positivo para HER4 (Figura 55A).

Al agrupar las muestras en dos grupos: “tinción” (incluye las muestras identificadas como 1+/2+/3+) y “no tinción” (aquellas cuya tinción es negativa, muy débil o débil) y someter los grupos a un test de Pearson ( $\chi^2$ ) se obtuvieron resultados significativamente diferentes para mujeres jóvenes, con un p-valor de 0,011 (Figura 55B). Se observó además que las muestras procedentes de mujeres jóvenes tenían mayor intensidad que las de mujeres mayores, alcanzando valores de intensidad de 2+ o 3+, mientras que las procedentes de mujeres mayores alcanzaron una tinción de 1+ (2 muestras) y 2+ (1 muestra). Aun así, la mayoría de las muestras en ambos grupos mostraron tinción negativa (Figura 55C)

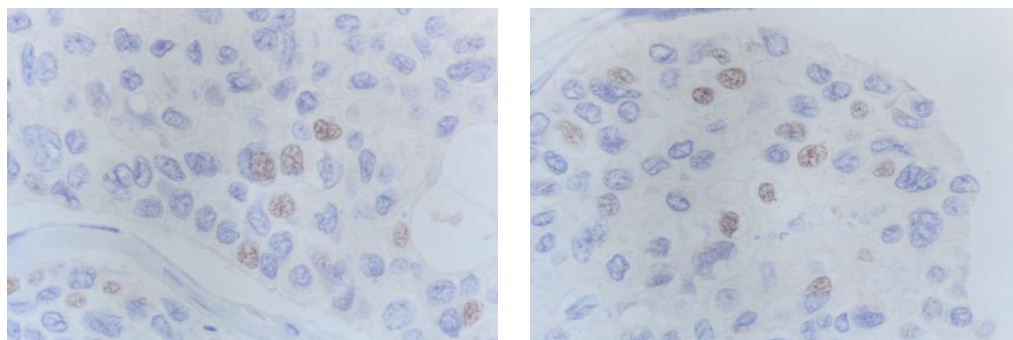


**Figura 54. Tinción positiva de membrana mediante IHC anti-HER4 de tejido FFPE tumoral mamario. A,** tinciones categorizadas como 1+. **B,** tinciones categorizadas como 2+. **C,** tinciones categorizadas como 3+. Las tinciones se realizaron empleando desmascaramiento antigénico por presión en autoclave, seguida de incubación 60' del anticuerpo primario y 30' del anticuerpo secundario, la tinción se visualizó mediante el método EnVision, con tinción de DAB al 5%.



**Figura 55. Intensidad y localización de la tinción del anticuerpo anti-HER4 en muestras procedentes de mujeres jóvenes comparadas con mujeres mayores.** A) Distribución de la tinción en mujeres mayores (morado) y jóvenes (rosa). B) Comparación entre mujeres mayores y jóvenes en la presencia o ausencia de tinción, considerando como tinción positiva aquellas tinciones categorizadas como 1+ (+), 2+ (++) o 3+ (+++). C) Intensidad de tinción de HER4 en ambos grupos de edad, codificada como: 0, no tinción; 1, tinción muy débil; 2, tinción débil; 3, tinción identificada como 1+; 4, tinción 2+; 5, tinción identificada como 3+. D) Distribución de la localización de la tinción en la muestra según los dos grupos de edad.

Se observan diferencias en la localización de la tinción entre mujeres mayores y mujeres jóvenes. Mientras que las mujeres mayores mostraron un patrón de tinción nuclear prácticamente en su totalidad (15 muestras a excepción de una muestra, cuya tinción fue citoplasmática), las muestras procedentes de mujeres jóvenes mostraron mayor diversidad, cuatro de ellas tenían tinción nuclear y nueve de membrana (Figuras 55D y 56). Al excluir la muestra con tinción citoplasmática y comparar aquellas con tinción nuclear o de membrana mediante un test de Pearson ( $\chi^2$ ) entre mujeres jóvenes y mayores, se obtuvo un p-valor inferior a  $10^{-4}$ , indicando que existen diferencias significativas en la localización de la tinción de HER4 entre mujeres jóvenes y mayores, observándose una mayor proporción de tinción de membrana en tumores procedentes de mujeres jóvenes.



**Figura 56. Tinción nuclear mediante IHC anti-HER4 de tejido FFPE tumoral mamario.** Se muestra la tinción presente en una muestra de tejido tumoral mamario de una paciente mayor de 45 años. Las tinciones se realizaron empleando desenmascaramiento antigénico por presión en autoclave, seguida de incubación 60' del anticuerpo primario y 30' del anticuerpo secundario, la tinción se visualizó mediante el método EnVision, con tinción de DAB al 5%.

# DISCUSIÓN



## PARTE 1: VIABILIDAD DEL TEJIDO TUMORAL BIOPSIADO CONSERVADO EN PARAFINA PARA ESTUDIOS MOLECULARES

---

Las muestras frescas congeladas constituyen el material idóneo de partida para llevar a cabo estudios moleculares a gran escala, sin embargo, este tipo de material no siempre está disponible. Así pues, a menudo resulta necesario trabajar con muestras en otras condiciones de conservación, como son los bloques de parafina, que permiten llevar a cabo estudios retrospectivos (220) (Klopfleisch 2011). Con el fin de evaluar la posible utilidad de las muestras conservadas en bloques de parafina en estudios moleculares, se realizó un análisis comparativo de muestras fijadas en formol e incluidas en parafina y muestras congeladas en condiciones óptimas.

En primer lugar se llevó a cabo el análisis de la integridad del RNA total y del RNA de pequeño tamaño. Ambos análisis mostraron que las muestras parafinadas están altamente degradadas. El perfil electroforético de las muestras parafinadas presentó una acumulación de fragmentos de RNA de un tamaño inferior a 200 nucleótidos, debido a la degradación derivada del proceso de fijación y extracción de RNA al que se someten este tipo de muestras. En una muestra parafinada, el proceso de fijación en formol e inclusión en parafina provoca una degradación enzimática y una degradación química del RNA, además, produce entrecruzamientos entre las moléculas de RNA y las proteínas, haciendo al RNA más resistente a la extracción (156, 159, 161). La degradación de las muestras viene acompañada de la formación de pequeños fragmentos de RNA que puede conducir a una sobrestimación de la cantidad de miRNAs.

Así pues, de acuerdo con el trabajo de Becker y colaboradores, este método no resulta adecuado para la cuantificación exacta de miRNAs en muestras degradadas. Es prácticamente imposible cuantificar la cantidad real exacta de miRNAs presentes en una muestra biológica con los métodos actualmente disponibles (221).

Es por eso que se planteó la realización de qRT-PCR para evaluar la calidad del material recuperado y su utilidad de cara a realizar estudios moleculares a gran escala. Para ello se seleccionaron microRNAs, mRNAs y snoRNAs y se evaluaron en muestras parafinadas y en sus correspondientes muestras pareadas congeladas.

Los resultados muestran que los miRNAs *hsa-miR-125b* y *hsa-miR-191* y el snoRNA *RNU6B* en muestras parafinadas se detectan mejor (a un menor Ct) que en muestras congeladas. En el caso de los mRNAs, *GAPDH* y *HPRT*, su detección en muestras fue a un Ct muy superior al obtenido cuando se emplearon muestras congeladas óptimas, de acuerdo a trabajos anteriores en los que se describe que a mayor tamaño del amplicón, mayor diferencia entre el Ct al que se detecta un gen en una muestra óptima respecto a su correspondiente parafina degradada (222). Debido a los entrecruzamientos del RNA, la cantidad de RNA que puede extraerse de una muestra parafinada representa sólo una fracción de la que puede obtenerse de tejido fresco congelado, además cuanto mayor es la molécula de RNA entrecruzada, más difícil resulta su extracción, de modo que, las moléculas más pequeñas son extraídas con mayor facilidad que las grandes, dando como resultado una mayor representación de moléculas de pequeño tamaño en muestras parafinadas que en congeladas (222).

La comparación entre muestras pareadas determinó una correlación estadísticamente significativa con  $r=0,7312$  entre las muestras congeladas y sus correspondientes parafinas cuando

se analizaron las seis moléculas de RNA estudiadas. Considerando sólo el análisis de pequeños RNAs (miRNAs y snoRNA), la correlación entre los congelados y sus parafinas mejoró considerablemente ( $r$  pasó a 0,9248). Estudios previos demuestran esta sorprendente buena correlación en el análisis de miRNAs entre muestras parafinadas y congeladas (Li, Doleshal, Xi). De hecho, dado su pequeño tamaño, su similitud estructural y su asociación a proteínas del complejo RISC, los miRNAs constituyen un tipo de moléculas menos susceptibles a la degradación (164, 220).

Se debe tener en cuenta que en este trabajo no se han considerado algunos factores, como el tipo y la duración de la anestesia, la medicación del paciente, el procedimiento quirúrgico y los tiempos de isquemia, que podrían conducir a la alteración de la expresión de determinados genes y miRNAs, influyendo en el análisis de calidad e integridad del RNA.

Los miRNAs han demostrado ser una herramienta útil como marcadores moleculares, demostrando su valor diagnóstico y pronóstico, además permiten clasificar a las pacientes de cáncer de mama de manera eficaz. Su alta estabilidad y fácil recuperación a partir de diferentes fuentes de material, ofrece un acceso a mayor número de muestras, facilitando los estudios moleculares y la obtención de una clasificación molecular más rentable y, por tanto, reproducible y accesible por una gran mayoría de la comunidad científica y médica. Permiten, además, su detección como biomarcador en suero, facilitando la obtención de biopsias líquidas y minimizando el uso de métodos invasivos de seguimiento (223, 224). Por todo ello, los miRNAs constituyen un potente marcador de estudio, con valor pronóstico, diagnóstico y exploratorio, ya que permiten detectar una desregulación epigenética que aporte información adicional sobre los procesos biológicos que desencadenan la formación de tumores, proporcionando una nueva vía de acción terapéutica, mediante el control de la desregulación de los miRNAs (86).

Los resultados obtenidos sugieren que los miRNAs constituyen el marcador molecular idóneo y más robusto cuando se dispone de muestras parafinadas, dada su alta estabilidad y su resistencia a la degradación (220).



## PARTE 2: ANÁLISIS DEL PERFIL DE EXPRESIÓN DE MIRNAS EN CÁNCER DE MAMA DE MUJERES JÓVENES

---

### *2.1. Las mujeres jóvenes, menores de 35 años, como entidad clínica y biológica diferenciada*

El cáncer de mama se clasifica atendiendo principalmente a su subtipo molecular, el tratamiento recibido se determina en base a esta clasificación motivada por una mayor eficacia y, por tanto, una mayor tasa de supervivencia final. Existe, pues, una fragmentación del cáncer de mama en entidades diferentes, con diferentes pronóstico y comportamiento general. La segmentación en base a la edad no es inaudita, habiéndose establecido dos tipos de cáncer de mama, basados en dos picos de edad de diagnóstico, 65 años y 50 años, denominados post-menopáusicos y pre-menopáusicos (aunque en algunos casos pueda ser peri-menopáusicos (167). Es posible que estemos obviando un tercer grupo, basándonos únicamente en su incidencia. Si bien el cáncer de mama en mujeres menores de 35 años es poco frecuente en la población, atiende a otros factores, no menos importantes, como son la baja supervivencia de este grupo de pacientes, y las consideraciones especiales que este grupo de edad requiere, principalmente de cara a garantizar la fertilidad y a minimizar secuelas laborales, sociales y psicológicas (54, 65). Además, se trata del grupo de edad cuyo aumento en la mortalidad por cáncer de mama ha sido mayor desde el año 1970.

Usualmente las mujeres jóvenes presentan características tumorales más agresivas que desembocan en un peor pronóstico, por lo que se han llevado a cabo numerosos estudios sobre las características clínicas particulares del cáncer de mama en mujeres jóvenes (225, 226) Existe controversia sobre si las mujeres jóvenes con cáncer de mama son una entidad biológica diferenciada dentro del cáncer de mama general o si, por el contrario, su comportamiento particular responde a una sobre-representación de subtipos agresivos, debida principalmente al fallo en la detección precoz de tumores mamarios en este grupo de edad.

En la literatura no existen estudios previos que analicen la expresión de miRNAs en cáncer de mama de mujeres muy jóvenes, sin embargo, sí existen trabajos que evidencian una diferencia molecular en este grupo de pacientes. Anders y colaboradores (2008) apuntaron a una diferencia en la expresión de genes en mujeres jóvenes (57). En este trabajo sacan dos conclusiones relevantes, en primer lugar que en el caso de mujeres jóvenes, la edad es un factor pronóstico *per se*, independientemente del estatus hormonal, contando con peor pronóstico las pacientes más jóvenes. Además, el efecto pronóstico de la edad, no se muestra en pacientes mayores, enfatizando la influencia de pertenecer al grupo de mujeres muy jóvenes en el pronóstico del cáncer de mama. Este resultado apoya la existencia de diferencias moleculares exclusivas y asociadas a las pacientes más jóvenes, como demuestra nuestro trabajo. En segundo lugar, el estudio de Anders identifica dos rutas oncogénicas con mayor probabilidad de desregulación en tumores de mujeres jóvenes, comparado con mujeres mayores. En Azim et al 2012, también se concluye que, en efecto, existe una complejidad biológica añadida en pacientes de cáncer de mama jóvenes (227).

Un trabajo publicado recientemente, apoya las existencia de un perfil de expresión de genes asociado a mujeres jóvenes (menores de 45 años), remarcando la importancia, en este caso, de

rutas asociadas a la respuesta inmune, concretamente, del gen CXCL, cuya expresión se encuentra sobre-expresada en mujeres jóvenes (228).

Existe una cierta arbitrariedad en la definición del término “joven” o “muy joven”, habiendo estudios que clasifican a las mujeres jóvenes por debajo de 45, 40, 35 e incluso 30 años (52, 56, 174, 227, 229-234). Para poder establecer comparaciones entre estudios y definir eficazmente la biología molecular del cáncer de mama en mujeres jóvenes es necesario establecer una edad límite estándar que atienda a criterios biológicos y clínicos específicos (65). Con esto en mente, Han y colaboradores (54) realizaron un estudio de las características clínicas de pacientes jóvenes de cáncer de mama por grupos de 5 años de edad desde 30 hasta 45 años. Concluyen que existe una diferencia más evidente en la supervivencia de mujeres menores de 35 años, proponiendo como edad límite 35 años (54). Es por esto que el presente estudio se ha centrado en las pacientes menores de 35 años de edad, en contraposición con las mujeres mayores de 65 años, que es la edad común de aparición del cáncer de mama (167).

Si se parte de la base de que, en efecto las mujeres jóvenes constituyen un grupo homogéneo dentro del cáncer de mama, su mayor agresividad, bien sea causa o efecto de un comportamiento biológico diferente, dejará una huella molecular. Por ello se planteó el análisis de la expresión de miRNAs a partir de muestras de tejido parafinado del Hospital Clínico de Valencia.

Una parte importante del cáncer de mama en mujeres jóvenes se debe a la mutación en los genes BRCA1 y BRCA2, pero este grupo constituye un grupo diferente dentro del cáncer de mama, ya que se engloba dentro del cáncer de mama familiar (174). De hecho, un estudio reciente ha demostrado la diferencia molecular existente entre mujeres menores de 35 años con cáncer de mama esporádico y con cáncer de mama familiar debido a alteraciones en los genes BRCA1/2. En este trabajo, Bastos y colaboradores destacan que ambos grupos de pacientes presentaban las mismas características clínicas, sin embargo, existían diferencias moleculares, siendo el perfil de miRNAs identificado capaz de diferenciar las muestras según pertenecían a cáncer de mama familiar o esporádico (229). Por este motivo, en el presente trabajo se han seleccionado únicamente pacientes de cáncer de mama esporádico sin mutación en BRCA1/2, lo que aporta homogeneidad al grupo de mujeres muy jóvenes y robustez al perfil obtenido.

El porcentaje de celularidad tumoral en la muestra puede influir en el perfil molecular obtenido, ya que la matriz extracelular cuenta con un perfil molecular propio, pudiendo añadir variabilidad entre los grupos al estar una muestra sobre-representada en adipocitos. Además, las mujeres muy jóvenes poseen una densidad mamaria mayor, lo que también podría alterar el perfil molecular (235). Para minimizar esta variabilidad se emplearon en todos los casos muestras que superasen el 60% de celularidad tumoral.

Debido a la baja incidencia del cáncer de mama en mujeres jóvenes, el acceso a muestras tumorales de este grupo de pacientes es limitado. Como se observa en la Figura 8, se partió de una base de 5381 pacientes donde se esperaba que 296 (un 5%) fuesen menores de 35 años. Un total de 215 eran menores de 35 años, se seleccionaron 180 por no tener mutación en BRCA1/2, de acuerdo con los criterios de selección establecidos, de las cuales se disponía de biopsia de 148. Tras filtrar las muestras disponibles por motivos de disponibilidad de material parafinado y calidad y cantidad del material genético extraído, se seleccionaron finalmente 43

muestras procedentes de mujeres jóvenes que no habían recibido tratamiento antes de la biopsia, además, se seleccionó una muestra representativa de todos los subtipos.

La exhaustiva selección de pacientes jóvenes llevada a cabo dota al estudio de robustez y fiabilidad en los resultados obtenidos. En trabajos previos, se han obtenido perfiles de expresión de miRNAs asociados a subtipo tumoral y a otras características tumorales, por lo que las muestras de mujeres mayores, seleccionadas en base a los mismos criterios que en las pacientes jóvenes, fueron pareadas en función del subtipo y características tumorales para minimizar el posible sesgo introducido debido a las características tumorales agresivas, más representadas en el grupo de pacientes jóvenes.

Se emplearon, además, tres muestras control de mujeres sanas jóvenes, para minimizar los efectos externos, como son los cambios moleculares propios de la edad o el estilo de vida. Se controló además, en la medida de lo posible, que estas mujeres presentasen parámetros clínicos y demográficos similares al de las pacientes jóvenes (rango de edad, IMC, nacionalidad, etnia) (35, 236).

El desarrollo de nuevas técnicas de alto rendimiento para la detección de la expresión génica han permitido el análisis a gran escala de la expresión de marcadores moleculares, como son los miRNAs, pudiendo establecer perfiles de expresión asociados a determinadas enfermedades, incluyendo tumores malignos. Cada vez más, los miRNAs se emplean como marcadores de diagnóstico y pronóstico, y permitiendo una selección eficaz de los pacientes de cara a un tratamiento personalizado y eficaz.

Debido a la idoneidad de los miRNAs como marcador molecular en muestras parafinadas descrita en este trabajo, y al bajo porcentaje de muestras de mujeres jóvenes disponibles, se decidió emplear los miRNAs como marcador molecular. Además el estudio de miRNAs resulta menos costoso que las técnicas actuales que emplean la expresión de los genes, en el ámbito del diagnóstico, clasificación y pronóstico, alzándose como un marcador más adecuado, principalmente por su gran estabilidad frente a la degradación, gran especificidad y reproducibilidad de los perfiles obtenidos (75, 79).

La plataforma Affymetrix permite el análisis mediante el chip de hibridación GeneChip® miRNA 2.0, por muestra, de la expresión de 15000 RNAs de pequeño tamaño no codificante, incluyendo snoRNA, miRNAs y sus precursores, entre los que se destacan los 1100 miRNAs humanos maduros, analizados en este estudio. Se trata de una tecnología ampliamente usada y respaldada por la comunidad científica, por lo que fue la plataforma elegida en este trabajo de Tesis Doctoral. Para minimizar la variabilidad entre las muestras, ya que los ensayos se realizan individual e independientemente, se necesitan métodos de normalización de los valores adecuados, siendo el más recomendado el método RMA, empleado en este trabajo (237).

Se ha demostrado la tremenda utilidad de los estudios genómicos a gran escala, sin embargo, éstos generan una gran cantidad de datos para analizar, por lo que el uso de herramientas de análisis bioinformático se hace indispensable. Es por eso que la comunidad científica pone a disposición recursos públicos online, que facilitan el análisis de grandes volúmenes de datos, minimizando el sesgo obtenido y permitiendo la reproducibilidad de los resultados al estandarizar y automatizar el procedimiento de análisis. Otro problema derivado del análisis a gran escala es la generación de falsos positivos, por lo que se han aplicado en todo momento estadísticos conservadores y métodos de corrección por múltiples comparaciones.

## 2.2. Existencia de un perfil molecular característico de cáncer de mama

Empleando los microRNAs como biomarcador de estudio, debido a su estabilidad y reproducibilidad en muestras FFPE (más accesibles) (79), en este trabajo se ha presentado un análisis exhaustivo de la expresión diferencial de miRNAs en cáncer de mama de mujeres jóvenes, pretendiendo elucidar los mecanismos moleculares de los que se vale la célula para adquirir características tumorales agresivas (238). Estos resultados, además de aplicables para mujeres jóvenes, pueden extrapolarse a otros tumores con características agresivas particulares.

Un 8,3% de los miRNAs analizados, 91 miRNAs, mostraron expresión diferente en mujeres muy jóvenes; de ellos 43 presentaron una expresión mayor en mujeres jóvenes y 48, menor. La exhaustiva selección de las muestras sugiere que estos resultados apoyan la existencia de diferencias moleculares en tumores de mujeres jóvenes. No es probable que estas diferencias sean debidas a factores relacionados con la edad, ya que las muestras mayores presentan un perfil de expresión más similar a mujeres sanas jóvenes. Por esta misma razón, se puede descartar, además, que el perfil obtenido sea debido a una mayor densidad del estroma mamario, típica de mujeres jóvenes, si bien es digno de estudio, ya que trabajos recientes demuestran que el estroma contribuye a crear un entorno más óptimo para el desarrollo tumoral que en mujeres mayores, pudiendo aumentar su agresividad (235).

Para comprobar que no existe sesgo debido a diferencias en las características tumorales, de manera adicional al pareado de los grupos, se analizó la expresión de los miRNAs atendiendo a las características tumorales (Tabla 12). Ninguno de los miRNAs significativamente diferentes según la edad mostró una expresión asociada a características tumorales, por lo que, basándonos en un modelo de mediación, se puede descartar que el perfil obtenido esté debido únicamente a diferencias en las características tumorales de los grupos estudiados. De hecho, un trabajo reciente llevado a cabo sobre datos de 4467 pacientes de cáncer de mama, refuerza el papel pronóstico de la edad *per se*, independientemente del subtipo al que pertenezca, encontrando una notable menor supervivencia en mujeres menores de 40 años, incluso tras corregir por subtipo tumoral (232).

Existen evidencias de un aumento del riesgo a padecer cáncer de mama en los cinco años que siguen a un embarazo llevado a término y a la lactancia (7). Esto es debido al cambio molecular que sufre la glándula mamaria tras la involución. Algunos autores han apuntado a varios genes desregulados en la involución mamaria como responsables de proporcionar un entorno propicio para el desarrollo de tumores, NF- $\kappa$ B, TGF- $\beta$ , MMPs y calpaínas como Cpn1 y Cpn2 (239). Este riesgo aumentado podría ser el motivo de la agresividad el cáncer de mama en mujeres muy jóvenes, al proporcionar el microambiente favorable para la progresión tumoral. Sin embargo, el perfil detectado no parece estar indicando la desregulación propia del evento de involución mamaria, ya que, tras analizar la muestra de mujeres muy jóvenes, se observó que sólo 3 de las 22 pacientes jóvenes había llevado un embarazo a término en los 5 años previos al diagnóstico. Por otro lado, el riesgo derivado de la lactancia es mayor en mujeres cuyo primer embarazo llevado a término ha sido tras los 35 años, por lo que este efecto se observaría aún más en mujeres de 35 a 45 años, no obstante, la supervivencia es menor en mujeres menores de 35 años que en el resto de grupos de edad (54).

Una vez establecido un perfil molecular de expresión de miRNAs característico de tumor mamario en mujeres muy jóvenes, se planteó la búsqueda de los posibles procesos biológicos responsables de la iniciación y progresión tumoral en este grupo de pacientes.

Los miRNAs poseen numerosas dianas hipotéticas, pudiéndose unir aún sin coincidencia perfecta de la secuencia, además, se unen a los pseudogenes transcritos y a isoformas del mismo modo que a sus genes diana, lo que complica el análisis de la regulación epigenética mediante miRNAs (72, 73). Además, se descubren nuevos miRNAs, del mismo modo que se excluyen de la lista, por tratarse estos en realidad de fragmentos de otros tipos de RNA (como tRNAs), por lo que el análisis real y global de la expresión de miRNAs es infinitamente complejo. Los miRNAs mantienen un equilibrio delicado con sus dianas, sin embargo, son los genes y su expresión lo que ejerce, en última instancia, la acción sobre la célula, en este caso, proporcionando características moleculares particulares y confiriendo a las células tumorales de mayor agresividad. Por ello, es importante el análisis de los genes diana y las rutas posiblemente desreguladas.

Para ello se realizó el análisis de enriquecimiento de rutas, mediante el cual se obtiene una visión global de los procesos que pueden resultar alterados de la desregulación de los miRNAs mediante el mayor o menor silenciamiento de sus genes diana. Tras el agrupamiento jerárquico realizado con los miRNAs que componen el perfil, se obtuvieron ocho grupos de miRNAs (nodos) cuya expresión presenta una mayor correlación en todas las muestras, es decir, se expresan de manera similar, y se planteó la posibilidad de que cada uno de estos grupos de miRNAs representase unas rutas metabólicas diferentes (Figura 23 y Tabla 13). Por esto se realizó un análisis de enriquecimiento de cada nodo por separado (Anexos VI y VII y Figuras 31 - 37). La mayoría de los nodos apuntaron a genes diana relacionados en mayor o menor medida con rutas metabólicas asociadas principalmente a adhesión y movilidad celular, a procesos neuronales y a rutas de señalización proliferativas; se encontraron también varias rutas relacionadas con varios tipos de cáncer (Figura 30).

Entre las rutas relacionadas principalmente con adhesión y movilidad celular se pueden destacar *Biosíntesis de condroitín sulfato* y *Biosíntesis de heparán sulfato*, ambas moléculas son componentes esenciales de la matriz extracelular. Particularmente el heparán sulfato es un elemento decisivo en la formación del tejido mamario en las fases de desarrollo. Estudios recientes ligan la expresión de proteoglicanos (formados por cadenas de heparán sulfato y condroitín sulfato), como el syndecan-2, a la progresión e invasión tumoral, habiendo encontrado una mayor expresión en tejido de pacientes de cáncer de mama (240). También encontramos varias rutas de adhesión como son *Adhesiones focales*, *Uniones adherentes* y *Uniones estrechas*, así como la ruta *Regulación del citoesqueleto de actina*. La desregulación de todas estas rutas favorece la invasión tumoral, la migración y, por tanto, la capacidad metastásica (241-246).

También se encontraron rutas desreguladas relacionadas con señalización y proliferación celular. Entre ellas destaca la ruta de *Señalización de MAPK*, (247-249). Algunos estudios apuntan a una disminución de los niveles de ER inducida por MAPK (250). También aparece enriquecida la ruta de *Señalización de PI3K/Akt/mTOR*. Además, se encuentran muchas rutas de señalización iniciadas por RTK, como la familia de EGFR, IGFR, PDGFR. En concreto, encontramos la ruta de *Señalización de ERBB* como posiblemente desregulada, habiéndose hallado tanto en el análisis de enriquecimiento de rutas mediante dianas teóricas como en el de dianas validadas. Estas rutas

se encontrarían sobre-expresadas en mujeres jóvenes, debido a una menor expresión de los miRNAs que silencian a sus genes diana. Estos resultados concuerdan con los obtenidos en trabajos previos de la expresión molecular en mujeres jóvenes, que apuntan a rutas relacionadas con proliferación como Myc, MAPK y PI3K ([56](#), [57](#), [227](#)).

El perfil obtenido descubre numerosas rutas relacionadas con procesos y sinapsis neuronales, aunque esto pudiera parecer extraño en un principio, un análisis más exhaustivo reveló que estas rutas se encuentran acopladas a la ruta de *Señalización de calcio*, vital para la actividad y proliferación celular y son asociadas a características de células madre. De hecho, en varios estudios previos de la expresión del perfil de miRNAs se encuentra una desregulación (tras analizar el enriquecimiento de rutas) de genes diana relacionados con procesos neuronales y diferenciación celular asociados a cáncer pancreático y cáncer de mama ([251](#), [252](#)).

Entre esas rutas encontramos *Guía de axones*, que está acoplada al control del citoesqueleto de actina y de la promoción de la angiogénesis ([253](#)) y que se encuentra enriquecida en miRNAs desregulados en varios tipos de cáncer ([254-256](#)). También destacamos la ruta de *Señalización por neurotrofina*, así como numerosas rutas relacionadas con la formación de sinapsis (*Sinapsis glutamatérgica*, *Sinapsis colinérgica*, *Sinapsis serotoninérgica*, *Sinapsis dopaminérgica* y *Sinapsis GABAérgica*). Destacamos también en este grupo a las rutas *Potenciación a largo plazo* y *Depresión a largo plazo*, desreguladas en un alto número de nodos y relacionadas con la plasticidad sináptica y formación de memoria, encontrándose, además estrechamente relacionadas con la actividad quina de la célula ([257](#)).

Como hemos visto, el perfil obtenido indica rutas metabólicas relacionadas, directa o indirectamente, con proliferación celular, migración y metástasis, y una pobre diferenciación celular, características que pueden conferir a los tumores mayor agresividad. No obstante, el estudio de la acción de los miRNAs es más complejo, ya que una desregulación de los miRNAs puede o no traducirse en un cambio en la represión de los mRNAs diana. Un mismo gen posee sitios de unión en su región 3'UTR a varios miRNAs, por lo que, el aumento en la represión debida a un miRNA concreto puede equilibrarse con el descenso en la expresión de otro miRNA con la misma diana ([71](#)). Teniendo en cuenta esto se planteó la estimación de un valor de represión en cada una de las dianas obtenidas, para tener una idea de la desregulación global sobre los genes por parte de los miRNAs desregulados en mujeres jóvenes.

Teniendo en cuenta sólo las dianas cuyos valores de represión o sobre-expresión se encontraban por encima del 75%, se realizó un análisis de enriquecimiento de rutas tanto en dianas validadas, como en dianas teóricas. Existen numerosas dianas teóricas de los miRNAs, teniendo alguno del orden de 700 dianas, por lo que es complicado saber si todos los sitios de unión son efectivos o no. Para ello se filtraron las dianas teóricas en base a un valor umbral de confianza del 0,9 (se considera un valor aceptable por encima de 0,7), para reducir el número de dianas y asegurar la fiabilidad de los resultados. Las dianas validadas ofrecen una mayor confianza a la hora de asegurar las dianas, sin embargo, pueden introducir sesgos, ya que se validan mediante diferentes técnicas, en diferentes estudios, siendo, además, un reflejo subjetivo del interés científico. Por eso se decidió realizar el análisis tanto en dianas teóricas como en validadas y comparar los resultados obtenidos (Anexo XI). Se obtuvieron rutas relacionadas con los mismos procesos que los obtenidos tras el enriquecimiento por nodos.

Se seleccionaron las rutas más representadas, tanto por su p-valor como por la cantidad de miRNAs asociados a esa ruta. Las rutas candidatas fueron: *Síntesis del condroitín sulfato*, *Adhesiones focales*, *Regulación del citoesqueleto de actina*, *Guía de axones*, *Potenciación a largo plazo*, *Sinapsis GABAérgica*, *Sinapsis dopaminérgica*, *Señalización por neurotrofina*, *Señalización por TGF $\beta$* , *Señalización por PI3K*, *Señalización por ERBB*, *Señalización por mTOR* y *Señalización por MAPK*. Todas estas rutas tienen como resultado una mayor movilidad, auto-mantenimiento y proliferación, propiedades que confieren a una célula características tumorales.

El análisis de interacciones de los genes diana afectados por la desregulación de miRNAs sobre estas rutas apuntó a genes *hub* (es decir, aquellos que cuentan con un alto número de interacciones dentro de una red), ampliamente relacionados con iniciación y progresión tumoral, entre los que destacan varias MAPquinasas, *Akt*, mTOR, varias fosfatasa y tirosina cinasas, relacionadas con activación y proliferación celular (258, 259). Además, numerosas proteínas RAS se encontrarían sobre-expresadas, favoreciendo la activación de la ruta de señalización MAPK, entre otras, y favoreciendo la proliferación (260, 261). Se destacan genes asociados con el control de la apoptosis, diferenciación y ciclo celular, como Bcl, JUN, SMAD, TGF- $\beta$ . PTEN muestra una mayor represión por parte de los miRNAs que en mujeres mayores, por lo que no ejercería su papel supresor de tumores (262, 263). AKT1, al contrario que el resto de cinasas, se encuentra muy reprimida por los miRNAs, pero esto podría ser un hecho derivado de la represión de PTEN, al haber menos PTEN, este no inactiva a la proteína AKT1, por lo que hay más cantidad y, por tanto, se necesitarían más miRNAs para silenciarla (264). Destacamos también aquellas dianas relacionadas con movilidad, como actininas, integrinas, lamininas, caveolinas y colágeno. Receptores de membrana RTK, como ERBB2, ERBB3, ERBB4, PDGFR, IGFR, VGFR, se muestran menos reprimidos por los miRNAs, facilitando la iniciación de las rutas de transducción de señales y, por tanto, la proliferación (265).

### 2.3. Validación del perfil característico de cáncer de mama en mujeres muy jóvenes

Las líneas celulares constituyen un modelo idóneo para el estudio del comportamiento tumoral, se pueden realizar estudios de proliferación celular, y estudiar el efecto que tiene la desregulación de los marcadores moleculares en las células tumorales, así como el efecto de diversos tratamientos. Por ello es importante identificar de base una línea celular que constituya un buen modelo para la realización de estudios funcionales en cáncer de mama de mujeres jóvenes.

Aunque existen numerosas líneas celulares de cáncer de mama comercializadas, pocas proceden de mujeres menores de 35 años. Para estudiar la adecuación de las líneas celulares como modelo para estudios funcionales se seleccionaron 4 líneas procedentes de mujeres menores de 35 años: Hs500(B).t., HBL100, HCC1500 y HCC1937; y se compararon con tres líneas ampliamente empleadas en el estudio del cáncer de mama: MCF7, MDA MB 231 y MDA MB 468. Sin embargo, la línea Hs500(B) presentó un crecimiento muy lento e irregular, siendo descartada de inmediato. A las tres líneas restantes se les realizó un análisis de la expresión de miRNAs, del mismo modo que el realizado sobre muestras tumorales. Las líneas se sembraron por triplicado y no pasaron más de 4 pases desde su cultivo hasta la recogida del pellet de células para la extracción de material genético; esto se hizo para controlar el sesgo introducido por la

acumulación de mutaciones debida a la inestabilidad génica típica de las líneas celulares de cáncer (266, 267).

Tras el análisis de expresión, se obtuvieron 143 miRNAs desregulados en líneas de mujeres jóvenes, un número superior al obtenido a partir de muestras tumorales de pacientes, que puede ser debido a que se hayan producido mutaciones puntuales, a pesar de las precauciones tomadas.

Tras el agrupamiento jerárquico de líneas, las tres réplicas presentaron valores de correlación entre sí muy buenos, cercanos a 0,9 (Figura 42), y las líneas procedentes de mujeres jóvenes se agruparon juntas, confirmando la existencia de un perfil en mujeres jóvenes. Al realizar el agrupamiento teniendo en cuenta tanto las líneas como las muestras previamente estudiadas, tanto las líneas como las muestras procedentes de mujeres jóvenes se agruparon juntas, a excepción de HBL100, que no sigue el mismo patrón de expresión que las líneas de mujeres jóvenes, siendo más similar a las procedentes de mujeres mayores, de hecho, esta línea ha sido recientemente descatalogada debido a que posee un cromosoma Y (Figuras 41 y 43). La línea HCC1937 fue la que mostró una expresión más similar a la de pacientes menores de 35 años, sin embargo, HCC1500, también muy similar en expresión a mujeres jóvenes, presentó una mayor correlación entre sus réplicas, por lo que puede considerarse que presenta una mayor estabilidad genómica.

Las líneas celulares HCC1500 y HCC1937 constituyen, pues, un modelo acertado de cara a realizar futuros estudios de proliferación, estudios funcionales y de resistencia o sensibilidad a fármacos, ya que se puede reproducir el perfil de expresión obtenido para pacientes en un gran número de miRNAs.

Gracias a los repositorios online que facilitan la accesibilidad a datos genómicos, se permite la validación de resultados en una muestra de datos que, de otro modo, sería muy costosa, tanto a nivel de infraestructura como de tiempo y esfuerzo. Por ello, las bases de datos online que ponen a disposición los resultados masivos de otros estudios proporcionan un sustrato de estudio valioso y de otro modo inaccesible. Debido al tamaño muestral empleado, se planteó realizar el mismo análisis en una muestra independiente de pacientes con mayor número de muestras, accesible gracias a la base de datos EGA, que contaba con la información clínica necesaria, las muestras empleadas en el estudio sobre el cáncer de mama METABRIC (177).

El análisis de los datos de expresión de miRNAs del estudio METABRIC y su papel en cáncer de mama de mujeres jóvenes permitió validar la existencia de un perfil molecular distintivo de mujeres jóvenes, del mismo modo que se obtuvo en muestras procedentes de pacientes del Hospital Clínico de Valencia y en líneas celulares de tumor mamario de mujeres jóvenes. Se obtuvieron 25 miRNAs sobre-expresados y 31 reprimidos, los cuales, tras el análisis de rutas, apuntaron a rutas muy similares a las obtenidas en el análisis de enriquecimiento obtenido en el estudio de muestras de paciente, principalmente rutas relacionadas con proliferación y adhesión celular, incluyendo los receptores de la familia ERBB. Sin embargo, los miRNAs desregulados no coinciden con aquellos desregulados en muestras y en líneas celulares, debido, principalmente, a que las plataformas empleadas fueron diferentes, Affymetrix en los estudios realizados en muestras de pacientes y en líneas celulares y EXIQON en el estudio METABRIC, lo que, además de generar un sesgo debido a la metodología, implica que, de base, los miRNAs analizados fueron diferentes, y muchos de los que se encontraban desregulados en el primer estudio, no se incluían



en el estudio METABRIC (ANEXO XIII). No obstante, los perfiles de muestras y de líneas celulares comparten 43 miRNAs significativamente diferentes en mujeres jóvenes, lo que indica la presencia de un *pool* de miRNAs compartido (Figuras 46).

Los resultados obtenidos en los tres estudios realizados (mostrados en la Figura 47 y en el Anexo XIII) indican que tres miRNAs permanecen significativos en todos ellos, hsa-miR-30c, hsa-miR-489 y hsa-miR-663, este último, además, obtuvo el p-valor más elevado al realizar el meta-análisis mediante Fisher ( $p\text{-valor} < 1 \times 10^{-12}$ ). El miRNA hsa-miR-663 se ha encontrado asociado a cáncer de próstata, leucemia y otros cánceres, siendo tratado como un supresor de tumores, entre sus dianas validadas encontramos efectores de la ruta de TGF- $\beta$ , HSPG2, eEF1 y p21. Además, un trabajo muestra una sobre-expresión de hsa-miR-489 en cáncer de mama de tipo inflamatorio, un tipo particularmente agresivo (268). En cuanto al hsa-miR-30c, la familia miR-30 se encuentra fuertemente reprimida en células madre de cáncer de mama, habiéndose demostrado que controla la expresión de *Ube9* y *AVEN*, entre otros genes (119). Es más, se han encontrado diferencias en la expresión de miRNAs que incluyen la familia miR30 en melanoma de pacientes jóvenes, menores de 30 años, al compararlo con pacientes mayores de 60 años (269).

Entre las rutas desreguladas se destacan en todo momento aquellas relacionadas con movilidad celular, con procesos sinápticos y con señalización celular. Como se ha comentado antes, todas estas rutas tienen como resultado una mayor movilidad, auto-mantenimiento y proliferación, propiedades que confieren a una célula características tumorales.

Si el cáncer de mama en mujeres muy jóvenes es, en efecto, una entidad única o no, ha sido cuestión de debate continuo. Sin embargo, como ya se ha comentado, la comunidad médica parece coincidir en que las mujeres muy jóvenes presentan unas características clínicas y tumorales particulares que merecen una atención especial. Dicha particularidad es un reflejo de una biológica única y característica; sin embargo, algunos autores atribuyen las posibles diferencias en la expresión génica a la mayor representación de subtipos agresivos mencionada (59). Se puede decir, no obstante, que la causa de la progresión tumoral, se debe, en última instancia, a cambios en la epigenética de las células tumorales. Es más, si no existiesen diferencias moleculares en ambos grupos de edad, estadísticamente no existirían diferencias en las frecuencias fenotípicas, ni las diferencias encontradas en cuanto a la expresión de miRNAs, más allá de las propias de la edad (no se debe descartar que éstas también podrían afectar a dicha agresividad, asociada a una mayor actividad celular y proliferación).

Si bien los cambios descritos son intrínsecos al grupo de mujeres jóvenes o el reflejo de una mayor representación de características agresivas, no influye en el hecho de que existen cambios epigenéticos que provocan tal agresividad en los tumores y que esos cambios, además, parecen tener mayor presencia en pacientes menores de 35 años. Identificar de qué mecanismos moleculares se está valiendo el tumor para desarrollar ese comportamiento nos aporta no solo conocimientos sobre la biología del cáncer de mama en mujeres jóvenes, si no del cáncer de mama en particular y del cáncer en general.

## PARTE 3: VALIDACIÓN DE LA EXPRESIÓN DE LOS miRNAs SELECCIONADOS EN UNA SERIE INDEPENDIENTE DE MUESTRAS MEDIANTE qRT-PCR

---

### 3.1. Selección de los miRNAs relevantes y representativos

Debido a los posibles sesgos introducidos en el análisis mediante el chip de hibridación de miRNAs (entre los que se encuentran posibles fluctuaciones en los detectores de intensidad de fluorescencia) es aconsejable validar los resultados obtenidos analizando la expresión de miRNAs mediante otra técnica (270). Por ello, los miRNAs representativos del perfil con resultados relevantes se validaron en una serie independiente de mujeres mediante una técnica diferente y más restrictiva, como es la qRT-PCR, una técnica empleada en la comunidad científica, asequible y reproducible. Además, la identificación de un número menor de marcadores que reproduzcan el perfil de expresión de miRNAs obtenidos, facilita la detección de los mismos, la reconstrucción del perfil y la aplicabilidad desde un punto traslacional, de cara a emplear los miRNAs detectados como marcador de pronóstico y diagnóstico.

Como se ha comentado, tras el análisis de enriquecimiento de rutas de los miRNAs que componen los principales nodos desregulados se seleccionaron aquellas rutas relacionadas con características tumorales agresivas, como son proliferación (señalización por PI3K, TGF- $\beta$ , MAPK, ERBB, mTOR, neurotrofina); adhesión celular y migración (*Biosíntesis del condroitín sulfato*, *Guía de axones*, *Sinapsis GABAérgica* y *Sinapsis dopaminérgica*, *Sdhensiones focales*, *Citoesqueleto de actina* y *Potenciación a largo plazo*); ya que, de entre las rutas enriquecidas, estas presentaron mayor representación de miRNAs y mayor p-valor asociado.

Se seleccionaron los miRNAs en función al análisis de enriquecimiento de miRNAs en las rutas seleccionadas, es decir, aquellos que mejor representaban a las rutas seleccionadas con mejor p-valor. Los miRNAs seleccionados fueron: hsa-miR125b, hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-149\*, hsa-miR-92b, hsa-miR-139-5p, hsa-miR-132, hsa-miR-379, hsa-miR-409-3p, y hsa-miR-433, hsa-miR-23a, hsa-miR-27b, hsa-miR-28-5p, hsa-miR-30c, hsa-miR-30e\* y hsa-miR-195-5p. Además, se incluyeron miRNAs relevantes que presentaron una expresión marcadamente diferente y significativa en mujeres jóvenes (como el caso de hsa-miR-3196 y hsa-miR-1228\*). Tras la selección de miRNAs, el análisis de enriquecimiento de rutas de los mismos reprodujo las rutas consideradas relevantes en un primer momento, lo que aseguró una buena elección de los miRNAs (Figura 37).

En el análisis de expresión de los miRNAs mediante qRT-PCR, se emplearon tres rangos de edad: pacientes menores de 35 años, de entre 45 y 65 años de edad y mayores de 65 años, en lugar de los dos rangos extremos utilizados en los análisis anteriores. Se introdujo la categoría intermedia para establecer si la desregulación de los miRNAs era identificativa de mujeres jóvenes o si, por el contrario, se trata de un suceso que ocurre en otros rangos de edad y lo que observamos es, en realidad, un perfil característico de mujeres mayores de 65 años.

Como se ha comentado anteriormente, los estudios de cáncer de mama en mujeres jóvenes han establecido varios cortes a la hora categorizar a un grupo como joven, siendo el umbral más alto 45 años (52, 230, 233, 234). Se puede deducir, pues, que a partir de 45 años se considera un

grupo homogéneo como control frente a mujeres jóvenes, por lo que se empleó como la edad de corte por encima de la cual no se consideran pacientes jóvenes de cáncer de mama.

Los resultados de la cuantitativa mostraron que la expresión de todos los miRNAs evaluados, el grupo de muestras de pacientes entre 45 y 65 años se comportan del mismo modo que las mujeres mayores de 65 años, descartando que las diferencias que veamos sean exclusivas del grupo de pacientes post-menopáusicas, y confirmando también la distinción del cáncer de mama en mujeres muy jóvenes de aquellas pre-menopáusicas (con pico de aparición a los 50 años) (167).

### 3.3. *Los miRNAs son validados mediante qRT-PCR*

De los 18 miRNAs seleccionados para la validación, 12 mostraron resultados significativos con la misma tendencia de expresión que en la primera fase, sin embargo, solo seis de estos aguantaron corrección por Bonferroni para múltiples comparaciones. No obstante, la corrección por Bonferroni es un método estricto, además, el escaso tamaño muestral reduce el poder estadístico, por lo que se necesitan más muestras para obtener la significación. El hecho de que consigamos validar la expresión de seis miRNAs y que, además, el resto de ellos mantenga la tendencia a la significación, otorga robustez a los resultados obtenidos en la primera fase del estudio, confirmando la existencia de un perfil molecular propio de mujeres jóvenes y garantizando su detección mediante una tecnología asequible como es la qRT-PCR.

### 3.4. *Relevancia de los miRNAs validados*

De los miRNAs validados mediante qRT-PCR, se destacan los seis que aguantan corrección por Bonferroni, ya que sus valores de significación son muy robustos: hsa-miR-30c, hsa-miR-92b, hsa-miR-139, hsa-miR-1228, hsa-miR-1275, hsa-miR-3196.

La expresión del miRNA miR30c se ha encontrado asociada a la expresión de teniendo un papel supresor de tumores (271, 272). Se encuentra reprimido de manera dependiente de TGF- $\beta$  en lesiones hepáticas (Roy 2015). Reprime la expresión de IER2, una proteína que promueve la metástasis (273) (Wu 2015). Se ha descrito su desregulación en cáncer de mama hereditario sin mutación en *BRCA1/2*, habiéndose visto su asociación con la expresión de *KRAS* (175).

Sin embargo, algunos estudios evidencian el papel oncogénico del miR30c, siendo de peor pronóstico en cáncer de mama debido al silenciamiento de *NOV/CCN3*, que resulta en un aumento de la invasión en células MDA-MB-231, o mediante el silenciamiento del gen supresor de tumores p53 (271, 274). El papel que desempeña la familia miR-30 en el inicio, progresión e invasión del cáncer no está claro, debido a la gran cantidad de genes diana que regula, por lo que se necesitan más estudios de esta familia de miRNAs para determinar su papel particular en cáncer de mama y, concretamente, en cáncer de mama de mujeres muy jóvenes, dónde se muestra principalmente reprimida.

**Tabla 18.** Genes diana de los miRNAs validados y su implicación en cáncer.

miRNAs	Gen diana	Función	Referencias
<b>miRNAs sobre-expresados en mujeres jóvenes</b>			
miR1228*	<i>MOAP1</i>	Homólogo de <i>Bcl</i> , su represión permite a las células escapar de la apoptosis.	Yan et al. 2012 (275)
miR3196	<i>PAX2</i> , <i>THTPA</i> , <i>PIK3R2</i> , <i>BBC3</i>	<i>PAX2</i> está implicado en la supresión de la traducción de proteínas (mediante WT1). Se asocia con bajo riesgo de cáncer de mama, su represión promueve tumores más agresivos. <i>THTPA</i> es un gen supresor de metástasis tumoral. <i>PIK3R2</i> es un gen anti-apoptótico implicado en rutas de proliferación. <i>BBC3</i> es un gen pro-apoptótico asociado a tamaño tumoral.	Beauchemin et al. 2011. (276) Kovacevic et al. 2008. (277) Tajnik et al. 2012. (278) Wong et al. 2012. (279)
miR1275	<i>IGF1</i> , <i>NFIX</i> , <i>Claudin11</i>	<i>IGF1</i> está relacionado con proliferación tumoral. <i>NFIX</i> se encuentra hipermetilado en líneas celulares de cáncer de mama. <i>Claudin 11</i> , molécula de adhesión celular cuya represión está asociada con invasión y capacidad metastásica.	Castaño et al. 2013. (280) Awsare et al. 2011. (281) Webb et al. 2013. (282) Özata et al. 2011. (283) Lian et al. 2012. (284) Katsushima et al. 2012. (285)
<b>miRNAs reprimidos en mujeres jóvenes</b>			
miR139	<i>RAP1B</i> , <i>c-FOS</i> , <i>TOP2A</i> , <i>CXCR4</i>	<i>RAP1B</i> , familia RAS (oncogén). <i>FOS</i> está implicada en ciclo celular, diferenciación y transformación celular, su represión induce un aumento en la apoptosis y mayor diferenciación celular. <i>TOP2A</i> , topoisomerasa asociada a proliferación tumoral. <i>CXCR4</i> , se encuentra asociado con progresión y metástasis en cáncer colorrectal, regulado por HER2-CD44 vía miR-139.	Guo et al. 2012. (286) Milde-Langosch et al. 2013. (287) Shen et al. 2012. (288) Bao et al. 2011. (289)
miR92b	<i>PSMD10</i> , <i>FOX2</i>	<i>PSMD10</i> . Su inhibición ralentiza el crecimiento tumoral en hepatocarcinoma. Altamente expresado en glioma. <i>FOX2</i> . Proteína que regula el procesado alternativo.	Leidner et al. 2013. (290) Qian et al. 2012. (291)
miR30c	<i>p53</i> , <i>KRAS</i> , TGF- $\beta$ , <i>SNAIL</i> , <i>IER2</i>	<i>KRAS</i> , actúa como oncogén. Se requiere miR-30c para inducir la apoptosis mediada por TGF- $\beta$ . <i>SNAIL</i> es un factor de transcripción que promueve la represión de la adhesión celular, habiéndose encontrado desregulado en varios cánceres, incluyendo glioma y de próstata, en cáncer de mama favorece la transición epitelio-mesénquima. <i>IER2</i> , promueve la metástasis.	Li et al. 2010. (274) Shen et al. 2009. (292) Wu et al. 2015. (273) Tanic et al. 2012. (175) Wang et al. 2013. (293) Dong et al. 2014. (294) Roy et al. 2015. (295)

El papel oncogénico del policistrón 17/92 se ha descrito en varios tumores como por ejemplo linfomas (296), retinoblastoma (297), cáncer colorrectal (298), neuroblastomas (299), cáncer pancreático (300, 301) y cáncer de pulmón (302) entre otros. En especial se ha visto sobreexpresado en tumores triples negativos y en aquellos tumores de cáncer de mama metastásicos (303-306). Sin embargo, se ha encontrado una elevada expresión de hsa-miR-92b en líneas celulares de cáncer de mama luminal en comparación con el resto de subtipos en un estudio de 55 líneas celulares de cáncer de mama, siendo indicativo de buen pronóstico (307). Además se sabe que diversos miembros de este policistrón tienen genes diana supresores de

tumores como *PTEN*, la familia de *E2F1-3*, *TGF- $\beta$* , *Smad2-4* o *BCL2L11*. Los resultados en datos de METABRIC indican valores sobre-expresados de este policistrón, mientras que en muestras procedentes del Hospital Clínico y en líneas celulares se encuentra muy sobre-expresado el miRNA miR92b\*, sin embargo miR-92a y miR-92b mostraron una expresión menor en mujeres jóvenes, por lo que no se puede concluir su implicación en cáncer de mama de mujeres jóvenes.

En cuanto a miR139, se ha demostrado su papel como supresor de tumores en glioma, cáncer de ovario, cáncer colorrectal y de vejiga, y en carcinoma esofágico y de células renales (272, 294, 308-310). Se ha demostrado su papel como regulador de la metástasis en cáncer de mama (311). Entre sus genes diana destaca *RAP1B*, un oncogén de la familia RAS, para el que existen variantes de procesamiento de mRNA y varios pseudogenes (Tabla 18).

Es interesante comentar que miR125a, cuya expresión es muy diferente en los grupos de edad, aunque no aguantó corrección por Bonferroni, se encuentra marcadamente reprimido en células de cáncer de mama, siendo esta represión más marcada cuando se sobre-expresan HER2 y/o HER3. Además, favorece que el tumor muestre características más agresivas, especialmente una mayor invasión tumoral (87, 94, 244, 312, 313). Esta asociación se describe también en cáncer gástrico y de esófago (293, 314).

En cuanto a los miRNAs que se encuentran sobre-expresados en mujeres jóvenes, como se resume en la tabla 19, apuntan a genes diana relacionados principalmente con rutas que promueven la apoptosis y la diferenciación, por lo que una mayor represión de estas rutas por parte de los miRNAs conduce a una mayor proliferación y características tumorales más agresivas. Además el miRNA miR-1228 promueve el crecimiento tumoral y metástasis en cáncer de mama (218).

El miRNA miR1207-5p, aunque no se mantuvo significativo tras aplicar Bonferroni, silencia a *DHCR24*, cuya expresión se encuentra muy reprimida en cáncer de próstata. Además, entre sus genes diana destacan numerosas claudinas, proteínas asociadas con movilidad celular, invasión tumoral y metástasis (282, 315).

### 3.5. *MicroRNAs como marcadores de supervivencia*

En la práctica clínica, los marcadores de supervivencia suponen una herramienta muy útil de cara a establecer un pronóstico y administrar un tratamiento concreto (200). La estabilidad de los miRNAs y su posible detección en suero, los sitúan como unos marcadores de supervivencia ideales, cuya asociación merece ser evaluada.

Los análisis de supervivencia se llevaron a cabo sobre los datos obtenidos del consorcio METABRIC, que contaban con los datos de supervivencia. En la muestra procedente de pacientes del Hospital Clínico no se obtuvieron resultados concluyentes debido al tamaño muestral reducido. Se analizaron los miRNAs validados mediante qRT-PCR y cuya expresión se había analizado en la plataforma EXIQON en el estudio de METABRIC: miR30c, miR92b, miR125a y miR139. Se incluyó además, miR663, cuya expresión se validó mediante el meta-análisis, además de encontrarse sobre-expresado en mujeres jóvenes tanto en el estudio llevado a cabo sobre muestras, como en el de líneas celulares y el realizado con datos del estudio METABRIC.

Tras los análisis de supervivencia se confirmó una menor supervivencia global en mujeres menores de 35 años respecto a mujeres mayores de 65. Además, esta tendencia se refuerza al separar por mujeres ER positivas y ER negativas, lo que confirma trabajos previos (54). Al estudiar la relación entre la desregulación de determinados miRNAs y la supervivencia, las pacientes mostraron una peor supervivencia asociada a la represión de los miRNAs miR92b y miR139. Otros miRNAs mostraron cierta asociación a la supervivencia, sin embargo los resultados no alcanzaron la significación, es el caso de miR30c y miR27b. Al separar la muestra por edad, las mujeres menores de 35 años, aunque mostraron una curva de supervivencia muy diferente según la expresión del miRNA, no se pudo concluir ninguna asociación estadística, debido al tamaño muestral (33 mujeres jóvenes) y a su fragmentación tras separar por grupos de expresión.

Los resultados obtenidos sugieren un papel discreto como marcador de la supervivencia de los miRNAs miR92b y miR139, aunque se necesitan más estudios con una población de muestras mayor.

El estudio de la expresión de miRNAs ha llevado a identificar una serie de miRNAs susceptibles de emplearse como marcadores pronóstico, destacando la asociación del miR-92b con una mayor supervivencia. Por otro lado, el miR-30c, asociado previamente a cáncer de mama, se encuentra muy reprimido en cáncer de mama de mujeres jóvenes y, aunque se necesitan de más estudios dirigidos, puede ser un marcador de buen pronóstico. Los miRNAs hsa-miR-1228, hsa-miR-1275 y hsa-miR-3196, se encuentran sobre-expresados en mujeres jóvenes, y su posible detección en suero los sitúa como marcadores de pronóstico candidatos en biopsias líquidas, menos invasivas.

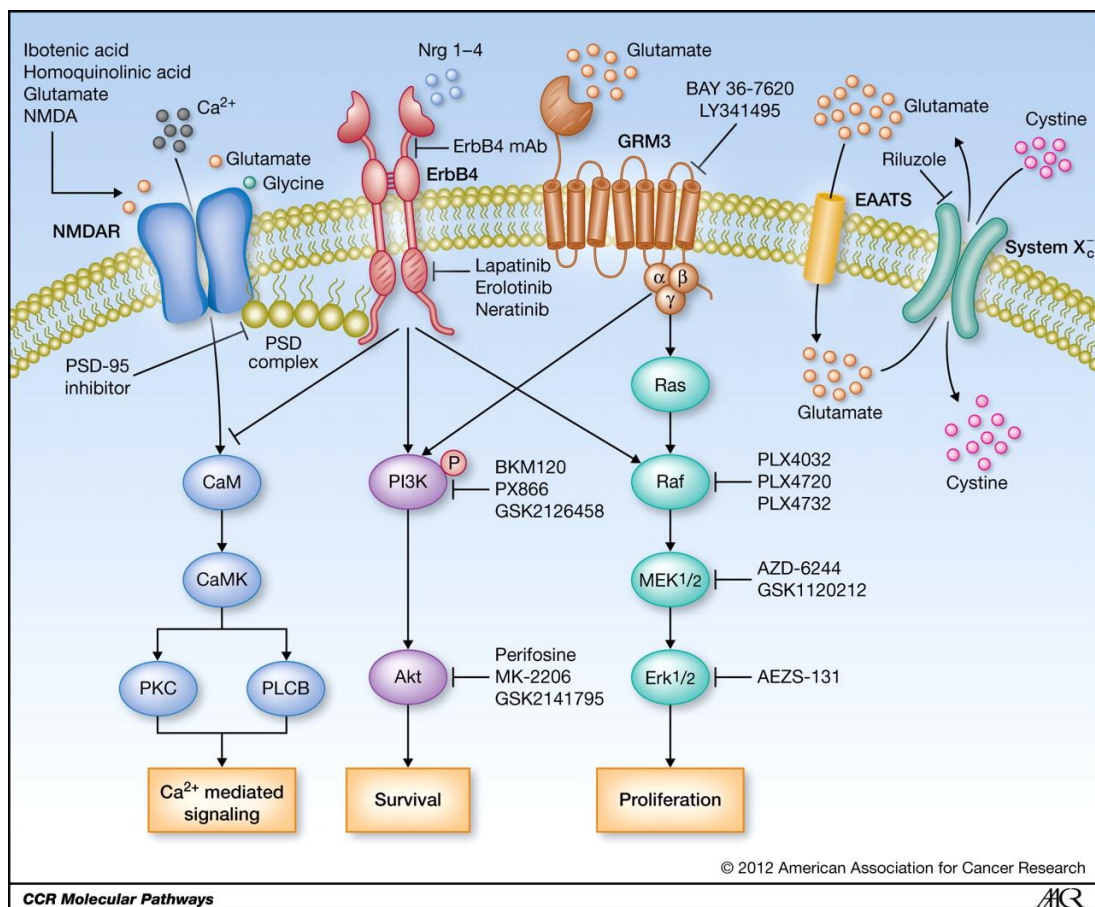
## PARTE 4: VALIDACIÓN DE LA IMPLICACIÓN DE LOS GENES SELECCIONADOS EN CÁNCER DE MAMA EN MUJERES JÓVENES

### 4.1. Selección de los receptores de la ruta de NMDAR como genes candidatos

**Correlación de los miRNAs desregulados con la ruta de EGFR y su acoplamiento con otras rutas de proliferación.**

La desregulación de rutas obtenida indica, en efecto, una mayor proliferación y capacidad invasiva, que podría valerse de mecanismos similares a los que emplean las células madre neuronales (con las que las células del epitelio mamario comparten origen ectodérmico) para provocar una mayor agresividad de las células tumorales en mujeres jóvenes.

Debido a la gran cantidad de rutas neuronales desreguladas, la atención se centró sobre la ruta del glutamato, una ruta que inicia una cascada de señalización que conlleva una mayor proliferación y que se encuentra acoplada a rutas de señalización como MAPK, PI3K/Akt/mTOR y señalización de calcio, todas ellas se muestran desreguladas en los análisis realizados (Figura 57).



**Figura 57. Receptores iniciadores de cascadas de señalización en la célula relacionados con la ruta del glutamato.** Se representan los receptores de membrana asociados a las rutas de proliferación y sus ligandos, así como algunos de los fármacos que alteran las rutas en alguno de sus puntos. Figura original de Prickett, 2012.

La ruta del glutamato parece estar relacionada con la iniciación de procesos cancerosos, habiendo aumentado la atención sobre ella en las últimas décadas. El estudio de la implicación de la ruta de señalización glutamatérgica en la homeostasis del cáncer ha aumentado exponencialmente en la última década, sugiriendo que juega un papel importante en la transformación celular (316). La ruta del glutamato está íntimamente ligada con la ruta de señalización de calcio, relevante para los procesos de desarrollo de tumores y proliferación, siendo las microcalcificaciones en mama un síntoma de alta actividad y, por tanto, un factor de riesgo (317, 318). Recientemente, se han asociado receptores de esta ruta, como NMDAR y otros mGluRs, a procesos cancerosos particularmente agresivos, incluyendo la transformación maligna de células mamarias (319-321).

No se debe descartar la implicación de otras rutas metabólicas, como las relacionadas con movilidad y adhesión celular, sin embargo, la inter-relación presente entre los tres tipos de rutas analizados, neuronales, movilidad y matriz extracelular y rutas de señalización responsable de la proliferación celular, indican que se avanza en la dirección correcta. Esto abre las puertas de cara a la búsqueda de nuevas dianas y estrategias terapéuticas, que permitan una calidad de vida y una reincorporación social satisfactoria para este grupo de pacientes.

Junto con la ruta de señalización del glutamato, la ruta de señalización de ERBB4 se presentó como posible candidata debido al enriquecimiento de los miRNAs desregulados en mujeres jóvenes, y a su interconexión con la ruta del glutamato, así como con otras rutas también enriquecidas importantes, como son las rutas de señalización PI3K/Akt y MAPK. Además tiene cierta relevancia en procesos neuronales como son potenciación y depresión a largo plazo, sinapsis glutamatérgica, etc; rutas generalizadamente enriquecidas en los miRNAs de los nodos seleccionados (322) (Figura 57).

Aunque los miRNAs desregulados presentan genes diana en todos los niveles de señalización de las rutas, los receptores de membrana son los responsables de iniciar la transducción de señales, su papel es de máxima relevancia, siendo, además, susceptibles de emplearse como dianas terapéuticas, debido a la presencia de antagonistas de sus ligandos.

Todos los genes seleccionados como candidatos son diana de miRNAs que se encuentran reprimidos en mujeres jóvenes, por lo tanto, los mensajeros de estos genes deben de encontrarse en mayor cantidad en mujeres jóvenes, lo que conduce a la sobreexpresión de la proteína a la que codifican (Anexo XI). Dicha sobre-expresión conduciría a un aumento del metabolismo del calcio, mayor proliferación y supervivencia celular, procesos clave en la progresión tumoral.

La ruta del glutamato se inicia mediante el receptor de membrana NMDAR, por lo que seleccionamos como genes candidatos aquellos que codifican para dos de sus subunidades GRIN2B y GRIN3A, ya que se ha revelado recientemente su posible papel en la proliferación de tumores sólidos, como el de cáncer de mama (323). El trabajo de Li y colaboradores (2013) indica una mayor actividad de la subunidad GRIN2B, o NR2B, del receptor NMDAR en los bordes de varios tipos de tumores sólidos, incluyendo tumores mamarios, lo que sugiere que podría estar indicando el camino de expansión del tumor, del mismo modo que ocurre cuando ejerce su acción normal en neuronas facilitando la plasticidad neuronal (323, 324). De hecho, la presencia de la subunidad NR2B es mayor durante las etapas del desarrollo embrionario y hasta la tercera semana de desarrollo post-natal, por lo que está asociada a períodos de alta actividad proliferativa (325-327).



Por otro lado, GRIN3A, o NR3, otra subunidad del receptor del glutamato NMDAR, se piensa que actúa suprimiendo la actividad de NMDAR, además, aumenta su presencia con la edad, siendo responsable de la pérdida de plasticidad neuronal y capacidad de aprendizaje asociada a la edad(325). No existen referencias acerca de la expresión de este gen asociada a cáncer, sin embargo, sería útil considerar la relación entre GRIN2B y GRIN3A y su desequilibrio como posible mecanismo de acción tumorigénica.

También relacionado con esta ruta, se seleccionó el gen del receptor metabotrópico del glutamato (*GRM3*), que está asociado con migración y proliferación de células tumorales (320, 322, 328). Se ha descrito que un aumento en los niveles del receptor GRM3, o mGluR3, ha sido asociado a varios cánceres, como son glioma, cáncer de tiroides, melanoma y de esófago (320, 328-330). Un estudio reciente demuestra que, en efecto, un bloqueo de GRM3 inhibe la proliferación y promueve la diferenciación en glioma (331). Interesantemente, se han detectado altos niveles de los receptores mGluR3 y mGluR5 en células madre neuronales, permitiendo el mantenimiento de la población de células madre embrionarias (332). El parecido entre las células madre neuronales y las células transformadas que originan procesos cancerosos, sugiere que estos receptores juegan un papel importante en la transformación epitelio-mesénquima y en la adquisición de características de célula madre por parte de las células tumorales, debido al origen ectodérmico de ambos tipos celulares (333).

Se seleccionó, por último, el gen *ERBB4*, como candidato para realizar estudios de expresión y confirmar su desregulación en mujeres jóvenes. Aunque *ERBB2* y *ERBB3* se presentaron como dianas validadas sobre-expresadas de los miRNAs desregulados, conocemos el estado de HER2 en las muestras, ya que se analiza en la rutina clínica como marcador de pronóstico, su implicación en cáncer de mama está ampliamente estudiada y su desregulación no explica los casos HER2-dentro de las mujeres jóvenes. *ERBB3*, por otro lado codifica para el receptor de membrana HER3, cuya actividad depende directamente de otros receptores de la familia de EGFR, ya que no posee dominio extracelular y, para ser activo, necesita dimerizar con HER2 o (141).

La ruta de HER, en concreto HER2 y HER3, ha sido ampliamente estudiada en relación al cáncer de mama, siendo marcadores de pronóstico decisivos y jugando un papel clave a la hora de dictar un tratamiento (334). Recientemente, HER4 ha sido ampliamente estudiado, al presentarse como una nueva diana terapéutica o como factor pronóstico (152, 335-337). Sin embargo, los autores no coinciden en si se trata de un marcador de buen pronóstico o de mal pronóstico, habiendo trabajos que apoyan ambas hipótesis. La actividad de HER4 es necesaria para la formación de la glándula mamaria, habiéndose observado un aumento en la fosforilación del receptor (necesaria para su activación) en etapas tardías del embarazo y la lactancia, además la eliminación del gen *ERBB4* impide la progresión del TEB y una correcta formación de la glándula mamaria para la lactancia, parece ser que, además, HER4 es necesario para la expresión mediante STAT5A de las proteínas lácteas  $\beta$ -caseína y WAP durante la lactancia (336).

#### 4.2. La expresión de los mRNAs en tejido tumoral mamario según la edad

Se midió la expresión de *ERBB4* mediante qRT-PCR, sin embargo, los resultados obtenidos no fueron significativos, no pudiendo concluir que la expresión de este gen en mujeres jóvenes sea significativamente mayor que en mujeres mayores. Se observa, sin embargo una tendencia a una mayor expresión, por lo que los resultados no son del todo contradictorios. Se observa una acumulación de valores cercanos a cero, como se ha demostrado anteriormente, la obtención de

material genético para el estudio de RNA de gran tamaño a partir de tejido FFPE no es óptima, pudiendo encontrarse el RNA degradado en la muestra. Por ello se pretende destacar el papel importante de los miRNAs como marcador molecular, debido a, como se ha comentado antes, su alta estabilidad y recuperación a partir de material parafinado, en el cual el RNA suele encontrarse degradado, sobre todo en estudios retrospectivos en los que las muestras tienen una cierta antigüedad.

La expresión de *GRIN2B*, sin embargo, sí resultó significativamente mayor en mujeres jóvenes, confirmando una mayor presencia del receptor NMDAR en las células tumorales de mujeres jóvenes con cáncer de mama. Sin embargo, los resultados de la qRT-PCR mostraron que la subunidad NR2B no se encuentra presente, o a niveles muy bajos en algunas muestras, lo que podría indicar que las células tumorales se sirven de este mecanismo sólo en algunos casos. O bien, que esta sobre-expresión es dependiente de la localización sobre el tumor, ya que se ha localizado preferentemente en los bordes tumorales; por lo tanto es posible que, aquellas muestras en las que no se observa, correspondan a cortes procedentes del centro del tumor (323).

Se ha descrito que el receptor NMDAR es un heterodímero compuesto por NR1 y alguna de los subtipos de NR2 (NR2A-NR2D). NR2 juega además un papel esencial durante el desarrollo neuronal, habiéndose asociado diferentes patrones de expresión de sus subtipos con la edad (326, 338). Se debería, pues, estudiar la presencia del resto de subunidades en estas muestras para confirmar la presencia de esta subunidad en estas mujeres.

De manera llamativa, la expresión de *GRIN3A* también resultó positiva sólo en algunas muestras, sin mostrar, como en el caso de *GRIN2B*, asociación aparente a ninguna característica clínica. Sin embargo, en todos los casos en los que *GRIN2B* no estaba expresado, tampoco lo estaba *GRIN3A*, a excepción de uno. Si consideramos que *GRIN3A* actúa como inhibidor, en cierta manera, de NMDAR, este hecho podría indicar que, efectivamente, ninguna de las subunidades de NR2 se encuentra presente, por lo que tampoco es necesaria su regulación mediante *GRIN3A*.

En el caso del receptor *GRM3* (mGluR3) no se encontró expresión en ninguna muestra y podría tratarse de un fallo en la detección o en la sensibilidad de la técnica. Sin embargo, el hecho de que tampoco se detecte proteína en tejido de cáncer de mama en la base de datos The Human Protein Atlas (<http://www.proteinatlas.org/>), parece indicar que este gen no se expresa en tumores mamarios; aunque, al ser la mayoría de tejidos analizados procedentes de mujeres de edad avanzada, a excepción de una de 40 años, no se puede concluir que así sea.

Por otro lado, y como se ha comentado antes, sí se ha visto presencia de *GRM3* en tejido tumoral de diferentes tipos de cáncer, por lo que se necesitan más estudios en esta dirección para concluir la implicación de este gen, aunque, en cualquier caso, no parece ser un factor relevante en la iniciación del cáncer de mama en mujeres muy jóvenes (321).

#### *4.3. Valoración de la cantidad de receptor in situ mediante IHC en mujeres muy jóvenes*

Se sabe que los miRNAs silencian a los genes mediante su unión a la región 3'UTR del mRNA, pero su silenciamiento no siempre tiene como final la degradación del mRNA. Cuando

no existe un apareamiento completo del miRNA con la secuencia diana, el mRNA es simplemente “secuestrado” por el miRNA y el complejo proteico RISC, que impiden la unión al ribosoma y su traducción a proteína. Sin embargo, en estos casos el mRNA se mantiene en la célula, posiblemente como un repositorio rápido de proteína en caso de necesidad inmediata (70, 71, 339). Los miRNAs maduros necesitan únicamente una secuencia semilla de 8 nucleótidos para poder comenzar la unión a su diana, por lo que determinar si el silenciamiento de la proteína está desembocando en su degradación o no resulta complicado (340).

Durante la extracción de material genético del tejido, las proteínas son degradadas, liberando todo el RNA que se encontrase en la célula en ese momento. No se puede concluir mediante qRT-PCR si ese RNA se encontraba silenciado por miRNAs o no, y no se puede establecer que existan diferencias entre mujeres jóvenes y mayores a este nivel. En el trabajo de Wu y colaboradores (2015) llegan a la misma conclusión al evaluar la acción del miRNA hsa-miR30c sobre el mRNA de *IERS2*, éste no cambia los niveles del mensajero, pero sí de la proteína final (273). Este hecho acentúa la importancia de los miRNAs como marcadores moleculares, fácilmente detectables y que indican una variabilidad no detectable mediante el análisis de la expresión de genes por qRT-PCR de sus mensajeros.

La expresión del mRNA de HER4 no resultó ser diferente de mujeres mayores en pacientes jóvenes, sin embargo, como se ha comentado, esto no indica necesariamente que no exista un efecto sobre el receptor. Por ello debe medirse la expresión de la misma, es decir, la cantidad final de proteína que se encuentra en tejido, para comprobar el efecto que los miRNAs están ejerciendo sobre la proteína.

Con este fin se realizó tinción inmunohistoquímica sobre cortes de tejido tumoral FFPE de mujeres jóvenes y mujeres mayores, así como de tejido sano. Se emplearon anticuerpos anti-HER4 para detectar la presencia del receptor HER4. Los resultados mostraron una expresión mayor en mujeres jóvenes, habiendo en este grupo mayor representación de tinción positiva y, por tanto, de receptor HER4. Se obtuvo, además, un mayor número de muestras con tinción muy intensa en este grupo de muestras de mujeres jóvenes, mientras que en mujeres mayores la mayoría presentó tinción débil o muy débil.

Estos resultados apoyan la hipótesis formulada, si los miRNAs que regulan la ruta de señalización mediante ERBB se encuentran reprimidos en mujeres jóvenes, sus genes diana, entre los que se encuentra HER4, estarán sobre-expresados y habrá más cantidad de proteína; aunque no aparece como diana validada, sí es una diana teórica (Anexo XI). Las células tumorales podrían valerse de HER4, tanto en aquellas que co-expresan HER2 como en las que no, para activar las rutas de proliferación provocando el crecimiento tumoral.

Los estudios realizados sobre HER4 se muestran contradictorios en cuanto a su valor pronóstico en cáncer de mama, mostrándose en unos trabajos como de buen pronóstico al promover la apoptosis celular, mientras que en otros indica mal pronóstico, algunos autores sugieren que esto depende en gran medida del anticuerpo empleado en el trabajo (341-346). Se ha observado que la co-expresión de HER4 junto con HER2 mejora la supervivencia tras tratamiento con Trastuzumab. Sin embargo, la co-expresión de HER4 con EGFR ha demostrado un peor pronóstico en varios tipos de cáncer (347, 348).

La falta de acuerdo entre los resultados puede deberse a la complejidad de HER4, que cuenta con varias isoformas y, además, sufre escisión de su dominio intracelular, pudiendo adoptar diversas localizaciones en la célula. La mayoría de trabajos concluyen que es necesaria la liberación del dominio intracelular de HER4 (4ICD) y su consecuente transporte al núcleo para que ejerza su acción supresora de tumores, ya que activa la transcripción de genes que promueven la diferenciación del epitelio mamario, siendo indicativo de buen pronóstico (347, 349, 350).

En 2009, Thor y colaboradores ponen de manifiesto que la localización del dominio intracelular es relevante a la hora de establecer un pronóstico (152). De hecho, Fujiwara y colaboradores (2014) en un trabajo reciente destacan que tanto la isoforma como su localización son decisivas a la hora de determinar un buen pronóstico, asociando la isoforma CYT-2 junto con la alta presencia de 4ICD en el núcleo a un mejor pronóstico (151). Del mismo modo, Wali y colaboradores (2014) asocian una sobre-expresión de la forma CYT-1 a la aparición de lesiones neoplásicas en modelos murinos, sugiriendo una función oncogénica para esta isoforma (351). De hecho, CYT-2 carece de una región de 16 residuos que incluye un dominio de unión a PI3K, y muestra un patrón de localización preferentemente nuclear (352, 353).

Es posible, pues que, no solo la expresión sea mayor en tumor de mama de mujeres jóvenes, si no que la localización juegue un papel importante como factor pronóstico, siendo su internalización al núcleo lo que ejerce su acción supresora de tumores promoviendo la apoptosis. En mujeres jóvenes, al encontrarse principalmente en membrana, no ejercería esta acción, dimerizando con otros receptores y activando las rutas PI3K y MAPK a las que se encuentra asociado el receptor, lo que llevaría a un aumento de la proliferación celular, independientemente del estado de HER2 en estas pacientes. Se debe destacar que el anticuerpo empleado detecta la tirosina fosforilada, pTyr1162, que se encuentra en el dominio intracelular 4ICD, que posteriormente es escindido y transportado al núcleo.

La región 3'UTR de ERBB4 contiene sitios de unión teóricos para 411 miRNAs (incluyendo los miRNAs con expresión desregulada validada en mujeres jóvenes, hsa-miR-1207-5p, hsa-miR-30c-5p y hsa-miR-125a que, además, se encuentra conservado entre especies). Los diferentes transcritos de ERBB4 varían entre sí considerablemente en esta región, por lo que es posible que la desregulación de miRNAs afecte de manera diferente a ambos transcritos del gen ERBB4, silenciándolos de manera diferencial. Es probable, pues, que la desregulación de miRNAs en cáncer de mama de mujeres muy jóvenes silencie de manera preferente a una de las dos isoformas expresadas en tejido mamario (CYT-1 y CYT-2). Es también una posible razón de que no se hayan observado diferencias en la expresión del mRNA de *ERBB4* debido a la incapacidad de la sonda empleada en diferenciar entre ambas isoformas. Futuros estudios son necesarios para esclarecer si el cáncer de mama en mujeres jóvenes se encuentra sobre-representado en la isoforma CYT-1, lo que podría conllevar un peor pronóstico.

Ya que en el presente trabajo se han encontrado rutas relacionadas con la plasticidad neuronal desreguladas en tumor mamario de cáncer de mama, es interesante destacar que, en tejido neuronal, HER4 se encuentra muy expresado en las etapas que van desde el desarrollo embrionario hasta la edad adulta, siendo indicativo de plasticidad neuronal y capacidad de aprendizaje y, por tanto, facilitando la reordenación y formación de sinapsis neuronales. Cuando se alcanza una cierta edad, se produce la desensibilización de HER4, que deja de ser activo en neuronas mediante la escisión de su dominio extracelular (354).

Se ha de tener en cuenta que en este estudio se han empleado pocas muestras, por lo que cualquier conclusión es especulativa. Se requieren de subsecuentes estudios para poder arrojar resultados que apoyen o contradigan este estudio exploratorio. En cualquier caso merece un estudio más exhaustivo, debido a la creciente importancia del estado de HER4 como marcador de pronóstico, tanto en mujeres jóvenes como en mayores, y la consideración de que los resultados obtenidos en HER4 deben tomarse con precaución a la hora de extrapolarse sobre este grupo particular de pacientes, ya que podría no actuar en la misma línea que en pacientes mayores.

Se han encontrado diferentes patrones de metilación en genes asociados a características de células madre, incluyendo *GRIN2B*, *ERBB4* y su ligando *NRG1*, en diferentes subtipos de cáncer de mama, por lo que un patrón de metilación diferente podría ser el responsable último de la desregulación molecular en cáncer de mama de mujeres jóvenes, siendo la desregulación de miRNAs un efecto derivado de esta metilación (355-357).

La sobreexpresión del mensajero de *GRIN2B*, así como de la proteína HER4, en tejido tumoral de cáncer de mama en mujeres jóvenes, sugieren que, en efecto, estas rutas se encuentran expresadas en mayor medida en tumores de mujeres menores de 35 años que en mujeres de más edad. Además, proponen nuevas aproximaciones terapéuticas para este grupo de pacientes con consideraciones especiales y características clínicas particulares.

#### 4.4. Posibles aplicaciones terapéuticas

La existencia de fármacos que actúan como inhibidores a diferentes niveles de la ruta del glutamato permiten abrir la posibilidad a nuevas aproximaciones de tratamiento en mujeres jóvenes mediante la co-administración de fármacos para los diferentes receptores desregulados (HER4 y NMDAR), o su desacople de las rutas de proliferación (Figura 57), (322). En el trabajo de Li y colaboradores (2013), se obtienen resultados alentadores con el fármaco MK801, antagonista de NMDAR, que revierte la proliferación e invasión tumoral en cultivos celulares y mostró efectos terapéuticos en ratones RIP1-Tag2, con alta proporción de tumores sólidos, disminuyendo el tamaño y número de tumores (323). Existen varios antagonistas de NMDAR que se emplean en el tratamiento de desórdenes mentales, como son Riluzole, Memantine, MK-801 y AP5, que, presentando una citotoxicidad débil, reducen notablemente la proliferación tumoral en leucemia (319). Estudios preclínicos también muestran una sensibilización a la radioterapia en metástasis de cerebro derivada de melanoma tras el uso de Riluzole (358).

En cuanto a la aplicación de HER4 como diana terapéutica, existen fármacos inhibidores pan-HER, es decir, se unen a todos los receptores de la familia de EGFR, incluyendo HER4, como el lapatinib, un fármaco ampliamente usado en cáncer de mama para tratar tumores HER2 positivos resistentes a trastuzumab, habiéndose demostrado el mismo mecanismo de inhibición para HER4 (359). Recientemente, se ha desarrollado un nuevo tipo de inhibidores pan-HER, que se une de manera irreversible, disminuyendo notablemente la fosforilación y, por tanto, la activación de las rutas de señalización iniciadas por los receptores de la familia EGFR. De ellos, el Neratinib ha pasado la fase pre-clínica y se encuentra en varios ensayos clínicos. Neratinib muestra resultados muy buenos en aquellas pacientes que sobre-expresan HER2, pero, además, también existen resultados preliminares alentadores, que muestran respuesta patológica en tumores HER2 negativos cuando se tratan con Neratinib junto con quimioterapia, aunque los resultados son

inconstantes en tumores triple negativos, poniendo de manifiesto la heterogeneidad de los tumores y la necesidad de estratificar y seleccionar adecuadamente a los pacientes (360, 361).

Actualmente, la terapia anti-HER se administra en la práctica clínica rutinaria a pacientes con tumores HER2 positivos, sin embargo, sería recomendable establecer también el estado de otros receptores como HER4, especialmente en mujeres jóvenes, ya que muestran una mayor expresión de HER4 independientemente del subtipo tumoral que presenten.

La existencia de un perfil molecular diferente en cáncer de mama de mujeres jóvenes remarca la necesidad de tratar a este grupo de pacientes con consideraciones especiales, así como la necesidad de ensayos clínicos para este grupo, o una mayor representación de estas pacientes en los mismos, incluyendo la edad como variable, ya que, la poca incidencia del cáncer de mama en mujeres jóvenes ocasiona que los efectos sobre este grupo se diluyan en los resultados globales.

En este trabajo se ha presentado, por primera vez, un análisis exhaustivo de la desregulación de miRNAs en mujeres jóvenes con cáncer de mama, mostrando un perfil molecular distintivo de tumores de mama de mujeres menores de 35 años. Se han identificado rutas metabólicas posiblemente desreguladas y evaluado la implicación de genes diana candidatos, sugiriendo así el posible mecanismo molecular que lleva a la progresión tumoral. Además, se han señalado posibles marcadores moleculares de supervivencia y posibles dianas terapéuticas. Se necesitan futuros estudios que amplíen el conocimiento acerca del cáncer de mama en mujeres jóvenes y confirmen los resultados obtenidos en el presente trabajo de Tesis Doctoral, sin embargo, los resultados apuntan a que el cáncer de mama en mujeres jóvenes es una entidad biológica diferente, con una etiología clínica y molecular específica, que merece ser tratada como tal.

# CONCLUSIONES





La conclusión principal que se extrae de los resultados obtenidos en este trabajo de Tesis Doctoral es que el cáncer de mama en mujeres muy jóvenes parece constituir una entidad biológica diferenciada dentro del cáncer de mama general.

1 El análisis de integridad del RNA no resulta efectivo para evaluar la idoneidad del RNA de pequeño tamaño para estudios moleculares. Mediante qRT-PCR se ha observado que los miRNAs son más estables y se recuperan mejor que RNA de mayor tamaño en muestras FFPE. En estudios moleculares a partir de material parafinado, el marcador biológico más adecuado son los miRNAs.

2 Se ha puesto de manifiesto la existencia de un perfil molecular distintivo de cáncer de mama en mujeres menores de 35 años. Un 8,3% de los miRNAs analizados (91 de 1100 miRNAs humanos maduros) mostraron una expresión diferente en cáncer de mama de mujeres muy jóvenes. De ellos 43 miRNAs tuvieron una expresión mayor que en mujeres mayores y 48 miRNAs, menor.

3 El análisis de enriquecimiento de rutas en dianas teóricas y validadas, apuntó a rutas relacionadas con movilidad celular y adhesión, con rutas de señalización relacionadas con proliferación y receptores RTK, con procesos neuronales y rutas asociadas a “sinapsis” y con varios tipos de cánceres. Se seleccionaron los receptores HER4, GRIN2B, GRIN3A y GRM3.

4 El perfil de expresión diferencial de miRNAs se validó en una muestra independiente mayor de pacientes de cáncer de mama usando datos de un estudio previo (METABRIC), obteniéndose 51 miRNAs desregulados (25 reprimidos y 31 sobre-expresados), además, estos miRNAs desregulados apuntaban a rutas similares a las obtenidas en nuestro estudio. De entre las rutas desreguladas se destacaron aquellas relacionadas con ERBB4, MAPK y PI3K/Akt/mTOR y guía de axones.

5 Se reconstruyó el perfil de miRNAs desregulado en mujeres jóvenes identificado en pacientes, mostrando la misma variabilidad que en las muestras procedentes de pacientes. Las líneas celulares HCC1937 y HCC1500 mostraron una expresión muy similar al de las pacientes jóvenes y una buena correlación entre sus réplicas, constituyendo un buen modelo para la realización de estudios funcionales.

6 Los perfiles obtenidos mediante los tres estudios (muestras, líneas celulares y base de datos online) mostraron 3 miRNAs comunes: miR30c, miR489 y miR663. Además, el meta-análisis reveló 22 miRNAs con expresión diferente en cáncer de mama de mujeres jóvenes.

7 De los 18 miRNAs seleccionados, 12 fueron validados y 6 mantuvieron corrección por Bonferroni en una serie independiente de muestras procedentes de pacientes de cáncer de mama: miR30c, miR92b, miR139, miR1228-5p, miR1275-3p y miR3196. Todos los miRNAs analizados mostraron una expresión con la misma tendencia que la obtenida en el análisis de expresión de miRNAs mediante chip de Affymetrix. Además los miRNAs miR92b y miR139 se asociaron a una menor supervivencia global.

8 De los genes diana seleccionados, se validó *GRIN2B*, cuya expresión de mRNA se encuentra sobre-expresada en mujeres jóvenes, concordando con la expresión de los miRNAs. La presencia de la proteína HER4 en tejido tumoral fue significativamente mayor en mujeres jóvenes. Además, la tinción en éstas fue principalmente de membrana, en contraposición a la tinción nuclear encontrada en mujeres mayores.

# BIBLIOGRAFÍA

---



BIBLIOGRAFÍA CONSULTADA

---

1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
3. Radisky DC, Hartmann LC. Mammary involution and breast cancer risk: transgenic models and clinical studies. *Journal of mammary gland biology and neoplasia*. 2009;14(2):181-91.
4. Russo J, Russo IH. Development of the human breast. *Maturitas*. 2004;49(1):2-15.
5. Watson CJ, Khaled WT. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development*. 2008;135(6):995-1003.
6. Sternlicht MD. Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. *Breast cancer research : BCR*. 2006;8(1):201.
7. Schedin P, O'Brien J, Rudolph M, Stein T, Borges V. Microenvironment of the involuting mammary gland mediates mammary cancer progression. *Journal of mammary gland biology and neoplasia*. 2007;12(1):71-82.
8. Baxter FO, Neoh K, Tevendale MC. The beginning of the end: death signaling in early involution. *Journal of mammary gland biology and neoplasia*. 2007;12(1):3-13.
9. Lund VE, Makinen JI. Ectopic pregnancy and seasonality. *Human reproduction*. 1996;11(3):683-4.
10. La Vecchia C. Oral contraceptives, menopause hormone replacement therapy, and risk of stroke. *Maturitas*. 2004;47(4):265-8.
11. Stefanick ML, Anderson GL, Margolis KL, Hendrix SL, Rodabough RJ, Paskett ED, et al. Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *Jama*. 2006;295(14):1647-57.
12. La Vecchia C, Altieri A, Franceschi S, Tavani A. Oral contraceptives and cancer: an update. *Drug safety*. 2001;24(10):741-54.
13. Key TJ, Allen NE. Hormones and breast cancer. *IARC scientific publications*. 2002;156:273-6.
14. McDaniel SM, Rumer KK, Biroc SL, Metz RP, Singh M, Porter W, et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. *The American journal of pathology*. 2006;168(2):608-20.
15. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nature reviews Cancer*. 2006;6(4):281-91.
16. Dzhandzhalia MT. [Risk of the development of breast cancer during gynecological pathology]. *Georgian medical news*. 2006(137):32-4.
17. Hanf V, Hanf D. Reproduction and breast cancer risk. *Breast care*. 2014;9(6):398-405.

18. Anderson WF, Jatoi I, Sherman ME. Qualitative age interactions in breast cancer studies: mind the gap. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2009;27(32):5308-11.
19. Colditz GA. NHS breast screening: a model national programme. *Journal of medical screening*. 2006;13(2):58.
20. Key T, Appleby P, Barnes I, Reeves G, Endogenous H, Breast Cancer Collaborative G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *Journal of the National Cancer Institute*. 2002;94(8):606-16.
21. Veronesi U, Goldhirsch A, Boyle P, Orecchia R, Viale G. Breast cancer. *Discovery medicine*. 2005;5(27):271-7.
22. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., et al. Alcohol, tobacco and breast cancer-collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British journal of cancer*. 2002;87(11):1234-45.
23. Key TJ, Allen NE, Spencer EA, Travis RC. The effect of diet on risk of cancer. *Lancet*. 2002;360(9336):861-8.
24. Michels KB, Mohllajee AP, Roset-Bahmanyar E, Beehler GP, Moysich KB. Diet and breast cancer: a review of the prospective observational studies. *Cancer*. 2007;109(12 Suppl):2712-49.
25. Le Marchand L, Kolonel LN, Earle ME, Mi MP. Body size at different periods of life and breast cancer risk. *American journal of epidemiology*. 1988;128(1):137-52.
26. Peacock SL, White E, Daling JR, Voigt LF, Malone KE. Relation between obesity and breast cancer in young women. *American journal of epidemiology*. 1999;149(4):339-46.
27. Peter Boyle BL. World cancer report. Cancer IAFRo, editor. Lyon: International Agency for Research on Cancer (IARC); 2008.
28. Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegelé C, Young EL, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast cancer research : BCR*. 2014;16(3):R58.
29. Heikkinen K, Rapakko K, Karppinen SM, Erkkö H, Knuutila S, Lundan T, et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis*. 2006;27(8):1593-9.
30. Litman R, Peng M, Jin Z, Zhang F, Zhang J, Powell S, et al. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCF. *Cancer cell*. 2005;8(3):255-65.
31. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(\*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature genetics*. 2002;31(1):55-9.
32. Siegel MB, Van Swearingen AE, Anders CK. Approaches for optimal drug development and clinical trial design for breast cancer brain metastasis. *Oncology*. 2014;28(7):579, 84-5.

33. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015;65(2):87-108.
34. Pollan M, Pastor-Barriuso R, Ardanaz E, Arguelles M, Martos C, Galceran J, et al. Recent changes in breast cancer incidence in Spain, 1980-2004. *Journal of the National Cancer Institute*. 2009;101(22):1584-91.
35. Pollan M. Epidemiology of breast cancer in young women. *Breast cancer research and treatment*. 2010;123 Suppl 1:3-6.
36. Pollan M, Michelena MJ, Ardanaz E, Izquierdo A, Sanchez-Perez MJ, Torrella A, et al. Breast cancer incidence in Spain before, during and after the implementation of screening programmes. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2010;21 Suppl 3:iii97-102.
37. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of surgical oncology*. 2010;17(6):1471-4.
38. Cornejo KM, Kandil D, Khan A, Cosar EF. Theranostic and molecular classification of breast cancer. *Archives of pathology & laboratory medicine*. 2014;138(1):44-56.
39. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-52.
40. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. *Nature reviews Cancer*. 2007;7(9):659-72.
41. Galanina N, Bossuyt V, Harris LN. Molecular predictors of response to therapy for breast cancer. *Cancer journal*. 2011;17(2):96-103.
42. Mehta R, Jain RK, Badve S. Personalized medicine: the road ahead. *Clinical breast cancer*. 2011;11(1):20-6.
43. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2010;16(21):5222-32.
44. 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. *The Journal of clinical investigation*. 2011;121(7):2750-67.
45. Criscitiello C, Giuliano M, Curigliano G, De Laurentiis M, Arpino G, Carlomagno N, et al. Surgery of the primary tumor in de novo metastatic breast cancer: To do or not to do? *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*. 2015.
46. Jatoi I. Breast cancer screening: past, present and future. *Future oncology*. 2015.
47. Zelnak AB, O'Regan RM. Optimizing Endocrine Therapy for Breast Cancer. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2015;13(8):e56-64.
48. Baselga J, Mendelsohn J. Type I receptor tyrosine kinases as targets for therapy in breast cancer. *Journal of mammary gland biology and neoplasia*. 1997;2(2):165-74.

49. Baselga J, Seidman AD, Rosen PP, Norton L. HER2 overexpression and paclitaxel sensitivity in breast cancer: therapeutic implications. *Oncology*. 1997;11(3 Suppl 2):43-8.
50. Cardoso F, Loibl S, Pagani O, Graziottin A, Panizza P, Martincich L, et al. The European Society of Breast Cancer Specialists recommendations for the management of young women with breast cancer. *European journal of cancer*. 2012;48(18):3355-77.
51. Kothari AS, Fentiman IS. 11. Breast cancer in young women. *International journal of clinical practice*. 2002;56(3):184-7.
52. Ruddy KJ, Gelber S, Tamimi RM, Schapira L, Come SE, Meyer ME, et al. Breast cancer presentation and diagnostic delays in young women. *Cancer*. 2014;120(1):20-5.
53. de la Rochefordiere A, Asselain B, Campana F, Scholl SM, Fenton J, Vilcoq JR, et al. Age as prognostic factor in premenopausal breast carcinoma. *Lancet*. 1993;341(8852):1039-43.
54. Han W, Kang SY, Korean Breast Cancer S. Relationship between age at diagnosis and outcome of premenopausal breast cancer: age less than 35 years is a reasonable cut-off for defining young age-onset breast cancer. *Breast cancer research and treatment*. 2010;119(1):193-200.
55. Canello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2010;21(10):1974-81.
56. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JA, et al. Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. *PLoS one*. 2008;3(1):e1373.
57. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(20):3324-30.
58. Canello G, Montagna E. Treatment of breast cancer in young women: do we need more aggressive therapies? *Journal of thoracic disease*. 2013;5 Suppl 1:S47-54.
59. Colleoni M, Anders CK. Debate: The biology of breast cancer in young women is unique. *The oncologist*. 2013;18(4):e13-5.
60. Goldhirsch A, Glick JH, Gelber RD, Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. *Journal of the National Cancer Institute*. 1998;90(21):1601-8.
61. Thuerlimann B. International consensus meeting on the treatment of primary breast cancer 2001, St. Gallen, Switzerland. *Breast cancer*. 2001;8(4):294-7.
62. Rosenberg SM, Partridge AH. Management of breast cancer in very young women. *Breast*. 2015.
63. Freedman RA, Partridge AH. Management of breast cancer in very young women. *Breast*. 2013;22 Suppl 2:S176-9.
64. Azim HA, Jr., Azim H. Breast cancer arising at a young age: do we need to define a cut-off? *Breast*. 2013;22(6):1007-8.



65. Goldhirsch A, Gelber RD, Yothers G, Gray RJ, Green S, Bryant J, et al. Adjuvant therapy for very young women with breast cancer: need for tailored treatments. *Journal of the National Cancer Institute Monographs*. 2001(30):44-51.
66. Pritchard KI. Adjuvant therapy for breast cancer: we are solving the puzzle, but pieces are still missing. *Journal of the National Cancer Institute*. 2007;99(7):494-5.
67. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998;391(6669):806-11.
68. Hammond SM. RNAi, microRNAs, and human disease. *Cancer chemotherapy and pharmacology*. 2006;58 Suppl 1:s63-8.
69. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993;75(5):843-54.
70. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281-97.
71. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. *Nature reviews Molecular cell biology*. 2009;10(2):126-39.
72. Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional regulation by pseudogenes. *Biochimie*. 2011;93(11):1916-21.
73. Rigoutsos I, Furnari F. Gene-expression forum: Decoy for microRNAs. *Nature*. 2010;465(7301):1016-7.
74. Hammond SM. MicroRNAs as oncogenes. *Current opinion in genetics & development*. 2006;16(1):4-9.
75. Di Leva G, Croce CM. miRNA profiling of cancer. *Current opinion in genetics & development*. 2013;23(1):3-11.
76. Garofalo M, Croce CM. microRNAs: Master regulators as potential therapeutics in cancer. *Annual review of pharmacology and toxicology*. 2011;51:25-43.
77. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nature reviews Cancer*. 2006;6(11):857-66.
78. Reddy KB. MicroRNA (miRNA) in cancer. *Cancer cell international*. 2015;15:38.
79. Peiro-Chova L, Pena-Chilet M, Lopez-Guerrero JA, Garcia-Gimenez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. *Virchows Archiv : an international journal of pathology*. 2013;463(6):765-74.
80. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. *Nature biotechnology*. 2008;26(4):462-9.
81. Adams BD, Kasinski AL, Slack FJ. Aberrant regulation and function of microRNAs in cancer. *Current biology : CB*. 2014;24(16):R762-76.
82. Cheng WC, Chung IF, Tsai CF, Huang TS, Chen CY, Wang SC, et al. YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research. *Nucleic acids research*. 2015;43(Database issue):D862-7.

83. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends in molecular medicine*. 2014;20(8):460-9.
84. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435(7043):834-8.
85. Li J, Tan S, Kooger R, Zhang C, Zhang Y. MicroRNAs as novel biological targets for detection and regulation. *Chemical Society reviews*. 2014;43(2):506-17.
86. van Rooij E, Kauppinen S. Development of microRNA therapeutics is coming of age. *EMBO molecular medicine*. 2014;6(7):851-64.
87. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer research*. 2005;65(16):7065-70.
88. Gasparini P, Cascione L, Fassan M, Lovat F, Guler G, Balci S, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. *Oncotarget*. 2014;5(5):1174-84.
89. Iorio MV, Casalini P, Piovan C, Di Leva G, Merlo A, Triulzi T, et al. microRNA-205 regulates HER3 in human breast cancer. *Cancer research*. 2009;69(6):2195-200.
90. Adams BD, Furneaux H, White BA. The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. *Molecular endocrinology*. 2007;21(5):1132-47.
91. Hossain A, Kuo MT, Saunders GF. Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. *Molecular and cellular biology*. 2006;26(21):8191-201.
92. Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M, et al. A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *The Journal of cell biology*. 2008;182(3):509-17.
93. Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK, et al. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Molecular cancer*. 2006;5:24.
94. Scott GK, Goga A, Bhaumik D, Berger CE, Sullivan CS, Benz CC. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *The Journal of biological chemistry*. 2007;282(2):1479-86.
95. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature cell biology*. 2008;10(5):593-601.
96. Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, et al. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell*. 2009;138(3):592-603.
97. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. *Cell*. 2005;120(5):635-47.
98. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, et al. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell*. 2007;131(6):1109-23.

99. Christophersen NS, Helin K. Epigenetic control of embryonic stem cell fate. *The Journal of experimental medicine*. 2010;207(11):2287-95.
100. Sun F, Fu H, Liu Q, Tie Y, Zhu J, Xing R, et al. Downregulation of CCND1 and CDK6 by miR-34a induces cell cycle arrest. *FEBS letters*. 2008;582(10):1564-8.
101. Welch C, Chen Y, Stallings RL. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene*. 2007;26(34):5017-22.
102. Wu F, Zhu S, Ding Y, Beck WT, Mo YY. MicroRNA-mediated regulation of Ubc9 expression in cancer cells. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(5):1550-7.
103. Valastyan S, Reinhardt F, Benaich N, Calogrias D, Szasz AM, Wang ZC, et al. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell*. 2009;137(6):1032-46.
104. Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang Q, Bos PD, et al. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*. 2008;451(7175):147-52.
105. Tsuchiya Y, Nakajima M, Takagi S, Taniya T, Yokoi T. MicroRNA regulates the expression of human cytochrome P450 1B1. *Cancer research*. 2006;66(18):9090-8.
106. Zhang J, Du YY, Lin YF, Chen YT, Yang L, Wang HJ, et al. The cell growth suppressor, mir-126, targets IRS-1. *Biochemical and biophysical research communications*. 2008;377(1):136-40.
107. Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science*. 2008;322(5908):1695-9.
108. Sachdeva M, Zhu S, Wu F, Wu H, Walia V, Kumar S, et al. p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(9):3207-12.
109. Bhaumik D, Scott GK, Schokrpur S, Patil CK, Campisi J, Benz CC. Expression of microRNA-146 suppresses NF-kappaB activity with reduction of metastatic potential in breast cancer cells. *Oncogene*. 2008;27(42):5643-7.
110. Wu H, Zhu S, Mo YY. Suppression of cell growth and invasion by miR-205 in breast cancer. *Cell research*. 2009;19(4):439-48.
111. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, Lund AH. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *The Journal of biological chemistry*. 2008;283(2):1026-33.
112. Qi L, Bart J, Tan LP, Platteel I, Sluis T, Huitema S, et al. Expression of miR-21 and its targets (PTEN, PDCD4, TM1) in flat epithelial atypia of the breast in relation to ductal carcinoma in situ and invasive carcinoma. *BMC cancer*. 2009;9:163.
113. Zhu S, Wu H, Wu F, Nie D, Sheng S, Mo YY. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell research*. 2008;18(3):350-9.
114. Kong W, Yang H, He L, Zhao JJ, Coppola D, Dalton WS, et al. MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA. *Molecular and cellular biology*. 2008;28(22):6773-84.

115. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*. 2007;449(7163):682-8.
116. Huang Q, Gumireddy K, Schrier M, le Sage C, Nagel R, Nair S, et al. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nature cell biology*. 2008;10(2):202-10.
117. Mertens-Talcott SU, Chintharlapalli S, Li X, Safe S. The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells. *Cancer research*. 2007;67(22):11001-11.
118. Miller TE, Ghoshal K, Ramaswamy B, Roy S, Datta J, Shapiro CL, et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *The Journal of biological chemistry*. 2008;283(44):29897-903.
119. Schwarzenbacher D, Balic M, Pichler M. The role of microRNAs in breast cancer stem cells. *International journal of molecular sciences*. 2013;14(7):14712-23.
120. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318(5858):1917-20.
121. Sabio G, Davis RJ. TNF and MAP kinase signalling pathways. *Seminars in immunology*. 2014;26(3):237-45.
122. Chang L, Karin M. Mammalian MAP kinase signalling cascades. *Nature*. 2001;410(6824):37-40.
123. Wilkinson B, Kaye J. Requirement for sustained MAPK signaling in both CD4 and CD8 lineage commitment: a threshold model. *Cellular immunology*. 2001;211(2):86-95.
124. Song K, Wang H, Krebs TL, Danielpour D. Novel roles of Akt and mTOR in suppressing TGF-beta/ALK5-mediated Smad3 activation. *The EMBO journal*. 2006;25(1):58-69.
125. Zheng L, Zhang D, Zhang Y, Wen Y, Wang Y. mTOR signal transduction pathways contribute to TN-C FNIII A1 overexpression by mechanical stress in osteosarcoma cells. *Molecules and cells*. 2014;37(2):118-25.
126. Pignata C, Romano R. In this issue: FOX genes and the immune response. *International reviews of immunology*. 2014;33(2):81-2.
127. Zhou BP, Liao Y, Xia W, Zou Y, Spohn B, Hung MC. HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nature cell biology*. 2001;3(11):973-82.
128. Paplomata E, O'Regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Therapeutic advances in medical oncology*. 2014;6(4):154-66.
129. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science*. 1999;284(5415):770-6.
130. Grego-Bessa J, Luna-Zurita L, del Monte G, Bolos V, Melgar P, Arandilla A, et al. Notch signaling is essential for ventricular chamber development. *Developmental cell*. 2007;12(3):415-29.

131. Moellering RE, Cornejo M, Davis TN, Del Bianco C, Aster JC, Blacklow SC, et al. Direct inhibition of the NOTCH transcription factor complex. *Nature*. 2009;462(7270):182-8.
132. Ranganathan P, Weaver KL, Capobianco AJ. Notch signalling in solid tumours: a little bit of everything but not all the time. *Nature reviews Cancer*. 2011;11(5):338-51.
133. Dorsam RT, Gutkind JS. G-protein-coupled receptors and cancer. *Nature reviews Cancer*. 2007;7(2):79-94.
134. Sharma G, Prossnitz ER. Mechanisms of estradiol-induced insulin secretion by the G protein-coupled estrogen receptor GPR30/GPER in pancreatic beta-cells. *Endocrinology*. 2011;152(8):3030-9.
135. Luttrell LM, Daaka Y, Lefkowitz RJ. Regulation of tyrosine kinase cascades by G-protein-coupled receptors. *Current opinion in cell biology*. 1999;11(2):177-83.
136. Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. *Oncogene*. 2000;19(49):5548-57.
137. Zwick E, Bange J, Ullrich A. Receptor tyrosine kinase signalling as a target for cancer intervention strategies. *Endocrine-related cancer*. 2001;8(3):161-73.
138. Cohen S. The epidermal growth factor (EGF). *Cancer*. 1983;51(10):1787-91.
139. Thompson DM, Gill GN. The EGF receptor: structure, regulation and potential role in malignancy. *Cancer surveys*. 1985;4(4):767-88.
140. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. 2010;141(7):1117-34.
141. Roskoski R, Jr. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacological research*. 2014;79:34-74.
142. Roskoski R, Jr. ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. *Pharmacological research*. 2014;87:42-59.
143. Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN, Ueno NT. Role of epidermal growth factor receptor in breast cancer. *Breast cancer research and treatment*. 2012;136(2):331-45.
144. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, et al. Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer discovery*. 2013;3(2):224-37.
145. Chiu CG, Masoudi H, Leung S, Voduc DK, Gilks B, Huntsman DG, et al. HER3 overexpression is prognostic of reduced breast cancer survival: a study of 4046 patients. *Annals of surgery*. 2010;251(6):1107-16.
146. Aceto N, Duss S, MacDonald G, Meyer DS, Roloff TC, Hynes NE, et al. Co-expression of HER2 and HER3 receptor tyrosine kinases enhances invasion of breast cells via stimulation of interleukin-8 autocrine secretion. *Breast cancer research : BCR*. 2012;14(5):R131.
147. Koutras AK, Fountzilas G, Kalogeras KT, Starakis I, Iconomou G, Kalofonos HP. The upgraded role of HER3 and HER4 receptors in breast cancer. *Critical reviews in oncology/hematology*. 2010;74(2):73-8.

148. Haskins JW, Nguyen DX, Stern DF. Neuregulin 1-activated ERBB4 interacts with YAP to induce Hippo pathway target genes and promote cell migration. *Science signaling*. 2014;7(355):ra116.
149. Kiuchi T, Ortiz-Zapater E, Monypenny J, Matthews DR, Nguyen LK, Barbeau J, et al. The ErbB4 CYT2 variant protects EGFR from ligand-induced degradation to enhance cancer cell motility. *Science signaling*. 2014;7(339):ra78.
150. Wali VB, Haskins JW, Gilmore-Hebert M, Platt JT, Liu Z, Stern DF. Convergent and divergent cellular responses by ErbB4 isoforms in mammary epithelial cells. *Molecular cancer research : MCR*. 2014;12(8):1140-55.
151. Fujiwara S, Hung M, Yamamoto-Ibusuk CM, Yamamoto Y, Yamamoto S, Tomiguchi M, et al. The localization of HER4 intracellular domain and expression of its alternately-spliced isoforms have prognostic significance in ER+ HER2- breast cancer. *Oncotarget*. 2014;5(11):3919-30.
152. Thor AD, Edgerton SM, Jones FE. Subcellular localization of the HER4 intracellular domain, 4ICD, identifies distinct prognostic outcomes for breast cancer patients. *The American journal of pathology*. 2009;175(5):1802-9.
153. Carpenter G, Liao HJ. Trafficking of receptor tyrosine kinases to the nucleus. *Experimental cell research*. 2009;315(9):1556-66.
154. Wang SC, Hung MC. Nuclear translocation of the epidermal growth factor receptor family membrane tyrosine kinase receptors. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(21):6484-9.
155. Lin SY, Makino K, Xia W, Matin A, Wen Y, Kwong KY, et al. Nuclear localization of EGF receptor and its potential new role as a transcription factor. *Nature cell biology*. 2001;3(9):802-8.
156. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic acids research*. 1999;27(22):4436-43.
157. Bresters D, Schipper ME, Reesink HW, Boeser-Nunnink BD, Cuypers HT. The duration of fixation influences the yield of HCV cDNA-PCR products from formalin-fixed, paraffin-embedded liver tissue. *Journal of virological methods*. 1994;48(2-3):267-72.
158. Cronin M, Pho M, Dutta D, Stephans JC, Shak S, Kiefer MC, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *The American journal of pathology*. 2004;164(1):35-42.
159. Korga A, Wilkolaska K, Korobowicz E. [Difficulties in using archival paraffin-embedded tissues for RNA expression analysis]. *Postepy higieny i medycyny doswiadczalnej*. 2007;61:151-5.
160. Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, et al. Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2002;15(9):979-87.
161. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American journal of pathology*. 2002;161(6):1961-71.

162. Hui AB, Shi W, Boutros PC, Miller N, Pintilie M, Fyles T, et al. Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Laboratory investigation; a journal of technical methods and pathology*. 2009;89(5):597-606.
163. Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nature methods*. 2004;1(2):155-61.
164. Tang GQ, Maxwell ES. *Xenopus* microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome research*. 2008;18(1):104-12.
165. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *Rna*. 2007;13(10):1668-74.
166. Lin CH, Lu YS, Huang CS, Kuo KT, Wang CC, You SL, et al. Prognostic molecular markers in women aged 35 years or younger with breast cancer: is there a difference from the older patients? *Journal of clinical pathology*. 2011;64(9):781-7.
167. Anderson WF, Matsuno R. Breast cancer heterogeneity: a mixture of at least two main types? *Journal of the National Cancer Institute*. 2006;98(14):948-51.
168. Anders CK, Deal AM, Miller CR, Khorram C, Meng H, Burrows E, et al. The prognostic contribution of clinical breast cancer subtype, age, and race among patients with breast cancer brain metastases. *Cancer*. 2011;117(8):1602-11.
169. Colak D, Nofal A, Albakheet A, Nirmal M, Jeprel H, Eldali A, et al. Age-specific gene expression signatures for breast tumors and cross-species conserved potential cancer progression markers in young women. *PloS one*. 2013;8(5):e63204.
170. Pastrello C, Polesel J, Della Puppa L, Viel A, Maestro R. Association between hsa-mir-146a genotype and tumor age-of-onset in BRCA1/BRCA2-negative familial breast and ovarian cancer patients. *Carcinogenesis*. 2010;31(12):2124-6.
171. Shen J, Ambrosone CB, DiCioccio RA, Odunsi K, Lele SB, Zhao H. A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis*. 2008;29(10):1963-6.
172. Song F, Zheng H, Liu B, Wei S, Dai H, Zhang L, et al. An miR-502-binding site single-nucleotide polymorphism in the 3'-untranslated region of the SET8 gene is associated with early age of breast cancer onset. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009;15(19):6292-300.
173. Nassar FJ, El Sabban M, Zgheib NK, Tfayli A, Boulos F, Jabbour M, et al. miRNA as potential biomarkers of breast cancer in the Lebanese population and in young women: a pilot study. *PloS one*. 2014;9(9):e107566.
174. Tanic M, Andres E, Rodriguez-Pinilla SM, Marquez-Rodas I, Cebollero-Presmanes M, Fernandez V, et al. MicroRNA-based molecular classification of non-BRCA1/2 hereditary breast tumours. *British journal of cancer*. 2013;109(10):2724-34.
175. Tanic M, Yanowsky K, Rodriguez-Antona C, Andres R, Marquez-Rodas I, Osorio A, et al. Deregulated miRNAs in hereditary breast cancer revealed a role for miR-30c in regulating KRAS oncogene. *PloS one*. 2012;7(6):e38847.

176. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-52.
177. Dvinge H, Git A, Graf S, Salmon-Divon M, Curtis C, Sottoriva A, et al. The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature*. 2013;497(7449):378-82.
178. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*. 1986;51 Pt 1:263-73.
179. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/technology*. 1993;11(9):1026-30.
180. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-8.
181. Hutzinger R, Mrazek J, Vorwerk S, Huttenhofer A. NcRNA-microchip analysis: a novel approach to identify differential expression of noncoding RNAs. *RNA biology*. 2010;7(5):586-95.
182. Bastien RR, Rodriguez-Lescure A, Ebbert MT, Prat A, Munarriz B, Rowe L, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC medical genomics*. 2012;5:44.
183. Cheang MC, Martin M, Nielsen TO, Prat A, Voduc D, Rodriguez-Lescure A, et al. Defining breast cancer intrinsic subtypes by quantitative receptor expression. *The oncologist*. 2015;20(5):474-82.
184. Sullivan LM, D'Agostino RB. Robustness and power of analysis of covariance applied to data distorted from normality by floor effects: homogeneous regression slopes. *Statistics in medicine*. 1996;15(5):477-96.
185. Sullivan LM, D'Agostino RB. Robustness of the t test applied to data distorted from normality by floor effects. *Journal of dental research*. 1992;71(12):1938-43.
186. Rochon J, Gondan M, Kieser M. To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC medical research methodology*. 2012;12:81.
187. Royston P. Estimating departure from normality. *Statistics in medicine*. 1991;10(8):1283-93.
188. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Statistics in medicine*. 1990;9(7):811-8.
189. Smyser CF, Cleverdon RC, Kulp WL. The use of mice in a modified Welch-Nuttall test for the isolation of *Clostridium Perfringens*. *American journal of clinical pathology*. 1952;22(10):1036-7.
190. Boiculescu LV, Azoicai D, Rezus E, Rezus C, Dorneanu O. [Comparing population means using the ANOVA method]. *Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi*. 2003;107(4):906-12.



191. Morrissey ER, Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic acids research*. 2009;37(Web Server issue):W581-6.
192. Fay MP, Proschan MA. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*. 2010;4:1-39.
193. Dunn OJ. Multiple Comparisons Using Rank Sums. *Technometrics*. 1964;6(3):11.
194. Pearson K. Determination of the Coefficient of Correlation. *Science*. 1909;30(757):23-5.
195. McLachlan GJ. On the EM algorithm for overdispersed count data. *Statistical methods in medical research*. 1997;6(1):76-98.
196. Kemp KE. Multiple comparisons: comparisonwise versus experimentwise Type I error rates and their relationship to power. *Journal of dairy science*. 1975;58(9):1374-8.
197. Ludbrook J. On making multiple comparisons in clinical and experimental pharmacology and physiology. *Clinical and experimental pharmacology & physiology*. 1991;18(6):379-92.
198. Tarone RE. A modified Bonferroni method for discrete data. *Biometrics*. 1990;46(2):515-22.
199. Assennato G, Bruzzi P. [Bonferroni in biomedical research]. *Giornale italiano di nefrologia : organo ufficiale della Societa italiana di nefrologia*. 2002;19(2):178-83.
200. Haybittle JL. Significance testing in the comparison of survival curves from clinical trials of cancer treatment. *European journal of cancer & clinical oncology*. 1986;22(11):1279-83.
201. Rodary C, Laplanche A, Com Nougue C, Flamant R. [Comparison of survival curves: Logrank test (author's transl)]. *Bulletin du cancer*. 1981;68(2):201-4.
202. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports Part 1*. 1966;50(3):163-70.
203. Brown M. A method for combining non-independent, one-sided tests of significance. *Biometrics*. 1975;31((4)):5.
204. Curtis D, Vine AE, Knight J. A simple method for assessing the strength of evidence for association at the level of the whole gene. *Advances and applications in bioinformatics and chemistry : AABC*. 2008;1:115-20.
205. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of evolutionary biology*. 2005;18(5):1368-73.
206. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-93.
207. Eisen MB, Brown PO. DNA arrays for analysis of gene expression. *Methods in enzymology*. 1999;303:179-205.
208. Lin CS. Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik*. 1982;62(3):277-80.

209. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*. 2013;14:128.
210. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, et al. Reactome: a knowledgebase of biological pathways. *Nucleic acids research*. 2005;33(Database issue):D428-32.
211. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, et al. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic acids research*. 2012;40(Web Server issue):W498-504.
212. Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome biology*. 2003;4(10):R70.
213. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*. 2011;39(Database issue):D691-7.
214. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1):27-30.
215. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*. 2009;19(1):92-105.
216. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lscy-6* and other microRNAs. *Nature structural & molecular biology*. 2011;18(10):1139-46.
217. Guo LH, Li H, Wang F, Yu J, He JS. The Tumor Suppressor Roles of miR-433 and miR-127 in Gastric Cancer. *International journal of molecular sciences*. 2013;14(7):14171-84.
218. Lin L, Liu D, Liang H, Xue L, Su C, Liu M. MiR-1228 promotes breast cancer cell growth and metastasis through targeting SCAI protein. *International journal of clinical and experimental pathology*. 2015;8(6):6646-55.
219. Weiner-Gorzal K, Dempsey E, Milewska M, McGoldrick A, Toh V, Walsh A, et al. Overexpression of the microRNA miR-433 promotes resistance to paclitaxel through the induction of cellular senescence in ovarian cancer cells. *Cancer medicine*. 2015;4(5):745-58.
220. Klopffleisch R, Weiss AT, Gruber AD. Excavation of a buried treasure-DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histology and histopathology*. 2011;26(6):797-810.
221. Becker C, Hammerle-Fickinger A, Riedmaier I, Pfaffl MW. mRNA and microRNA quality control for RT-qPCR analysis. *Methods*. 2010;50(4):237-43.
222. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC biotechnology*. 2007;7:36.
223. Bertoli G, Cava C, Castiglioni I. MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer. *Theranostics*. 2015;5(10):1122-43.

224. Duttagupta R, Jiang R, Gollub J, Getts RC, Jones KW. Impact of cellular miRNAs on circulating miRNA biomarker signatures. *PloS one*. 2011;6(6):e20769.
225. Adami HO, Malker B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. *The New England journal of medicine*. 1986;315(9):559-63.
226. Nixon AJ, Neuberger D, Hayes DF, Gelman R, Connolly JL, Schnitt S, et al. Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1994;12(5):888-94.
227. Azim HA, Jr., Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(5):1341-51.
228. Chen L, Huang Z, Yao G, Lyu X, Li J, Hu X, et al. The expression of CXCL13 and its relation to unfavorable clinical characteristics in young breast cancer. *Journal of translational medicine*. 2015;13:168.
229. Bastos EP, Brentani H, Pasini FS, Silva AR, Torres CH, Puga RD, et al. MicroRNAs discriminate familial from sporadic non-BRCA1/2 breast carcinoma arising in patients  $\leq 35$  years. *PloS one*. 2014;9(7):e101656.
230. Canello G, Maisonneuve P, Mazza M, Montagna E, Rotmensz N, Viale G, et al. Pathological features and survival outcomes of very young patients with early breast cancer: how much is "very young"? *Breast*. 2013;22(6):1046-51.
231. Chen W, Sonke JJ, Stroom J, Bartelink H, Verheij M, Gilhuijs K. The effect of age in breast conserving therapy: a retrospective analysis on pathology and clinical outcome data. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2015;114(3):314-21.
232. Liedtke C, Rody A, Gluz O, Baumann K, Beyer D, Kohls EB, et al. The prognostic impact of age in different molecular subtypes of breast cancer. *Breast cancer research and treatment*. 2015;152(3):667-73.
233. Rosenberg SM, Newman LA, Partridge AH. Breast Cancer in Young Women: Rare Disease or Public Health Problem? *JAMA oncology*. 2015.
234. Villarreal-Garza C, Aguila C, Magallanes-Hoyos MC, Mohar A, Bargallo E, Meneses A, et al. Breast cancer in young women in Latin America: an unmet, growing burden. *The oncologist*. 2013;18 Suppl:26-34.
235. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, et al. Relationship of mammographic density and gene expression: analysis of normal breast tissue surrounding breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2013;19(18):4972-82.
236. Pollan M, Lope V, Miranda-Garcia J, Garcia M, Casanova F, Sanchez-Contador C, et al. Adult weight gain, fat distribution and mammographic density in Spanish pre- and post-menopausal women (DDM-Spain). *Breast cancer research and treatment*. 2012;134(2):823-38.

237. Lopez-Romero P, Gonzalez MA, Callejas S, Dopazo A, Irizarry RA. Processing of Agilent microRNA array data. *BMC research notes*. 2010;3:18.
238. Pena-Chilet M, Martinez MT, Perez-Fidalgo JA, Peiro-Chova L, Oltra SS, Tormo E, et al. MicroRNA profile in very young women with breast cancer. *BMC cancer*. 2014;14:529.
239. Zaragoza R, Garcia-Trevijano ER, Lluch A, Ribas G, Vina JR. Involvement of Different networks in mammary gland involution after the pregnancy/lactation cycle: Implications in breast cancer. *IUBMB life*. 2015;67(4):227-38.
240. Lim HC, Mulhaupt HA, Couchman JR. Cell surface heparan sulfate proteoglycans control adhesion and invasion of breast carcinoma cells. *Molecular cancer*. 2015;14(1):15.
241. Hoskin V, Szeto A, Ghaffari A, Greer PA, Cote GP, Elliott BE. Ezrin regulates focal adhesion and invadopodia dynamics by altering calpain activity to promote breast cancer cell invasion. *Molecular biology of the cell*. 2015.
242. Kowalewski JM, Shafqat-Abbasi H, Jafari-Mamaghani M, Endrias Ganebo B, Gong X, Stromblad S, et al. Disentangling Membrane Dynamics and Cell Migration; Differential Influences of F-actin and Cell-Matrix Adhesions. *PloS one*. 2015;10(8):e0135204.
243. Lal S, Kersch C, Beeson KA, Wu YJ, Muldoon LL, Neuwelt EA. Interactions between alphav-Integrin and HER2 and Their Role in the Invasive Phenotype of Breast Cancer Cells In Vitro and in Rat Brain. *PloS one*. 2015;10(7):e0131842.
244. Tang J, Ahmad A, Sarkar FH. The role of microRNAs in breast cancer migration, invasion and metastasis. *International journal of molecular sciences*. 2012;13(10):13414-37.
245. Zhang P, Bai H, Liu G, Wang H, Chen F, Zhang B, et al. MicroRNA-33b, upregulated by EF24, a curcumin analog, suppresses the epithelial-to-mesenchymal transition (EMT) and migratory potential of melanoma cells by targeting HMGA2. *Toxicology letters*. 2015;234(3):151-61.
246. Zibara K, Awada Z, Dib L, El-Saghir J, Al-Ghadban S, Ibrik A, et al. Anti-angiogenesis therapy and gap junction inhibition reduce MDA-MB-231 breast cancer cell invasion and metastasis in vitro and in vivo. *Scientific reports*. 2015;5:12598.
247. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. *Oncogene*. 2007;26(22):3279-90.
248. Golfopoulos V, Pentheroudakis G, Goussia A, Siozopoulou V, Bobos M, Krikelis D, et al. Intracellular signalling via the AKT axis and downstream effectors is active and prognostically significant in cancer of unknown primary (CUP): a study of 100 CUP cases. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012;23(10):2725-30.
249. Roberts PJ, Der CJ. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*. 2007;26(22):3291-310.
250. Nakareangrit W, Thiantanawat A, Visitnonthachai D, Watcharasit P, Satayavivad J. Sodium arsenite inhibited genomic estrogen signaling but induced pERalpha (Ser118) via MAPK pathway in breast cancer cells. *Environmental toxicology*. 2015.
251. Han S, Cao C, Tang T, Lu C, Xu J, Wang S, et al. ROBO3 promotes growth and metastasis of pancreatic carcinoma. *Cancer letters*. 2015;366(1):61-70.

252. Lee YH, Kim JH, Song GG. Genome-wide pathway analysis of breast cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2014;35(8):7699-705.
253. Adams RH, Eichmann A. Axon guidance molecules in vascular patterning. *Cold Spring Harbor perspectives in biology*. 2010;2(5):a001875.
254. Chedotal A. Chemotropic axon guidance molecules in tumorigenesis. *Progress in experimental tumor research*. 2007;39:78-90.
255. ElHefnawi M, Soliman B, Abu-Shahba N, Amer M. An integrative meta-analysis of microRNAs in hepatocellular carcinoma. *Genomics, proteomics & bioinformatics*. 2013;11(6):354-67.
256. Iyer AN, Bellon A, Baudet ML. microRNAs in axon guidance. *Frontiers in cellular neuroscience*. 2014;8:78.
257. Bencsik N, Sziber Z, Liliom H, Tarnok K, Borbely S, Gulyas M, et al. Protein kinase D promotes plasticity-induced F-actin stabilization in dendritic spines and regulates memory formation. *The Journal of cell biology*. 2015;210(5):771-83.
258. Cargnello M, Roux PP. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiology and molecular biology reviews : MMBR*. 2011;75(1):50-83.
259. Pearson LL, Castle BE, Kehry MR. CD40-mediated signaling in monocytic cells: up-regulation of tumor necrosis factor receptor-associated factor mRNAs and activation of mitogen-activated protein kinase signaling pathways. *International immunology*. 2001;13(3):273-83.
260. Downward J. Targeting RAS signalling pathways in cancer therapy. *Nature reviews Cancer*. 2003;3(1):11-22.
261. Goodsell DS. The molecular perspective: the ras oncogene. *The oncologist*. 1999;4(3):263-4.
262. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*. 1997;275(5308):1943-7.
263. Stambolic V, Suzuki A, de la Pompa JL, Brothers GM, Mirtsos C, Sasaki T, et al. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*. 1998;95(1):29-39.
264. Simpson L, Parsons R. PTEN: life as a tumor suppressor. *Experimental cell research*. 2001;264(1):29-41.
265. Voudouri K, Berdiaki A, Tzardi M, Tzanakakis GN, Nikitovic D. Insulin-like growth factor and epidermal growth factor signaling in breast cancer cell growth: focus on endocrine resistant disease. *Analytical cellular pathology*. 2015;2015:975495.
266. Geraghty RJ, Capes-Davis A, Davis JM, Downward J, Freshney RI, Knezevic I, et al. Guidelines for the use of cell lines in biomedical research. *British journal of cancer*. 2014;111(6):1021-46.

267. Whang-Peng J, Lee EC, Kao-Shan CS, Seibert K, Lippman M. Cytogenetic studies of human breast cancer lines: MCF-7 and derived variant sublines. *Journal of the National Cancer Institute*. 1983;71(4):687-95.
268. Maltseva DV, Galatenko VV, Samatov TR, Zhikrivetskaya SO, Khaustova NA, Nechaev IN, et al. miRNome of inflammatory breast cancer. *BMC research notes*. 2014;7:871.
269. Jukic DM, Rao UN, Kelly L, Skaf JS, Drogowski LM, Kirkwood JM, et al. MicroRNA profiling analysis of differences between the melanoma of young adults and older adults. *Journal of translational medicine*. 2010;8:27.
270. Han B, Xie R, Wu S, Li L, Zhu L. A case study on the identification of confounding factors for gene disease association analysis. *Cancer biomarkers : section A of Disease markers*. 2015.
271. Dobson JR, Taipaleenmaki H, Hu YJ, Hong D, van Wijnen AJ, Stein JL, et al. hsa-mir-30c promotes the invasive phenotype of metastatic breast cancer cells by targeting NOV/CCN3. *Cancer cell international*. 2014;14:73.
272. Lavon I, Zrihan D, Granit A, Einstein O, Fainstein N, Cohen MA, et al. Gliomas display a microRNA expression profile reminiscent of neural precursor cells. *Neuro-oncology*. 2010;12(5):422-33.
273. Wu W, Zhang X, Liao Y, Zhang W, Cheng H, Deng Z, et al. miR-30c negatively regulates the migration and invasion by targeting the immediate early response protein 2 in SMMC-7721 and HepG2 cells. *American journal of cancer research*. 2015;5(4):1435-46.
274. Li J, Donath S, Li Y, Qin D, Prabhakar BS, Li P. miR-30 regulates mitochondrial fission through targeting p53 and the dynamin-related protein-1 pathway. *PLoS genetics*. 2010;6(1):e1000795.
275. Yan B, Zhao JL. miR-1228 prevents cellular apoptosis through targeting of MOAP1 protein. *Apoptosis : an international journal on programmed cell death*. 2012;17(7):717-24.
276. Beauchemin D, Lacombe C, Van Themsche C. PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype. *Molecular cancer*. 2011;10:148.
277. Kovacevic Z, Fu D, Richardson DR. The iron-regulated metastasis suppressor, NdrG-1: identification of novel molecular targets. *Biochimica et biophysica acta*. 2008;1783(10):1981-92.
278. Tajnik M, Strazisar M, Volavsek M, Bostjancic E, Glavac D. BBC3 is down-regulated with increased tumor size independently of p53 expression in head and neck cancer. *Cancer biomarkers : section A of Disease markers*. 2012;11(5):197-208.
279. Wong FY, Liem N, Xie C, Yan FL, Wong WC, Wang L, et al. Combination therapy with gossypol reveals synergism against gemcitabine resistance in cancer cells with high BCL-2 expression. *PloS one*. 2012;7(12):e50786.
280. Castano Z, Marsh T, Tadipatri R, Kuznetsov HS, Al-Shahrour F, Paktinat M, et al. Stromal EGF and igfI together modulate plasticity of disseminated triple-negative breast tumors. *Cancer discovery*. 2013;3(8):922-35.

281. Awsare NS, Martin TA, Haynes MD, Matthews PN, Jiang WG. Claudin-11 decreases the invasiveness of bladder cancer cells. *Oncology reports*. 2011;25(6):1503-9.
282. Webb PG, Spillman MA, Baumgartner HK. Claudins play a role in normal and tumor cell motility. *BMC cell biology*. 2013;14:19.
283. Ozata DM, Caramuta S, Velazquez-Fernandez D, Akcakaya P, Xie H, Hoog A, et al. The role of microRNA deregulation in the pathogenesis of adrenocortical carcinoma. *Endocrine-related cancer*. 2011;18(6):643-55.
284. Lian ZQ, Wang Q, Li WP, Zhang AQ, Wu L. Screening of significantly hypermethylated genes in breast cancer using microarray-based methylated-CpG island recovery assay and identification of their expression levels. *International journal of oncology*. 2012;41(2):629-38.
285. Katsushima K, Shinjo K, Natsume A, Ohka F, Fujii M, Osada H, et al. Contribution of microRNA-1275 to Claudin11 protein suppression via a polycomb-mediated silencing mechanism in human glioma stem-like cells. *The Journal of biological chemistry*. 2012;287(33):27396-406.
286. Guo H, Hu X, Ge S, Qian G, Zhang J. Regulation of RAP1B by miR-139 suppresses human colorectal carcinoma cell proliferation. *The international journal of biochemistry & cell biology*. 2012;44(9):1465-72.
287. Milde-Langosch K, Karn T, Muller V, Witzel I, Rody A, Schmidt M, et al. Validity of the proliferation markers Ki67, TOP2A, and RacGAP1 in molecular subgroups of breast cancer. *Breast cancer research and treatment*. 2013;137(1):57-67.
288. Shen K, Liang Q, Xu K, Cui D, Jiang L, Yin P, et al. MiR-139 inhibits invasion and metastasis of colorectal cancer by targeting the type I insulin-like growth factor receptor. *Biochemical pharmacology*. 2012;84(3):320-30.
289. Bao W, Fu HJ, Xie QS, Wang L, Zhang R, Guo ZY, et al. HER2 interacts with CD44 to up-regulate CXCR4 via epigenetic silencing of microRNA-139 in gastric cancer cells. *Gastroenterology*. 2011;141(6):2076-87 e6.
290. Leidner RS, Li L, Thompson CL. Dampening enthusiasm for circulating microRNA in breast cancer. *PloS one*. 2013;8(3):e57841.
291. Qian YW, Chen Y, Yang W, Fu J, Cao J, Ren YB, et al. p28(GANK) prevents degradation of Oct4 and promotes expansion of tumor-initiating cells in hepatocarcinogenesis. *Gastroenterology*. 2012;142(7):1547-58 e14.
292. Shen J, Ambrosone CB, Zhao H. Novel genetic variants in microRNA genes and familial breast cancer. *International journal of cancer Journal international du cancer*. 2009;124(5):1178-82.
293. Wang S, Huang J, Lyu H, Lee CK, Tan J, Wang J, et al. Functional cooperation of miR-125a, miR-125b, and miR-205 in entinostat-induced downregulation of erbB2/erbB3 and apoptosis in breast cancer cells. *Cell death & disease*. 2013;4:e556.
294. Zhang L, Dong Y, Zhu N, Tsoi H, Zhao Z, Wu CW, et al. microRNA-139-5p exerts tumor suppressor function by targeting NOTCH1 in colorectal cancer. *Molecular cancer*. 2014;13:124.

295. Roy S, Benz F, Vargas Cardenas D, Vucur M, Gautheron J, Schneider A, et al. miR-30c and miR-193 are a part of the TGF-beta dependent regulatory network controlling extracellular matrix genes in liver fibrosis. *Journal of digestive diseases*. 2015.
296. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer research*. 2004;64(9):3087-95.
297. Sage J, Ventura A. miR than meets the eye. *Genes & development*. 2011;25(16):1663-7.
298. Dews M, Homayouni A, Yu D, Murphy D, Sevignani C, Wentzel E, et al. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nature genetics*. 2006;38(9):1060-5.
299. Schulte JH, Schowe B, Mestdagh P, Kaderali L, Kalaghatgi P, Schlierf S, et al. Accurate prediction of neuroblastoma outcome based on miRNA expression profiles. *International journal of cancer Journal international du cancer*. 2010;127(10):2374-85.
300. Morimura R, Komatsu S, Ichikawa D, Takeshita H, Tsujiura M, Nagata H, et al. Novel diagnostic value of circulating miR-18a in plasma of patients with pancreatic cancer. *British journal of cancer*. 2011;105(11):1733-40.
301. Szafranska AE, Davison TS, John J, Cannon T, Sipos B, Maghnouj A, et al. MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. *Oncogene*. 2007;26(30):4442-52.
302. Osada H, Takahashi T. let-7 and miR-17-92: small-sized major players in lung cancer development. *Cancer science*. 2011;102(1):9-17.
303. Chang YY, Kuo WH, Hung JH, Lee CY, Lee YH, Chang YC, et al. Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing. *Molecular cancer*. 2015;14:36.
304. Fan M, Sethuraman A, Brown M, Sun W, Pfeiffer LM. Systematic analysis of metastasis-associated genes identifies miR-17-5p as a metastatic suppressor of basal-like breast cancer. *Breast cancer research and treatment*. 2014;146(3):487-502.
305. Farazi TA, Horlings HM, Ten Hoeve JJ, Mihailovic A, Halfwerk H, Morozov P, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer research*. 2011;71(13):4443-53.
306. Yu JX, Sieuwerts AM, Zhang Y, Martens JW, Smid M, Klijn JG, et al. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC cancer*. 2007;7:182.
307. Riaz M, van Jaarsveld MT, Hollestelle A, Prager-van der Smissen WJ, Heine AA, Boersma AW, et al. miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast cancer research : BCR*. 2013;15(2):R33.
308. Jia AY, Castillo-Martin M, Domingo-Domenech J, Bonal DM, Sanchez-Carbayo M, Silva JM, et al. A common MicroRNA signature consisting of miR-133a, miR-139-3p, and miR-142-3p clusters bladder carcinoma in situ with normal umbrella cells. *The American journal of pathology*. 2013;182(4):1171-9.
309. Liu R, Yang M, Meng Y, Liao J, Sheng J, Pu Y, et al. Tumor-suppressive function of miR-139-5p in esophageal squamous cell carcinoma. *PloS one*. 2013;8(10):e77068.



310. Miles GD, Seiler M, Rodriguez L, Rajagopal G, Bhanot G. Identifying microRNA/mRNA dysregulations in ovarian cancer. *BMC research notes*. 2012;5:164.
311. Krishnan K, Steptoe AL, Martin HC, Pattabiraman DR, Nones K, Waddell N, et al. miR-139-5p is a regulator of metastatic pathways in breast cancer. *Rna*. 2013;19(12):1767-80.
312. Guo X, Wu Y, Hartley RS. MicroRNA-125a represses cell growth by targeting HuR in breast cancer. *RNA biology*. 2009;6(5):575-83.
313. Lyu H, Huang J, Edgerton SM, Thor AD, He Z, Liu B. Increased erbB3 promotes erbB2/neu-driven mammary tumor proliferation and co-targeting of erbB2/erbB3 receptors exhibits potent inhibitory effects on breast cancer cells. *International journal of clinical and experimental pathology*. 2015;8(6):6143-56.
314. Fassan M, Pizzi M, Realdon S, Balistreri M, Guzzardo V, Zagonel V, et al. The HER2-miR125a5p/miR125b loop in gastric and esophageal carcinogenesis. *Human pathology*. 2013;44(9):1804-10.
315. Romanuik TL, Ueda T, Le N, Haile S, Yong TM, Thomson T, et al. Novel biomarkers for prostate cancer including noncoding transcripts. *The American journal of pathology*. 2009;175(6):2264-76.
316. Teh JL, Chen S. Glutamatergic signaling in cellular transformation. *Pigment cell & melanoma research*. 2012;25(3):331-42.
317. Marsden KC, Beattie JB, Friedenthal J, Carroll RC. NMDA receptor activation potentiates inhibitory transmission through GABA receptor-associated protein-dependent exocytosis of GABA(A) receptors. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2007;27(52):14326-37.
318. Takasu MA, Dalva MB, Zigmond RE, Greenberg ME. Modulation of NMDA receptor-dependent calcium influx and gene expression through EphB receptors. *Science*. 2002;295(5554):491-5.
319. Kamal T, Green TN, Morel-Kopp MC, Ward CM, McGregor AL, McGlashan SR, et al. Inhibition of glutamate regulated calcium entry into leukemic megakaryoblasts reduces cell proliferation and supports differentiation. *Cellular signalling*. 2015;27(9):1860-72.
320. Kunz M. Oncogenes in melanoma: an update. *European journal of cell biology*. 2014;93(1-2):1-10.
321. Teh J, Chen S. mGlu Receptors and Cancerous Growth. *Wiley interdisciplinary reviews Membrane transport and signaling*. 2012;1(2):211-20.
322. Prickett TD, Samuels Y. Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(16):4240-6.
323. Li L, Hanahan D. Hijacking the neuronal NMDAR signaling circuit to promote tumor growth and invasion. *Cell*. 2013;153(1):86-100.
324. Li B, Otsu Y, Murphy TH, Raymond LA. Developmental decrease in NMDA receptor desensitization associated with shift to synapse and interaction with postsynaptic density-95. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2003;23(35):11244-54.

325. Chatterton JE, Awobuluyi M, Premkumar LS, Takahashi H, Talantova M, Shin Y, et al. Excitatory glycine receptors containing the NR3 family of NMDA receptor subunits. *Nature*. 2002;415(6873):793-8.
326. Cull-Candy S, Brickley S, Farrant M. NMDA receptor subunits: diversity, development and disease. *Current opinion in neurobiology*. 2001;11(3):327-35.
327. Li JH, Wang YH, Wolfe BB, Krueger KE, Corsi L, Stocca G, et al. Developmental changes in localization of NMDA receptor subunits in primary cultures of cortical neurons. *The European journal of neuroscience*. 1998;10(5):1704-15.
328. Murugan AK, Yang C, Xing M. Mutational analysis of the GNA11, MMP27, FGD1, TRRAP and GRM3 genes in thyroid cancer. *Oncology letters*. 2013;6(2):437-41.
329. Chattopadhyay I, Phukan R, Singh A, Vasudevan M, Purkayastha J, Hewitt S, et al. Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering. *Oncology reports*. 2009;21(5):1135-46.
330. Prickett TD, Wei X, Cardenas-Navia I, Teer JK, Lin JC, Walia V, et al. Exon capture analysis of G protein-coupled receptors identifies activating mutations in GRM3 in melanoma. *Nature genetics*. 2011;43(11):1119-26.
331. Zhou K, Song Y, Zhou W, Zhang C, Shu H, Yang H, et al. mGlu3 receptor blockade inhibits proliferation and promotes astrocytic phenotype in glioma stem cells. *Cell biology international*. 2014;38(4):426-34.
332. Melchiorri D, Cappuccio I, Ciceroni C, Spinsanti P, Mosillo P, Sarichelou I, et al. Metabotropic glutamate receptors in stem/progenitor cells. *Neuropharmacology*. 2007;53(4):473-80.
333. Lowdon RF, Zhang B, Bilenky M, Mauro T, Li D, Gascard P, et al. Regulatory network decoded from epigenomes of surface ectoderm-derived cell types. *Nature communications*. 2014;5:5442.
334. Moasser MM, Krop IE. The Evolving Landscape of HER2 Targeting in Breast Cancer. *JAMA oncology*. 2015.
335. Fujiwara S, Ibusuki M, Yamamoto S, Yamamoto Y, Iwase H. Association of ErbB1-4 expression in invasive breast cancer with clinicopathological characteristics and prognosis. *Breast cancer*. 2014;21(4):472-81.
336. Jones FE. HER4 intracellular domain (4ICD) activity in the developing mammary gland and breast cancer. *Journal of mammary gland biology and neoplasia*. 2008;13(2):247-58.
337. Olsen DA, Ostergaard B, Bokmand S, Wamberg PA, Jakobsen EH, Jakobsen A, et al. HER1-4 protein concentrations in normal breast tissue from breast cancer patients are expressed by the same profile as in the malignant tissue. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2009;47(8):977-84.
338. Sans N, Petralia RS, Wang YX, Blahos J, 2nd, Hell JW, Wenthold RJ. A developmental change in NMDA receptor-associated proteins at hippocampal synapses. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2000;20(3):1260-71.
339. Kim S. A study of microRNAs in silico and in vivo. *The FEBS journal*. 2009;276(8):2139.

340. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120(1):15-20.
341. Lodge AJ, Anderson JJ, Gullick WJ, Haugk B, Leonard RC, Angus B. Type 1 growth factor receptor expression in node positive breast cancer: adverse prognostic significance of c-erbB-4. *Journal of clinical pathology*. 2003;56(4):300-4.
342. Naresh A, Long W, Vidal GA, Wimley WC, Marrero L, Sartor CI, et al. The ERBB4/HER4 intracellular domain 4ICD is a BH3-only protein promoting apoptosis of breast cancer cells. *Cancer research*. 2006;66(12):6412-20.
343. Sartor CI, Zhou H, Kozłowska E, Guttridge K, Kawata E, Caskey L, et al. Her4 mediates ligand-dependent antiproliferative and differentiation responses in human breast cancer cells. *Molecular and cellular biology*. 2001;21(13):4265-75.
344. Suo Z, Risberg B, Kalsson MG, Willman K, Tierens A, Skovlund E, et al. EGFR family expression in breast carcinomas. c-erbB-2 and c-erbB-4 receptors have different effects on survival. *The Journal of pathology*. 2002;196(1):17-25.
345. Witton CJ. Structure of HER receptors and intracellular localisation of downstream effector elements gives insight into mechanism of tumour growth promotion. *Breast cancer research : BCR*. 2003;5(4):206-7.
346. Witton CJ, Reeves JR, Going JJ, Cooke TG, Bartlett JM. Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. *The Journal of pathology*. 2003;200(3):290-7.
347. Portier BP, Minca EC, Wang Z, Lanigan C, Gruver AM, Downs-Kelly E, et al. HER4 expression status correlates with improved outcome in both neoadjuvant and adjuvant Trastuzumab treated invasive breast carcinoma. *Oncotarget*. 2013;4(10):1662-72.
348. Silva SD, Alaoui-Jamali MA, Hier M, Soares FA, Graner E, Kowalski LP. Cooverexpression of ERBB1 and ERBB4 receptors predicts poor clinical outcome in pN+ oral squamous cell carcinoma with extranodal spread. *Clinical & experimental metastasis*. 2014;31(3):307-16.
349. Gallo RM, Bryant IN, Mill CP, Kaverman S, Riese DJ, 2nd. Multiple Functional Motifs Are Required for the Tumor Suppressor Activity of a Constitutively-Active ErbB4 Mutant. *Journal of cancer research and therapeutic oncology*. 2013;1(1):10.
350. Muraoka-Cook RS, Sandahl M, Husted C, Hunter D, Miraglia L, Feng SM, et al. The intracellular domain of ErbB4 induces differentiation of mammary epithelial cells. *Molecular biology of the cell*. 2006;17(9):4118-29.
351. Wali VB, Gilmore-Hebert M, Mamillapalli R, Haskins JW, Kurppa KJ, Elenius K, et al. Overexpression of ERBB4 JM-a CYT-1 and CYT-2 isoforms in transgenic mice reveals isoform-specific roles in mammary gland development and carcinogenesis. *Breast cancer research : BCR*. 2014;16(6):501.
352. Sundvall M, Peri L, Maatta JA, Tvorogov D, Paatero I, Savisalo M, et al. Differential nuclear localization and kinase activity of alternative ErbB4 intracellular domains. *Oncogene*. 2007;26(48):6905-14.
353. Muraoka-Cook RS, Sandahl MA, Strunk KE, Miraglia LC, Husted C, Hunter DM, et al. ErbB4 splice variants Cyt1 and Cyt2 differ by 16 amino acids and exert opposing effects on the mammary epithelium in vivo. *Molecular and cellular biology*. 2009;29(18):4935-48.

354. Yamamoto H, Higa-Nakamine S, Noguchi N, Maeda N, Kondo Y, Toku S, et al. Desensitization by different strategies of epidermal growth factor receptor and ErbB4. *Journal of pharmacological sciences*. 2014;124(3):287-93.
355. Chua YL, Ito Y, Pole JC, Newman S, Chin SF, Stein RC, et al. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. *Oncogene*. 2009;28(46):4041-52.
356. Das PM, Thor AD, Edgerton SM, Barry SK, Chen DF, Jones FE. Reactivation of epigenetically silenced HER4/ERBB4 results in apoptosis of breast tumor cells. *Oncogene*. 2010;29(37):5214-9.
357. Park SY, Kwon HJ, Choi Y, Lee HE, Kim SW, Kim JH, et al. Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2012;25(2):185-96.
358. Yu LJ, Wall BA, Chen S. The current management of brain metastasis in melanoma: a focus on riluzole. *Expert review of neurotherapeutics*. 2015;15(7):779-92.
359. Qiu C, Tarrant MK, Choi SH, Sathyamurthy A, Bose R, Banjade S, et al. Mechanism of activation and inhibition of the HER4/ErbB4 kinase. *Structure*. 2008;16(3):460-7.
360. Carlson RW, Allred DC, Anderson BO, Burstein HJ, Carter WB, Edge SB, et al. Breast cancer. Clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2009;7(2):122-92.
361. Feldinger K, Kong A. Profile of neratinib and its potential in the treatment of breast cancer. *Breast Cancer (Dove Med Press)*. 2015;7:147-62.

# ANEXOS

---



## ANEXO I: Características de las sondas empleadas en el estudio.

Código NCBI	ID*	IDs previos	ID del ensayo	Secuencia/Tamaño del amplicón
<b>RNAs de pequeño tamaño (miRNAs)</b>				
MIMAT0000065	hsa-let-7d-5p	hsa-let-7d	000380	AGAGGUAGUAGGUUGCAUAGUU
MIMAT0000076	hsa-miR-21-5p	hsa-miR-21	000397	UAGCUUAUCAGACUGAUGUUGA
MIMAT0000078	hsa-miR-23a-3p	hsa-miR-23a	000399	AUCACAUUGCCAGGGAUUUCC
MIMAT0000419	hsa-miR-27b-3p	hsa-miR-27b	000409	UUCACAGUGGCUAAGUUCUGC
MIMAT0000085	hsa-miR-28-5p	hsa-miR-28	000411	AAGGAGCUCACAGUCUAUUGAG
MIMAT0000244	hsa-miR-30c-5p	hsa-miR-30c	000419	UGUAAACAUCUACACUCUCAGC
MIMAT0000693	hsa-miR-30e-3p	hsa-miR-30e*	000422	CUUUCAGUCGGAUGUUUACAGC
MIMAT0003658	hsa-miR-92b-3p	hsa-miR-92b	007028_mat	UAUUGCACUCGUCCCGGCCUC
MIMAT0004792	hsa-miR-92b-5p	hsa-miR-92b*	002343	AGGGACGGGACCGGUGCAGUG
MIMAT0000443	hsa-miR-125a-5p	hsa-miR-125a	002198	UCCUGAGACCCUUAACCUUGUGA
MIMAT0000423	hsa-miR-125b-5p	hsa-miR-125b	000449	UCCUGAGACCCUAAUCUUGUGA
MIMAT0000426	hsa-miR-132-3p	hsa-miR-132	000457	UAACAGUCUACAGCCAUGGUGG
MIMAT0000250	hsa-miR-139-5p	hsa-miR-139	001096	UCUACAGUGCAGGUGUCUCCAGU
MIMAT0000450	hsa-miR-149-5p	hsa-miR-149	000472	UCUGGCUCCGUGUCUUCACUCCC
MIMAT0000440	hsa-miR-191-5p	hsa-miR-191	002299	CAACGGAAUCCCAAAGCAGCUG
MIMAT0000461	hsa-miR-195-5p	hsa-miR-195	000494	UAGCAGCACAGAAUAUUGGC
MIMAT0000733	hsa-miR-379-5p	hsa-miR-379	000568	UGGUAGACUAUGGAACGUAGG
MIMAT0001639	hsa-miR-409-3p	hsa-miR-409	002332	GAAUGUUGCUCCGUGAACCCCU
MIMAT0001627	hsa-miR-433	hsa-miR-433	001028	AUCAUGAUGGGCUCCUCGGUGU
MIMAT0005871	hsa-miR-1207-5p	hsa-miR-1207	241060_mat	UGGCAGGGAGGCUCCGAGGGG
MIMAT0005929	hsa-miR-1228-5p	hsa-miR-1228*	002763	GUGGGGGAGAGGCUCUGC
MIMAT0005582	hsa-miR-1275	hsa-miR-1275	002840	GUGGGCGGGGCAGGUGUGUG
MIMAT0015080	hsa-miR-3196	hsa-miR-3196	241941_mat	CGGGGCGGCAGGGGCCUC
<b>RNAs de pequeño tamaño nucleolares (snoRNAs)</b>				
NR_002752	RNU6B	U6B	001093	CGCAAGGATGACACGCAAATT CGTGAAGCGTTCCATATTTTT
NR_002439	RNU43	U43	001095	GAACTTATTGACGGGCGGACAGAAAC TGTGTGCTGATTGTACAGTTCTGATT
<b>RNA mensajeros (mRNAs)</b>				
NM_133445	GRIN3A	<i>NR3A, GluN3A</i>	Hs00370290_m1	106 nt
NM_001042599	<i>ERBB4</i>	<i>HER4, ALS19</i>	Hs00955525_m1	62 nt
NM_000840	<i>GRM3</i>	<i>GPRC1C, GLUR3</i>	Hs00168260_m1	114 nt
NM_000834	<i>GRIN2B</i>	<i>NR2B, GluN2B</i>	Hs00168230_m1	78 nt
NM_001256799	<i>GAPDH</i>	<i>GAPD, G3PD</i>	Hs99999905_m1	122 nt
NM_000194	<i>HPRT1</i>	<i>HGPRT</i>	Hs01003267_m1	72 nt

\*Nomenclatura miRBase v.19

## ANEXO II: Procedimiento del análisis de enriquecimiento de rutas de dianas validadas y teóricas de los miRNAs desregulados en mujeres jóvenes.

```
setwd("/home/jfores/Desktop/EscritorioOrdenYEquilibrio/MariaMicros/")
options(stringsAsFactors = FALSE)
```

### DIANAS VALIDADAS

Partiendo de los microRNAs diferencialmente expresados determinamos sus dianas validadas utilizando la base de datos miRTarBase.

```
MM <- read.table(file="identificadores miRBase_v21_OK.txt", sep="\t", header = T)
validated_targets <- read.csv(file="hsa_MTI (2).csv", sep="\t")
table(MM$miRBase_v21 %in% validated_targets$miRNA)
```

```
##
## FALSE TRUE
## 24 64
```

De los 88 miRNAs diferencialmente expresados encontramos 64 con dianas descritas en esta base de datos.

```
MM_filt <- MM[MM$miRBase_v21 %in% validated_targets$miRNA,]
validated_targets <- validated_targets[,c(2,4)]
colnames(validated_targets) <- c("ID", "TARGET")
validated_targets_filtered <- validated_targets[validated_targets$ID %in% MM$miRBase_v21,]
table(validated_targets_filtered[,2] == "")
```

```
##
## FALSE TRUE
## 6955 1
```

Filtramos una fila de la tabla de dianas en la que observamos que el miRNA no tiene diana asociada.

```
validated_targets_filtered <- validated_targets_filtered[!validated_targets_filtered[,2] == "",]
```

Obtenemos la lista de genes diana para efectuar enriquecimiento en categorías funcionales.

```
write.csv(file="ValidatedTargets.csv", unique(as.character(validated_targets_filtered[,2])), row.names = 1)
```

### ENRIQUECIMIENTO KEGG & REACTOME A PARTIR DE TODOS LOS GENES DIANA

Efectuamos el enriquecimiento en categorías funcionales de Kegg y Reactome utilizando Enrichr:

<http://amp.pharm.mssm.edu/Enrichr/enrich>

Una vez efectuado el análisis de enriquecimiento leemos los resultados de los mismos.

```
KeggEnrichValidated <- read.csv(file="KEGG_2015_Validated.csv", sep="\t")
KeggEnrichValidatedSig <- KeggEnrichValidated[KeggEnrichValidated$Adjusted.P.value < 0.05,]
KeggEnrichValidatedSig[,c(1,2,3,4)]
```



```
##           Term Overlap      P.value Adjusted.P.value
## 1           ribosome  45/71 0.0003224431      0.03290143
## 2           cell cycle 59/104 0.0003717676      0.03290143
## 3 chronic myeloid leukemia 45/75 0.0007658903      0.04518753
```

```
ReactomeEnrichValidated <- read.csv(file="Reactome_2015_Validated.csv",sep="\t")
head(ReactomeEnrichValidated$Adjusted.P.value)
```

```
## [1] "2.85095616893118E-005" "0.000132492"      "0.0001149386"
## [4] "0.000132492"      "0.0001149386"      "0.0001149386"
```

```
ReactomeEnrichValidatedSig <- ReactomeEnrichValidated[ReactomeEnrichValidated$Adjusted.P.value < 0.05,]
ReactomeEnrichValidatedSig[,c(1)]
```

```
ReactomeEnrichValidatedSig[,c(2,3,4)]
```

## CÁLCULO DEL EFECTO COMBINADO SOBRE CADA GEN.

Generamos el fichero con los recuentos del efecto combinado ejercido por los microRNAs sobre cada gen diana.

```
validated_splited <- split(validated_targets_filtered,as.character(validated_targets_filtered[,1]))
```

Hacemos el recuento del número de dianas que tiene cada microRNA.

```
dfcounts <- cbind(names(lapply(validated_splited,nrow)),as.numeric(lapply(validated_splited,nrow)))
colnames(dfcounts) <- c("miRBase_v21","Count")
mergedData <- merge(MM_filt,dfcounts,by="miRBase_v21")
mergedData$Count <- as.numeric(mergedData$Count)
mergedData$RelStrg <- mergedData$Fold_change/mergedData$Count
rownames(mergedData) <- mergedData[,1]
```

Visualizamos las primeras líneas de la tabla con el efecto relativo ejercido por cada microRNA sobre cada gen que será función de su foldchange y del número de dianas.

Creamos el vector de genes que utilizaremos para almacenar el efecto final sobre cada gen.

```
genes <- as.character(unique(validated_targets_filtered[,2]))
vecCount <- rep(0,length(genes))
names(vecCount) <- genes
vecCountPru <- vecCount
```

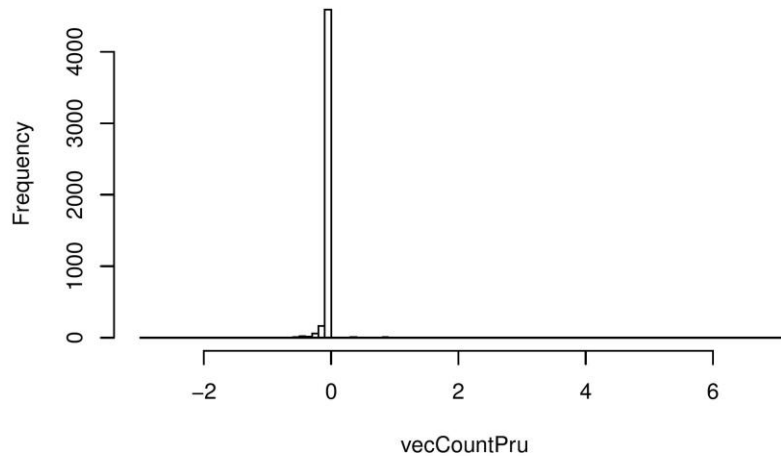
Calculamos el efecto sobre cada gen.

```
for(i in 1:length(validated_splited)){
  for (j in 1:nrow(validated_splited[[i]])){
    gen <- as.character(validated_splited[[i]][j,2])
    micro <- names(validated_splited)[i]
    effect <- mergedData[micro,5]
    #print("effect")
    #print(effect)
    #print("Previous value")
    #print(vecCountPru[gen])
    #print("Combined Value")
    #print(vecCountPru[gen] + effect)
    vecCountPru[gen] <- vecCountPru[gen] + effect
    #print("Final Value")
    #print(vecCountPru[gen])
  }
}
```

Mostramos la distribución de los valores de efecto. Y generamos el fichero de efectos definitivo. Probablemente será necesario modificar el anterior.

```
hist(vecCountPru,breaks = 100)
```

Histogram of vecCountPru



```
vecCountPru[vecCountPru > 0]
```

##	TCEAL1	TGIF2	BCL2L11	PTEN	RB1	E2F1	JUNB
##	0.7729721	1.5431765	0.8018298	1.5912964	2.4823874	1.3644442	0.8071498
##	JUND	AKT1	ARNTL	AMD1	HSPG2	CDK1	MYBL2
##	0.8071498	1.6040792	0.8648307	6.6328936	0.8071498	0.8044690	1.6029582
##	WASF2	GALK1	OSCP1	TMEM69	HIST2H3A	PDIA6	MPRIIP
##	1.2483592	2.4649302	4.6905260	1.9182460	1.8486618	3.2134395	3.2008271
##	COQ4	ATP6	COX2	POLR2I	FASN	GPI	WDR77
##	1.8038113	1.6660550	1.7884810	2.4387616	2.4096953	0.3020760	2.4664900
##	NUP62	TRAPPC1	ABI2	SH3BP2	AMPD2	KIAA1671	ACOX1
##	1.2499190	1.2467652	1.2336922	3.2930110	3.2930110	0.2502487	0.3831391
##	HIST1H2B0	HIST1H2AJ	NDE1	USP6NL	SRGAP1	DDT	CACNG8
##	0.3831391	0.3705267	7.1155666	0.3831391	0.3831391	0.3701114	0.3799852
##	DLST	SAMD4B	RNASEK	MTA2	HNRNPAO	TRAF4	
##	0.3159531	0.3831391	0.3831391	1.7632200	1.7506076	3.1209056	

```
dfsalida <- data.frame(names(vecCountPru),vecCountPru)
write.table(file = "recuentoValidated.csv",dfsalida,row.names = F,col.names = F,sep="\t",quote = F)
```

## ENRIQUECIMIENTO BASADO EN EL CONJUNTO DE GENES MÁS AFECTADO.

Realizamos de nuevo el enriquecimiento haciendo uso tan solo de los genes más afectados. Para ello dividimos en cuantiles y nos quedamos que los valores del superiores al cuantil 75%, considerando los valores absolutos del efecto.

```
quantile(abs(vecCountPru), probs = seq(0, 1, 0.25), na.rm = FALSE,names = TRUE, type = 7)
```

##	0%	25%	50%	75%	100%
##	0.001559780	0.002680808	0.012612357	0.031480237	7.115566590

```
quanfilt <- vecCountPru[vecCountPru > 0.031480237 | vecCountPru < -0.031480237]
dfsalida <- data.frame(names(quanfilt),quanfilt)
dim(dfsalida)
```

```
## [1] 1321 2
```

```
write.table(file = "recuentoValidatedQuantile.csv",dfsalida,row.names = F,col.names = F,sep="\t",quote :
```

Efectuamos el enriquecimiento para los genes de mayor efecto.

```
KeggEnrichValidatedQuantile <- read.csv(file="KEGG_2015_Validated_quantile.csv",sep="\t")
KeggEnrichValidatedQuantileSig <- KeggEnrichValidatedQuantile[KeggEnrichValidatedQuantile$Adjusted.P.va
KeggEnrichValidatedQuantileSig[,c(1,2,3,4)]
```

```
##          Term Overlap      P.value Adjusted.P.value
## 1      prostate cancer  28/86 1.400184e-06  0.0001375693
## 2          cell cycle  31/104 1.810122e-06  0.0001375693
## 3      pancreatic cancer  23/73 1.945300e-05  0.0009856191
## 4 chronic myeloid leukemia  23/75 2.782809e-05  0.0010574676
## 5          ribosome   22/71 3.652089e-05  0.0011102351
## 6 non small cell lung cancer  17/53 2.003630e-04  0.0050758631
## 7      small cell lung cancer  22/85 3.321702e-04  0.0061663382
## 8          glioma     18/62 3.651121e-04  0.0061663382
## 9      bladder cancer  14/40 3.537586e-04  0.0061663382
## 10 erbB signaling pathway  21/85 7.608790e-04  0.0115653608
## 11      focal adhesion  36/192 1.255793e-03  0.0146831211
## 12 renal cell carcinoma  18/69 1.059107e-03  0.0134153584
## 13 adherens junction  19/75 1.044094e-03  0.0134153584
## 14 colorectal cancer  20/84 1.487362e-03  0.0161484984
## 15 melanoma          17/71 3.145518e-03  0.0318745846
```

```
ReactomeEnrichValidatedQuantile <- read.csv(file="Reactome_2015_Validated_quantile.csv",sep="\t")
ReactomeEnrichValidatedQuantileSig <- ReactomeEnrichValidatedQuantile[ReactomeEnrichValidatedQuantile$A,
ReactomeEnrichValidatedQuantileSig[,c(1)]
```

Al efectuar el enriquecimiento de este modo obtenemos muchas mas enriquecimiento en categorías funcionales. Además el enriquecimiento observado es más coherente con los resultados previos.

## DIANAS TEÓRICAS

Partiendo de los microRNAs diferencialmente expresados obtenemos sus dianas teóricas. La información sobre las dianas proviene fueron computadas mediante el algoritmo de predicción de dianas microT y fueron descargadas desde:

<http://www.micorna.gr/microT-CDS>

El algoritmo genera una lista de dianas con un score de calidad asociado.

```
MM <- read.table(file="identificadores miRBase_v21_OK.txt",sep="\t",header = T)
predicted_targets <- read.csv(file="microT_CDS_dataFilt.csv",sep="\t",header = F)
colnames(predicted_targets) <- c("ID","TARGET","SCORE")
dim(predicted_targets)
```

```
## [1] 7337705      3
```

En primer lugar filtramos las dianas en función de su Score de calidad y nos quedamos con aquellas dianas que presenten un score mayor a 0.9.

```
predicted_targets_filtered <- predicted_targets[predicted_targets$SCORE > 0.9,]
dim(predicted_targets_filtered)
```

```
## [1] 306022      3
```

```
predicted_targets_filtered <- predicted_targets_filtered[predicted_targets_filtered$TARGET != "",]
table(MM$miRBase_v21 %in% predicted_targets_filtered$ID)
```

```
##
## FALSE TRUE
##      8   80
```

En el caso de las dianas teóricas encontramos dianas para 80 de los 88 miRNAs diferencialmente expresados. Nos quedamos solo con las dianas de los microRNAs diferencialmente expresados.

```
predicted_targets_filtered <- predicted_targets_filtered[predicted_targets_filtered$ID %in% MM$miRBase_v21,]
dim(predicted_targets_filtered)
```

```
## [1] 12558      3
```

```
MM_filt<- MM[MM$miRBase_v21 %in% predicted_targets_filtered$ID,]
rownames(MM_filt) <- MM_filt[,1]
```

Como en el caso de las dianas validadas en primer lugar efectuamos el análisis de enriquecimiento con todos los genes diana.

```
write.csv(file="PredictedTargets09.csv",unique(as.character(predicted_targets_filtered[,2])),row.names =
```

```
KeggEnrichPredicted <- read.csv(file="KEGG_2015_Predicted09.csv",sep="\t")
KeggEnrichPredictedSig <- KeggEnrichPredicted[KeggEnrichPredicted$Adjusted.P.value < 0.05,]
KeggEnrichPredictedSig[,c(1,2,3,4)]
```

```
##           Term Overlap      P.value Adjusted.P.value
## 1 axon guidance  82/126 6.734815e-06    0.001185327
## 2 mapk signaling pathway 126/248 2.215054e-04    0.019492474
## 3 long term potentiation 45/67 4.996209e-04    0.029311095
```

```
ReactomeEnrichPredicted <- read.csv(file="Reactome_2015_Predicted09.csv",sep="\t")
ReactomeEnrichPredictedSig <- ReactomeEnrichPredicted[ReactomeEnrichPredicted$Adjusted.P.value < 0.05,]
ReactomeEnrichPredictedSig[,c(1)]
```

```
## [1] "Axon guidance"      "Signalling by NGF"
```

```
ReactomeEnrichPredictedSig[,c(2,3,4)]
```

```
## Overlap      P.value Adjusted.P.value
## 1 187/327 6.643600e-08    9.075158e-05
## 2 154/287 1.446012e-05    9.876261e-03
```

Utilizando las dianas teóricas sin filtrar previamente encontramos muy poca señal biológicas. Muy poco enriquecimiento en categorías funcionales. Computamos el efecto de los micros sobre los genes de las dianas predichas.

```
predicted_targets_filtered <- predicted_targets_filtered[,c(1,2)]
predicted_splited <- split(predicted_targets_filtered,as.character(predicted_targets_filtered[,1]))
```

Efectuamos el recuento de los genes regulados por cada microRNA.

```
dfcounts <- cbind(names(lapply(predicted_splited,nrow)),as.numeric(lapply(predicted_splited,nrow)))
rownames(dfcounts) <- dfcounts[,1]
colnames(dfcounts) <- c("miRBase_v21","Count")
mergedData <- merge(MM_filt,dfcounts,by="miRBase_v21")
mergedData$Count <- as.numeric(mergedData$Count)
mergedData$RelStrg <- mergedData$Fold_change/mergedData$Count
rownames(mergedData) <- mergedData[,1]
```

```
genes <- as.character(unique(predicted_targets_filtered[,2]))
vecCount <- rep(0,length(genes))
names(vecCount) <- genes
vecCountPru <- vecCount
```

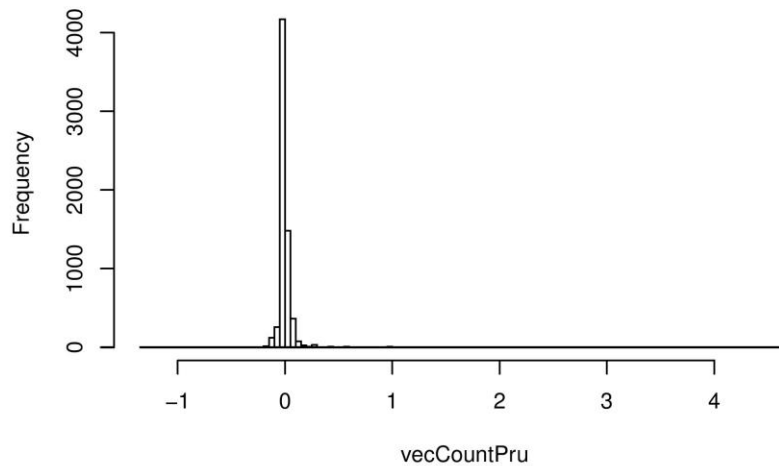
Efectuamos el computo del efecto.

```
for(i in 1:length(predicted_splited)){
  for (j in 1:nrow(predicted_splited[[i]])){
    gen <- as.character(predicted_splited[[i]][j,2])
    #print(gen)
    micro <- names(predicted_splited)[i]
    #print(micro)
    effect <- mergedData[micro,5]
    #print(effect)
    #print("effect")
    #print(effect)
    #print("Previous value")
    #print(vecCountPru[gen])
    #print("Combined Value")
    #print(vecCountPru[gen] + effect)
    vecCountPru[gen] <- vecCountPru[gen] + effect
    #print("Final Value")
    #print(vecCountPru[gen])
  }
}
```

```
hist(vecCountPru,breaks = 100)
```



Histogram of vecCountPru



```
dfsalida <- data.frame(names(vecCountPru),vecCountPru)
write.table(file = "recuentoPredicted.csv",dfsalida,row.names = F,col.names = F,sep="\t",quote = F)
```

Repetimos el enriquecimiento siguiendo el método de los quintiles.

```
quantile(abs(vecCountPru), probs = seq(0, 1, 0.25), na.rm = FALSE,names = TRUE, type = 7)
```

```
##          0%          25%          50%          75%          100%
## 3.085223e-05 9.661726e-03 1.838153e-02 3.396862e-02 4.690526e+00
```

```
quanfilt <- vecCountPru[vecCountPru > 3.396862e-02 | vecCountPru < -3.396862e-02]
dfsalida <- data.frame(names(quanfilt),quanfilt)
dim(dfsalida)
```

```
## [1] 1660  2
```

```
write.table(file = "recuentoPredictedQuantile.csv",dfsalida,row.names = F,col.names = F,sep="\t",quote :
```

Leemos las tablas finales.

```
KeggEnrichPredictedQuantile <- read.csv(file="KEGG_2015_Predicted09_quantile.csv",sep="\t")
KeggEnrichPredictedQuantileSig <- KeggEnrichPredictedQuantile[KeggEnrichPredictedQuantile$Adjusted.P.va
KeggEnrichPredictedQuantileSig[,c(1,2,3,4)]
```

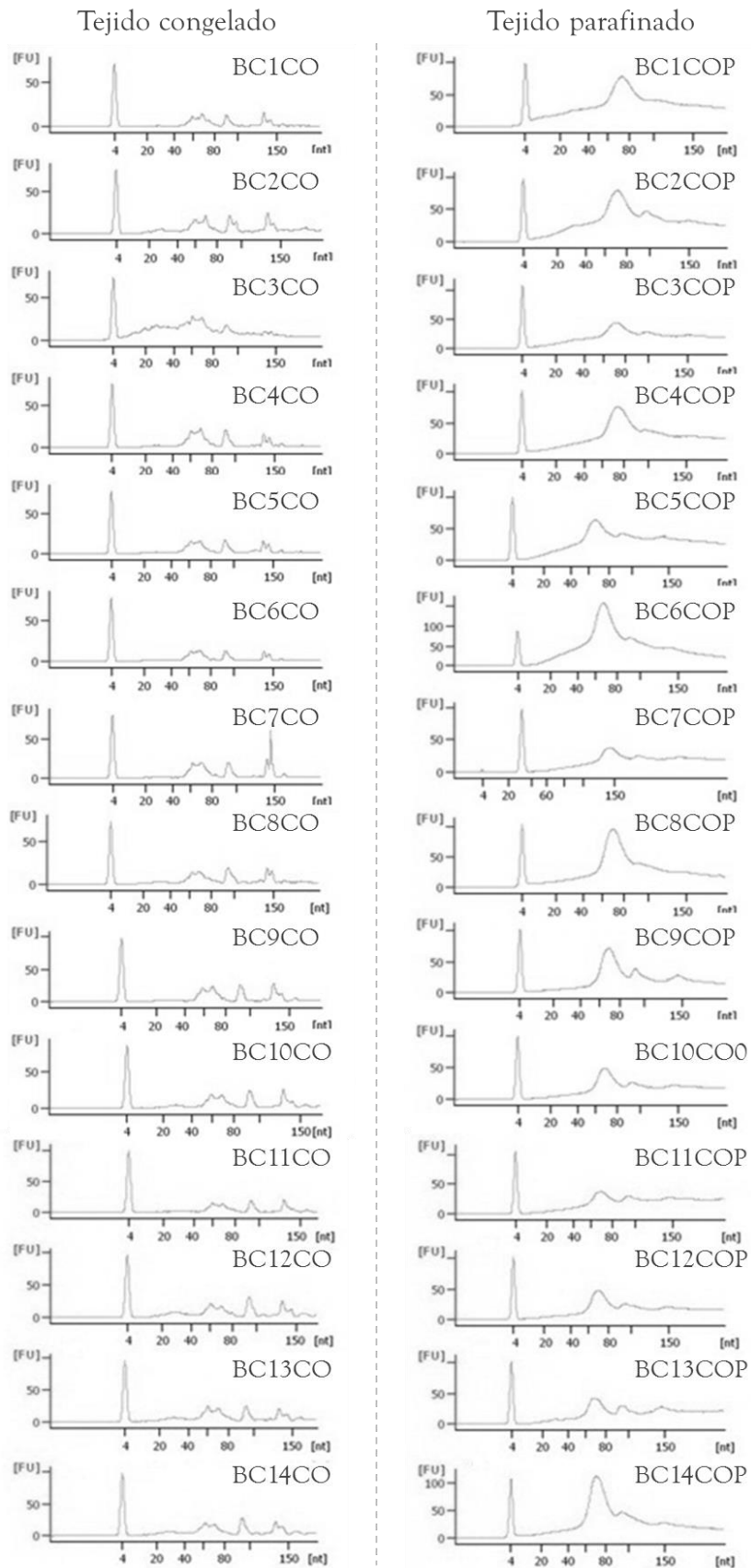
```
##          Term Overlap      P.value Adjusted.P.value
## 1 chronic myeloid leukemia 18/75 0.0001237370      0.01602986
## 2 mapk signaling pathway 37/248 0.0004361867      0.01602986
## 3 renal cell carcinoma 16/69 0.0004100940      0.01602986
## 4 long term potentiation 16/67 0.0003089662      0.01602986
```

```
ReactomeEnrichPredictedQuantile <- read.csv(file="Reactome_2015_Predicted09_quantile.csv",sep="\t")
ReactomeEnrichPredictedQuantileSig <- ReactomeEnrichPredictedQuantile[ReactomeEnrichPredictedQuantile$A
ReactomeEnrichPredictedQuantileSig[,c(1)]
```

```
## [1] "Signalling by NGF"
## [2] "NGF signalling via TRKA from the plasma membrane"
## [3] "Signaling by FGFR in disease"
## [4] "Signaling by FGFR"
## [5] "Signaling by PDGF"
## [6] "Downstream signal transduction"
## [7] "Downstream signaling of activated FGFR"
## [8] "Signaling by EGFRvIII in Cancer"
## [9] "Signaling by Overexpressed Wild-Type EGFR in Cancer"
## [10] "Signaling by EGFR"
## [11] "Signaling by Ligand-Responsive EGFR Variants in Cancer"
## [12] "Signaling by EGFR in Cancer"
## [13] "Transcriptional activation of mitochondrial biogenesis"
```

```
ReactomeEnrichPredictedQuantileSig[,c(2,3,4)]
```

**ANEXO III:** Perfil electroforético de las muestras descongeladas y de sus correspondientes parafinas de RNA de pequeño tamaño.



**ANEXO IV:** Diferencia del Ct medio de cada muestra congelada y su correspondiente muestra parafinada (DCt = Ct medio muestra parafinada - Ct medio muestra congelada).

Muestra	hsa-miR-21			hsa-miR-125b			hsa-miR-191		
	Ct medio		Diferencia	Ct medio		Diferencia	Ct medio		Diferencia
	CO	COP	COP-CO	CO	COP	COP-CO	CO	COP	COP-CO
BC1	18,04	18,86	0,819	18,37	18,44	0,069	17,89	17,84	-0,058
BC2	17,9	18,18	0,278	18,97	18,07	-0,897	17,96	16,87	-1,094
BC3	17,74	17,27	-0,466	18,84	18,67	-0,176	19,39	18,27	-1,126
BC4	17,68	18,08	0,404	17,62	17,99	0,368	17,99	17,69	-0,299
BC5	17,37	18,57	1,204	16,79	18,16	1,371	17,79	16,82	-0,967
BC6	17,85	16,92	-0,93	20,28	18,15	-2,125	18,93	17,21	-1,718
BC7	17,89	18,2	0,31	20,26	18,11	-2,147	18,39	17,87	-0,525
BC8	18,85	17,06	-1,79	21,86	18,19	-3,667	18,38	17,2	-1,175
BC9	18,04	16,87	-1,166	20,68	17,54	-3,137	18,17	17,06	-1,118
BC10	18,55	18,38	-0,175	20,21	17,98	-2,231	18,74	17,79	-0,945
BC11	18,24	18,3	0,058	20,66	17,83	-2,828	18,82	18,65	-0,167
BC12	18,32	19,42	1,091	19,82	18,14	-1,678	18,26	18,03	-0,229
BC13	18,28	19,46	1,181	18,87	17,61	-1,26	17,84	17,64	-0,209
BC14	17,94	17,8	-0,137	21,07	18,71	-2,36	19,52	18,08	-1,435
Muestra	<i>RNU6B</i>			<i>GAPDH</i>			<i>HPRT1</i>		
	Ct medio		Diferencia	Ct medio		Diferencia	Ct medio		Diferencia
	CO	COP	COP-CO	CO	COP	COP-CO	CO	COP	COP-CO
BC1	25,98	23,22	-2,753	22,25	35,14	12,897	27,29	36,67	9,378
BC2	24,9	22,32	-2,577	22,24	30,03	7,794	26,76	33,52	6,757
BC3	25,59	22,44	-3,145	22,42	29,6	7,18	26,71	32,21	5,501
BC4	26,25	23,03	-3,226	22,19	31,63	9,44	26,41	33,1	6,689
BC5	25,92	22,49	-3,439	23,14	30	6,854	28,67	33,29	4,615
BC6	26,8	22,28	-4,519	23,09	32,31	9,215	27,6	33,93	6,331
BC7	26,61	22,32	-4,284	22,89	29,51	6,613	27,54	32,14	4,597
BC8	26,75	22,15	-4,596	23,12	31,26	8,136	26,98	33,47	6,492
BC9	27,54	22,18	-5,36	23,41	28,01	4,6	27,54	31,01	3,466
BC10	26,72	22,77	-3,951	22,92	27,8	4,879	27,76	30,94	3,174
BC11	27,21	21,9	-5,312	24,04	30,05	6,002	27,45	31,83	4,374
BC12	26,22	23,08	-3,142	24,68	29,57	4,886	30,06	32,2	2,143
BC13	25,21	22,18	-3,026	23,69	26,59	2,899	28,5	29,77	1,268
BC14	25,8	22,72	-3,075	21,3	28,48	7,183	26,73	31,76	5,034

Ct: ciclo umbral; CO: congelado óptimo, CP: parafina.

**ANEXO V:** Tablas con los miRNAs con expresión diferente estadísticamente significativa en pacientes jóvenes con cáncer de mama.

miRNAs desregulados en el estudio de muestras de pacientes del Hospital Clínico de Valencia

miRNAs reprimidos				miRNAs sobreexpresados			
miRNA	p-valor	FDR	Expresión	miRNA	p-valor	FDR	Expresión
hsa-miR-132	4,99E-05	6,31E-04	-5,572	hsa-miR-3196	4,99E-05	6,31E-04	6,745
hsa-miR-125a-5p	4,99E-05	6,31E-04	-4,764	hsa-miR-762	4,99E-05	6,31E-04	6,646
hsa-miR-28-3p	4,99E-05	6,31E-04	-4,425	hsa-miR-939	4,99E-05	6,31E-04	6,586
hsa-miR-92b	2,50E-04	1,93E-03	-4,605	hsa-miR-1909	4,99E-05	6,31E-04	6,427
hsa-miR-23b	2,50E-04	1,93E-03	-4,541	hsa-miR-149*	4,99E-05	6,31E-04	6,423
hsa-miR-23a	2,50E-04	1,93E-03	-4,478	hsa-miR-1228*	4,99E-05	6,31E-04	6,016
hsa-miR-27b	3,00E-04	2,18E-03	-3,637	hsa-miR-1275	4,99E-05	6,31E-04	5,887
hsa-miR-30c	3,50E-04	2,47E-03	-3,424	hsa-miR-3197	4,99E-05	6,31E-04	5,768
hsa-miR-574-3p	4,00E-04	2,66E-03	-4,306	hsa-miR-1908	4,99E-05	6,31E-04	5,411
hsa-miR-205	4,49E-04	2,92E-03	-3,334	hsa-miR-3141	4,99E-05	6,31E-04	5,295
hsa-miR-26a	5,49E-04	3,47E-03	-3,771	hsa-miR-1202	4,99E-05	6,31E-04	5,226
hsa-miR-30a*	9,99E-04	5,85E-03	-3,776	hsa-miR-3162	4,99E-05	6,31E-04	5,121
hsa-miR-149	1,10E-03	5,99E-03	-4,026	hsa-miR-1268	4,99E-05	6,31E-04	5,013
hsa-miR-183	1,10E-03	5,99E-03	-3,907	hsa-miR-92b*	4,99E-05	6,31E-04	4,981
hsa-miR-148a	1,15E-03	6,13E-03	-3,884	hsa-miR-150*	4,99E-05	6,31E-04	4,841
hsa-miR-708	1,80E-03	9,18E-03	-3,492	hsa-miR-638	4,99E-05	6,31E-04	4,691
hsa-miR-28-5p	1,90E-03	9,49E-03	-3,137	hsa-miR-4299	9,99E-05	9,99E-04	5,776
hsa-miR-30e*	2,25E-03	1,04E-02	-3,490	hsa-miR-1225-5p	9,99E-05	9,99E-04	4,933
hsa-let-7b	2,25E-03	1,04E-02	-3,116	hsa-miR-3175	9,99E-05	9,99E-04	4,867
hsa-miR-181d	2,30E-03	1,04E-02	-3,500	hsa-miR-2861	9,99E-05	9,99E-04	4,355
hsa-miR-22	2,55E-03	1,11E-02	-2,446	hsa-miR-663	9,99E-05	9,99E-04	4,036
hsa-miR-500*	3,30E-03	1,41E-02	-3,287	hsa-miR-1224-5p	1,50E-04	1,44E-03	5,000
hsa-miR-455-3p	3,65E-03	1,53E-02	-3,159	hsa-miR-1207-5p	2,00E-04	1,78E-03	4,861
hsa-miR-195	4,84E-03	1,97E-02	-2,457	hsa-miR-1308	2,00E-04	1,78E-03	4,592
hsa-miR-139-5p	5,54E-03	2,22E-02	-3,131	hsa-miR-4281	2,50E-04	1,93E-03	4,608
hsa-miR-1180	5,74E-03	2,26E-02	-3,114	hsa-miR-665	3,00E-04	2,18E-03	4,471
hsa-miR-100	6,64E-03	2,57E-02	-2,801	hsa-miR-602	4,00E-04	2,66E-03	4,084
hsa-miR-152	7,49E-03	2,78E-02	-2,949	hsa-miR-1915	5,99E-04	3,69E-03	3,836
hsa-miR-29b-2*	7,54E-03	2,78E-02	-2,988	hsa-miR-378b	9,49E-04	5,69E-03	3,704
hsa-miR-361-5p	8,84E-03	3,12E-02	-2,862	hsa-miR-1973	1,10E-03	5,99E-03	3,948
hsa-miR-486-5p	9,09E-03	3,12E-02	-2,914	hsa-miR-4270	1,30E-03	6,77E-03	3,801
hsa-miR-181a-2*	9,29E-03	3,14E-02	-2,921	hsa-miR-494	1,95E-03	9,54E-03	3,459
hsa-miR-342-3p	1,12E-02	3,68E-02	-2,777	hsa-miR-4284	2,05E-03	9,83E-03	3,143
hsa-miR-29c	1,32E-02	4,22E-02	-2,740	hsa-miR-548a-3p	2,55E-03	1,11E-02	3,526
hsa-miR-339-5p	1,32E-02	4,22E-02	-2,729	hsa-miR-3180-3p	3,89E-03	1,61E-02	3,326
hsa-miR-500	1,35E-02	4,26E-02	-2,723	hsa-miR-1972	7,14E-03	2,72E-02	2,972
hsa-miR-502-3p	1,44E-02	4,50E-02	-2,721	hsa-miR-1469	7,84E-03	2,85E-02	2,771
hsa-miR-181a	1,50E-02	4,55E-02	-2,630	hsa-miR-3178	8,44E-03	3,02E-02	2,800
hsa-miR-26b	1,51E-02	4,55E-02	-2,561	hsa-miR-720	9,09E-03	3,12E-02	2,897
hsa-miR-422a	1,52E-02	4,55E-02	-2,686	hsa-miR-1246	9,69E-03	3,23E-02	2,739
hsa-miR-99b*	1,57E-02	4,59E-02	-2,672	hsa-miR-885-3p	1,54E-02	4,56E-02	2,641
hsa-miR-182	1,61E-02	4,59E-02	-2,508	hsa-miR-675	1,68E-02	4,59E-02	2,604
hsa-miR-489	1,62E-02	4,59E-02	-2,602	hsa-miR-1274b	1,72E-02	4,59E-02	2,567
hsa-miR-409-3p	1,65E-02	4,59E-02	-2,624				
hsa-miR-92a	1,71E-02	4,59E-02	-2,547				
hsa-miR-324-3p	1,71E-02	4,59E-02	-2,611				
hsa-miR-30c-2*	1,72E-02	4,59E-02	-2,564				
hsa-miR-99b	1,74E-02	4,60E-02	-2,594				



miRNAs desregulados en el estudio de líneas celulares de cáncer de mama procedentes de mujeres jóvenes

miRNAs reprimidos							
miRNA	p-valor	FDR p-valor	Expresión	miRNA	p-valor	FDR p-valor	Expresión
hsa-miR-99a	1,50E-04	1,28E-03	-10,902	hsa-miR-29	1,35E-03	5,35E-03	-4,127
hsa-miR-27a*	1,50E-04	1,28E-03	-9,746	hsa-miR-21	1,45E-03	5,59E-03	-4,169
hsa-miR-19b	1,50E-04	1,28E-03	-9,723	hsa-miR-43	1,70E-03	6,20E-03	-4,235
hsa-miR-19a	1,50E-04	1,28E-03	-9,704	hsa-miR-20	1,70E-03	6,20E-03	-4,212
hsa-let-7a	1,50E-04	1,28E-03	-7,355	hsa-miR-11	1,70E-03	6,20E-03	-4,176
hsa-miR-23a	1,50E-04	1,28E-03	-7,105	hsa-miR-27	2,25E-03	7,60E-03	-3,673
hsa-miR-1979	1,50E-04	1,28E-03	-7,067	hsa-miR-30	2,30E-03	7,68E-03	-3,178
hsa-miR-421	1,50E-04	1,28E-03	-6,737	hsa-miR-18	2,45E-03	8,08E-03	-4,000
hsa-miR-660	1,50E-04	1,28E-03	-6,625	hsa-miR-27	2,85E-03	8,67E-03	-3,786
hsa-let-7g	1,50E-04	1,28E-03	-6,244	hsa-miR-33	2,95E-03	8,84E-03	-3,801
hsa-miR-501-5p	1,50E-04	1,28E-03	-5,026	hsa-miR-12	3,05E-03	8,84E-03	-3,661
hsa-miR-128	1,50E-04	1,28E-03	-4,697	hsa-miR-23	3,10E-03	8,84E-03	-3,562
hsa-let-7f	1,50E-04	1,28E-03	-4,695	hsa-miR-7	3,10E-03	8,84E-03	-2,526
hsa-miR-301a	2,00E-04	1,44E-03	-6,306	hsa-miR-48	3,10E-03	8,84E-03	-2,510
hsa-miR-27b	2,00E-04	1,44E-03	-5,283	hsa-miR-19	3,79E-03	1,04E-02	-3,412
hsa-miR-15a	2,00E-04	1,44E-03	-4,856	hsa-miR-30	4,04E-03	1,08E-02	-3,510
hsa-miR-454	2,00E-04	1,44E-03	-3,378	hsa-miR-36	4,39E-03	1,13E-02	-3,433
hsa-miR-98	2,50E-04	1,59E-03	-5,076	hsa-miR-19	5,09E-03	1,25E-02	-3,237
hsa-miR-30b	2,50E-04	1,59E-03	-5,053	hsa-miR-10	5,59E-03	1,34E-02	-3,310
hsa-miR-320e	2,50E-04	1,59E-03	-4,773	hsa-miR-43	6,59E-03	1,54E-02	-3,211
hsa-miR-26b	2,50E-04	1,59E-03	-4,229	hsa-miR-29	8,24E-03	1,85E-02	-3,201
hsa-miR-320d	3,00E-04	1,82E-03	-4,487	hsa-miR-62	8,59E-03	1,91E-02	-3,136
hsa-miR-18b	4,00E-04	2,28E-03	-5,292	hsa-miR-94	1,00E-02	2,22E-02	-2,975
hsa-miR-891b	4,00E-04	2,28E-03	-3,941	hsa-miR-19	1,19E-02	2,56E-02	-2,853
hsa-miR-21	4,49E-04	2,51E-03	-4,660	hsa-miR-18	1,43E-02	2,97E-02	-2,676
hsa-miR-503	5,49E-04	2,74E-03	-5,019	hsa-let-7i	1,44E-02	2,97E-02	-2,653
hsa-miR-148b	1,10E-03	4,56E-03	-4,082	hsa-miR-12	1,69E-02	3,44E-02	-2,588
hsa-miR-92a-1*	1,20E-03	4,90E-03	-4,036	hsa-miR-19	1,69E-02	3,44E-02	-2,499
hsa-miR-424*	1,30E-03	5,23E-03	-4,279				

miRNAs sobreexpresados							
miRNA	p-valor	FDR p-valor	Expresión	miRNA	p-valor	FDR p-valor	Expresión
hsa-miR-31	1,50E-04	1,28E-03	47,978	hsa-miR-378	1,95E-03	6,84E-03	4,014
hsa-miR-4298	1,50E-04	1,28E-03	13,546	hsa-miR-1469	2,15E-03	7,45E-03	4,299
hsa-miR-4270	1,50E-04	1,28E-03	12,423	hsa-miR-2861	2,25E-03	7,60E-03	4,245
hsa-miR-1207-5p	1,50E-04	1,28E-03	11,542	hsa-miR-1228*	2,55E-03	8,31E-03	3,896
hsa-miR-155	1,50E-04	1,28E-03	10,376	hsa-miR-205	2,60E-03	8,37E-03	3,989
hsa-miR-574-5p	1,50E-04	1,28E-03	8,045	hsa-miR-1308	2,70E-03	8,59E-03	4,020
hsa-miR-4281	1,50E-04	1,28E-03	7,627	hsa-miR-130b	2,75E-03	8,65E-03	3,619
hsa-miR-3175	1,50E-04	1,28E-03	7,480	hsa-miR-150*	2,80E-03	8,67E-03	4,048
hsa-miR-3141	1,50E-04	1,28E-03	7,251	hsa-miR-1181	2,85E-03	8,67E-03	3,700
hsa-miR-744	1,50E-04	1,28E-03	6,863	hsa-miR-574-3	3,05E-03	8,84E-03	2,705
hsa-miR-28-5p	1,50E-04	1,28E-03	6,837	hsa-miR-663	3,50E-03	9,77E-03	4,003
hsa-miR-762	1,50E-04	1,28E-03	6,683	hsa-miR-1909	3,50E-03	9,77E-03	3,664
hsa-miR-1268	1,50E-04	1,28E-03	6,637	hsa-miR-1275	3,60E-03	9,95E-03	3,578
hsa-miR-92b*	1,50E-04	1,28E-03	6,333	hsa-miR-1225-	3,89E-03	1,06E-02	3,583
hsa-miR-874	1,50E-04	1,28E-03	5,948	hsa-miR-1246	4,04E-03	1,08E-02	3,817
hsa-miR-654-5p	1,50E-04	1,28E-03	5,913	hsa-miR-885-5	4,09E-03	1,08E-02	3,419
hsa-miR-663b	1,50E-04	1,28E-03	5,812	hsa-miR-572	4,29E-03	1,12E-02	3,568
hsa-miR-25*	1,50E-04	1,28E-03	5,555	hsa-miR-1291	4,34E-03	1,12E-02	3,214
hsa-miR-1307	1,50E-04	1,28E-03	5,384	hsa-miR-379	4,44E-03	1,13E-02	2,183
hsa-miR-3162	2,00E-04	1,44E-03	5,378	hsa-miR-943	4,69E-03	1,17E-02	3,502
hsa-miR-665	2,00E-04	1,44E-03	3,431	hsa-miR-370	4,69E-03	1,17E-02	2,525
hsa-miR-224*	2,50E-04	1,59E-03	4,180	hsa-miR-1975	4,99E-03	1,23E-02	3,363
hsa-miR-1826	3,00E-04	1,82E-03	4,637	hsa-let-7b	5,59E-03	1,34E-02	3,263
hsa-miR-196a	3,50E-04	2,08E-03	6,957	hsa-miR-3188	6,09E-03	1,45E-02	3,103
hsa-miR-324-5p	4,99E-04	2,74E-03	4,534	hsa-miR-939	6,49E-03	1,53E-02	3,269
hsa-miR-1910	5,49E-04	2,74E-03	5,536	hsa-miR-152	7,54E-03	1,74E-02	3,173
hsa-miR-149*	5,49E-04	2,74E-03	5,162	hsa-miR-339-3	7,54E-03	1,74E-02	3,092
hsa-miR-1224-5p	5,49E-04	2,74E-03	4,381	hsa-miR-1290	8,14E-03	1,85E-02	3,476
hsa-miR-432	5,49E-04	2,74E-03	2,362	hsa-miR-664*	8,24E-03	1,85E-02	3,232
hsa-miR-324-3p	5,99E-04	2,93E-03	4,387	hsa-miR-1908	1,04E-02	2,29E-02	3,050
hsa-miR-3196	6,49E-04	3,12E-03	5,425	hsa-miR-487a	1,16E-02	2,53E-02	2,158
hsa-miR-1202	6,99E-04	3,30E-03	4,757	hsa-miR-423-5	1,20E-02	2,58E-02	2,627
hsa-miR-1281	7,99E-04	3,71E-03	4,618	hsa-miR-602	1,25E-02	2,66E-02	2,834
hsa-miR-3197	8,49E-04	3,81E-03	4,478	hsa-miR-935	1,29E-02	2,73E-02	2,830
hsa-miR-4284	8,49E-04	3,81E-03	4,121	hsa-miR-675	1,40E-02	2,93E-02	2,832
hsa-miR-1909*	8,99E-04	3,97E-03	4,425	hsa-miR-1973	1,72E-02	3,46E-02	2,803
hsa-miR-3185	9,49E-04	4,13E-03	4,860	hsa-miR-339-5	1,73E-02	3,46E-02	2,712
hsa-miR-346	9,99E-04	4,21E-03	4,666	hsa-miR-3195	1,82E-02	3,61E-02	2,665
hsa-miR-423-3p	9,99E-04	4,21E-03	3,325	hsa-miR-3187	1,84E-02	3,63E-02	2,654
hsa-miR-151-5p	1,40E-03	5,47E-03	4,533	hsa-miR-378c	1,88E-02	3,68E-02	2,662
hsa-miR-638	1,55E-03	5,89E-03	4,556	hsa-miR-3180-	2,17E-02	4,21E-02	2,564
hsa-miR-2277	1,85E-03	6,66E-03	4,313	hsa-miR-615-3	2,36E-02	4,55E-02	2,514
hsa-miR-1915	1,95E-03	6,84E-03	4,215	hsa-miR-4299	2,56E-02	4,91E-02	2,424

miRNAs desregulados en el estudio de datos procedentes de METABRIC

miRNAs reprimidos				miRNAs sobreexpresados			
miRNA	p-valor	FDR p-valor	Expresión	miRNA	p-valor	FDR p-valor	Expresión
<b>hsa-miR-29c*</b>	5,00E-06	9,16E-04	-8,660	<b>hsa-miR-17</b>	4,00E-05	2,29E-03	4,672
<b>hsa-miR-29c</b>	5,00E-06	9,16E-04	-7,501	hsa-miR-130b	1,30E-04	5,95E-03	4,254
<b>hsa-miR-33b</b>	1,00E-05	9,16E-04	-5,903	hsa-miR-362-5p	1,85E-04	6,52E-03	4,763
<b>hsa-miR-375</b>	1,00E-05	9,16E-04	-5,470	hsa-miR-19b	2,05E-04	6,56E-03	4,257
<b>hsa-miR-148b</b>	1,00E-05	9,16E-04	-5,301	hsa-miR-19a	2,15E-04	6,56E-03	3,931
<b>hsa-miR-30c</b>	2,00E-05	1,53E-03	-4,927	<b>hsa-miR-17*</b>	4,90E-04	1,07E-02	4,489
<b>hsa-miR-338-3p</b>	3,00E-05	1,96E-03	-5,385	hsa-miR-18a	4,90E-04	1,07E-02	4,469
hsa-miR-489	7,00E-05	3,56E-03	-5,748	hsa-miR-24	4,90E-04	1,07E-02	3,813
<b>hsa-miR-193a-3p</b>	1,55E-04	6,45E-03	-4,218	hsa-miR-92a	1,02E-03	1,73E-02	3,866
<b>hsa-miR-30a*</b>	1,85E-04	6,52E-03	-4,183	hsa-miR-505*	1,06E-03	1,74E-02	3,773
hsa-miR-10a	3,20E-04	9,16E-03	-4,232	hsa-miR-503	1,19E-03	1,86E-02	4,039
hsa-miR-29b	3,60E-04	9,70E-03	-4,085	hsa-miR-18b	1,26E-03	1,86E-02	4,058
hsa-miR-592	4,85E-04	1,07E-02	-5,079	hsa-miR-130a	1,52E-03	2,00E-02	3,460
hsa-miR-30b	5,35E-04	1,11E-02	-3,767	hsa-miR-505	1,65E-03	2,11E-02	3,399
<b>hsa-let-7b</b>	6,10E-04	1,19E-02	-3,884	hsa-miR-93	2,29E-03	2,77E-02	3,414
hsa-miR-744	6,25E-04	1,19E-02	-4,333	hsa-miR-663	2,69E-03	3,07E-02	3,352
hsa-miR-200a	7,90E-04	1,45E-02	-3,953	hsa-miR-31	2,75E-03	3,07E-02	3,572
hsa-miR-190b	9,55E-04	1,68E-02	-5,285	hsa-miR-301b	3,25E-03	3,31E-02	3,543
<b>hsa-miR-342-5p</b>	1,22E-03	1,86E-02	-3,475	hsa-miR-125b	5,26E-03	4,67E-02	2,977
hsa-miR-29b-2*	1,35E-03	1,94E-02	-4,833				
<b>hsa-miR-631</b>	1,41E-03	1,96E-02	-4,267				
hsa-miR-34a	1,46E-03	1,97E-02	-3,457				
hsa-miR-30e*	1,74E-03	2,16E-02	-3,519				
hsa-miR-365	2,59E-03	3,04E-02	-3,199				
hsa-miR-499-5p	2,88E-03	3,14E-02	-3,893				
hsa-miR-629	3,00E-03	3,20E-02	-3,208				
hsa-let-7a*	3,25E-03	3,31E-02	-3,452				
hsa-miR-331-3p	3,59E-03	3,58E-02	-3,104				
<b>hsa-miR-342-3p</b>	3,67E-03	3,58E-02	-3,075				
hsa-miR-148a	4,03E-03	3,85E-02	-3,076				
hsa-miR-921	4,65E-03	4,35E-02	-3,910				
hsa-miR-193b	5,04E-03	4,62E-02	-2,978				
hsa-miR-34a*	5,36E-03	4,67E-02	-2,965				
hsa-miR-10a*	5,40E-03	4,67E-02	-5,056				

**ANEXO VI:** Rutas enriquecidas en genes diana por nodos de miRNAs desregulados en cáncer de mama en mujeres jóvenes mediante DIANA miRPath (*genes union*).

Nodo 60

ID KEGG	Ruta KEGG	FDR	Nº genes	Nº miRNAs
hsa05032	Adicción a la morfina	$6,48 \times 10^{-7}$	11	4
hsa04713	Adecuación circadiana	$1,31 \times 10^{-4}$	10	3
hsa04723	Señalización retrógrada del endocanabioide	$1,31 \times 10^{-4}$	10	4
hsa04725	Sinapsis colinérgica	$4,48 \times 10^{-4}$	11	4
hsa05221	Leucemia mieloides aguda	$1,65 \times 10^{-3}$	7	2
hsa04727	Sinapsis GABAérgica	$2,88 \times 10^{-3}$	9	4
hsa03013	Transporte de RNA	$4,56 \times 10^{-3}$	12	4
hsa04721	Ciclo vesicular de las sinapsis	$5,12 \times 10^{-3}$	7	4
hsa04726	Sinapsis serotoninérgica	$5,12 \times 10^{-3}$	10	4
hsa04962	Reabsorción de agua regulada por vasopresina	$7,09 \times 10^{-3}$	6	2
hsa04914	Maduración del oocito mediada por progesterona	$8,19 \times 10^{-3}$	8	3
hsa04724	Sinapsis glutamatérgica	$8,23 \times 10^{-3}$	10	4
hsa04310	Señalización Wnt	$9,95 \times 10^{-3}$	11	4
hsa04728	Sinapsis dopaminérgicas	$1,15 \times 10^{-2}$	10	4
hsa00532	Biosíntesis del condroitín sulfato	$1,70 \times 10^{-2}$	2	1
hsa04666	Fagocitosis mediada por Fc gamma R	$1,70 \times 10^{-2}$	8	4
hsa00450	Metabolismo del selenocomponente	$2,01 \times 10^{-2}$	3	2
hsa00310	Degradación de lisina	$2,40 \times 10^{-2}$	4	3
hsa04012	Señalización ErbB	$2,40 \times 10^{-2}$	8	4
hsa04916	Melanogénesis	$2,40 \times 10^{-2}$	8	4
hsa00920	Metabolismo del azufre	$2,58 \times 10^{-2}$	3	1

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 18

ID KEGG	Ruta KEGG	FDR	Nº genes	Nº miRNAs
hsa04360	Guía de axones	$2,95 \times 10^{-9}$	36	6
hsa04010	Señalización MAPK	$1,32 \times 10^{-7}$	53	8
hsa05166	Infección por HTLV-I	$3,89 \times 10^{-7}$	55	9
hsa04340	Señalización Hedgehog	$4,68 \times 10^{-7}$	16	7
hsa05217	Carcinoma de célula basal	$1,37 \times 10^{-6}$	16	6
hsa04151	Señalización PI3K/Akt	$4,36 \times 10^{-5}$	61	9
hsa05215	Cáncer de próstata	$4,74 \times 10^{-5}$	21	5
hsa04510	Adhesiones focales	$2,09 \times 10^{-4}$	39	9
hsa00591	Metabolismo del ácido linoleico	$7,02 \times 10^{-4}$	11	6
hsa05200	Rutas en cáncer	$8,83 \times 10^{-4}$	57	8
hsa04310	Señalización Wnt	$9,44 \times 10^{-4}$	33	6
hsa04916	Melanogénesis	$1,45 \times 10^{-3}$	22	8
hsa05202	Desregulación transcripcional en cáncer	$2,85 \times 10^{-3}$	37	7
hsa04066	Señalización HIF-1	$3,31 \times 10^{-3}$	23	7
hsa04810	Regulación del citoesqueleto de actina	$4,28 \times 10^{-3}$	38	8
hsa04725	Sinapsis colinérgica	$4,62 \times 10^{-3}$	24	8
hsa04020	Señalización del calcio	$4,62 \times 10^{-3}$	32	9
hsa04150	Señalización mTOR	$8,85 \times 10^{-3}$	14	7
hsa04728	Sinapsis dopaminérgicas	$8,85 \times 10^{-3}$	26	7
hsa04270	Contracción del músculo vascular liso	$8,85 \times 10^{-3}$	24	8
hsa04666	Fagocitosis mediada por Fc gamma R	$1,08 \times 10^{-2}$	19	7
hsa00592	Metabolismo del ácido alfa-linoleico	$1,81 \times 10^{-2}$	8	5
hsa04721	Ciclo de vesículas sinápticas	$1,90 \times 10^{-2}$	15	7
hsa04370	Señalización por VEGF	$2,07 \times 10^{-2}$	14	5
hsa04130	Interacciones SNARE	$2,42 \times 10^{-2}$	9	5
hsa04710	Ritmo circadiano	$2,47 \times 10^{-2}$	8	4
hsa00310	Degradación de lisina	$2,94 \times 10^{-2}$	10	7
hsa04720	Potenciación a largo plazo	$3,67 \times 10^{-2}$	14	6
hsa05161	Hepatitis B	$3,73 \times 10^{-2}$	26	6

FDR se refiere al pvalor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 80

ID KEGG	Ruta metabólica	p-valor	Nº genes	Nº miRNAs
hsa04120	Proteólisis mediada por ubiquitina	$6,73 \times 10^{-6}$	26	5
hsa04722	Señalización de neurotrofina	$6,00 \times 10^{-5}$	22	3
hsa04010	Señalización MAPK	$1,07 \times 10^{-4}$	38	5
hsa05200	Rutas en cáncer	$1,00 \times 10^{-3}$	45	5
hsa00534	Biosíntesis de heparán sulfato	$1,59 \times 10^{-3}$	6	3
hsa04720	Potenciación a largo plazo	$2,10 \times 10^{-3}$	13	4
hsa05142	Enfermedad de Chagas	$2,10 \times 10^{-3}$	18	4
hsa05211	Carcinoma de células renales	$2,10 \times 10^{-3}$	13	4
hsa04310	Señalización Wnt	$3,71 \times 10^{-3}$	24	4
hsa04350	Señalización TGF-beta	$3,71 \times 10^{-3}$	12	4
hsa03015	Síntesis de mRNA	$3,83 \times 10^{-3}$	15	4
hsa04150	Señalización mTOR	$3,83 \times 10^{-3}$	12	4
hsa04530	Uniones estrechas	$3,83 \times 10^{-3}$	21	5
hsa04012	Señalización ErbB	$5,13 \times 10^{-3}$	15	3
hsa05222	Cáncer de pulmón	$6,42 \times 10^{-3}$	14	4
hsa04620	Señalización por receptores tipo Toll	$7,15 \times 10^{-3}$	16	4
hsa04144	Endocitosis	$7,74 \times 10^{-3}$	27	5
hsa00471	Metabolismo del D-Glutamato	$7,88 \times 10^{-3}$	2	2
hsa03013	Transporte de RNA	$1,10 \times 10^{-2}$	21	4
hsa05161	Hepatitis B	$1,15 \times 10^{-2}$	21	4
hsa04115	Señalización p53	$1,32 \times 10^{-2}$	12	3
hsa05212	Cáncer de páncreas	$1,32 \times 10^{-2}$	12	4
hsa04370	Señalización VEGF	$1,79 \times 10^{-2}$	11	3
hsa04971	Secreción de ácido gástrico	$2,61 \times 10^{-2}$	12	4
hsa00650	Metabolismo del butanoato	$4,75 \times 10^{-2}$	6	2
hsa04622	Señalización por receptores tipo RIG-I	$4,75 \times 10^{-2}$	11	3
hsa04664	Señalización Fc epsilon RI	$4,75 \times 10^{-2}$	11	3
hsa05133	Tos ferina	$4,75 \times 10^{-2}$	11	3
hsa04520	Uniones adherentes	$4,75 \times 10^{-2}$	12	4

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 62

ID KEGG	Ruta metabólica	FDR	Nº de genes	Nº de miRNAs
hsa04350	Señalización TGF-beta	$1,11 \times 10^{-8}$	20	7
hsa04120	Proteólisis mediada por ubiquitina	$2,20 \times 10^{-7}$	28	7
hsa04062	Señalización por quemoquinas	$1,36 \times 10^{-5}$	32	7
hsa04960	Reabsorción de sodio regulada por aldosterona	$1,90 \times 10^{-5}$	11	8
hsa04151	Señalización PI3K/Akt	$3,31 \times 10^{-5}$	49	7
hsa05131	Shigelosis	$1,03 \times 10^{-4}$	15	7
hsa04530	Uniones estrechas	$2,72 \times 10^{-4}$	24	6
hsa04510	Adhesiones focales	$3,53 \times 10^{-4}$	31	7
hsa05220	Leucemia mieloide crónica	$6,32 \times 10^{-4}$	14	6
hsa04728	Sinapsis dopaminérgicas	$6,32 \times 10^{-4}$	22	7
hsa04360	Guía de axones	$6,76 \times 10^{-4}$	23	7
hsa04722	Señalización de neurotrofina	$1,06 \times 10^{-3}$	21	7
hsa04114	Meiosis del oocito	$1,23 \times 10^{-3}$	21	7
hsa04144	Endocitosis	$1,23 \times 10^{-3}$	30	7
hsa04910	Señalización por insulina	$1,23 \times 10^{-3}$	22	7
hsa00780	Metabolismo de biotina	$1,33 \times 10^{-3}$	1	1
hsa05223	Cáncer de pulmón no de células pequeñas	$1,83 \times 10^{-3}$	10	7
hsa04920	Señalización de adipocitokina	$1,86 \times 10^{-3}$	13	6
hsa04914	Maduración del oocito mediada por progesterona	$3,55 \times 10^{-3}$	15	7
hsa05166	Infección por HTLV-I	$3,55 \times 10^{-3}$	36	7
hsa04012	Señalización ErbB	$5,66 \times 10^{-3}$	15	7
hsa04150	Señalización mTOR	$6,28 \times 10^{-3}$	12	7
hsa05212	Cáncer de páncreas	$7,14 \times 10^{-3}$	13	7
hsa04540	Uniones GAP	$1,06 \times 10^{-2}$	12	6
hsa04520	Uniones adherentes	$1,06 \times 10^{-2}$	14	7
hsa04971	Secreción de ácido gástrico	$1,27 \times 10^{-2}$	13	6
hsa04720	Potenciación a largo plazo	$1,28 \times 10^{-2}$	12	5
hsa04662	Señalización de receptor de células B	$1,44 \times 10^{-2}$	13	7
hsa04912	Señalización de GnRH	$1,76 \times 10^{-2}$	14	6
hsa04725	Sinapsis colinérgica	$1,96 \times 10^{-2}$	18	6
hsa04310	Señalización Wnt	$2,13 \times 10^{-2}$	23	7
hsa05215	Cáncer de próstata	$2,73 \times 10^{-2}$	14	6
hsa03013	Transporte de RNA	$2,73 \times 10^{-2}$	20	7
hsa00310	Degradación de lisina	$3,18 \times 10^{-2}$	8	4
hsa04723	Señalización retrógrada del endocanabioide	$3,18 \times 10^{-2}$	17	6
hsa05213	Cáncer de endometrio	$3,81 \times 10^{-2}$	9	6
hsa05222	Cáncer de pulmón	$3,81 \times 10^{-2}$	13	6
hsa05211	Carcinoma de células renales	$3,81 \times 10^{-2}$	12	5
hsa04115	Señalización p53	$4,16 \times 10^{-2}$	11	6
hsa04666	Fagocitosis mediada por Fc gamma R	$4,16 \times 10^{-2}$	14	7
hsa05161	Hepatitis B	$4,16 \times 10^{-2}$	21	7
hsa05214	Glioma	$4,16 \times 10^{-2}$	12	7
hsa04141	Procesado de proteínas en el retículo	$4,17 \times 10^{-2}$	23	7
hsa04010	Señalización MAPK	$4,82 \times 10^{-2}$	32	7
hsa05142	Enfermedad de Chagas	$4,87 \times 10^{-2}$	16	7

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 78

ID KEGG	Ruta metabólica	FDR	Nº genes	Nº miRNAs
hsa04512	Interacción EMC-receptor	5,02 x 10 <sup>-57</sup>	23	5
hsa04974	Digestión y absorción de proteínas	1,93 x 10 <sup>-20</sup>	27	5
hsa05020	Enfermedades por priones	1,31 x 10 <sup>-15</sup>	3	4
hsa04510	Adhesiones focales	1,31 x 10 <sup>-15</sup>	40	5
hsa05146	Amebiosis	8,65 x 10 <sup>-15</sup>	22	5
hsa04151	Señalización PI3K/Akt	2,53 x 10 <sup>-12</sup>	52	5
hsa05222	Cáncer de pulmón	2,57 x 10 <sup>-10</sup>	20	5
hsa00310	Degradación de lisina	6,94 x 10 <sup>-09</sup>	11	5
hsa05214	Glioma	4,38 x 10 <sup>-07</sup>	15	5
hsa05215	Cáncer de próstata	2,43 x 10 <sup>-05</sup>	16	5
hsa05202	Desregulación transcripcional en cáncer	8,78 x 10 <sup>-05</sup>	27	5
hsa05223	Cáncer de pulmón no de células pequeñas	3,23 x 10 <sup>-04</sup>	11	5
hsa04728	Sinapsis dopaminérgicas	1,66 x 10 <sup>-03</sup>	19	5
hsa05218	Melanoma	7,20 x 10 <sup>-03</sup>	12	5
hsa05410	Cardiomiopatía hipertrofica	7,46 x 10 <sup>-03</sup>	13	5
hsa00270	Metabolismo de la cisteína y metionina	7,65 x 10 <sup>-03</sup>	8	5
hsa04010	Señalización MAPK	7,65 x 10 <sup>-03</sup>	31	5
hsa04722	Señalización de neurotrofina	7,65 x 10 <sup>-03</sup>	17	5
hsa04730	Depresión a largo plazo	7,65 x 10 <sup>-03</sup>	10	5
hsa00534	Biosíntesis de heparán sulfato	7,80 x 10 <sup>-03</sup>	5	3
hsa04724	Sinapsis glutamatérgicas	9,08 x 10 <sup>-03</sup>	16	5
hsa05200	Rutas en cáncer	9,94 x 10 <sup>-03</sup>	39	5
hsa04120	Proteólisis mediada por ubiquitina	1,53 x 10 <sup>-02</sup>	18	5
hsa05211	Carcinoma de células renales	1,86 x 10 <sup>-02</sup>	11	5
hsa05033	Adición a la nicotina	3,32 x 10 <sup>-02</sup>	7	5
hsa04310	Señalización Wnt	3,35 x 10 <sup>-02</sup>	18	5
hsa04720	Potenciación a largo plazo	3,75 x 10 <sup>-02</sup>	10	5
hsa04713	Ritmo circadiano	4,20 x 10 <sup>-02</sup>	14	5
hsa04130	Interacciones SNARE del transporte	4,63 x 10 <sup>-02</sup>	6	5

FDR se refiere al pvalor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.



## Nodo 48a

ID KEGG	Ruta metabólica	FDR	Nº genes	Nº miRNAs
hsa04510	Adhesiones focales	2,90 x 10 <sup>-35</sup>	88	12
hsa04151	Señalización PI3K/Akt	1,45 x 10 <sup>-32</sup>	128	12
hsa04150	Señalización mTOR	6,30 x 10 <sup>-20</sup>	32	12
hsa05222	Cáncer de pulmón	2,75 x 10 <sup>-19</sup>	39	10
hsa05215	Cáncer de próstata	3,28 x 10 <sup>-19</sup>	40	10
hsa04115	Señalización p53	6,42 x 10 <sup>-19</sup>	34	12
hsa04010	Señalización MAPK	1,36 x 10 <sup>-18</sup>	95	12
hsa05200	Rutas en cáncer	7,18 x 10 <sup>-18</sup>	120	12
hsa04512	Interacción EMC-receptor	3,60 x 10 <sup>-17</sup>	38	10
hsa04012	Señalización ErbB	2,49 x 10 <sup>-15</sup>	39	12
hsa04722	Señalización de neurotrofina	3,28 x 10 <sup>-15</sup>	50	12
hsa05146	Amebiosis	1,97 x 10 <sup>-14</sup>	41	10
hsa04310	Señalización Wnt	3,68 x 10 <sup>-13</sup>	58	12
hsa04350	Señalización TGF-beta	9,23 x 10 <sup>-13</sup>	34	9
hsa04810	Regulación del citoesqueleto de actina	1,37 x 10 <sup>-12</sup>	75	12
hsa04910	Señalización por insulina	2,93 x 10 <sup>-12</sup>	51	12
hsa04974	Absorción y digestión de proteínas	7,19 x 10 <sup>-12</sup>	36	10
hsa04662	Señalización de receptor de células B	1,40 x 10 <sup>-10</sup>	31	12
hsa04660	Señalización de receptor de células T	1,85 x 10 <sup>-10</sup>	41	12
hsa04120	Proteólisis mediada por ubiquitina	3,26 x 10 <sup>-10</sup>	50	11
hsa05218	Melanoma	4,29 x 10 <sup>-10</sup>	30	10
hsa05223	Cáncer de pulmón no de células pequeñas	8,31 x 10 <sup>-10</sup>	24	11
hsa05210	Cáncer colorrectal	9,50 x 10 <sup>-10</sup>	27	12
hsa04110	Ciclo celular	4,50 x 10 <sup>-09</sup>	48	11
hsa04720	Potenciación a largo plazo	5,81 x 10 <sup>-09</sup>	28	11
hsa04360	Guía de axones	1,86 x 10 <sup>-08</sup>	51	12
hsa05220	Leucemia mieloide crónica	5,52 x 10 <sup>-08</sup>	30	10
hsa00061	Biosíntesis de ácidos grasos	7,75 x 10 <sup>-08</sup>	2	2
hsa04664	Señalización Fc epsilon RI	7,75 x 10 <sup>-08</sup>	28	11
hsa05214	Glioma	9,41 x 10 <sup>-08</sup>	29	10
hsa04666	Fagocitosis mediada por Fc gamma R	1,09 x 10 <sup>-07</sup>	34	12
hsa05211	Carcinoma de células renales	1,78 x 10 <sup>-07</sup>	29	11
hsa05221	Leucemia mieloide aguda	3,20 x 10 <sup>-07</sup>	23	11
hsa04144	Edocitosis	1,30 x 10 <sup>-06</sup>	63	12
hsa05014	Esclerosis lateral amiotrófica	1,37 x 10 <sup>-06</sup>	22	9
hsa05212	Cáncer de páncreas	2,91 x 10 <sup>-06</sup>	27	11
hsa04141	Procesado de proteínas en el retículo endoplasmático	4,09 x 10 <sup>-06</sup>	55	12
hsa05213	Cáncer de endometrio	4,09 x 10 <sup>-06</sup>	21	12
hsa05161	Hepatitis B	4,40 x 10 <sup>-06</sup>	51	12
hsa05410	Cardiomiopatía hipertrófica	7,88 x 10 <sup>-06</sup>	30	11
hsa04370	Señalización por VEGF	2,08 x 10 <sup>-05</sup>	24	11
hsa00310	Degradación de lisina	2,34 x 10 <sup>-05</sup>	19	10
hsa04380	Diferenciación del osteoclasto	3,23 x 10 <sup>-05</sup>	42	11
hsa05202	Desregulación transcripcional en cáncer	4,75 x 10 <sup>-05</sup>	58	12
hsa04728	Sinapsis dopaminérgicas	7,00 x 10 <sup>-05</sup>	42	12
hsa03015	Síntesis de mRNA	1,23 x 10 <sup>-04</sup>	30	11
hsa04912	Señalización de GnRH	1,59 x 10 <sup>-04</sup>	30	12
hsa04730	Depresión a largo plazo	1,92 x 10 <sup>-04</sup>	25	10

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 48b

ID KEGG	Ruta metabólica	FDR	Nº genes	Nº miRNAs
hsa05020	Enfermedades por priones	2,11 x 10 <sup>-04</sup>	6	8
hsa04540	Uniones GAP	2,73 x 10 <sup>-04</sup>	28	11
hsa04914	Maduración del oocito mediada por progesterona	3,07 x 10 <sup>-04</sup>	28	12
hsa05166	Infección por HTLV-I	4,41 x 10 <sup>-04</sup>	73	12
hsa05131	Shigelosis	7,45 x 10 <sup>-04</sup>	21	12
hsa04320	Formación del eje dorso-ventral	1,11 x 10 <sup>-03</sup>	10	7
hsa05414	Cardiomiopatía dilatada	1,11 x 10 <sup>-03</sup>	29	11
hsa04724	Sinapsis glutamatérgica	2,12 x 10 <sup>-03</sup>	35	12
hsa04960	Reabsorción de sodio regulada por aldosterona	2,52 x 10 <sup>-03</sup>	14	10
hsa04930	Diabetes mellitus tipo II	4,39 x 10 <sup>-03</sup>	16	10
hsa04066	Señalización HIF-1	5,96 x 10 <sup>-03</sup>	33	10
hsa05203	Carcinogénesis viral	6,23 x 10 <sup>-03</sup>	57	12
hsa04725	Sinapsis colinérgica	8,15 x 10 <sup>-03</sup>	36	12
hsa04971	Secreción de ácido gástrico	8,32 x 10 <sup>-03</sup>	23	11
hsa00562	Sistema de señalización por fosfatidil-inositol	1,16 x 10 <sup>-02</sup>	20	11
hsa03060	Exportación de proteínas	1,65 x 10 <sup>-02</sup>	9	8
hsa05145	Toxoplasmosis	1,86 x 10 <sup>-02</sup>	35	11
hsa03013	Transporte de RNA	2,12 x 10 <sup>-02</sup>	43	12
hsa05142	Enfermedad de Chagas	2,16 x 10 <sup>-02</sup>	29	11
hsa04920	Señalización de adipocitokina	2,88 x 10 <sup>-02</sup>	20	11
hsa04916	Melanogénesis	2,88 x 10 <sup>-02</sup>	29	12
hsa00533	Biosíntesis del ketarán sulfato	4,13 x 10 <sup>-02</sup>	6	5
hsa05162	Rubeola	4,19 x 10 <sup>-02</sup>	40	10
hsa05100	Invasión bacteriana de células epiteliales	4,52 x 10 <sup>-02</sup>	22	11
hsa04210	Apoptosis	4,59 x 10 <sup>-02</sup>	27	10

FDR se refiere al pvalor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 51

ID KEGG	Ruta metabólica	FDR	Nº genes	Nº miRNAs
hsa05130	Infección por <i>Escherichia coli</i>	1,32 x 10 <sup>-5</sup>	10	5
hsa04120	Proteólisis mediada por ubiquitina	1,55 x 10 <sup>-4</sup>	16	5
hsa04110	Ciclo celular	7,15 x 10 <sup>-4</sup>	14	5
hsa05202	Desregulación transcripcional en cáncer	3,20 x 10 <sup>-3</sup>	17	3
hsa05014	Esclerosis lateral amiotrófica	3,20 x 10 <sup>-3</sup>	8	4
hsa04910	Señalización por insulina	8,76 x 10 <sup>-3</sup>	13	5
hsa04930	Diabete mellitus ttipo II	1,19 x 10 <sup>-2</sup>	7	4
hsa05222	Cáncer de pulmón	1,97 x 10 <sup>-2</sup>	9	4
hsa00510	Biosíntesis de N-glicano	2,23 x 10 <sup>-2</sup>	6	5
hsa05220	Leucemia mieloide crónica	2,28 x 10 <sup>-2</sup>	8	4
hsa00270	Metabolismo de cisteína y metionina	4,04 x 10 <sup>-2</sup>	5	3

FDR se refiere al pvalor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

## Nodo 72

ID KEGG	Ruta metabólica	FDR	Nº genes	Nº miRNAs
hsa04720	Potenciación a largo plazo	$7.50 \times 10^{-13}$	21	4
hsa04971	Secreción de ácido gástrico	$3.01 \times 10^{-10}$	20	4
hsa04012	Señalización ErbB	$7.33 \times 10^{-8}$	20	4
hsa04540	Uniones GAP	$7.33 \times 10^{-8}$	20	4
hsa04961	Reabsorción de calcio mediada por sistema endocrino	$1.56 \times 10^{-7}$	14	4
hsa04723	Señalización retrógrada del endocanabioide	$6.74 \times 10^{-7}$	23	4
hsa04912	Señalización de GnRH	$1.03 \times 10^{-6}$	19	4
hsa04150	Señalización mTOR	$1.41 \times 10^{-6}$	15	4
hsa04724	Sinapsis glutamatérgicas	$3.28 \times 10^{-6}$	22	4
hsa04120	Proteólisis mediada por ubiquitina	$1.03 \times 10^{-5}$	25	4
hsa04728	Sinapsis dopaminérgicas	$2.31 \times 10^{-5}$	24	4
hsa05202	Desregulación transcripcional en cáncer	$2.31 \times 10^{-5}$	30	4
hsa04010	Señalización MAPK	$3.04 \times 10^{-5}$	38	4
hsa04722	Señalización de neurotrofina	$4.26 \times 10^{-5}$	22	4
hsa04910	Señalización por insulina	$5.86 \times 10^{-5}$	23	4
hsa04970	Secreción salivar	$6.24 \times 10^{-5}$	17	4
hsa04115	Señalización p53	$4.65 \times 10^{-4}$	14	4
hsa05221	Leucemia mieloide aguda	$5.22 \times 10^{-4}$	11	4
hsa04916	Melanogénesis	$1.12 \times 10^{-3}$	16	4
hsa04727	Sinapsis GABAérgicas	$1.71 \times 10^{-3}$	17	4
hsa05143	Tripanosomiasis africana	$1.82 \times 10^{-3}$	8	4
hsa04972	Secreción pancreática	$2.40 \times 10^{-3}$	16	4
hsa04070	Sistema de señalización por fosfatidil-inositol	$4.21 \times 10^{-3}$	14	4
hsa04725	Sinapsis colinérgica	$4.53 \times 10^{-3}$	19	4
hsa04726	Sinapsis serotoninérgicas	$4.53 \times 10^{-3}$	17	4
hsa05142	Enfermedad de Chagas	$4.71 \times 10^{-3}$	17	4
hsa04730	Depresión a largo plazo	$4.74 \times 10^{-3}$	12	4
hsa04810	Regulación del citoesqueleto de actina	$4.87 \times 10^{-3}$	29	4
hsa04020	Señalización de calcio	$5.16 \times 10^{-3}$	25	4
hsa03015	Síntesis de mRNA	$5.51 \times 10^{-3}$	15	4
hsa05110	Infección por <i>Vibrio cholerae</i>	$5.94 \times 10^{-3}$	10	3
hsa05146	Amebiosis	$8.48 \times 10^{-3}$	17	4
hsa04310	Señalización Wnt	$1.19 \times 10^{-2}$	23	4
hsa05214	Glioma	$1.19 \times 10^{-2}$	12	4
hsa05222	Cáncer de pulmón	$2.01 \times 10^{-2}$	13	4
hsa04210	Apoptosis	$2.03 \times 10^{-2}$	14	4
hsa04914	Maduración del oocito mediada por progesterona	$2.25 \times 10^{-2}$	13	4
hsa04114	Meiosis del oocito	$2.44 \times 10^{-2}$	19	4
hsa04530	Uniones estrechas	$2.44 \times 10^{-2}$	19	4
hsa04614	Sistema renina-angiotensina	$2.48 \times 10^{-2}$	4	3
hsa04630	Señalización por Jak-STAT	$2.77 \times 10^{-2}$	20	4
hsa00562	Metabolismo del inositol fosfato	$3.58 \times 10^{-2}$	10	4
hsa00650	Metabolismo del butanoato	$3.68 \times 10^{-2}$	6	4
hsa05215	Cáncer de próstata	$4.13 \times 10^{-2}$	13	4
hsa04960	Reabsorción de sodio regulada por aldosterona	$4.39 \times 10^{-2}$	7	4
hsa05200	Rutas en cáncer	$4.92 \times 10^{-2}$	40	4

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg.

ANEXO VII: Rutas enriquecidas en miRNAs desregulados en cáncer de mama en mujeres jóvenes mediante DIANA miRPath (*pathway union*)

Nodo 15			
KEGG pathway	FDR	Nº genes	Nº miRNAs
<b>Cell adhesion molecules (CAMs) (hsa04514)</b>	1,35E-07	14	3
<b>Morphine addiction (hsa05032)</b>	1,55E-07	9	4
<b>GABAergic synapse (hsa04727)</b>	3,08E-06	7	3
<b>Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)</b>	3,41E-06	2	2
<b>Axon guidance (hsa04360)</b>	7,39E-06	26	2
Retrograde endocannabinoid signaling(hsa04723)	2,93E-04	9	3
Circadian entrainment (hsa04713)	1,43E-03	7	3
Calcium signaling pathway (hsa04020)	1,56E-03	19	3
Dilated cardiomyopathy (hsa05414)	3,82E-03	8	4
PI3K-Akt signaling pathway (hsa04151)	6,56E-03	31	3
Regulation of actin cytoskeleton (hsa04810)	1,01E-02	29	3
HTLV-I infection (hsa05166)	1,29E-02	23	2
MAPK signaling pathway (hsa04010)	1,65E-02	47	3
Basal cell carcinoma (hsa05217)	2,49E-02	7	3
Synaptic vesicle cycle (hsa04721)	3,33E-02	11	3
Glutamatergic synapse (hsa04724)	3,74E-02	4	2
Lysine degradation (hsa00310)	3,80E-02	6	3

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 18			
KEGG pathway	FDR	Nº genes	Nº miRNAs
<b>Drug metabolism - cytochrome P450 (hsa00982)</b>	<1 x 10 <sup>-16</sup>	1	1
<b>Glycosphingolipid biosynthesis - lacto and neolacto series (hsa00601)</b>	3,91 x 10 <sup>-6</sup>	2	1
<b>Morphine addiction (hsa05032)</b>	8,19 x 10 <sup>-6</sup>	9	4
<b>Cell adhesion molecules (CAMs) (hsa04514)</b>	1,09 x 10 <sup>-5</sup>	14	3
<b>GABAergic synapse (hsa04727)</b>	1,11 x 10 <sup>-4</sup>	7	3
<b>Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)</b>	1,45 x 10 <sup>-4</sup>	2	2
<b>Axon guidance (hsa04360)</b>	3,14 x 10 <sup>-4</sup>	26	2
Calcium signaling pathway (hsa04020)	1,14 x 10 <sup>-3</sup>	23	4
Retrograde endocannabinoid signaling(hsa04723)	2,79 x 10 <sup>-3</sup>	9	3
HTLV-I infection (hsa05166)	3,68 x 10 <sup>-3</sup>	34	3
Hedgehog signaling pathway (hsa04340)	9,56 x 10 <sup>-3</sup>	10	3
Circadian entrainment (hsa04713)	1,67 x 10 <sup>-2</sup>	7	3
Dilated cardiomyopathy (hsa05414)	2,13 x 10 <sup>-2</sup>	8	4
Lysine degradation (hsa00310)	2,17 x 10 <sup>-2</sup>	7	4
Melanogenesis (hsa04916)	3,30 x 10 <sup>-2</sup>	15	3
Basal cell carcinoma (hsa05217)	3,45 x 10 <sup>-2</sup>	11	4
PI3K-Akt signaling pathway (hsa04151)	4,36 x 10 <sup>-2</sup>	31	3

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 48			
KEGG pathway	FDR	Nº genes	Nº miRNAs
Protein digestion and absorption (hsa04974)	<1e-16	23	1
Fatty acid biosynthesis (hsa00061)	<1e-16	2	2
ECM-receptor interaction (hsa04512)	<1e-16	22	2
Focal adhesion (hsa04510)	<1e-16	76	5
Ubiquitin mediated proteolysis (hsa04120)	<1e-16	38	6
PI3K-Akt signaling pathway (hsa04151)	<1e-16	127	10
Amoebiasis (hsa05146)	9,99E-16	19	1
Prion diseases (hsa05020)	2,89E-15	1	1
Prostate cancer (hsa05215)	2,89E-15	29	5
Pathways in cancer (hsa05200)	5,93E-14	90	7
Melanoma (hsa05218)	8,75E-13	21	5
p53 signaling pathway (hsa04115)	1,27E-12	26	6
Glioma (hsa05214)	1,70E-12	27	6
Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)	5,80E-12	2	2
MAPK signaling pathway (hsa04010)	2,08E-10	72	7
Wnt signaling pathway (hsa04310)	2,60E-10	57	7
Small cell lung cancer (hsa05222)	6,66E-10	25	5
Axon guidance (hsa04360)	1,09E-09	24	3
ErbB signaling pathway (hsa04012)	1,81E-08	27	4
mTOR signaling pathway (hsa04150)	2,21E-08	28	6
Insulin signaling pathway (hsa04910)	2,34E-08	39	4
Neurotrophin signaling pathway (hsa04722)	6,99E-08	41	6
Non-small cell lung cancer (hsa05223)	2,07E-07	16	4
Long-term potentiation (hsa04720)	2,93E-07	21	5
B cell receptor signaling pathway (hsa04662)	9,04E-07	13	2
TGF-beta signaling pathway (hsa04350)	3,60E-06	29	5
Cell cycle (hsa04110)	2,67E-05	41	6
Protein processing in endoplasmic reticulum(hsa04141)	5,10E-05	25	3
T cell receptor signaling pathway (hsa04660)	8,70E-05	24	3
Hepatitis B (hsa05161)	9,55E-05	25	4
Chronic myeloid leukemia (hsa05220)	1,09E-04	16	3
Long-term depression (hsa04730)	1,79E-04	16	4
Endometrial cancer (hsa05213)	1,84E-04	14	4
Endocytosis (hsa04144)	2,04E-04	28	3
Regulation of actin cytoskeleton (hsa04810)	4,56E-04	46	4
Hypertrophic cardiomyopathy (HCM) (hsa05410)	6,76E-04	20	4
Dopaminergic synapse (hsa04728)	7,09E-04	30	4
Acute myeloid leukemia (hsa05221)	1,79E-03	8	2
Glycosaminoglycan biosynthesis - heparan sulfate(hsa00534)	2,41E-03	5	3
Colorectal cancer (hsa05210)	3,03E-03	8	2
Arrhythmogenic right ventricular cardiomyopathy (hsa05412)	9,10E-03	11	2
Amyotrophic lateral sclerosis (ALS) (hsa05014)	1,32E-02	9	2
Lysine degradation (hsa00310)	1,47E-02	11	2
Pancreatic cancer (hsa05212)	1,59E-02	10	2
Renal cell carcinoma (hsa05211)	2,85E-02	15	2
Dilated cardiomyopathy (hsa05414)	4,22E-02	15	3
Transcriptional misregulation in cancer(hsa05202)	4,25E-02	19	2
Fc epsilon RI signaling pathway (hsa04664)	4,49E-02	12	1

Nodo 51			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>Phagosome (hsa04145)</b>	1,14E-06	4	1
<b>Amphetamine addiction (hsa05031)</b>	6,25E-05	7	2
<b>Pathogenic Escherichia coli infection (hsa05130)</b>	1,13E-04	9	3
<b>Dopaminergic synapse (hsa04728)</b>	1,49E-04	8	2
<b>Amyotrophic lateral sclerosis (ALS) (hsa05014)</b>	2,80E-04	6	2
<b>Transcriptional misregulation in cancer(hsa05202)</b>	1,43E-03	14	2
Long-term potentiation (hsa04720)	0,0110	5	2
Lysine degradation (hsa00310)	0,0116	4	2
Collecting duct acid secretion (hsa04966)	0,0148	2	1
Ubiquitin mediated proteolysis (hsa04120)	0,0168	11	2
ErbB signaling pathway (hsa04012)	0,0169	8	2
Axon guidance (hsa04360)	0,0194	7	1
Synaptic vesicle cycle (hsa04721)	0,0209	2	1
Chronic myeloid leukemia (hsa05220)	0,0335	5	1
One carbon pool by folate (hsa00670)	0,0348	3	1
Glioma (hsa05214)	0,0351	5	1
S+L3:L19almonella infection (hsa05132)	0,0439	5	2

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 60			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)</b>	1,90 x 10 <sup>-5</sup>	2	1
<b>Morphine addiction (hsa05032)</b>	1,35 x 10 <sup>-4</sup>	6	2
<b>mRNA surveillance pathway (hsa03015)</b>	2,30 x 10 <sup>-4</sup>	6	2
<b>RNA transport (hsa03013)</b>	3,44 x 10 <sup>-4</sup>	9	2
<b>Nicotine addiction (hsa05033)</b>	1,37 x 10 <sup>-3</sup>	1	1
<b>Acute myeloid leukemia (hsa05221)</b>	4,37 x 10 <sup>-3</sup>	5	1
Retrograde endocannabinoid signaling(hsa04723)	5,40 x 10 <sup>-3</sup>	6	2
Wnt signaling pathway (hsa04310)	7,09 x 10 <sup>-3</sup>	7	2
Synaptic vesicle cycle (hsa04721)	7,71 x 10 <sup>-3</sup>	2	1
Circadian entrainment (hsa04713)	1,27 x 10 <sup>-2</sup>	8	2
Sphingolipid metabolism (hsa00600)	1,86 x 10 <sup>-2</sup>	2	2
Other types of O-glycan biosynthesis (hsa00514)	1,89 x 10 <sup>-2</sup>	1	1
Serotonergic synapse (hsa04726)	2,07 x 10 <sup>-2</sup>	7	2
Cholinergic synapse (hsa04725)	2,13 x 10 <sup>-2</sup>	5	1
Sulfur metabolism (hsa00920)	2,37 x 10 <sup>-2</sup>	3	1
Vasopressin-regulated water reabsorption(hsa04962)	2,71 x 10 <sup>-2</sup>	6	2
GABAergic synapse (hsa04727)	3,14 x 10 <sup>-2</sup>	4	1
Hedgehog signaling pathway (hsa04340)	3,29 x 10 <sup>-2</sup>	4	2
ErbB signaling pathway (hsa04012)	4,00 x 10 <sup>-2</sup>	5	2

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 62			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>Aldosterone-regulated sodium reabsorption(hsa04960)</b>	4,30E-07	6	3
<b>Biotin metabolism (hsa00780)</b>	1,52E-06	1	1
<b>TGF-beta signaling pathway (hsa04350)</b>	1,58E-06	15	3
<b>Ubiquitin mediated proteolysis (hsa04120)</b>	4,20E-05	17	3
<b>Neurotrophin signaling pathway (hsa04722)</b>	2,39E-04	20	3
<b>Non-small cell lung cancer (hsa05223)</b>	4,06E-04	9	2
<b>Chronic myeloid leukemia (hsa05220)</b>	1,00E-03	11	2
Insulin signaling pathway (hsa04910)	1,86E-03	14	2
Shigellosis (hsa05131)	2,98E-03	11	2
<b>PI3K-Akt signaling pathway (hsa04151)</b>	3,02E-03	38	3
B cell receptor signaling pathway (hsa04662)	3,06E-03	10	2
Endometrial cancer (hsa05213)	5,85E-03	9	2
Lysine degradation (hsa00310)	7,13E-03	5	2
Renal cell carcinoma (hsa05211)	8,10E-03	10	1
Long-term potentiation (hsa04720)	9,72E-03	10	2
Chemokine signaling pathway (hsa04062)	1,04E-02	18	2
Pancreatic cancer (hsa05212)	1,27E-02	10	2
Axon guidance (hsa04360)	1,31E-02	14	2
mRNA surveillance pathway (hsa03015)	1,38E-02	8	1
VEGF signaling pathway (hsa04370)	2,02E-02	7	1
mTOR signaling pathway (hsa04150)	2,24E-02	11	2
Adherens junction (hsa04520)	2,25E-02	7	2
Focal adhesion (hsa04510)	2,71E-02	15	1
Wnt signaling pathway (hsa04310)	2,82E-02	15	1
Acute myeloid leukemia (hsa05221)	3,08E-02	8	2
ErbB signaling pathway (hsa04012)	3,46E-02	11	2
Hepatitis B (hsa05161)	3,91E-02	12	1
HTLV-I infection (hsa05166)	4,34E-02	17	1
GnRH signaling pathway (hsa04912)	4,73E-02	9	2

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 72			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>Long-term potentiation (hsa04720)</b>	1,31E-07	21	4
<b>Gastric acid secretion (hsa04971)</b>	7,23E-06	12	2
<b>ErbB signaling pathway (hsa04012)</b>	1,39E-05	18	3
<b>Acute myeloid leukemia (hsa05221)</b>	2,35E-05	9	2
<b>Insulin signaling pathway (hsa04910)</b>	3,81E-05	18	3
<b>Transcriptional misregulation in cancer(hsa05202)</b>	4,15E-05	10	1
<b>mTOR signaling pathway (hsa04150)</b>	4,78E-05	15	4
<b>Endocrine and other factor-regulated calcium reabsorption (hsa04961)</b>	6,36E-05	9	2
<b>MAPK signaling pathway (hsa04010)</b>	8,11E-05	21	2
<b>Ubiquitin mediated proteolysis (hsa04120)</b>	8,55E-05	20	2
<b>Neurotrophin signaling pathway (hsa04722)</b>	1,60E-04	21	3
<b>Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)</b>	2,87E-04	3	1
<b>Gap junction (hsa04540)</b>	3,32E-04	11	2
<b>Dopaminergic synapse (hsa04728)</b>	6,94E-04	15	2
<b>Serotonergic synapse (hsa04726)</b>	7,64E-04	14	2
<b>GABAergic synapse (hsa04727)</b>	1,01E-03	12	2
<b>Retrograde endocannabinoid signaling(hsa04723)</b>	1,15E-03	8	1
<b>GnRH signaling pathway (hsa04912)</b>	1,52E-03	6	1
<b>Glutamatergic synapse (hsa04724)</b>	3,68E-03	7	1
<b>Regulation of actin cytoskeleton (hsa04810)</b>	3,92E-03	13	1
<b>Glioma (hsa05214)</b>	5,14E-03	5	1
<b>Cholinergic synapse (hsa04725)</b>	5,17E-03	9	1
p53 signaling pathway (hsa04115)	5,82E-03	12	2
Phosphatidylinositol signaling system (hsa04070)	6,37E-03	13	3
Melanogenesis (hsa04916)	6,79E-03	9	2
Wnt signaling pathway (hsa04310)	9,02E-03	16	2
Oocyte meiosis (hsa04114)	1,40E-02	14	2
mRNA surveillance pathway (hsa03015)	1,48E-02	6	1
Salivary secretion (hsa04970)	2,29E-02	9	2
PI3K-Akt signaling pathway (hsa04151)	2,56E-02	10	1
Alanine, aspartate and glutamate metabolism(hsa00250)	3,05E-02	6	2
Endometrial cancer (hsa05213)	3,12E-02	7	2
Prostate cancer (hsa05215)	4,28E-02	4	1
Circadian entrainment (hsa04713)	4,92E-02	6	1

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.



Nodo 78			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>ECM-receptor interaction (hsa04512)</b>	<1e-16	18	1
<b>Protein digestion and absorption (hsa04974)</b>	<1e-16	23	1
<b>Amoebiasis (hsa05146)</b>	<1e-16	19	1
<b>Prion diseases (hsa05020)</b>	<1e-16	1	2
<b>Focal adhesion (hsa04510)</b>	5,55E-16	26	1
<b>PI3K-Akt signaling pathway (hsa04151)</b>	4,64E-13	39	3
<b>Small cell lung cancer (hsa05222)</b>	4,87E-10	14	1
<b>Prostate cancer (hsa05215)</b>	1,02E-08	13	3
<b>Transcriptional misregulation in cancer (hsa05202)</b>	1,49E-08	21	4
<b>Glioma (hsa05214)</b>	1,92E-08	14	4
<b>Lysine degradation (hsa00310)</b>	2,08E-06	9	2
<b>Renal cell carcinoma (hsa05211)</b>	3,63E-06	5	2
<b>One carbon pool by folate (hsa00670)</b>	6,32E-06	3	2
<b>Glutamatergic synapse (hsa04724)</b>	6,43E-06	10	1
Non-small cell lung cancer (hsa05223)	0,00207043	8	3
Melanoma (hsa05218)	0,00659443	8	1
Pathways in cancer (hsa05200)	0,00715289	21	1
Nicotine addiction (hsa05033)	0,00958978	5	1
Neurotrophin signaling pathway (hsa04722)	0,02301084	7	1
Glycosaminoglycan biosynthesis - heparan sulfate / heparin (hsa00534)	0,0269272	3	1
Endometrial cancer (hsa05213)	0,04043012	3	1

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

Nodo 80			
KEGG pathway	p-value	Nº genes	Nº miRNAs
<b>Glycosaminoglycan biosynthesis - heparan sulfate / heparin (hsa00534)</b>	3,81E-10	5	2
<b>Drug metabolism - cytochrome P450 (hsa00982)</b>	3,36E-07	2	1
<b>Lysine degradation (hsa00310)</b>	3,95E-07	4	1
<b>Ubiquitin mediated proteolysis (hsa04120)</b>	5,51E-06	21	3
<b>Renal cell carcinoma (hsa05211)</b>	0,00030292	9	2
<b>Cell adhesion molecules (CAMs) (hsa04514)</b>	0,00482962	1	1
<b>Glycosaminoglycan biosynthesis - chondroitin sulfate (hsa00532)</b>	0,00508517	3	1
MAPK signaling pathway (hsa04010)	0,00569197	31	3
<b>Neurotrophin signaling pathway (hsa04722)</b>	0,00658001	22	3
Chagas disease (American trypanosomiasis)(hsa05142)	0,00999257	13	2
Pathways in cancer (hsa05200)	0,01256902	17	2
Endocytosis (hsa04144)	0,0127887	15	2
ErbB signaling pathway (hsa04012)	0,01902716	11	1
TGF-beta signaling pathway (hsa04350)	0,02577822	2	1
Long-term potentiation (hsa04720)	0,02828379	12	2
Tight junction (hsa04530)	0,03033858	12	2
Metabolism of xenobiotics by cytochrome P450(hsa00980)	0,03295383	2	1

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg. En negrita se indican las rutas que se mantienen significativas tras aplicar estadísticos muy conservadores.

**ANEXO VIII:** Rutas enriquecidas en los miRNAs seleccionados como representativos del perfil de cáncer de mama en mujeres jóvenes mediante DIANA miRPath (*pathway union*) empleando estadísticos muy conservadores

Ruta KEGG	p-valor	Nº genes	Nº miRNAs
Enfermedad por priones	<10 <sup>-16</sup>	1	1
Biosíntesis de ácidos grasos	1,22E-15	2	1
Proteólisis mediada por ubiquitina	1,60E-10	45	5
Señalización PI3K/Akt	6,80E-08	111	7
Señalización por neurotrofina	1,93E-07	53	6
Biosíntesis de condroitín sulfato	2,36E-06	5	3
Señalización MAPK	9,55E-06	83	6
TGF-beta	2,90E-05	28	5
Adhesiones focales	4,72E-05	70	5
Señalización ERBB	1,57E-04	28	3
Cáncer de próstata	3,01E-04	29	4
Guía de axones	3,47E-04	29	3
Potenciación a largo plazo	8,26E-04	31	6
Señalización Wnt	8,93E-04	44	3
Desregulación transcripcional en cáncer	9,80E-04	47	5
Señalización por insulina	2,48E-03	37	4
Señalización p53	3,26E-03	24	5
Carcinoma de células renales	9,83E-03	18	3
Cáncer de pulmón de células no pequeñas	1,54E-02	16	3
Glioma	1,59E-02	23	4
Sinapsis GABAérgica	2,29E-02	7	4
Infección por HTLV-I	2,43E-02	59	4
Endocitosis	2,91E-02	45	4
Sinapsis dopaminérgica	3,69E-02	42	5
Señalización mTOR	3,92E-02	24	4
Señalización del receptor de células B	4,41E-02	16	2

FDR se refiere al p-valor obtenido tras corrección por múltiples comparaciones de Benjamini & Hochberg mediante estadísticos conservadores.

## ANEXO IX: Valores estimados de represión debida a miRNAs desregulados en pacientes jóvenes de cáncer de mama en comparación con mayores.

Dianas sobre-expresadas con valores de represión debida a miRNAs desregulados en mujeres jóvenes negativas

Diana	Represión	Diana	Represión	Diana	Represión	Diana	Represión	Diana	Represión
ARID1A	-2,937	MGAT5	-0,340	ACTN4	-0,215	IL24	-0,152	ATP5A1	-0,131
PTPN11	-2,677	TGFA	-0,330	DDX5	-0,210	IL32	-0,152	SLC7A6	-0,130
USP31	-1,464	ECHS1	-0,328	CDKN1A	-0,208	TP73	-0,152	TUBA1A	-0,130
CD40	-1,457	R3HCC1L	-0,328	SMAP2	-0,207	LRRK2	-0,152	VEZT	-0,130
TMED1	-1,457	HS6ST2	-0,328	HIST2H2B	-0,207	AGO1	-0,151	ZNF418	-0,130
MKRN1	-1,435	MARCKSL1	-0,326	RNF41	-0,204	WDR82	-0,151	RSRC1	-0,130
IGLON5	-1,435	BCL2	-0,320	AKT2	-0,197	TNRC6A	-0,150	TNFRSF10	-0,130
IGF1R	-1,145	RPS3	-0,317	INPPL1	-0,197	SP1	-0,148	AGGF1	-0,130
RPS26	-0,804	VEGFA	-0,310	KRT7	-0,197	C1QBP	-0,146	EXOSC2	-0,130
FOS	-0,783	UBXN1	-0,307	RQCD1	-0,195	RPL19	-0,145	PAIP2	-0,130
TGS1	-0,783	RAN	-0,305	SF3B3	-0,193	PDAP1	-0,145	FBLN1	-0,130
CYP7A1	-0,674	GNB2L1	-0,299	RPS2	-0,191	NRD1	-0,145	EDC3	-0,127
CYP8B1	-0,672	CDK6	-0,284	EIF3C	-0,190	DZIP1	-0,144	ND6	-0,126
RPS7	-0,672	CYR61	-0,283	RANBP2	-0,186	GALNT2	-0,144	KHSRP	-0,126
MLF1	-0,672	LNPEP	-0,280	SEC14L1	-0,186	NUP98	-0,143	MAPK1	-0,126
SNX1	-0,593	CHMP2A	-0,274	ERBB2	-0,186	LRRC42	-0,143	CS	-0,124
GAPDH	-0,583	LAMB1	-0,270	YES1	-0,186	PSAP	-0,143	FLNA	-0,122
RTFDC1	-0,575	TRIM37	-0,269	KMT2C	-0,185	HIST1H3D	-0,143	CSDE1	-0,121
COMMD4	-0,575	TUBB	-0,267	PRKCE	-0,185	PBX2	-0,143	HIST1H1C	-0,120
SFXN1	-0,548	KLHL15	-0,266	CD44	-0,184	CHMP4C	-0,142	ZBTB10	-0,117
ZC3H18	-0,521	EEF2	-0,254	TMEM88	-0,184	LONP2	-0,142	LBR	-0,117
COX3	-0,516	KDM1A	-0,253	SMG5	-0,184	TIMP3	-0,142	FOXO3	-0,115
DHX40	-0,498	NFKBIA	-0,253	FAM92A1	-0,184	SEC63	-0,142	SLC25A5	-0,115
TRPC5	-0,498	FADS2	-0,252	C14orf93	-0,184	RAC3	-0,142	DYRK1A	-0,112
UROS	-0,498	TBC1D9B	-0,252	VPS29	-0,184	SLC7A2	-0,142	CCND3	-0,112
CDK13	-0,498	RBBP7	-0,252	NRCAM	-0,184	TMX3	-0,142	CCND2	-0,110
HIC2	-0,492	PPP2R2B	-0,249	TOP2A	-0,184	PITPNB	-0,142	SKI	-0,107
INTS2	-0,492	FAM47E	-0,249	CPSF3	-0,184	SMU1	-0,142	RAB36	-0,105
YDJC	-0,492	DGCR8	-0,249	LAMC1	-0,181	AP3S1	-0,142	E2F6	-0,105
C3orf17	-0,492	SERPINE2	-0,249	ARID3B	-0,180	VAC14	-0,142	OTUB1	-0,105
VPS53	-0,492	ANKRD27	-0,247	ALMS1	-0,178	FDPS	-0,142	CENPV	-0,105
CYB5B	-0,492	SNRPD3	-0,247	SRC	-0,178	SYK	-0,141	IMPDH1	-0,105
FGB	-0,464	HMGB1	-0,243	ENO1	-0,177	USMG5	-0,141	ZNF106	-0,105
FGA	-0,464	PLEKHB2	-0,240	EEF1A1	-0,174	KIRREL	-0,141	VAPB	-0,104
FGG	-0,464	FDXR	-0,236	HSPA1B	-0,174	PPP5C	-0,141	CD276	-0,104
PHF10	-0,464	CCDC47	-0,235	CTGF	-0,172	ZNF793	-0,141	SREBF2	-0,104
DNMT1	-0,455	TRIM49C	-0,235	PDHX	-0,172	BRCA1	-0,141	NRN1	-0,103
YWHAE	-0,450	SOAT1	-0,235	HSP90AB1	-0,171	TMEM97	-0,141	SSRP1	-0,103
COX1	-0,442	PAK2	-0,235	THBS1	-0,169	HRAS	-0,140	ANKRD52	-0,102
IFNG	-0,439	FKBP4	-0,233	MGRN1	-0,169	MGMT	-0,140	TEX261	-0,101
ANG	-0,437	HUWE1	-0,229	CPSF7	-0,169	C20orf173	-0,140	N4BP1	-0,101
CLSTN1	-0,410	PKNOX1	-0,227	MED1	-0,168	BORA	-0,140	MPL	-0,101
RPL27A	-0,385	MOV10L1	-0,227	BAZ2A	-0,165	FCHO2	-0,140	ZCCHC3	-0,101
RPS27	-0,381	CPNE1	-0,227	NAP1L1	-0,164	DHRS4	-0,140	TULP4	-0,101
SGSH	-0,380	EXTL3	-0,227	RPS14	-0,164	NID2	-0,140	AAAS	-0,101
HLA-G	-0,376	SLC36A1	-0,227	CA2	-0,163	GAB1	-0,140	ZBTB47	-0,101
IRS1	-0,376	ZNF227	-0,227	MEF2A	-0,163	C10orf35	-0,140	UNG	-0,101
EP300	-0,372	APP	-0,227	CAMKV	-0,162	TUFM	-0,140	CNN3	-0,101
DCTN2	-0,353	NFXL1	-0,227	PGAM1	-0,161	TPM4	-0,139	KLHL12	-0,101
ZSWIM6	-0,353	MCM2	-0,227	E2F5	-0,155	TMEM2	-0,135	TLK1	-0,101
ACTG1	-0,348	NONO	-0,224	SIGMAR1	-0,155	TRAP1	-0,135	SETD7	-0,101
SETDB1	-0,343	CTC1	-0,224	RPS4X	-0,154	GEMIN4	-0,135	SMYD1	-0,101
PITRM1	-0,343	PFN1	-0,223	LRP1	-0,152	SLC2A3	-0,134	SIK1	-0,101
DENND4B	-0,342	ERBB3	-0,221	ZEB1	-0,152	THOC5	-0,134	TPI1	-0,101
ASXL2	-0,340	YKT6	-0,216	ZEB2	-0,152	RPS5	-0,132	EZR	-0,101
								HMG2	-0,100

Dianas reprimidas con valores de represión debida a miRNAs desregulados en mujeres jóvenes positivos

Diana	Represión	Diana	Represión	Diana	Represión	Diana	Represión	Diana	Represión
NDE1	7,115	GALK1	2,465	AKT1	1,604	PTEN	1,051	SAMD4B	0,383
AMD1	6,633	POLR2I	2,439	MYBL2	1,603	JUNB	0,807	RNASEK	0,383
OSCP1	4,691	FASN	2,407	TGIF2	1,541	JUND	0,807	CACNG8	0,380
SH3BP2	3,293	TMEM69	1,918	E2F1	1,428	HSPG2	0,807	HIST1H2A	0,371
AMPD2	3,293	HIST2H3A	1,848	NUP62	1,250	CDK1	0,804	DDT	0,370
PDIA6	3,213	COQ4	1,804	WASF2	1,248	TCEAL1	0,771	DLST	0,315
MPRIIP	3,201	COX2	1,788	TRAPPC1	1,247	ACOX1	0,383	GPI	0,302
TRAF4	3,121	MTA2	1,763	ABI2	1,234	HIST1H2B	0,383	KIAA1671	0,250
RB1	2,499	HNRNPA0	1,751	ARNTL	1,153	USP6NL	0,383		
WDR77	2,466	ATP6	1,661	BCL2L11	1,087	SRGAP1	0,383		

Se muestran los genes con valor absoluto de represión > 0,1.

## ANEXO X: Rutas enriquecidas en genes diana validados por el total de miRNAs desregulados

#	Categoría	Ruta	p-valor	p-valor ajustado	Z-score	Score combinado
1	RNA	Translation	2,10E-08	2,85E-05	-2,01	21,02
2	otros	Influenza Life Cycle	6,61E-07	1,15E-04	-2,02	18,29
3	RNA	SRP-dependent cotranslational protein targeting to membrane	5,70E-07	1,15E-04	-1,99	18,03
4	RNA	Cap-dependent Translation Initiation	7,07E-07	1,15E-04	-1,99	18,02
5	RNA	Eukaryotic Translation Initiation	7,07E-07	1,15E-04	-1,97	17,86
6	RNA	Nonsense-Mediated Decay (NMD)	7,51E-07	1,15E-04	-1,95	17,71
7	RNA	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	7,51E-07	1,15E-04	-1,95	17,71
8	RNA	3'-UTR-mediated translational regulation	7,63E-07	1,15E-04	-1,92	17,44
9	otros	L13a-mediated translational silencing of Ceruloplasmin expression	7,63E-07	1,15E-04	-1,90	17,25
10	Ciclo celular	S Phase	1,01E-06	1,32E-04	-2,11	18,83
11	Ciclo celular	Mitotic G1-G1/S phases	1,08E-06	1,32E-04	-2,04	18,20
12	otros	Influenza Infection	1,25E-06	1,41E-04	-1,95	17,29
13	RNA	GTP hydrolysis and joining of the 60S ribosomal subunit	1,67E-06	1,74E-04	-1,88	16,30
14	otros	Influenza Viral RNA Transcription and Replication	2,01E-06	1,95E-04	-1,93	16,52
15	RNA	Formation of a pool of free 40S subunits	2,42E-06	2,18E-04	-1,87	15,75
16	RNA	Eukaryotic Translation Elongation	3,28E-06	2,78E-04	-1,87	15,28
17	RNA	Peptide chain elongation	4,44E-06	3,54E-04	-1,81	14,38
18	RNA	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	5,62E-06	4,23E-04	-1,85	14,35
19	RNA	DNA Replication	7,72E-06	5,23E-04	-1,99	15,03
20	RNA	Eukaryotic Translation Termination	7,65E-06	5,23E-04	-1,77	13,38
21	RNA	Synthesis of DNA	8,66E-06	5,59E-04	-1,99	14,88
22	Ciclo celular	G1/S Transition	1,18E-05	7,27E-04	-1,95	14,12
23	otros	Viral mRNA Translation	1,30E-05	7,68E-04	-1,73	12,40
24	RNA	Processing of Capped Intron-Containing Pre-mRNA	1,74E-05	9,20E-04	-1,78	12,43
25	RNA	mRNA Splicing	1,76E-05	9,20E-04	-1,74	12,15
26	RNA	mRNA Splicing - Major Pathway	1,76E-05	9,20E-04	-1,74	12,15
27	Ciclo celular	Cell Cycle Checkpoints	4,98E-05	2,50E-03	-1,89	11,35
28	Ciclo celular	Cellular responses to stress	6,87E-05	3,32E-03	-2,01	11,46
29	Ciclo celular	Regulation of DNA replication	8,93E-05	4,17E-03	-1,91	10,47
30	Ciclo celular	M/G1 Transition	1,03E-04	4,51E-03	-1,95	10,54
31	Ciclo celular	DNA Replication Pre-Initiation	1,03E-04	4,51E-03	-1,92	10,36
32	otros	HIV Infection	1,39E-04	5,88E-03	-2,09	10,74
33	Ciclo celular	Assembly of the pre-replicative complex	1,66E-04	6,80E-03	-1,90	9,47
34	otros	Removal of licensing factors from origins	1,99E-04	7,73E-03	-1,86	9,03
35		Orcl removal from chromatin	2,05E-04	7,73E-03	-1,85	9,01
36	Ciclo celular	Switching of origins to a post-replicative state	2,05E-04	7,73E-03	-1,82	8,87
37	Ciclo celular	M Phase	2,20E-04	8,07E-03	-2,00	9,62
38	RNA	Activation of the mRNA upon binding of the cap-binding complex and eIFs and subsequent binding to 43S	2,42E-04	8,62E-03	-1,61	7,65
39	RNA	Chromatin organization	2,62E-04	8,86E-03	-1,83	8,64
40	RNA	Chromatin modifying enzymes	2,62E-04	8,86E-03	-1,83	8,63
41	Ciclo celular	Cyclin E associated events during G1/S transition	2,84E-04	9,40E-03	-1,76	8,24
42	RNA	Translation initiation complex formation	3,20E-04	1,03E-02	-1,55	7,08
43	Ciclo celular	Cyclin A:Cdk2-associated events at S phase entry	3,61E-04	1,14E-02	-1,74	7,81
44	Ciclo celular	G1/S DNA Damage Checkpoints	3,94E-04	1,21E-02	-1,80	7,96
45	RNA	Transcription	4,34E-04	1,31E-02	-1,79	7,77
46	Ciclo celular	p53-Dependent G1/S DNA damage checkpoint	5,30E-04	1,53E-02	-1,72	7,17
47	Ciclo celular	p53-Dependent G1 DNA Damage Response	5,30E-04	1,53E-02	-1,72	7,17
48	RNA	Ribosomal scanning and start codon recognition	5,44E-04	1,53E-02	-1,47	6,12
49	apoptosis	Apoptosis	5,86E-04	1,62E-02	-1,64	6,76
50	PI3K	SCF(Skp2)-mediated degradation of p27/p21	6,96E-04	1,89E-02	-1,56	6,19
51	Wnt	Signaling by Wnt	8,05E-04	2,11E-02	-1,95	7,51
52	Ciclo celular	Regulation of mitotic cell cycle	8,27E-04	2,11E-02	-1,65	6,37
53	Ciclo celular	APC/C-mediated degradation of cell cycle proteins	8,27E-04	2,11E-02	-1,63	6,28
54	otros	Host Interactions of HIV factors	8,46E-04	2,12E-02	-1,71	6,58
55	Wnt	XAV939 inhibits tankyrase stabilizing AXIN	1,18E-03	2,70E-02	-1,76	6,37
56	Wnt	TCF dependent signaling in response to WNT	1,18E-03	2,70E-02	-1,77	6,39

#	Categoría	Ruta	p-valor	p-valor ajustado	Z-score	Score combinado
57	union	misspliced LRP5 mutants have enhanced beta-catenin-dependent	1,18E-03	2,70E-02	-1,76	6,37
58	wnt	RNF mutants show enhanced WNT signaling and proliferation	1,18E-03	2,70E-02	-1,76	6,34
59	RNA	CDT1 association with the CDC6:ORC:origin complex	1,12E-03	2,70E-02	-1,58	5,71
60	p53	Stabilization of p53	1,23E-03	2,78E-02	-1,49	5,33
61		Degradation of GLI2 by the proteasome	1,40E-03	3,01E-02	-1,54	5,41
62	Ciclo celular	Mitotic G2-G2/M phases	1,40E-03	3,01E-02	-1,53	5,34
63	RNA	Regulation of mRNA stability by proteins that bind AU-rich	1,46E-03	3,09E-02	-1,59	5,52
64	Ciclo celular	Regulation of APC/C activators between G1/S and early	1,48E-03	3,09E-02	-1,47	5,10
65	apoptosis	Programmed Cell Death	1,63E-03	3,26E-02	-1,51	5,16
66	Ciclo celular	Activation of APC/C and APC/C:Cdc20 mediated degradation	1,64E-03	3,26E-02	-1,46	5,00
67		Autodegradation of the E3 ubiquitin ligase COP1	1,61E-03	3,26E-02	-1,35	4,63
68	wnt	AXIN mutants destabilize the destruction complex activating	2,46E-03	3,37E-02	-1,33	4,51
69	wnt	Signaling by WNT in cancer	2,01E-03	3,37E-02	-1,68	5,68
70	wnt	deletions in the AXIN genes in hepatocellular carcinoma result in elevated WNT signaling	2,46E-03	3,37E-02	-1,49	5,04
71	union	S33 mutants of beta-catenin aren't phosphorylated	2,46E-03	3,37E-02	-1,48	5,03
72		degradation of DVL	1,89E-03	3,37E-02	-1,43	4,83
73		TCF7L2 mutants don't bind CTBP	2,46E-03	3,37E-02	-1,41	4,80
74	union	AMER1 mutants destabilize the destruction complex	2,46E-03	3,37E-02	-1,41	4,76
75	union	Degradation of beta-catenin by the destruction complex	2,46E-03	3,37E-02	-1,39	4,71
76	Ciclo celular	APC/C:Cdc20 mediated degradation of mitotic proteins	2,09E-03	3,37E-02	-1,38	4,69
77	union	S37 mutants of beta-catenin aren't phosphorylated	2,46E-03	3,37E-02	-1,38	4,67
78	Ciclo celular	G2/M Transition	2,15E-03	3,37E-02	-1,37	4,64
79	Ciclo celular	truncations of AMER1 destabilize the destruction complex	2,46E-03	3,37E-02	-1,36	4,61
80	Ciclo celular	deletions in the AMER1 gene destabilize the destruction complex	2,46E-03	3,37E-02	-1,36	4,60
81	Ciclo celular	APC truncation mutants are not K63 polyubiquitinated	2,46E-03	3,37E-02	-1,34	4,53
82	union	T41 mutants of beta-catenin aren't phosphorylated	2,46E-03	3,37E-02	-1,34	4,53
83	Ciclo celular	truncated APC mutants destabilize the destruction complex	2,46E-03	3,37E-02	-1,33	4,50
84	union	phosphorylation site mutants of CTNNB1 are not targeted to the proteasome by the destruction complex	2,46E-03	3,37E-02	-1,33	4,50
85	DNA	p53-Independent DNA Damage Response	2,04E-03	3,37E-02	-1,32	4,48
86	PI3K	SCF-beta-TrCP mediated degradation of Emi1	1,96E-03	3,37E-02	-1,32	4,47
87	wnt	AXIN missense mutants destabilize the destruction complex	2,46E-03	3,37E-02	-1,32	4,47
88	union	S45 mutants of beta-catenin aren't phosphorylated	2,46E-03	3,37E-02	-1,29	4,36
89	apoptosis	p53-Independent G1/S DNA damage checkpoint	2,04E-03	3,37E-02	-1,28	4,34
90		Ubiquitin Mediated Degradation of Phosphorylated Cdc25A	2,04E-03	3,37E-02	-1,28	4,34
91	Hedgehog	Degradation of GLI1 by the proteasome	2,26E-03	3,37E-02	-1,28	4,33
92	Hedgehog	GLI3 is processed to GLI3R by the proteasome	2,26E-03	3,37E-02	-1,28	4,33
93		misspliced GSK3beta mutants stabilize beta-catenin	2,46E-03	3,37E-02	-1,21	4,11
94	Wnt	APC truncation mutants have impaired AXIN binding	2,46E-03	3,37E-02	-1,20	4,06
95	Ciclo celular	Ubiquitin-dependent degradation of Cyclin D1	2,11E-03	3,37E-02	-1,17	3,96
96	Ciclo celular	Ubiquitin-dependent degradation of Cyclin D	2,11E-03	3,37E-02	-1,16	3,94
97	Ciclo celular	CDK-mediated phosphorylation and removal of Cdc6	2,11E-03	3,37E-02	-1,12	3,81
98	RNA	mRNA Splicing - Minor Pathway	2,38E-03	3,37E-02	-0,92	3,10
99	ER	ER-Phagosome pathway	2,54E-03	3,43E-02	-1,05	3,55
100		Vif-mediated degradation of APOBEC3G	2,55E-03	3,43E-02	-0,93	3,14
101	DNA	DNA Repair	2,66E-03	3,53E-02	-1,13	3,78
102		Regulation of activated PAK-2p34 by proteasome mediated	2,76E-03	3,63E-02	-0,75	2,50
103	immune	Downstream signaling events of B Cell Receptor (BCR)	2,93E-03	3,82E-02	-1,35	4,40
104	Ciclo celular	Mitotic Metaphase and Anaphase	3,11E-03	3,98E-02	-1,25	4,03
105	apoptosis	Cellular Senescence	3,09E-03	3,98E-02	-1,21	3,91
106	RNA	AUF1 (hnRNP D0) destabilizes mRNA	3,18E-03	4,03E-02	-0,87	2,78
107	Ciclo celular	APC:Cdc20 mediated degradation of cell cycle proteins prior to	3,39E-03	4,25E-02	-0,97	3,08
108	apoptosis	Regulation of Apoptosis	3,46E-03	4,30E-02	-0,75	2,36
109	Ciclo celular	Mitotic Anaphase	3,72E-03	4,59E-02	-1,20	3,71

## ANEXO XI: Rutas enriquecidas en genes diana validados más afectados por los miRNAs desregulados

#	Categoría	Ruta	p-valor	p-valor ajustado	Z-score	combinado
1	otros	Influenza Life Cycle	1,87E-11	2,18E-08	-2,033	35,867
2	otros	Influenza Infection	5,27E-11	3,08E-08	-2,035	35,196
3	otros	Influenza Viral RNA Transcription and Replication	2,62E-10	1,02E-07	-2,027	32,628
4	RNA	Eukaryotic Translation Elongation	4,04E-10	1,18E-07	-1,979	31,569
5	RNA	Peptide chain elongation	1,23E-09	2,88E-07	-1,929	29,048
6	RNA	Translation	3,51E-09	6,83E-07	-1,966	27,917
7	RNA	Nonsense-Mediated Decay (NMD)	6,11E-09	8,92E-07	-1,944	27,072
8	RNA	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction	6,11E-09	8,92E-07	-1,943	27,066
9	RNA	Formation of a pool of free 40S subunits	1,24E-08	1,32E-06	-1,906	25,803
10	RNA	Eukaryotic Translation Termination	1,24E-08	1,32E-06	-1,869	25,302
11	otros	Viral mRNA Translation	1,24E-08	1,32E-06	-1,857	25,142
12	RNA	Eukaryotic Translation Initiation	2,03E-08	1,79E-06	-1,906	25,221
13	RNA	Cap-dependent Translation Initiation	2,03E-08	1,79E-06	-1,906	25,221
14	RNA	SRP-dependent cotranslational protein targeting to membrane	2,15E-08	1,79E-06	-1,880	24,871
15	RNA	3'-UTR-mediated translational regulation	3,36E-08	2,46E-06	-1,851	23,909
16		L13a-mediated translational silencing of Ceruloplasmin expression	3,36E-08	2,46E-06	-1,848	23,872
17	RNA	Nonsense Mediated Decay (NMD) independent of the Exon Junction	3,75E-08	2,58E-06	-1,857	23,896
18	RNA	GTP hydrolysis and joining of the 60S ribosomal subunit	4,09E-08	2,66E-06	-1,839	23,609
19	Estrés	Cellular responses to stress	9,20E-08	5,66E-06	-2,089	25,242
20	RNA	Processing of Capped Intron-Containing Pre-mRNA	6,67E-07	3,90E-05	-1,812	18,399
21	ciclo celular	Mitotic G1-G1/S phases	2,44E-06	1,36E-04	-1,951	17,377
22	RNA	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	4,93E-06	2,62E-04	-1,811	14,934
23	RNA	Formation of the ternary complex, and subsequently, the 43S complex	7,46E-06	3,79E-04	-1,809	14,251
24	RNA	Ribosomal scanning and start codon recognition	1,20E-05	5,59E-04	-1,770	13,252
25	RNA	Translation initiation complex formation	1,20E-05	5,59E-04	-1,762	13,198
26	RNA	mRNA Splicing	1,81E-05	7,81E-04	-1,738	12,438
27	RNA	mRNA Splicing - Major Pathway	1,81E-05	7,81E-04	-1,720	12,305
28	immune	TRAF6 Mediated Induction of proinflammatory cytokines	2,66E-05	1,11E-03	-1,775	12,079
29	apoptosis	Cellular Senescence	4,63E-05	1,87E-03	-1,878	11,799
30	MAPK	Toll Like Receptor 10 (TLR10) Cascade	5,45E-05	1,98E-03	-1,797	11,187
31	immune	Toll Like Receptor 5 (TLR5) Cascade	5,45E-05	1,98E-03	-1,766	10,997
32		MyD88 cascade initiated on plasma membrane	5,45E-05	1,98E-03	-1,766	10,995
33	Estrés	Oxidative Stress Induced Senescence	5,59E-05	1,98E-03	-1,742	10,845
34	immune	Toll Like Receptor 9 (TLR9) Cascade	6,49E-05	2,17E-03	-1,819	11,159
35	MAPK	TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation	6,37E-05	2,17E-03	-1,768	10,848
36	immune	Toll Like Receptor 7/8 (TLR7/8) Cascade	8,63E-05	2,72E-03	-1,780	10,510
37		MyD88 dependent cascade initiated on endosome	8,63E-05	2,72E-03	-1,780	10,510
38	ciclo celular	Mitotic G2-G2/M phases	1,40E-04	4,30E-03	-1,752	9,546
39		MyD88:Mal cascade initiated on plasma membrane	2,32E-04	6,31E-03	-1,756	8,898
40	immune	Toll Like Receptor TLR1:TLR2 Cascade	2,32E-04	6,31E-03	-1,747	8,849
41	immune	Toll Like Receptor 2 (TLR2) Cascade	2,32E-04	6,31E-03	-1,728	8,756
42	immune	Toll Like Receptor TLR6:TLR2 Cascade	2,32E-04	6,31E-03	-1,719	8,709
43	RNA	Transcriptional regulation by small RNAs	2,30E-04	6,31E-03	-1,646	8,341
44	egfr	Signaling by ERBB4	2,56E-04	6,81E-03	-1,809	9,027
45	MAPK	Downstream signal transduction	2,63E-04	6,82E-03	-1,837	9,164
46	otros	Interactions of Rev with host cellular proteins	2,71E-04	6,88E-03	-1,458	7,259
47	egfr	Signaling by ERBB2	3,20E-04	7,90E-03	-1,864	9,022
48	MAPK	MAP kinase activation in TLR cascade	3,25E-04	7,90E-03	-1,528	7,399
49	otros	HIV Life Cycle	3,53E-04	8,41E-03	-1,716	8,199
50	otros	Nuclear import of Rev protein	4,24E-04	9,90E-03	-1,279	5,902
51	RNA	Regulatory RNA pathways	4,42E-04	1,01E-02	-1,563	7,182
52	ciclo celular	G2/M Transition	5,08E-04	1,14E-02	-1,613	7,216
53	otros	HIV Infection	5,36E-04	1,16E-02	-1,884	8,395
54	apoptosis	Oncogene Induced Senescence	5,34E-04	1,16E-02	-1,288	5,741
55	pdgfr	Signaling by PDGF	6,18E-04	1,28E-02	-1,780	7,762
56	ciclo celular	G1/S Transition	6,02E-04	1,28E-02	-1,652	7,204
57	ciclo celular	G1 Phase	6,45E-04	1,28E-02	-1,303	5,681

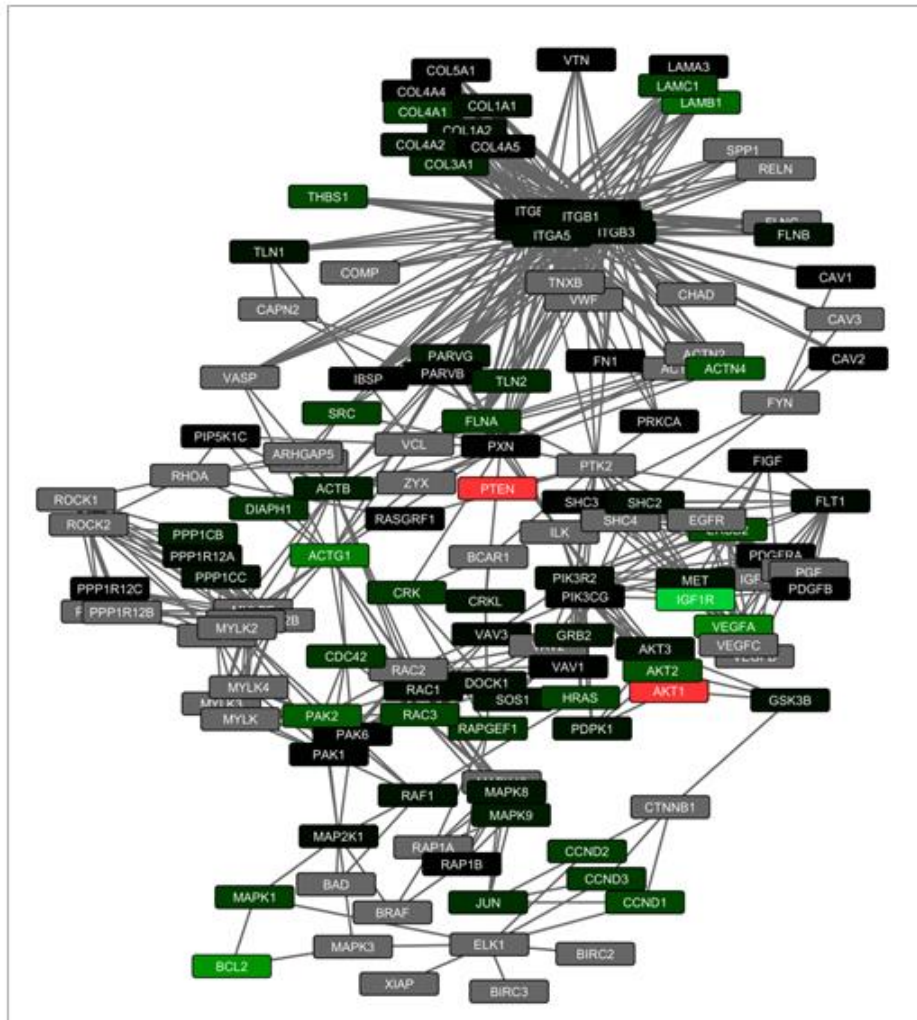


#	Categorfa	Ruta	p-valor	p-valor ajustado	Z-score	Score combinad
58	ciclo celular	Cyclin D associated events in G1	6,45E-04	1,28E-02	-1,303	5,681
59	apoptosis	Intrinsic Pathway for Apoptosis	6,45E-04	1,28E-02	-1,218	5,309
60	ciclo celular	M Phase	7,24E-04	1,37E-02	-1,775	7,614
61		Signaling by SCF-KIT	7,21E-04	1,37E-02	-1,663	7,136
62	ciclo celular	S Phase	7,39E-04	1,37E-02	-1,654	7,096
63	otros	Late Phase of HIV Life Cycle	7,39E-04	1,37E-02	-1,599	6,862
64	RNA	Chromatin organization	7,66E-04	1,38E-02	-1,610	6,898
65	RNA	Chromatin modifying enzymes	7,66E-04	1,38E-02	-1,609	6,896
66	egfr	Signaling by Ligand-Responsive EGFR Variants in Cancer	8,78E-04	1,42E-02	-1,675	7,123
67	egfr	Signaling by EGFRvIII in Cancer	8,78E-04	1,42E-02	-1,666	7,082
68	egfr	Signaling by EGFR in Cancer	8,78E-04	1,42E-02	-1,657	7,043
69	immune	Toll Like Receptor 3 (TLR3) Cascade	8,46E-04	1,42E-02	-1,504	6,393
70		MyD88-independent cascade	8,46E-04	1,42E-02	-1,502	6,384
71		TRIF-mediated TLR3/TLR4 signaling	8,46E-04	1,42E-02	-1,476	6,275
72	RNA	NEP/NS2 Interacts with the Cellular Export Machinery	8,39E-04	1,42E-02	-0,870	3,699
73	otros	Glucose transport	9,56E-04	1,53E-02	-0,994	4,157
74	otros	Host Interactions of HIV factors	1,01E-03	1,59E-02	-1,531	6,341
75	fgfr	Signaling by FGFR	1,06E-03	1,61E-02	-1,570	6,480
76	otros	Export of Viral Ribonucleoproteins from Nucleus	1,05E-03	1,61E-02	-0,726	2,998
77		Activation of BH3-only proteins	1,04E-03	1,61E-02	-0,543	2,240
78	fgfr	Signaling by FGFR in disease	1,14E-03	1,71E-02	-1,586	6,450
79	otros	Gluconeogenesis	1,31E-03	1,93E-02	-0,672	2,654
80	egfr	Signaling by Overexpressed Wild-Type EGFR in Cancer	1,35E-03	1,95E-02	-1,556	6,128
81	egfr	Signaling by EGFR	1,35E-03	1,95E-02	-1,556	6,128
82		GAB1 signalosome	1,49E-03	2,12E-02	-1,399	5,389
83	RNA	Transport of Mature Transcript to Cytoplasm	1,53E-03	2,15E-02	-1,002	3,846
84		Rev-mediated nuclear export of HIV RNA	1,61E-03	2,24E-02	-0,703	2,671
85	int	Integrin alphaIIb beta3 signaling	1,66E-03	2,28E-02	-0,461	1,743
86	RNA	Metabolism of non-coding RNA	1,74E-03	2,33E-02	-0,945	3,553
87	RNA	snRNP Assembly	1,74E-03	2,33E-02	-0,918	3,451
88	ciclo celular	Mitotic Prophase	1,85E-03	2,45E-02	-1,218	4,516
89	RNA	Nuclear Pore Complex (NPC) Disassembly	1,97E-03	2,58E-02	-0,676	2,471
90	RNA	Transport of Mature mRNA derived from an Intron-Containing	2,04E-03	2,61E-02	-0,902	3,289
91	Estrés	Cellular response to heat stress	2,04E-03	2,61E-02	-0,828	3,018
92		Activated TLR4 signalling	2,11E-03	2,67E-02	-1,317	4,770
93	fgfr	Downstream signaling of activated FGFR	2,17E-03	2,71E-02	-1,391	5,020
94	immune	Toll-Like Receptors Cascades	2,18E-03	2,71E-02	-1,333	4,810
95	RNA	Regulation of DNA replication	2,21E-03	2,72E-02	-1,182	4,260
96	otros	Interactions of Vpr with host cellular proteins	2,39E-03	2,90E-02	-0,632	2,235
97		Fc epsilon receptor (FCERI) signaling	2,51E-03	3,03E-02	-1,395	4,878
98	RNA	Transport of Ribonucleoproteins into the Host Nucleus	2,55E-03	3,04E-02	-0,240	0,839
99	immune	Toll Like Receptor 4 (TLR4) Cascade	2,60E-03	3,06E-02	-1,296	4,518
100		Hexose transport	2,72E-03	3,18E-02	-0,715	2,467
101	dna	DNA Damage/Telomere Stress Induced Senescence	2,75E-03	3,18E-02	-0,852	2,938
102	ciclo celular	Cell Cycle Checkpoints	3,11E-03	3,56E-02	-1,231	4,105
103	ciclo celular	G0 and Early G1	3,28E-03	3,72E-02	0,077	-0,255
104		Translocation of GLUT4 to the plasma membrane	3,62E-03	4,06E-02	-0,721	2,309
105	immune	Downstream signaling events of B Cell Receptor (BCR)	3,72E-03	4,13E-02	-1,339	4,266
106		Vpr-mediated nuclear import of PICs	3,79E-03	4,17E-02	-0,236	0,750
107	pi3k	PI3K/AKT Signaling in Cancer	3,99E-03	4,18E-02	-1,145	3,634
108	egfr	PI3K events in ERBB2 signaling	3,99E-03	4,18E-02	-1,135	3,604
109	pi3k	PI-3K cascade	3,99E-03	4,18E-02	-1,125	3,571
110	egfr	PI3K events in ERBB4 signaling	3,99E-03	4,18E-02	-1,090	3,459
111	pi3k	PIP3 activates AKT signaling	3,99E-03	4,18E-02	-1,085	3,446
112	ciclo celular	Regulation of PLK1 Activity at G2/M Transition	4,01E-03	4,18E-02	-0,922	2,927
113		DAP12 signaling	4,11E-03	4,21E-02	-1,221	3,869
114	pdgfr	Platelet Aggregation (Plug Formation)	4,10E-03	4,21E-02	-0,314	0,996
115		SOS-mediated signalling	4,17E-03	4,23E-02	0,386	-1,221
116	egfr	SHC1 events in ERBB2 signaling	4,56E-03	4,56E-02	-0,388	1,199
117	vegfr	VEGFR2 mediated cell proliferation	4,56E-03	4,56E-02	-0,372	1,150
118	RNA	HuR stabilizes mRNA	4,87E-03	4,82E-02	4,251	-12,889
119	otros	Nuclear Envelope Breakdown	4,96E-03	4,84E-02	-0,510	1,544
120	ciclo celular	G2/M Checkpoints	4,97E-03	4,84E-02	-0,355	1,075
121	igfr	IRS-related events triggered by IGF1R	5,17E-03	4,99E-02	-0,945	2,833

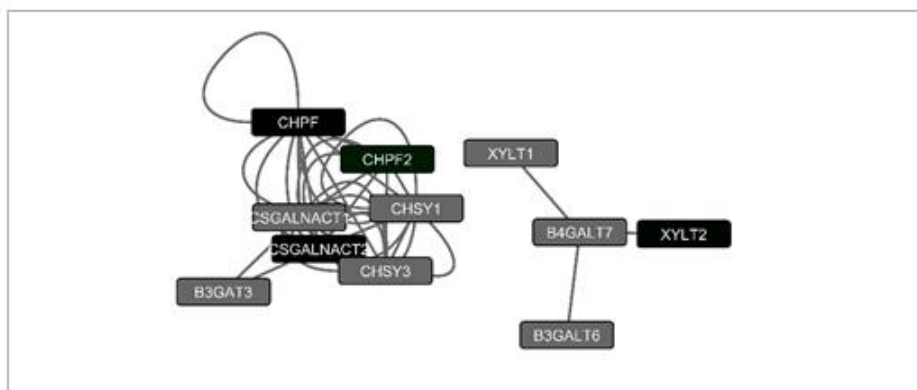


ANEXO XII: Interacciones en las rutas seleccionadas de los genes diana validados y teóricos más afectados por los miRNAs desregulados en mujeres jóvenes

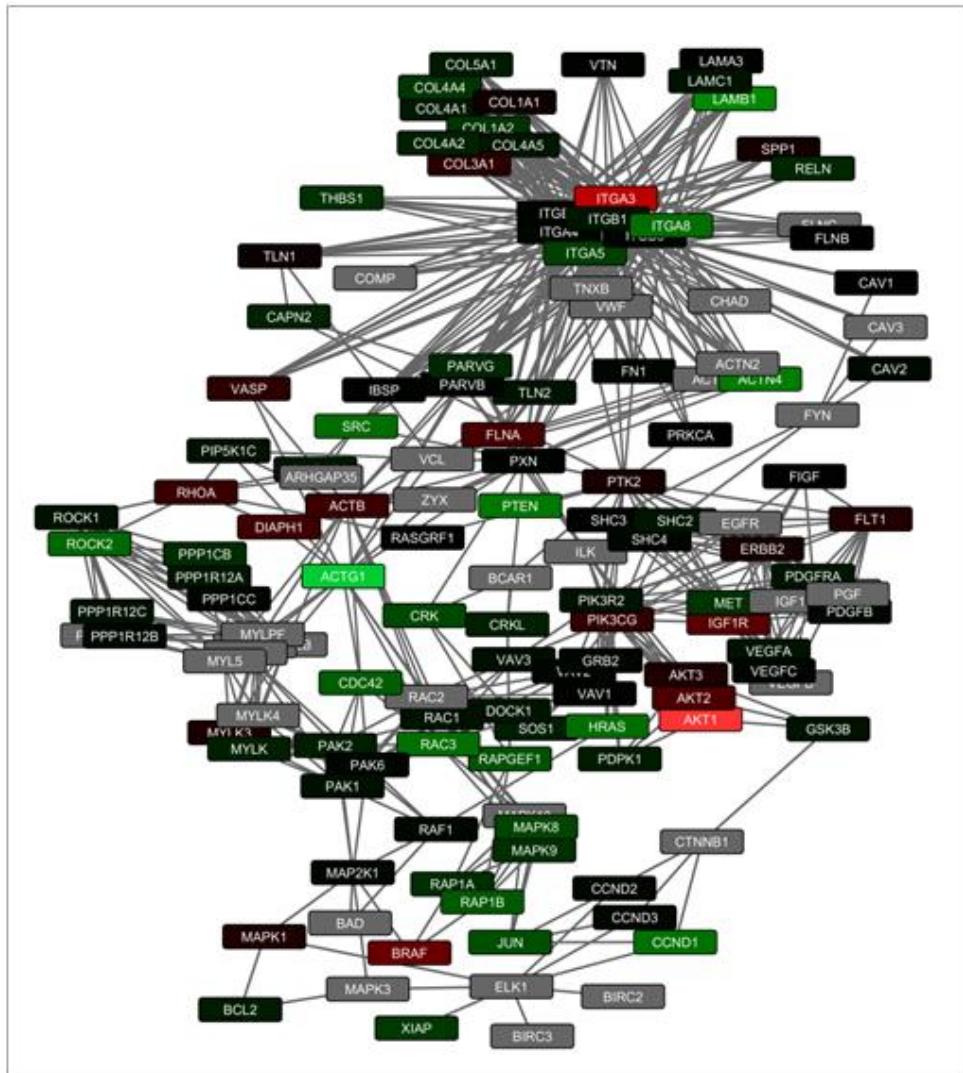
Dianas validadas Adhesiones focales (hsa04510)



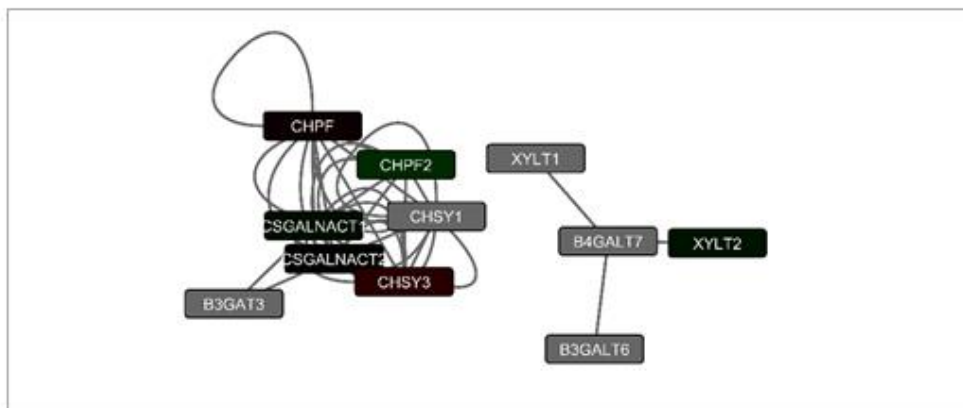
Dianas validadas Biosíntesis de condroitín sulfato (hsa00532)



Dianas teóricas Adhesiones focales (hsa04510)



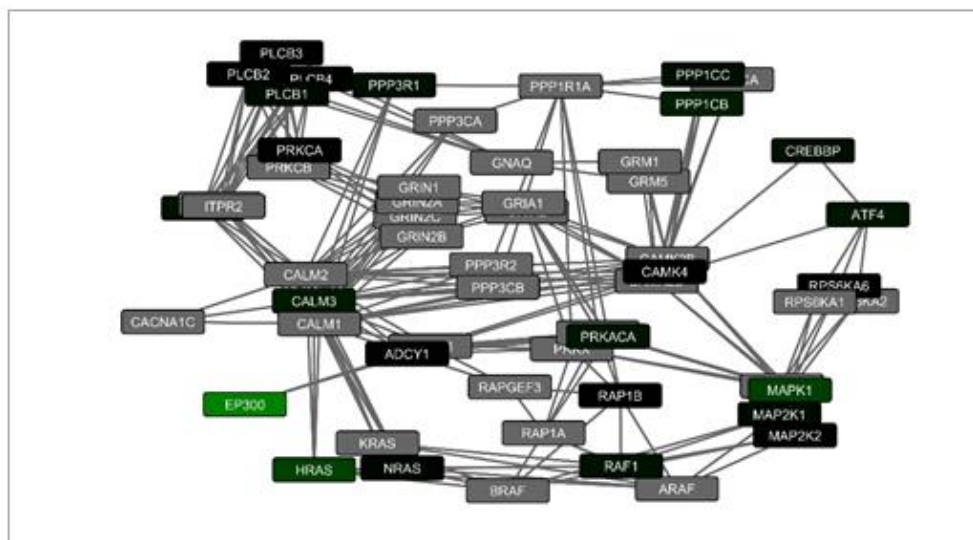
Dianas teóricas Biosíntesis de condroitín sulfato (hsa00532)



Dianas validadas Regulación del citoesqueleto de actina (hsa04810)



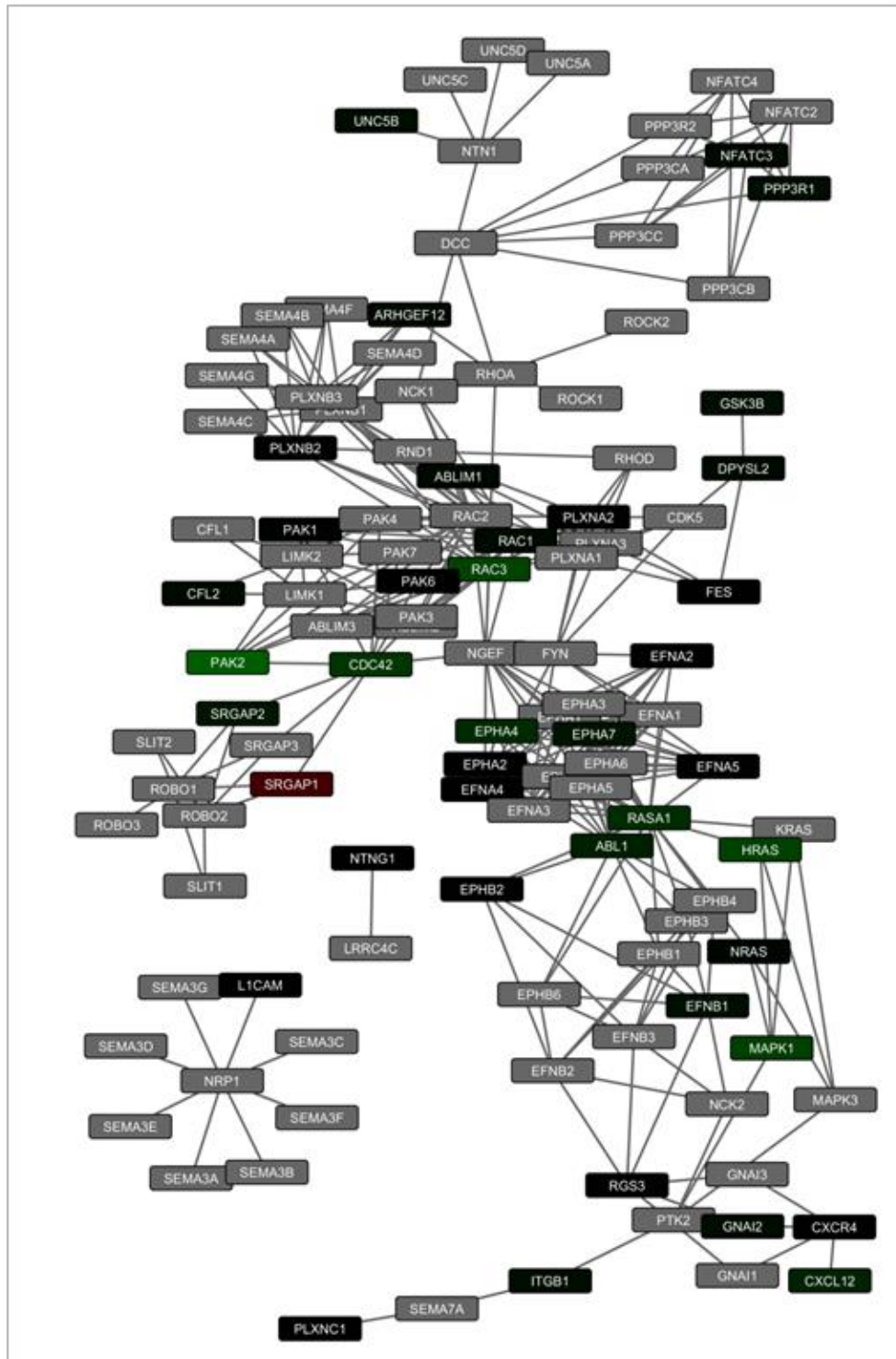
Dianas validadas Potenciación a largo plazo (hsa04720)



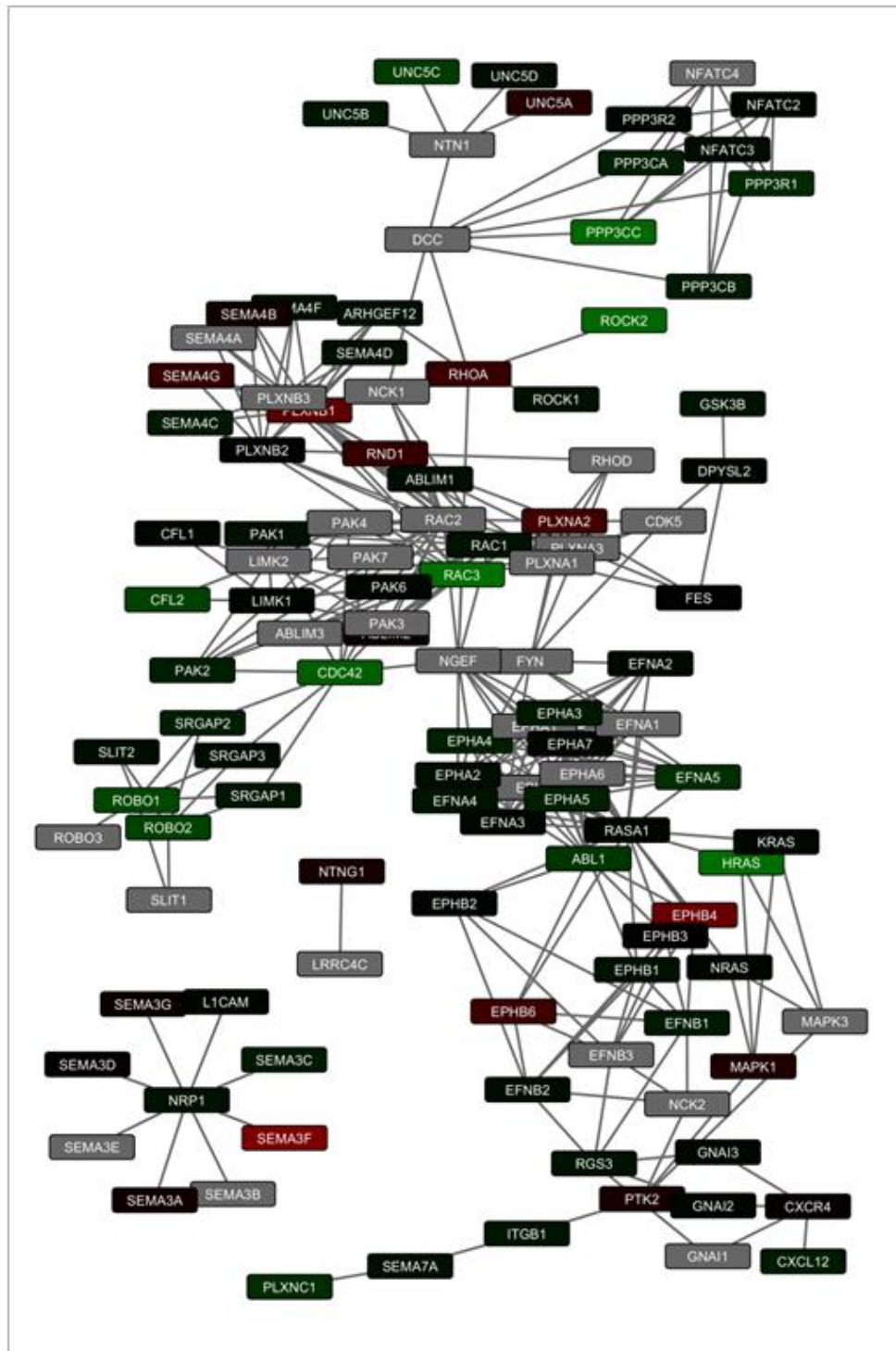




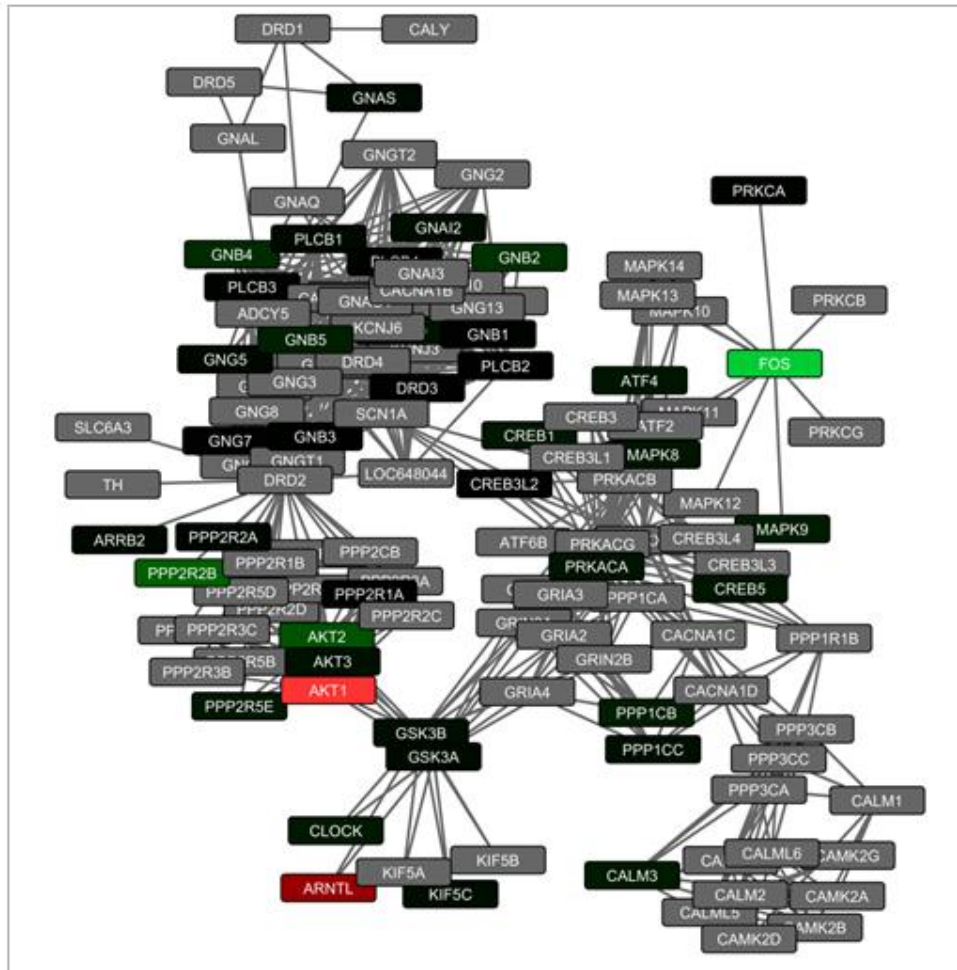
## Dianas validadas Guía de axones (hsa04360)



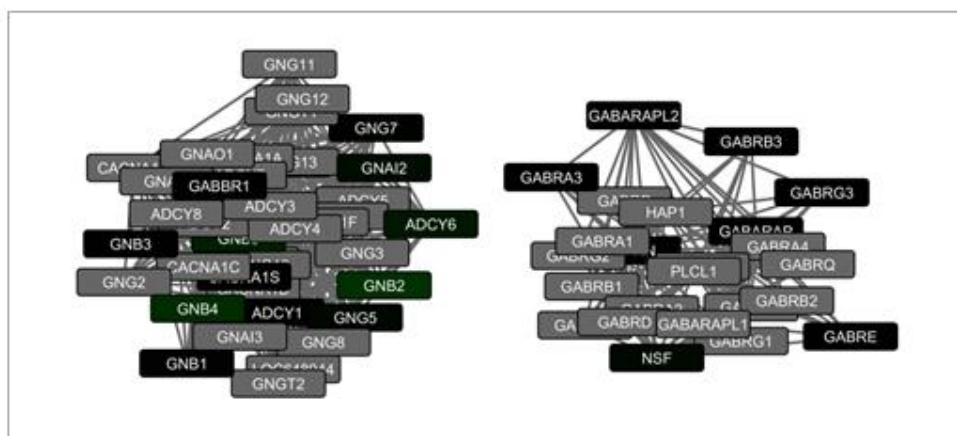
Dianas teóricas Guía de axones (hsa04360)



Dianas validadas Sinapsis dopaminérgica (hsa04728)

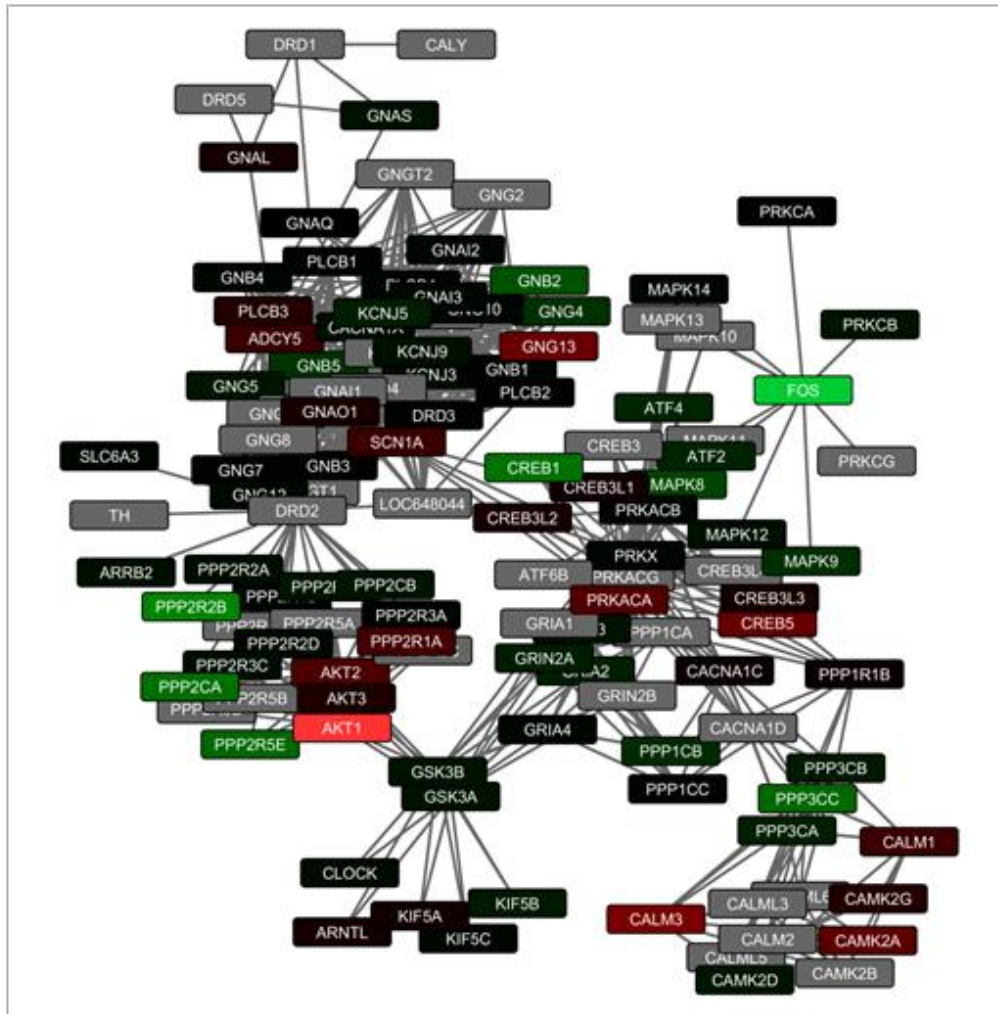


Dianas validadas Sinapsis GABAérgica (hsa04727)

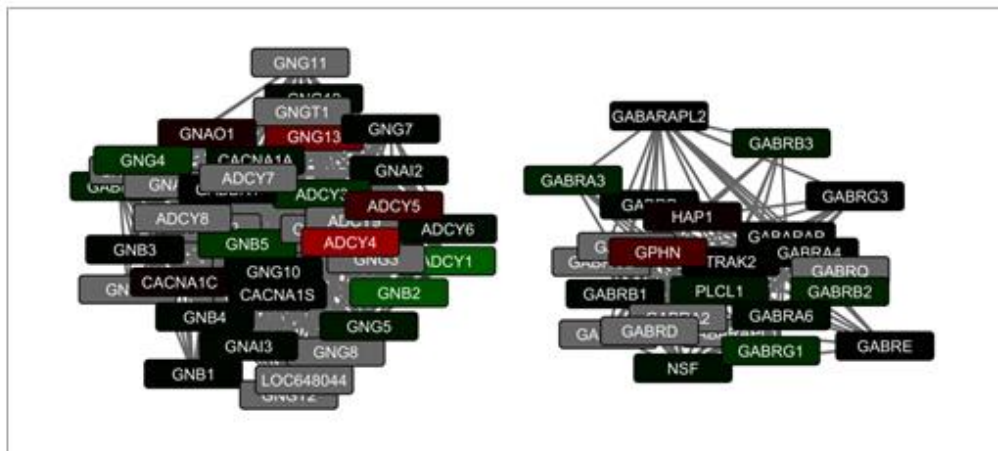




Dianas teóricas Sinapsis dopaminérgica (hsa04728)

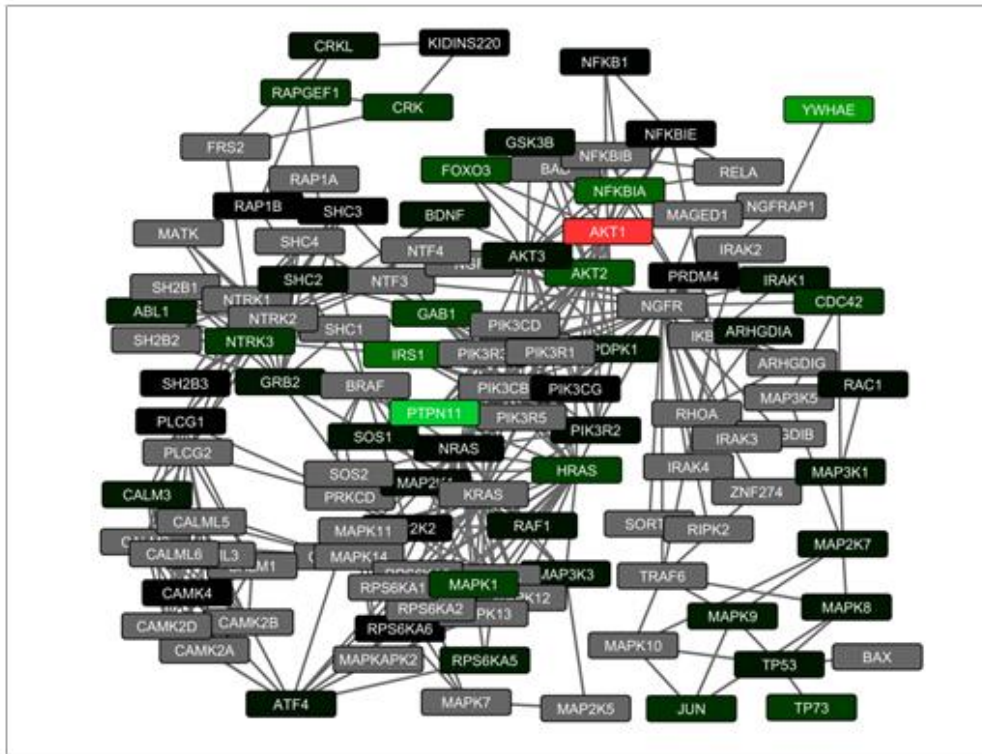


Dianas teóricas Sinapsis GABAérgica (hsa04727)

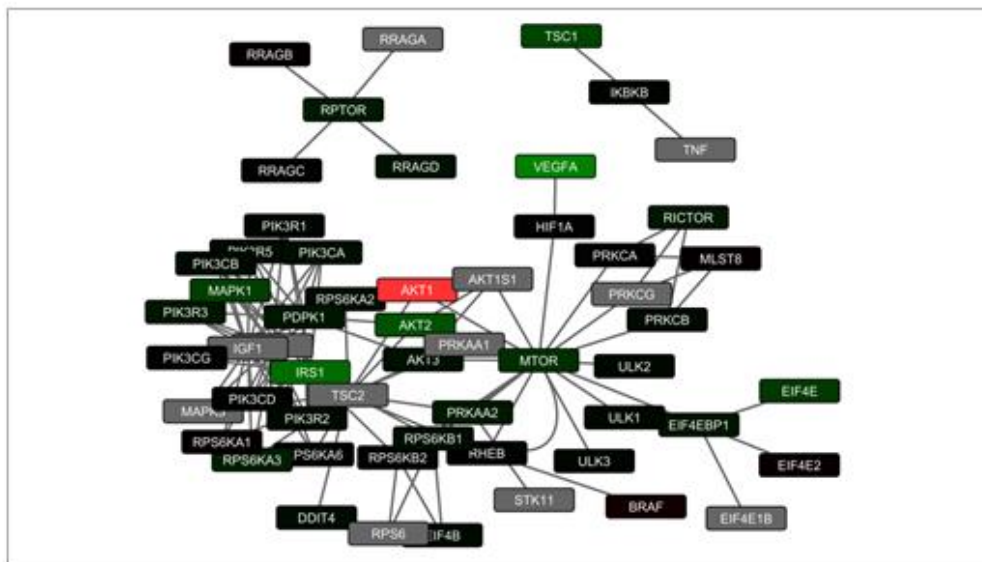




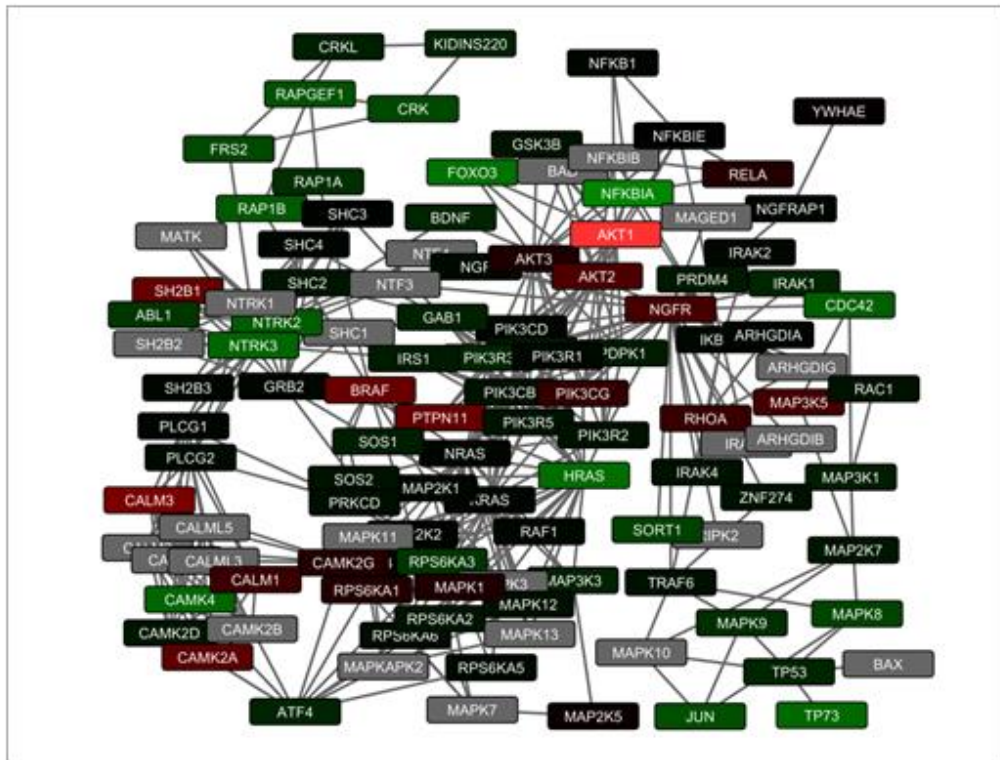
Dianas validadas Señalización por neurotrofina (hsa04722)



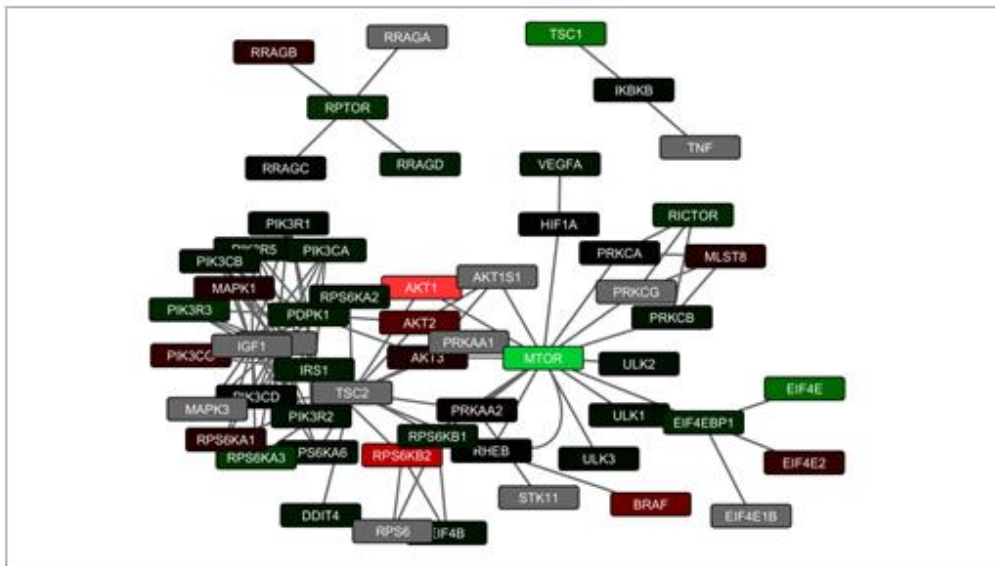
Dianas validadas Señalización mTOR (hsa04150)



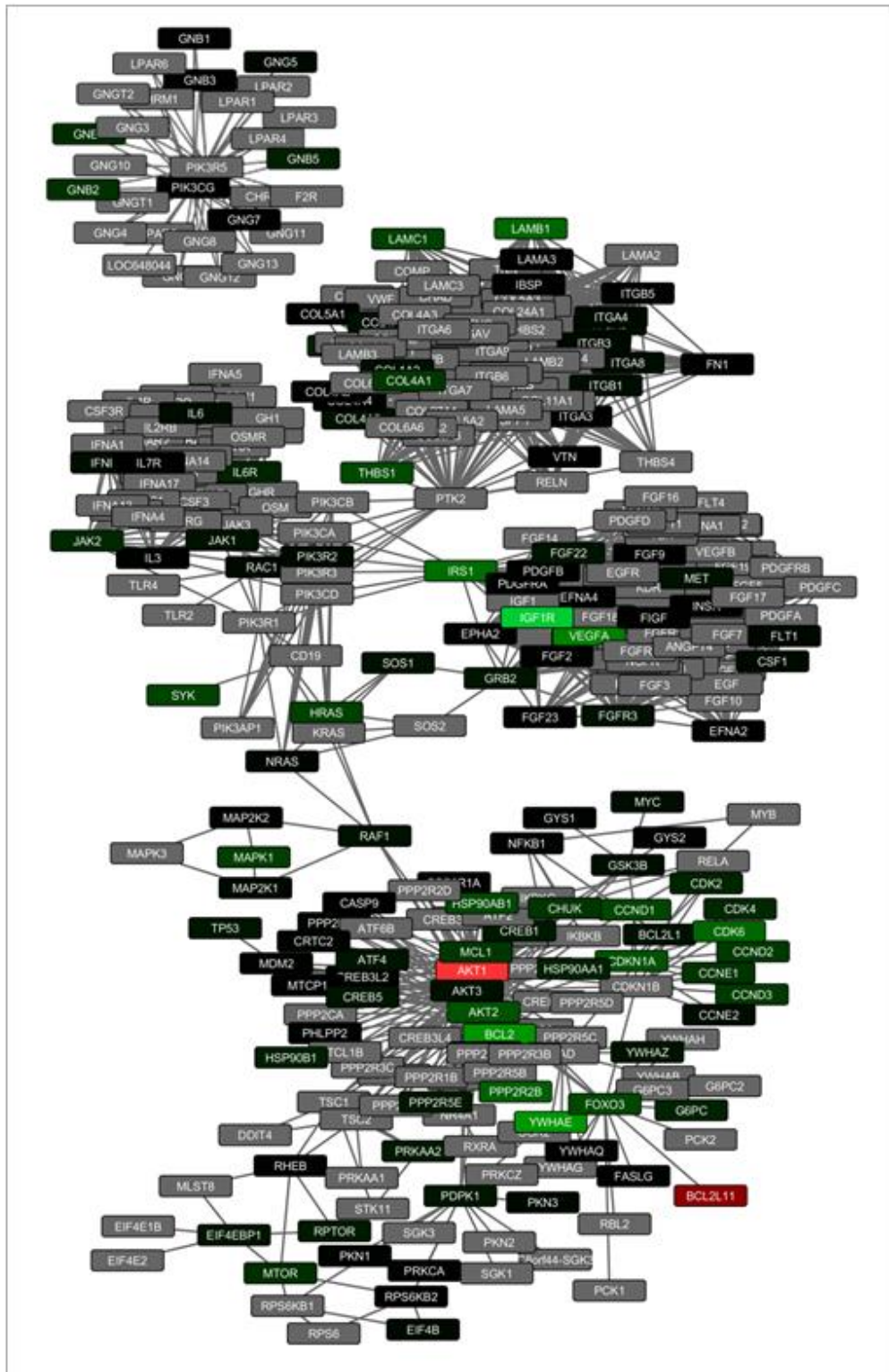
Dianas teóricas Señalización por neurotrofina (hsa04722)



Dianas teóricas Señalización mTOR (hsa04150)



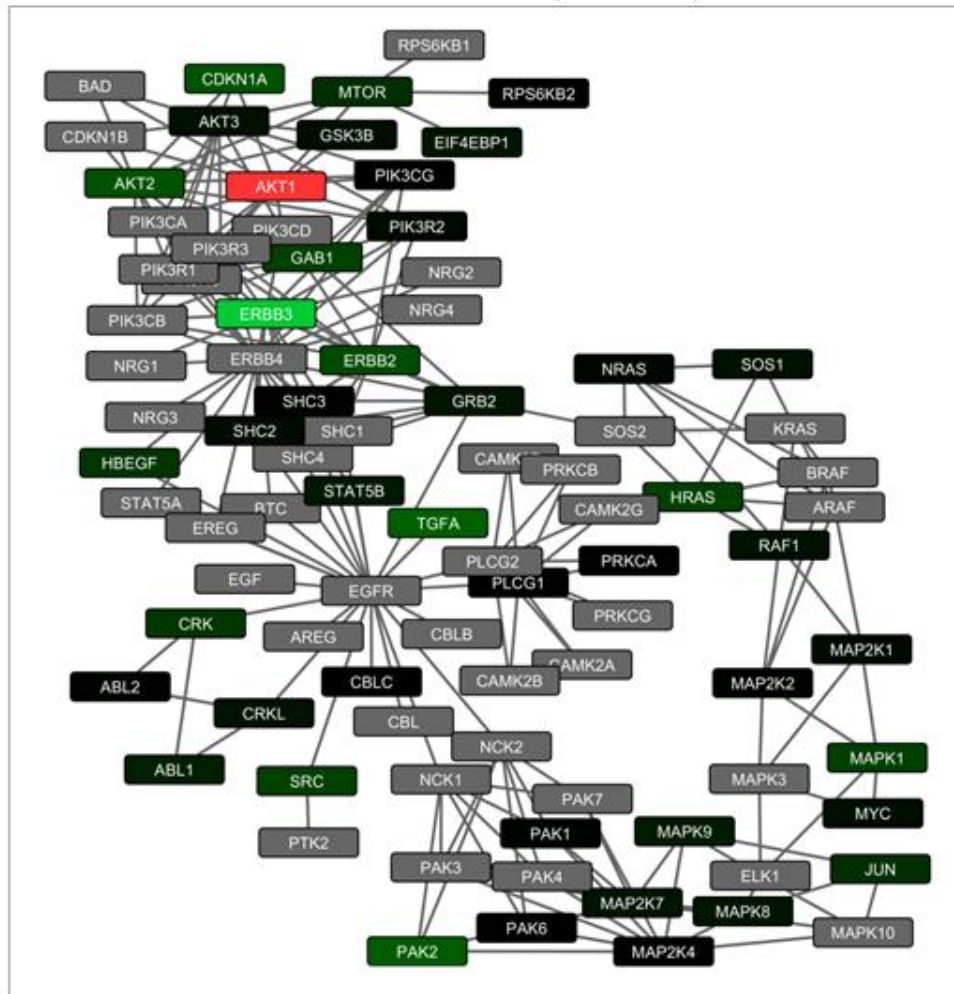
Dianas validadas Señalización PI3K (hsa04151)



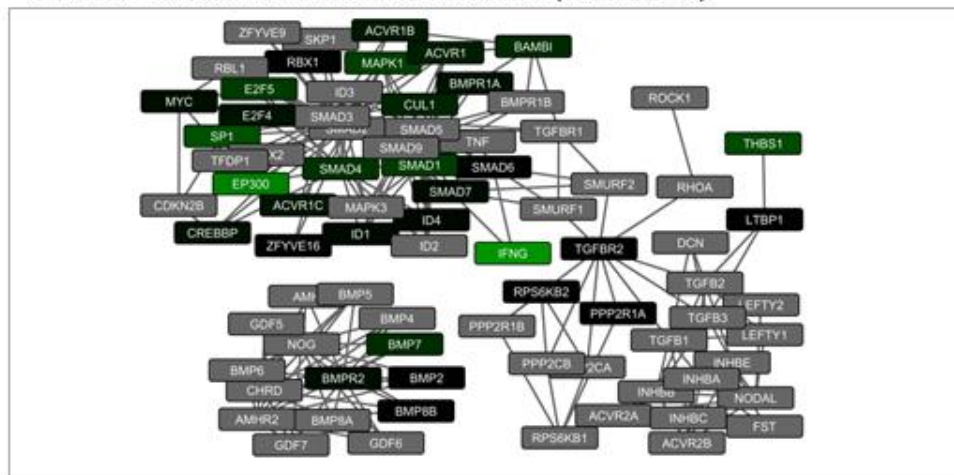




## Dianas validadas Señalización ERBB (hsa04012)



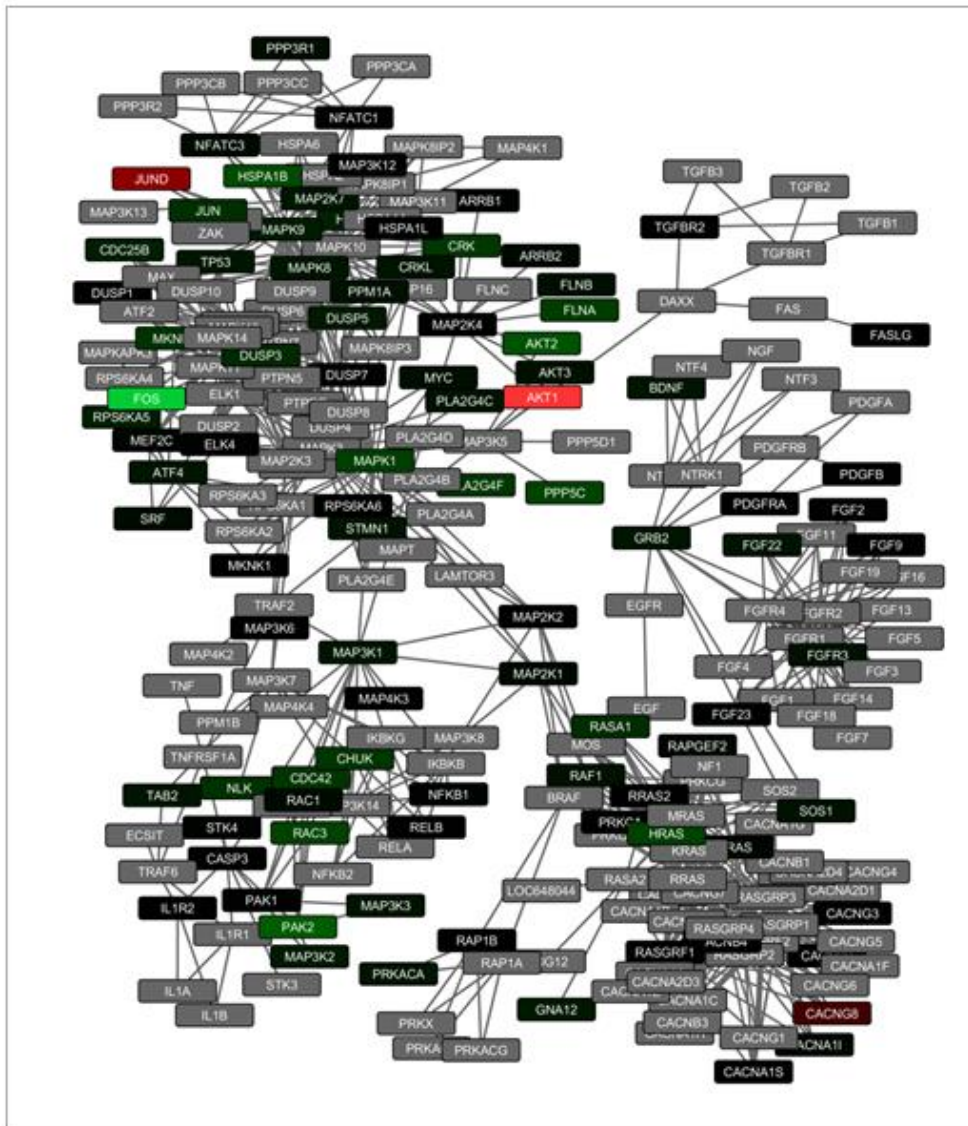
## Dianas validadas Señalización TGFβ (hsa04350)







Dianas validadas Señalización MAPK (hsa04010)







**ANEXO XIII:** Valores de significación y de expresión de los miRNAs con expresión significativamente diferente entre mujeres jóvenes y mayores en, al menos, dos de los tres estudios realizados.

miRNA	BCA		CLA		MBA	
	FDR	Expresión	FDR	Expresión	FDR	Expresión
<b>BCA+CLA+MBD</b>						
hsa-miR-30c	0,002	-3,424	0,008	-3,178	0,002	-4,927
hsa-miR-489	0,046	-2,602	0,009	-2,510	0,004	-5,748
hsa-miR-663	0,001	4,036	0,010	4,003	0,031	3,352
<b>BCA+CLA</b>						
hsa-miR-4270	0,007	3,801	0,001	12,423	-	-
hsa-miR-1207-5p	0,002	4,861	0,001	11,542	-	-
hsa-miR-4281	0,002	4,608	0,001	7,627	-	-
hsa-miR-3175	0,001	4,867	0,001	7,480	-	-
hsa-miR-3141	0,001	5,295	0,001	7,251	-	-
hsa-miR-23a	0,002	-4,478	0,001	-7,105	0,784	0,501
hsa-miR-762	0,001	6,646	0,001	6,683	-	-
hsa-miR-1268	0,001	5,013	0,001	6,637	-	-
hsa-miR-92b*	0,001	4,981	0,001	6,333	-	-
hsa-miR-3162	0,001	5,121	0,001	5,378	-	-
hsa-miR-27b	0,002	-3,637	0,001	-5,283	0,597	-0,846
hsa-miR-665	0,002	4,471	0,001	3,431	0,174	-2,269
hsa-miR-26b	0,046	-2,561	0,002	-4,229	0,162	-2,220
hsa-miR-149*	0,001	6,423	0,003	5,162	0,873	0,276
hsa-miR-1224-5p	0,001	5,000	0,003	4,381	-	-
hsa-miR-3196	0,001	6,745	0,003	5,425	-	-
hsa-miR-1202	0,001	5,226	0,003	4,757	-	-
hsa-miR-3197	0,001	5,768	0,004	4,478	-	-
hsa-miR-4284	0,010	3,143	0,004	4,121	-	-
hsa-miR-638	0,001	4,691	0,006	4,556	0,860	-0,308
hsa-miR-1915	0,004	3,836	0,007	4,215	-	-
hsa-miR-1469	0,029	2,771	0,007	4,299	-	-
hsa-miR-2861	0,001	4,355	0,008	4,245	-	-
hsa-miR-181d	0,010	-3,500	0,008	-4,000	0,349	-1,519
hsa-miR-1228*	0,001	6,016	0,008	3,896	-	-
hsa-miR-1308	0,002	4,592	0,009	4,020	-	-
hsa-miR-150*	0,001	4,841	0,009	4,048	0,805	0,457
hsa-miR-1909	0,001	6,427	0,010	3,664	-	-
hsa-miR-1275	0,001	5,887	0,010	3,578	-	-
hsa-miR-1225-5p	0,001	4,933	0,011	3,583	-	-
hsa-miR-1246	0,032	2,739	0,011	3,817	-	-

miRNA	BCA		CLA		MBD	
	FDR	Expresión	FDR	Expresión	FDR	Expresión
<b>BCA+CLA</b>						
hsa-miR-939	0,001	6,586	0,015	3,269	0,918	0,190
hsa-miR-1908	0,001	5,411	0,023	3,050	-	-
hsa-miR-602	0,003	4,084	0,027	2,834	0,583	0,896
hsa-miR-675	0,046	2,604	0,029	2,832	-	-
hsa-miR-183	0,006	-3,907	0,030	-2,676	0,489	-1,096
hsa-miR-1973	0,006	3,948	0,035	2,803	-	-
hsa-miR-3180-3p	0,016	3,326	0,042	2,564	-	-
hsa-miR-4299	0,001	5,776	0,049	2,424	-	-
<b>BCA+MBD</b>						
hsa-miR-30a*	0,006	-3,776	0,687	-0,517	0,007	-4,183
hsa-miR-342-3p	0,037	-2,777	0,962	-0,120	0,036	-3,075
hsa-miR-148a	0,006	-3,884	-	-	0,038	-3,076
hsa-miR-30e*	0,010	-3,490	-	-	0,022	-3,519
hsa-miR-29b-2*	0,028	-2,988	-	-	0,019	-4,833
hsa-miR-29c	0,042	-2,740	-	-	0,001	-7,501
<b>CLA+MBD</b>						
hsa-miR-31	0,142	-1,838	0,001	47,978	0,031	3,572
hsa-miR-30b	0,104	-2,076	0,002	-5,053	0,011	-3,767
hsa-miR-148b	-	-	0,005	-4,082	0,001	-5,301
hsa-miR-29b	-	-	0,005	-4,127	0,010	-4,085
hsa-miR-130b	0,247	-1,431	0,009	3,619	0,006	4,254
hsa-miR-331-3p	0,796	-0,306	0,009	-3,801	0,036	-3,104
hsa-miR-193a-3p	-	-	0,034	-2,499	0,006	-4,218

FDR se refiere al p-valor corregido por múltiples comparaciones mediante el método de Benjamini & Hochberg. BCA: estudio sobre pacientes del Hospital Clínico de Valencia. CLA: estudio sobre líneas celulares de cáncer de mama. MBD: estudio sobre datos obtenidos a partir del estudio METABRICK depositado online.

**ANEXO XIV:** Características tumorales de las muestras de las que proceden los datos analizados del consorcio METABRIC en mujeres menores de 35 años y mayores de 65.

Grupo de edad	Menores de 35 n (%)	Mayores de 65 n (%)
Media de edad (SD)	31,41 (4,04)	73,3 (10,18)
<b>Grado histológico</b>		
I	0 (0)	55 (7,72)
II	6 (18,18)	290 (40,73)
III	27 (81,82)	327 (45,93)
No conocido	0 (0)	40 (5,62)
<b>Tamaño tumoral</b>		
< 2 cm	9 (27,27)	280 (39,33)
2-5 cm	24 (72,73)	432 (60,67)
<b>Afectación ganglionar</b>		
Positiva	12 (36,36)	382 (53,65)
Negativa	21 (63,63)	330 (46,34)
<b>Estadio</b>		
0	0 (0)	5 (0,7)
1	7 (21,21)	187 (26,26)
2	19 (57,58)	350 (49,16)
3	4 (12,12)	62 (8,71)
4	0 (0)	6 (0,84)
No conocido	3 (9,09)	102 (14,33)
<b>Receptores</b>		
ER+	12 (36,36)	603 (84,69)
ER-	21 (63,64)	109 (15,31)
PR+	8 (24,24)	394 (55,34)
PR-	25 (75,76)	318 (44,66)
HER2+	5 (15,15)	79 (11,1)
HER2-	28 (84,85)	633 (88,9)
<b>Subtipo molecular (PAM50)</b>		
Luminal A	4 (12,12)	326 (45,79)
Luminal B	7 (21,21)	239 (33,57)
<i>Basal-like</i>	18 (54,55)	57 (8,01)
HER2	3 (9,09)	55 (7,72)
<i>Normal-like</i>	1 (3,03)	30 (4,21)
No conocido	0	5 (0,7)

SD: desviación típica.

**ANEXO XV:** Meta-análisis mediante Fisher de los resultados de expresión de miRNAs obtenidos a partir de muestras de pacientes, líneas celulares y datos procedentes del estudio METABRIC.

miRNA ID	p-valor	-log10 p-valor	miRNA ID	p-valor	-log10 p-valor
hsa-mir-663	1,00E-12	12,000	hsa-mir-10b	5,75E-02	1,240
hsa-mir-148a	1,00E-05	5,000	hsa-mir-874	6,12E-02	1,213
hsa-mir-184	4,00E-05	4,398	hsa-mir-130b	6,33E-02	1,199
hsa-let-7c	1,00E-04	4,000	hsa-mir-145	6,39E-02	1,195
hsa-mir-449a	3,00E-04	3,523	hsa-mir-107	6,41E-02	1,193
hsa-mir-29c	3,20E-04	3,495	hsa-mir-143	6,44E-02	1,191
hsa-mir-99a	3,30E-04	3,481	hsa-mir-99b	6,86E-02	1,164
hsa-mir-132	3,80E-04	3,420	hsa-mir-193b	9,19E-02	1,037
hsa-mir-708	6,90E-04	3,161	hsa-mir-20a	1,24E-01	0,907
hsa-mir-22	9,90E-04	3,004	hsa-mir-345	1,24E-01	0,906
hsa-mir-20b	9,90E-04	3,004	hsa-mir-98	1,36E-01	0,866
hsa-mir-205	1,56E-03	2,807	hsa-mir-29a	1,39E-01	0,858
hsa-mir-665	1,72E-03	2,764	hsa-mir-146a	1,43E-01	0,845
hsa-mir-31	2,31E-03	2,636	hsa-mir-196b	1,57E-01	0,805
hsa-mir-30b	2,67E-03	2,573	hsa-mir-93	1,70E-01	0,771
hsa-mir-489	2,84E-03	2,547	hsa-mir-378	1,77E-01	0,753
hsa-mir-10a	3,96E-03	2,402	hsa-mir-487b	2,09E-01	0,680
hsa-mir-625	6,30E-03	2,201	hsa-let-7i	2,33E-01	0,632
hsa-mir-449b	6,37E-03	2,196	hsa-mir-203	2,37E-01	0,625
hsa-mir-23b	6,57E-03	2,182	hsa-mir-432	2,50E-01	0,603
hsa-mir-23a	6,61E-03	2,180	hsa-mir-877	2,50E-01	0,603
hsa-mir-629	8,20E-03	2,086	hsa-mir-27a	2,86E-01	0,543
hsa-mir-130a	8,80E-03	2,056	hsa-mir-370	3,03E-01	0,519
hsa-mir-17	8,92E-03	2,050	hsa-mir-106b	3,23E-01	0,491
hsa-mir-939	9,21E-03	2,036	hsa-mir-221	3,51E-01	0,455
hsa-mir-181d	1,01E-02	1,996	hsa-mir-30d	3,59E-01	0,445
hsa-mir-183	1,02E-02	1,994	hsa-mir-451	3,60E-01	0,444
hsa-let-7b	1,06E-02	1,974	hsa-mir-497	3,87E-01	0,412
hsa-mir-375	1,15E-02	1,939	hsa-mir-210	4,10E-01	0,387
hsa-mir-26b	1,15E-02	1,939	hsa-mir-382	4,14E-01	0,383
hsa-mir-200a	1,27E-02	1,898	hsa-mir-15b	4,28E-01	0,369
hsa-mir-222	1,46E-02	1,835	hsa-mir-433	4,33E-01	0,363
hsa-mir-494	1,50E-02	1,825	hsa-mir-660	4,42E-01	0,355
hsa-mir-92b	1,53E-02	1,817	hsa-mir-425	4,49E-01	0,347
hsa-mir-100	1,76E-02	1,755	hsa-let-7g	4,80E-01	0,319
hsa-mir-27b	1,82E-02	1,739	hsa-mir-134	5,11E-01	0,291
hsa-mir-379	2,12E-02	1,674	hsa-mir-421	5,28E-01	0,277
hsa-mir-152	2,19E-02	1,659	hsa-mir-141	5,39E-01	0,268
hsa-mir-150	2,27E-02	1,643	hsa-let-7d	5,44E-01	0,264
hsa-mir-30a	2,33E-02	1,633	hsa-mir-652	5,82E-01	0,235
hsa-mir-744	2,33E-02	1,633	hsa-mir-30e	5,82E-01	0,235
hsa-let-7e	2,59E-02	1,587	hsa-mir-200b	5,94E-01	0,226
hsa-mir-195	2,79E-02	1,554	hsa-mir-155	6,36E-01	0,197
hsa-mir-149	3,00E-02	1,522	hsa-mir-25	6,97E-01	0,157
hsa-mir-224	3,51E-02	1,455	hsa-mir-185	8,67E-01	0,062
hsa-mir-34a	3,65E-02	1,437	hsa-mir-126	9,10E-01	0,041
hsa-mir-18a	3,87E-02	1,413	hsa-mir-214	9,10E-01	0,041
hsa-mir-503	4,00E-02	1,398	hsa-mir-200c	9,27E-01	0,033
hsa-mir-182	4,26E-02	1,371	hsa-mir-21	1,00E+00	0,000
hsa-mir-15a	5,59E-02	1,253			

**ANEXO XVI:** Valores de expresión relativa (RQ) obtenidos mediante qRT-PCR de los miRNAs seleccionados para su validación

Muestra	miR-1228	miR-1275	miR-1207	miR-433	miR-92b	miR-132	miR-139	miR-149	miR-3196	miR-379	miR-409
<b>Mujeres menores de 35 años</b>											
BCVYV15	14,176	7,578	9,613	0,694	0,247	1,118	2,052	3,558	17,854	1,030	0,363
BCVYV3	15,491	4,665	7,726	0,935	0,276	0,803	0,086	7,517	17,816	0,993	0,802
BCVYV5	12,673	7,413	13,211	0,948	0,685	1,244	0,140	17,087	18,395	1,304	0,711
BCVYV11	12,899	9,972	0,021	0,889	1,559	0,605	0,295	9,261	20,451	0,380	0,507
BCVYV12	1,660	2,674	1,202	0,213	0,102	0,368	0,062	9,506	3,358	0,177	0,314
BCVY016	0,978	8,672	-	3,684	3,569	2,635	3,840	6,393	1,500	3,070	2,190
BCVY077	22,152	15,570	-	0,936	0,456	0,570	0,133	7,343	27,407	0,844	0,730
BCVYV13	7,803	7,701	8,083	1,227	0,733	1,180	0,324	1,110	14,046	1,127	1,305
BCVYV14	1,469	29,198	2,267	5,979	4,718	4,363	1,609	28,912	0,319	2,440	2,706
4DP	1,141	5,226	-	1,183	2,378	8,649	1,280	4,692	0,603	0,873	1,405
BCVY133	4,057	4,946	7,105	0,531	0,615	0,975	0,382	4,201	6,306	0,881	0,494
BCVY176	14,555	13,764	-	0,636	1,301	0,852	0,115	1,261	9,846	1,024	0,844
BCVYV1	15,547	9,612	15,017	0,435	0,318	0,532	0,116	3,928	24,138	0,453	0,396
BCVYV10	6,721	4,394	0,003	1,000	1,105	1,481	0,570	25,542	1,868	1,000	0,759
BCVYV2	4,999	4,616	7,390	0,310	0,094	0,513	0,315	8,419	8,625	0,386	0,278
BCVYV6	0,756	2,128	1,971	0,188	0,138	0,397	0,067	2,572	2,237	0,195	0,170
BCVYV7	16,695	16,876	19,904	0,328	0,389	0,537	0,170	4,089	37,807	0,384	0,090
BCVYV8	3,520	8,216	19,009	0,722	0,336	1,453	0,645	1,705	5,835	0,739	0,370
BCVYV9	6,186	6,392	0,001	2,238	0,280	5,575	1,099	2,592	7,496	2,299	2,954
BCVY182	0,264	1,352	-	0,406	0,871	0,506	0,110	3,939	0,203	0,290	0,547
BCVY189	1,875	3,182	3,271	0,230	1,561	0,870	0,083	0,314	2,226	0,531	0,259
BCVYV4	44,776	8,959	12,829	0,242	0,450	0,629	0,278	0,222	97,785	0,917	0,164
<b>Mujeres mayores de 65 años</b>											
BC10	2,113	7,199	2,348	0,278	0,819	0,653	0,192	7,395	4,036	0,390	0,155
BC31	1,210	5,755	-	1,649	1,858	2,579	4,299	25,149	1,825	1,815	1,276
BC37	2,169	6,005	1,864	4,207	2,461	3,276	3,153	1,016	0,936	4,496	4,052
<b>BC40</b>	<b>1016,78</b>	<b>0,002</b>	<b>32,495</b>	<b>1698,75</b>	<b>1784,91</b>	<b>2206,52</b>	<b>1296,83</b>	<b>6563,28</b>	<b>473,368</b>	<b>1930,12</b>	<b>19,052</b>
BC41	0,978	1,447	0,938	0,535	1,157	0,779	0,562	8,408	0,788	0,651	0,594
BC43	1,651	5,734	0,001	2,169	2,521	4,205	2,183	17,189	1,612	2,277	2,177
BC09	1,834	3,015	-	0,292	1,052	0,451	0,115	29,930	3,581	0,401	0,329
BC28	0,678	1,753	2,646	0,223	0,639	0,574	0,268	1,263	0,655	0,276	0,101
BC29	7,297	3,452	1,561	0,852	1,654	1,669	0,947	12,385	8,691	0,905	0,576
BC30	0,609	2,855	6,487	0,407	0,202	0,276	0,354	0,648	0,728	0,411	0,317
BC32	3,211	9,315	0,237	2,476	1,941	2,670	0,893	15,240	2,693	2,680	2,505
BC33	1,805	3,592	0,088	0,502	0,660	0,741	0,845	4,171	1,715	0,587	0,312
BC36	0,376	3,320	3,853	0,310	0,735	0,672	0,432	1,114	0,311	0,254	0,189
BC38	0,448	2,947	1,419	0,384	1,124	1,872	0,946	1,351	0,631	0,430	0,230
BC39	1,877	2,502	16,405	0,730	0,679	1,401	0,778	1,344	1,461	1,280	0,873
BC34	0,948	3,431	19,916	0,830	1,020	1,281	1,130	1,110	0,630	1,043	0,766
BC42	17,799	8,308	0,256	2,587	0,980	1,831	2,722	2,876	12,478	2,314	1,358
<b>Mujeres de entre 45 y 65 años</b>											
1DP	2,153	5,322	-	0,630	2,117	1,533	0,517	30,013	2,073	0,550	0,425
BC26	3,130	7,369	7,199	1,715	20,517	1,904	0,849	10,918	6,601	1,385	0,947
1COP	1,747	4,459	-	2,529	2,315	2,808	1,338	3,966	1,868	1,381	1,776
5COP	2,224	1,000	-	3,416	3,436	1,302	0,794	3,075	2,343	2,257	3,415
9COP	0,342	2,992	-	0,433	0,775	0,509	0,330	2,002	0,298	0,453	0,503
BC11	4,169	1,880	0,491	0,542	3,088	0,557	0,172	15,056	2,959	0,549	0,521
BC14	1,417	1,822	12,978	0,433	0,960	0,608	0,098	1,094	1,643	0,367	0,420
BC16	1,683	3,119	0,272	1,439	1,217	1,179	0,822	3,721	1,585	1,496	1,344
BC22	2,646	5,183	11,718	2,034	0,710	1,713	2,304	6,605	2,616	2,356	1,537
BC23	2,201	4,912	2,662	2,918	1,167	2,576	1,529	6,616	1,572	2,881	2,532
10COP	1,000	1,000	-	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
3COP	0,826	3,080	-	0,500	0,755	0,477	0,327	0,666	0,699	0,702	0,484
BC18	0,966	2,518	22,948	1,186	1,915	0,824	0,897	5,744	1,284	1,211	0,755
BC21	1,337	2,303	0,332	0,907	1,659	0,822	0,988	2,558	0,909	1,191	1,046
14COP	1,506	5,186	-	1,267	3,894	1,663	1,284	1,419	1,039	1,376	1,201
BC17	0,731	1,852	0,025	0,694	0,996	0,996	0,509	7,716	1,087	0,776	0,733
BC24	3,440	2,132	0,002	1,399	0,804	0,941	1,214	2,051	1,275	1,015	1,149

ID Muestra	miR-125a	miR-132	miR-195-5p	miR-23a	miR-27b-3p	miR-28-5p	miR-30c	miR-30e-3p
<b>Mujeres menores de 35 años</b>								
4DP	2,676	4,101	1,166	0,647	0,489	1,252	0,456	1,097
BCVY096	1,469	3,348	1,650	3,544	3,589	1,546	0,209	2,343
BCVY12681	2,770	7,501	3,354	6,293	3,803	4,685	1,891	5,774
BCVY133	1,623	3,298	2,042	2,316	2,794	1,507	0,645	1,933
BCVY14310	1,601	2,198	1,684	3,649	1,639	1,075	0,378	1,935
BCVY15243	2,210	2,188	1,264	6,490	1,771	3,698	1,277	3,342
BCVY15244	3,215	4,728	1,198	4,780	1,493	2,144	0,414	4,340
BCVY162	2,303	1,574	1,734	1,427	1,290	1,034	0,339	2,070
BCVY16963	4,604	8,372	2,272	4,964	1,326	1,867	1,424	2,381
BCVY176	1,708	1,722	0,912	4,505	3,688	1,239	0,206	1,916
BCVY18368	1,865	4,504	2,335	5,453	4,942	4,453	1,102	2,516
BCVY18577	1,619	1,948	0,844	1,373	1,111	1,258	0,199	1,097
BCVY189	0,643	1,424	1,236	2,762	0,879	1,839	0,925	1,219
BCVY-1920	1,078	3,034	1,551	5,358	0,777	0,907	0,337	0,633
BCVY1924	0,505	1,709	1,884	1,160	0,883	1,218	0,320	1,994
BCVY25503	0,790	1,427	1,296	2,514	1,381	1,020	0,314	1,809
BCVY2734	0,628	2,218	0,630	1,409	1,193	0,737	0,211	0,694
BCVY2906	1,655	5,888	2,254	6,137	2,952	3,348	0,966	4,805
BCVY4988	0,790	0,620	0,488	0,768	0,800	0,488	0,100	0,456
BCVY9649	0,987	3,469	1,205	1,038	0,707	0,867	0,430	1,133
BCVY10	6,473	3,252	3,147	6,074	5,329	3,293	1,532	6,142
BCVY12	0,919	0,810	0,140	1,326	0,087	0,678	0,187	1,086
BCVY15	2,170	6,712	2,561	3,829	3,093	1,045	0,365	5,708
BCVY2	0,421	0,928	0,486	0,780	0,430	0,486	0,141	0,749
BCVY3	1,016	1,821	0,971	9,432	8,241	4,152	0,983	3,868
BCVY4	0,594	2,977	0,965	1,769	1,183	1,717	0,212	4,499
BCVY5	1,359	2,535	1,843	3,188	4,290	1,406	0,277	2,832
BCVY6	0,557	0,984	0,825	1,284	0,815	0,646	0,429	1,403
BCVY7	1,375	1,241	0,536	0,727	0,599	0,471	0,105	0,940
BCVY8	1,402	3,434	1,561	4,052	1,022	0,957	0,400	3,837
BCVY9	1,309	12,882	4,156	1,985	2,088	2,045	0,810	2,319
<b>Mujeres mayores de 65 años</b>								
7DP	2,121	3,052	1,011	3,763	1,456	1,004	0,447	4,762
BC10	2,205	1,341	0,412	1,257	1,665	0,705	0,612	1,303
BC28	1,260	1,309	1,511	1,020	0,799	0,793	0,824	1,577
BC30	0,389	0,607	0,942	0,664	0,838	0,710	0,225	0,519
BC32	1,052	2,403	0,993	5,547	7,660	3,751	1,560	1,641
BC33	28,942	51,694	40,496	46,442	65,991	26,333	16,061	39,865
BC34	4,068	3,793	1,583	8,695	4,262	0,812	0,239	4,956
BC35	3,089	2,427	2,386	4,339	3,786	2,026	0,530	6,440
BC36	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
BC38	0,472	2,434	0,626	1,960	0,988	0,985	0,698	1,406
BC40	3,565	5,302	2,374	5,760	3,323	3,615	1,482	2,474
BC41	4,729	5,736	2,587	4,292	3,029	1,574	2,164	13,618
BC42	1,063	3,299	2,584	4,865	3,068	2,021	0,733	2,856
BC43	26,811	17,081	7,068	11,475	17,285	8,519	4,956	42,475
<b>Mujeres de entre 45 y 65 años</b>								
10COP	1,774	2,565	0,774	2,950	1,995	0,898	0,578	1,496
10DP	1,185	1,799	0,658	4,016	3,140	0,677	0,689	1,525
14COP	2,635	2,860	0,697	5,304	3,084	4,423	1,343	3,159
1COP	3,754	5,394	1,636	7,062	3,190	1,297	0,949	4,840
1DP	10,425	6,461	1,540	5,403	2,170	1,530	1,128	4,556
4COP	4,788	5,690	2,054	17,971	13,053	2,892	2,006	20,497
9COP	1,018	1,015	0,566	2,191	1,097	1,333	1,007	1,178
BC16	1,194	1,118	0,708	3,769	2,280	1,002	0,687	1,033
BC17	1,528	2,256	0,925	2,332	1,276	1,100	0,511	1,263
BC21	1,222	1,379	0,897	3,383	2,067	1,310	0,850	1,165
BC24	28,699	7,588	1,742	27,700	15,448	8,439	2,761	46,792
BC26	2,158	2,704	1,440	3,116	2,131	1,535	0,806	1,767

**ANEXO XVII:** Sitios altamente conservados en vertebrados de unión a miRNAs en la región 3'UTR de los genes diana seleccionados.

*ERBB4*

miRNA	sitios altamente conservados				sitios poco conservados				Valor contexto	P <sub>CT</sub> Agr
	Total	8mero	7mer-m8	7mer-1A	Total	8mero	7mer-m8	7mer-1A		
miR-193/193b/193a-3p	1	1	0	0	2	0	1	1	-0,71	0,42
miR-205/205ab	1	1	0	0	3	1	0	2	-0,29	0,76
miR-383	1	0	0	1	2	0	2	0	-0,2	0,21
miR-23abc/23b-3p	1	1	0	0	2	1	1	0	-0,2	0,79
miR-200bc/429/548a	1	0	1	0	2	0	0	2	-0,17	0,53
miR-22/22-3p	1	1	0	0	0	0	0	0	-0,15	0,68
miR-7/7ab	1	0	1	0	0	0	0	0	-0,15	0,14
miR-221/222/222ab/1928	1	1	0	0	1	0	0	1	-0,14	0,61
miR-101/101ab	1	0	0	1	3	1	0	2	-0,14	0,31
miR-145	1	1	0	0	2	1	1	0	-0,13	0,76
miR-199ab-5p	1	1	0	0	0	0	0	0	-0,13	0,92
miR-135ab/135a-5p	2	1	1	0	0	0	0	0	-0,12	0,95
miR-194	1	0	0	1	1	0	0	1	-0,12	0,49
miR-144	1	0	0	1	4	0	1	3	-0,12	0,31
miR-204/204b/211	0	0	0	0	3	0	2	1	-0,12	0,29
miR-130ac/301ab/301b/301b-3p/454/721/4295/3666	1	0	1	0	0	0	0	0	-0,11	0,91
miR-26ab/1297/4465	0	0	0	0	2	0	1	1	-0,1	0,34
miR-1ab/206/613	0	0	0	0	1	0	0	1	-0,1	0,18
miR-375	0	0	0	0	3	0	0	3	-0,1	0,29
miR-138/138ab	1	0	0	1	0	0	0	0	-0,08	0,74
miR-19ab	2	1	1	0	0	0	0	0	-0,08	0,96
miR-143/1721/4770	0	0	0	0	1	0	1	0	-0,08	0,11
miR-223	0	0	0	0	1	0	0	1	-0,07	0,19
miR-218/218a	1	0	0	1	0	0	0	0	-0,07	0,71
miR-129-5p/129ab-5p	0	0	0	0	4	0	1	3	-0,05	0,33
miR-125a-5p/125b-5p/351/670/4319	1	0	1	0	1	0	1	0	-0,05	0,86
miR-425/425-5p/489	0	0	0	0	2	0	1	1	-0,05	0,14
miR-93/93a/105/106a/291a-3p/294/295/302abcde/372/373/428/519a/520be/520acd-3p/1378/1420ac	1	0	0	1	0	0	0	0	-0,04	0,81
miR-455-5p	1	0	0	1	0	0	0	0	-0,04	0,43
miR-140/140-5p/876-3p/1244	1	0	1	0	0	0	0	0	-0,02	0,41
miR-137/137ab	1	0	1	0	0	0	0	0	-0,02	0,90
miR-203	0	0	0	0	2	0	0	2	-0,02	0,17
miR-21/590-5p	0	0	0	0	2	0	0	2	-0,02	0,11
miR-96/507/1271	1	0	0	1	0	0	0	0	-0,02	0,74
miR-141/200a	0	0	0	0	1	0	0	1	-0,01	0,11
miR-182	1	0	0	1	0	0	0	0	-0,01	0,74
miR-132/212/212-3p	0	0	0	0	1	0	0	1	-0,01	0,13

**GRIN2B**

miRNA family	sitios altamente conservados				sitios poco conservados				score de contexto	P <sub>CT</sub> Agregado
	Total	8mer	7mer-m8	7mer-1A	Total	8mer	7mer-m8	7mer-1A		
miR-135-5p	2	2	0	0	2	1	1	0	-0.74	> 0.99
miR-143-3p/4770/6088	0	0	0	0	5	0	2	3	-0.61	0.11
let-7-5p/98-5p/miR-4458/4500	2	0	0	2	1	0	0	1	-0.56	> 0.99
miR-129-3p	1	0	1	0	3	0	2	1	-0.56	0.49
miR-34-5p/449-5p	1	0	1	0	1	0	0	1	-0.54	0.67
miR-214-5p	0	0	0	0	5	0	2	3	-0.36	0.23
<b>miR-22-3p</b>	0	0	0	0	6	2	3	1	-0.32	0.59
miR-365-3p	0	0	0	0	3	0	2	1	-0.25	0.10
miR-146-5p/7153-5p	0	0	0	0	3	0	2	1	-0.24	0.16
miR-3064-5p/6504-5p	1	1	0	0	4	0	1	3	-0.21	0.51
miR-375	1	0	0	1	4	0	0	4	-0.18	0.51
<b>miR-182-5p</b>	0	0	0	0	2	0	0	2	-0.16	0.41
miR-9-5p	0	0	0	0	2	0	1	1	-0.16	0.25
<b>miR-148-3p/152-3p</b>	1	1	0	0	6	0	2	4	-0.14	0.54
<b>miR-29-3p</b>	1	0	0	1	3	0	2	1	-0.13	0.82
<b>miR-23-3p/130a-5p</b>	0	0	0	0	8	0	5	3	-0.13	0.24
miR-17-5p/20-5p/93-5p/106-5p/519-3p/526-3p	1	0	0	1	7	1	2	4	-0.12	0.56
miR-192-5p/215-5p	1	0	1	0	1	0	0	1	-0.12	0.28
<b>miR-15-5p/16-5p/195-5p/424-5p/497-5p/6838-5p</b>	0	0	0	0	4	1	2	1	-0.12	0.33
miR-7-5p	1	0	1	0	7	0	2	5	-0.12	0.67
miR-142-5p/5590-3p	1	1	0	0	4	0	3	1	-0.10	0.75
miR-218-5p	1	1	0	0	4	1	0	3	-0.09	0.60
miR-128-3p/216-3p/3681-3p	2	0	2	0	4	0	1	3	-0.09	0.59
<b>miR-489-3p</b>	0	0	0	0	5	1	2	2	-0.09	0.37
miR-21-5p/590-5p	0	0	0	0	6	0	3	3	-0.09	0.24
miR-96-5p/1271-5p	1	1	0	0	3	0	2	1	-0.08	0.71
miR-27-3p	0	0	0	0	5	0	3	2	-0.08	0.28
miR-338-3p	0	0	0	0	5	0	2	3	-0.08	0.30
miR-455-3p.1	0	0	0	0	4	0	3	1	-0.07	0.29
miR-138-5p	0	0	0	0	4	0	0	4	-0.06	0.50
miR-145-5p/5195-3p	0	0	0	0	3	1	1	1	-0.06	0.26
miR-425-5p	0	0	0	0	3	0	3	0	-0.06	0.17
miR-24-3p	0	0	0	0	2	0	2	0	-0.04	0.33
miR-200bc-3p/429	0	0	0	0	2	0	1	1	-0.03	0.24
miR-221-3p/222-3p	0	0	0	0	2	0	0	2	-0.02	0.12

**GRM3**

miRNA	sitios altamente conservados				sitios poco conservados				Valor de contexto	P <sub>CT</sub> Agr
	Total	8mero	7mer-m8	7mer-1A	Total	8mero	7mer-m8	7mer-1A		
miR-132/212/212-3p	1	1	0	0	0	0	0	0	-0,38	0,65
miR-30abcdef/30abe-5p/384-5p	1	1	0	0	0	0	0	0	-0,17	0,16
miR-129-5p/129ab-5p	0	0	0	0	2	0	0	2	-0,03	0,55



*GRIN3A*

miRNA	sitios altamente conservados				sitios poco conservados				Valor contexto	P <sub>CT</sub> Agr
	Total	8mero	7mer-m8	7mer-1A	Total	8mero	7mer-m8	7mer-1A		
miR-24/24ab/24-3p	0	0	0	0	1	0	1	0	-0,27	0,44
miR-10abc/10a-5p	1	0	0	1	0	0	0	0	-0,15	0,28
miR-124/124ab/506	1	0	1	0	0	0	0	0	-0,15	0,87
miR-130ac/301ab/301b/301b-3p/454/721/4295/3666	0	0	0	0	1	0	0	1	-0,14	0,27
miR-141/200a	1	0	1	0	1	0	1	0	-0,12	0,49
miR-155	0	0	0	0	2	0	0	2	-0,12	0,14
miR-15abc/16/16abc/195/322/424/497/1907	0	0	0	0	1	0	1	0	-0,08	0,43
miR-146ac/146b-5p	0	0	0	0	2	1	0	1	-0,06	0,12
miR-33ab/33-5p	1	0	0	1	1	0	1	0	-0,03	0,35
miR-143/1721/4770	0	0	0	0	1	0	1	0	-0,02	0,11
miR-204/204b/211	0	0	0	0	2	0	0	2	-0,02	0,27
miR-375	0	0	0	0	1	0	0	1	-0,01	0,36
let-7/98/4458/4500	1	0	0	1	0	0	0	0	-0,01	0,87
miR-216a	0	0	0	0	1	0	0	1	-0,01	0,10
miR-223	0	0	0	0	1	0	0	1	-0,01	0,19
miR-214/761/3619-5p	0	0	0	0	1	0	0	1	-0,01	0,25

Los sitios de unión conservados se han obtenido mediante TargetScan 6.0 ([http://www.targetscan.org/vert\\_61/](http://www.targetscan.org/vert_61/)). Se muestran los resultados con P<sub>CT</sub> agregada (P<sub>CT</sub> Agr) superior a 0,1.



## ANEXO XVIII

Publicaciones relacionadas  
con el trabajo de tesis doctoral



## High stability of microRNAs in tissue samples of compromised quality

Lorena Peiró-Chova · María Peña-Chilet · José Antonio López-Guerrero · José Luis García-Giménez · Elisa Alonso-Yuste · Octavio Burgues · Ana Lluch · Jaime Ferrer-Lozano · Gloria Ribas

Received: 25 June 2013 / Revised: 13 September 2013 / Accepted: 17 September 2013 / Published online: 3 October 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** Degradation of tissue samples limits performing RNA-based molecular studies, but little is known about the potential usefulness of samples of compromised quality for studies focused on miRNAs. In this work we analyze a series of cryopreserved tissue samples ( $n=14$ ), frozen samples that underwent a severe thawing process ( $n=10$ ), and their paired formalin-fixed paraffin-embedded (FFPE) tissue samples ( $n=24$ ) from patients with breast cancer obtained during primary surgical resection and collected in 2011. Quality and integrity analyses of the total and small fraction of RNA were carried out. Recovery of specific RNA molecules (miRNAs *hsa-miR-21*, *hsa-miR-125b*, and *hsa-miR-191*; snoRNA

*RNU6B*; and mRNAs *GAPDH* and *HPRT1*) was also analyzed by quantitative RT-PCR. Our results suggest that visualisation of the small RNA electrophoretic profiles obtained using the Agilent 2100 bioanalyzer makes it possible to differentiate between the three groups of samples (optimally frozen, thawed, and FFPE). We demonstrate that specific miRNA molecules can be similarly recovered from different tissue sample sources, which supports their high degree of stability. We conclude that miRNAs are robustly detected irrespective of the quality of the tissue sample. In this regard, a word of caution should be raised before degraded samples are discarded: although prior quality assessment of the biological material to be analyzed is recommended, our work demonstrates that degraded tissue samples are also suitable for miRNA studies.

L. Peiró-Chova · J. Ferrer-Lozano  
Biobank, INCLIVA Biomedical Research Institute, Avenida Blasco Ibáñez, 17, 46010 Valencia, Spain

M. Peña-Chilet · E. Alonso-Yuste · O. Burgues · A. Lluch · J. Ferrer-Lozano · G. Ribas (✉)  
Medical Oncology and Hematology, INCLIVA Biomedical Research Institute, Avenida Blasco Ibáñez, 17, Valencia, Spain  
e-mail: gribasdespuig@gmail.com

J. A. López-Guerrero  
Laboratory of Molecular Biology, Fundación Instituto Valenciano de Oncología, C/ Profesor Beltrán Bágüena, 8, Valencia, Spain

J. L. García-Giménez  
Center for Biomedical Network Research on Rare Diseases, Department of Physiology, School of Medicine and Dentistry, Universitat de Valencia. INCLIVA Biomedical Research Institute, Avenida Blasco Ibáñez, 15, Valencia, Spain

O. Burgues  
Department of Pathology, Hospital Clínico Universitario de Valencia, Avenida Blasco Ibáñez, 17, Valencia, Spain

A. Lluch  
Department of Medical Oncology and Hematology, Hospital Clínico Universitario de Valencia, Avenida Blasco Ibáñez, 17, Valencia, Spain

**Keywords** Biobank · Breast cancer · Compromised quality samples · microRNA

### Introduction

In recent years, biobanks have become a key resource for providing high-quality samples and annotated information for biomedical research to the scientific community [1]. Biological material must be stored under specific conditions that enable its long-term stabilization and preservation, thus guaranteeing a minimum quality for a broad use of the specimen for different purposes. However, this storage is not exempt of disastrous incidents that can compromise the integrity of this valuable material, i.e., an uncontrolled, undesired thawing process of frozen material. If this happens, biobanks and researchers should be prepared to analyze the damage and evaluate the potential usefulness of the material before discarding it.

The analysis of transcriptional profiles (mRNA) using formalin-fixed paraffin-embedded (FFPE) tissue samples is

technically challenging due to RNA degradation during the fixation process [2, 3] and deterioration over time [4]. However, microRNAs (miRNAs), a large family of short non-coding RNAs (17–25 nucleotides) [5], do not seem to be significantly affected during the fixation process. MicroRNAs are less susceptible to chemical and enzymatic degradation and can easily be isolated from FFPE samples due to their small size and high stability [6, 7].

There are ever more studies based on miRNAs because they can regulate many biological processes [8, 9], such as cancer, so that the study of these small molecule expression patterns may be useful in the development of new diagnostic and predictive markers [10–12]. A high correlation between matched frozen and FFPE samples has been described when analysing individual miRNAs by qRT-PCR and when comparing their miRNA expression profiles, but these correlations are not so high when analysing mRNA [13–16].

To our knowledge, the preservation and integrity of miRNAs in compromised quality samples other than FFPE tissue samples has hardly been addressed in the literature. In this study, we analyzed the potential usefulness of thawed tissue samples, compared with their matched FFPE tissues, for RNA-based molecular studies.

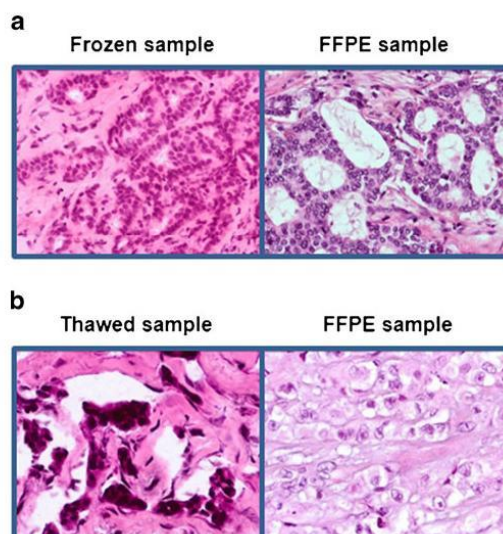
## Materials and methods

### Sample collection

Twenty-four matched frozen and FFPE tissue samples from patients with breast cancer, obtained during primary surgical resection, were collected in 2011 at INCLIVA Biobank. Written and informed consent was obtained from each patient, and the use of the material was approved by the Institutional Ethics Committee. Fresh samples were embedded in Optimal Cutting Temperature (OCT) freezing medium and frozen by direct immersion in pre-cooled isopentane, and then stored at  $-80^{\circ}\text{C}$  until their use. FFPE samples were made following routine formalin fixation and paraffin embedding protocols and stored at room temperature until use. Ten of the 24 OCT-frozen samples underwent a severe thawing process (30 h over  $0^{\circ}\text{C}$ ). These ten samples were embedded again in OCT, re-frozen by immersion in pre-cooled isopentane and restored at  $-80^{\circ}\text{C}$  (thawed samples).

### Histologic control

First, 3- $\mu\text{m}$ -thick paraffin sections and 5- $\mu\text{m}$ -thick cryosections (from optimal frozen and thawed samples) were stained with haematoxylin and eosin (H&E) and analyzed under the microscope in order to obtain a histologic control of all samples (Fig. 1). The vast majority of samples (87.5 %) contained at least 50 % of tumour cells, with only six samples containing



**Fig. 1** Histologic control. Haematoxylin and eosin (H&E) stain of matched samples: **a** a frozen tissue and its matched FFPE sample showing good morphological resemblance, and **b** a thawed tissue with distorted architecture and clumped cells that are hardly recognizable in its corresponding FFPE sample

less than this percentage (Table 1). Preservation of tumour architectural and cytologic features in cryosections was assessed and compared with those of matched paraffin samples.

### RNA isolation

About ten to twelve 10- $\mu\text{m}$ -thick cryosections were disrupted and homogenized with a 5-mm steel ball per sample using the TissueLyser LT (QIAGEN®, Hilden, Germany), according to the manufacturer's instructions. The mirVana™ miRNA Isolation Kit (Applied Biosystems by Life Technologies™, Carlsbad, CA, USA) was used to extract total RNA (including miRNAs) from frozen and thawed tissue samples.

Four 20- $\mu\text{m}$ -thick sections from FFPE tissue samples were deparaffined with 100 % xylol, washed twice with 100 % ethanol, and total RNA (including miRNAs) was extracted using the RecoverAll™ Total Nucleic Acid Isolation Kit (Applied Biosystems by Life Technologies™), according to the manufacturer's recommendations.

RNA concentration was measured at 260 and 280 nm using NanoDrop ND 2000 UV–vis Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

### RNA integrity

RNA integrity, determined by RNA integrity number (RIN) was assessed with the Agilent 2100 Bioanalyzer using the



**Table 1** Total RNA quantity, RIN value, and percentage of tumour material for each sample

Sample ID	[RNA] (ng/ $\mu$ l)	RIN value	%tumor material	Sample ID	[RNA] (ng/ $\mu$ l)	RIN value	% tumor material
1-T	511.3	2.4	60	1-FFPE(T)	350.6	2.3	60
2-T	162.9	2.6	70	2-FFPE(T)	76.4	1.5	60
3-T	147.4	2.0	70	3-FFPE(T)	120.0	2.2	70
4-T	183.9	2.4	60	4-FFPE(T)	69.4	2.1	60
5-T	131.1	1.8	50	5-FFPE(T)	111.1	2.1	60
6-T	103.1	1.7	30	6-FFPE(T)	73.3	1.9	40
7-T	110.3	1.9	60	7-FFPE(T)	50.9	1.0	30
8-T	74.2	1.3	40	8-FFPE(T)	79.0	2.0	40
9-T	335.7	N/A	50	9-FFPE(T)	369.9	2.4	60
10-T	136.2	2.1	60	10-FFPE(T)	170.7	2.3	60
1-F	166.8	8.3	70	1-FFPE(F)	53.1	2.1	60
2-F	212.6	7.8	50	2-FFPE(F)	79.1	1.9	65
3-F	257.3	8.8	55	3-FFPE(F)	191.3	2.4	60
4-F	155.0	9.7	80	4-FFPE(F)	79.5	1.2	60
5-F	116.5	9.0	70	5-FFPE(F)	120.0	2.4	55
6-F	219.5	9.3	60	6-FFPE(F)	57.0	2.5	60
7-F	361.2	9.6	70	7-FFPE(F)	142.6	1.7	60
8-F	161.1	9.2	70	8-FFPE(F)	184.9	2.4	70
9-F	391.0	9.4	80	9-FFPE(F)	137.1	2.0	60
10-F	79.3	9.2	90	10-FFPE(F)	112.3	1.3	70
11-F	141.7	8.7	50	11-FFPE(F)	137.3	2.2	50
12-F	64.6	7.9	60	12-FFPE(F)	67.0	1.1	50
13-F	67.6	8.7	60	13-FFPE(F)	87.5	1.4	35
14-F	65.2	9.0	50	14-FFPE(F)	102.1	2.4	50

*Sample ID* sample identification, *[RNA]* total RNA quantity, *RIN* RNA integrity number, *T* thawed sample, *FFPE(T)* formalin-fixed paraffin-embedded sample matched to thawed sample, *F* frozen sample, *FFPE(F)* formalin-fixed paraffin-embedded sample matched to frozen sample, *N/A* not analyzed

RNA 6000 Nano Assay (Agilent Technologies, Santa Clara, CA, USA). Small RNA integrity was assessed with the same microfluidics-based platform using the Small RNA Assay (Agilent Technologies). Agilent 2100 expert software (Agilent Technologies) was used to compare electropherograms.

#### cDNA synthesis

The reverse transcription of miRNA was performed using a RT-primer pool with the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems™ by Life Technologies™). We prepared a RT primer pool at a final concentration of 0.5X each primer (RNU6B ABI assay ID: 001093; has-miR-21 ABI assay ID: 000397; has-miR-125b ABI assay ID: 000449 and has-miR-191 ABI assay ID: 002299, TaqMan™ MicroRNA Assays, Applied Biosystems™ by Life Technologies™). Then, 200 ng total RNA were transcribed using MultiScribe™ Reverse Transcriptase (Applied Biosystems™ by Life Technologies™), according to the Procedure for multiplexing the RT step without preamplification while using TaqMan™ MicroRNA Assays (available at [http://tools.](http://tools.invitrogen.com/content/sfs/manuals/cms_094060.pdf)

[invitrogen.com/content/sfs/manuals/cms\\_094060.pdf](http://tools.invitrogen.com/content/sfs/manuals/cms_094060.pdf)). An RT negative control was included in each batch of reactions.

The reverse transcription of mRNA was performed from 200 ng total RNA using the High-Capacity Reverse Transcription Kit (Applied Biosystems™ by Life Technologies™), according to the manufacturer's instructions.

#### Real-time quantitative PCR (qRT-PCR)

The recovery of specific RNA species (miRNA, snoRNA and mRNA) was estimated by qRT-PCR and TaqMan™ technology (TaqMan™ Gene Expression Assays and TaqMan™ MicroRNA Assays, Applied Biosystems™ by Life Technologies™). The PCR reactions were carried out in a final volume of 10  $\mu$ l. Briefly, reactions consisted of 1  $\mu$ l of cDNA, 0.5  $\mu$ l of the corresponding TaqMan™ Probe 20 $\times$ , 5  $\mu$ l of the corresponding TaqMan™ Universal Master Mix 2 $\times$ , and 3.5  $\mu$ l of nuclease-free water. The PCR reactions were initiated with 2 min incubation at 50 °C followed by 10 min at 95 °C, and then, 40 cycles of 95 °C for 15 s and 60 °C for 60 s. All reactions were performed in triplicate on a 7900HT Fast

Real-Time PCR system and data analyzed with the 7900HT Fast system software SDS v.2.4 (Applied Biosystems™ by Life Technologies™).

#### Data analysis

Data were processed and graphically plotted using GraphPad Prism v5.00. Statistical significance was evaluated with the Newman–Keuls multiple comparison test ( $p < 0.01$ ) following the one-way analysis of variance (ANOVA). Correlation between paired samples was performed using Pearson's correlation coefficient ( $r$ ), with  $p < 0.0001$  and 95 % confidence interval (two-tailed test).

## Results

#### Histologic control of tissue samples

A section of each tissue sample analyzed was subjected to H&E staining in order to evaluate its morphology. As shown in Fig. 1, tumour architecture was well preserved in optimally frozen samples, and it was easily recognizable as matching in paired FFPE samples. Cellular features were also maintained (Fig. 1a). Conversely, the morphologic appearance of thawed tissues was severely affected, with distorted structures that hardly resembled tumour architecture, stromal retraction, or clumping of tumour cells. Such cells showed smudged nuclei and poorly preserved cytoplasmic detail, making them difficult to match their corresponding FFPE tissue (Fig. 1b).

#### Differences in total RNA electrophoretic profiles

Total RNA quantity from frozen and thawed tissue samples ranged from 64.6 to 511.3 ng/μl, mainly depending on the number of cells in each sample. This quantity ranged from 50.9 to 350.6 ng/μl in the case of FFPE samples (Table 1).

RNA quality was assessed by capillary electrophoresis (Agilent RNA Bioanalyzer, Agilent Technologies). The mean RIN value from the 14 optimally frozen samples was 8.9 ( $\pm 0.6$ , range 7.8–9.6) as expected from high quality samples. On the other hand, the mean RIN value of their corresponding FFPE samples was 1.9 ( $\pm 0.6$ , range 1.1–2.5) as expected from highly degraded samples. This value was 2.0 ( $\pm 0.4$ , range 1.3–2.6) in the case of the ten thawed samples and also 2.0 ( $\pm 0.4$ , range 1.0–2.4) for their corresponding FFPE samples (Table 1). As was expected, the RIN value indicated that thawed samples were as degraded as FFPE samples, but electrophoretic profiles showed differences between these two groups (Fig. 2). Ribosomal peaks, corresponding to the 18S and 28S rRNAs, were lost in both thawed and FFPE samples, but FFPE profiles showed an accumulation of RNA fragments with an average size of less than 200

nucleotides (nt), while this distribution of RNA fragments was extended to sizes greater than 1,000 nt in the case of thawed samples.

#### Characteristic small RNA profiles

In this process, 10 ng of total RNA from each sample was analyzed using the Agilent Small RNA Assay with the Agilent 2100 Bioanalyzer (Agilent Technologies). This assay shows the distribution of small RNA fragments ranging in size from 6 to 150 nt, thus allowing the detection of different regions: miRNAs between 10 and 40 nt; tRNAs in the range of 40–80 nt; 5S and 5.8 S rRNAs and pre-miRNAs between 90 and 155 nt; and snoRNAs in the range of 60–300 nt. This assay also allows quantification of miRNAs in the concentration range of 50–2,000 pg/μl. The average quantity of small RNAs and miRNAs and standard deviation for each group of samples are shown in Table 2, where the small RNA concentration refers to fragments ranging from 6 to 150 nt and the miRNA concentration refers to fragments ranging from 6 to 40 nt, although mature miRNA has been reported to be about 20 nt long.

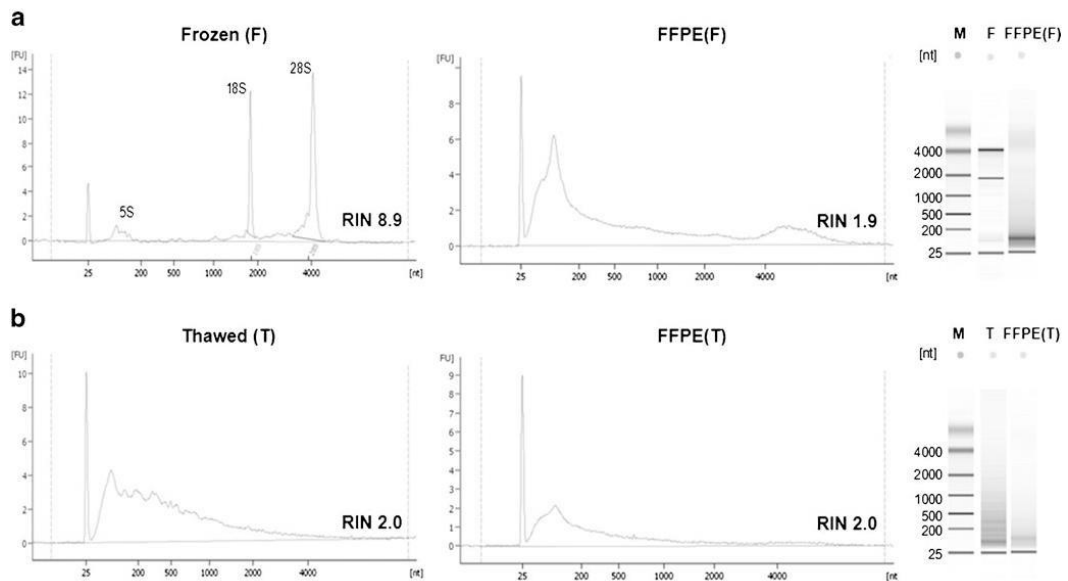
The quantity of small RNAs and miRNAs from degraded samples (thawed and FFPE) was much higher than that obtained from high quality samples (optimally frozen), since degradation of total RNA can produce fragments in the small RNA size range that could cause an overestimation in the proportion of its small sized fragments.

In addition, electrophoretic profiles of 10 ng of total RNA from each sample showed differences between the three groups of samples (Fig. 3). Hence, all high-quality samples showed a homogeneous profile, with the presence of three characteristic peaks: one in the range of 40–80 nt, another around 90 nt, and the last one near 150 nt, probably corresponding to tRNAs, highly conserved rRNAs, and certain snoRNAs. Thawed samples showed a jagged profile, probably due to the presence of small RNA fragments from degradation products of larger molecules. Finally, FFPE samples showed a characteristic profile with the presence of a unique characteristic peak in the range of 60–80 nt.

miRNAs are similarly recovered from all sample types

The recovery of specific RNA species was analyzed by qRT-PCR. We studied three specific miRNAs described in the literature that are involved in human breast cancer (*miR-21*, *miR-125b* and *miR-191*) [6, 17–19], one snoRNA frequently used as a normaliser in miRNA expression studies (*RNU6B*) [20], and two mRNAs usually used as endogenous control genes (*GAPDH* and *HPRT1*) [21], using TaqMan® technology and an equal input of total RNA for each sample (200 ng). Figure 4 shows the mean  $C_t$  values of triplicates for each of the samples (14 optimally frozen samples with their paired





**Fig. 2** Integrity analysis of total RNA. Electrophoretic profiles of total RNA obtained using the RNA 6000 Nano chip on the Agilent 2100 bioanalyzer from **a** a frozen tissue sample (*left*) and its paired FFPE sample (*middle*). Microcapillary electrophoresis image (*right*; M marker, F frozen tissue sample, FFPE(F) matched FFPE sample) is also shown. *Inset*: RIN values

Electrophoretic profiles of total RNA from **b** a thawed tissue sample (*left*) and its paired FFPE sample (*middle*). Microcapillary electrophoresis image (*right*; M marker, T thawed tissue sample, FFPE(T) matched FFPE sample) is also shown. *Inset*: RIN values

FFPE samples and ten thawed samples with their paired FFPE samples). It can be observed that miRNAs were detected in all types of samples at a similar  $C_t$  value, while for longer RNA molecules (snoRNA and mRNAs) greater differences in  $C_t$  value were observed between groups. Hence, the  $C_t$  value for detecting mRNAs was lower in optimally frozen samples than in degraded samples. But interestingly, snoRNA *RNU6B* was similarly recovered in optimally frozen and thawed samples.

#### Correlation in the detection of small RNAs between paired samples

To evaluate further the correlation between matched samples, Pearson's correlation ( $r$ ) of  $C_t$  values between paired samples (optimally frozen vs. FFPE and thawed vs. FFPE) was calculated (Fig. 5). Figure 5a shows the mean  $C_t$  values for mRNAs, snoRNA and, miRNAs of optimally frozen vs. FFPE samples.

**Table 2** Average quantity and standard deviation of small RNAs and miRNAs for each group of samples

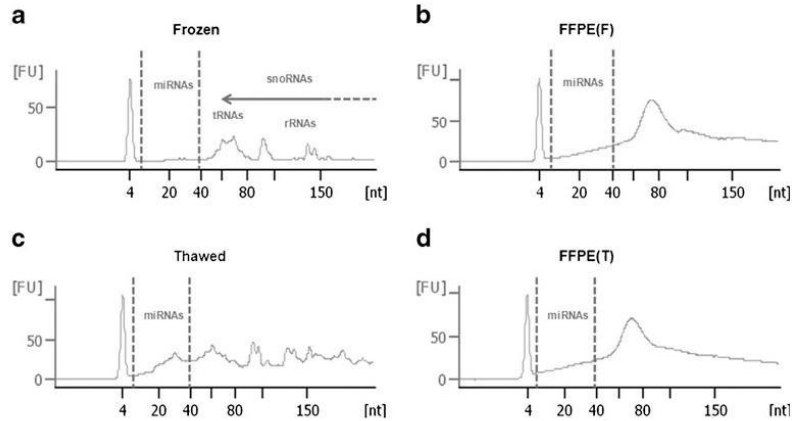
Sample group	[Small RNA] (pg/ $\mu$ l)	[miRNA] (pg/ $\mu$ l)	% miRNA
Thawed ( $n=10$ )	4,187.2 $\pm$ 1,888.4	1,866.7 $\pm$ 825.4	45 $\pm$ 4
Frozen ( $n=14$ )	966.1 $\pm$ 229.9	307.6 $\pm$ 132.5	31 $\pm$ 8
FFPE(F) ( $n=24$ )	3,837.4 $\pm$ 1,856.4	1,184.7 $\pm$ 685.5	29 $\pm$ 5

A high statistically significant correlation ( $r=0.9248$ ) was found between these two groups when only small RNAs (snoRNA and miRNAs) were considered and  $C_t$  values of mRNAs (*GAPDH* and *HPRT1*) were excluded. Interestingly, a high degree of correlation was found between thawed and their matched FFPE samples (Fig. 5b), with  $r=0.9461$ , when considering all the analyzed RNA species (miRNAs, snoRNA, and mRNAs).

#### Differences between sample types

Since it would be relevant to increase the information about the differences between the three groups of samples, a study comparing the mean  $C_t$  values for specific RNA species was performed using the ANOVA one-way analysis of variance. Figure 6 shows the mean  $C_t$  of triplicates for each of the samples, grouped in sample types (optimally frozen vs. thawed vs. FFPE). The analysis shows significant differences between the groups in all cases except for *miR-21*. Small RNA species (*miR-21*, *miR-191*, *miR-125b*, and *RNU6B*) were recovered similarly in optimally frozen and thawed samples, while the recovery of larger RNA molecules, like mRNAs *GAPDH* and *HPRT1*, was more similar in degraded samples (thawed and FFPE samples). Notably, the analysis of miRNA species showed fewer than 2-fold differences in mean

**Fig. 3** Analysis of small RNA. Electrophoretic profiles of small RNA obtained using the Small RNA Assay on the Agilent 2100 bioanalyzer from **a** a frozen tissue sample, **b** its paired FFPE sample, **c** a thawed tissue sample and **d** its matched FFPE sample. The regions corresponding to different small RNA species (miRNAs, tRNAs, rRNAs, and snoRNAs) are shown in **a**. The miRNA region considered by this assay (10–40 nt) is shown within dotted lines



$C_t$  values between frozen samples (optimally frozen and thawed) and FFPE samples, which underscores the high stability of these small RNA molecules.

**Discussion**

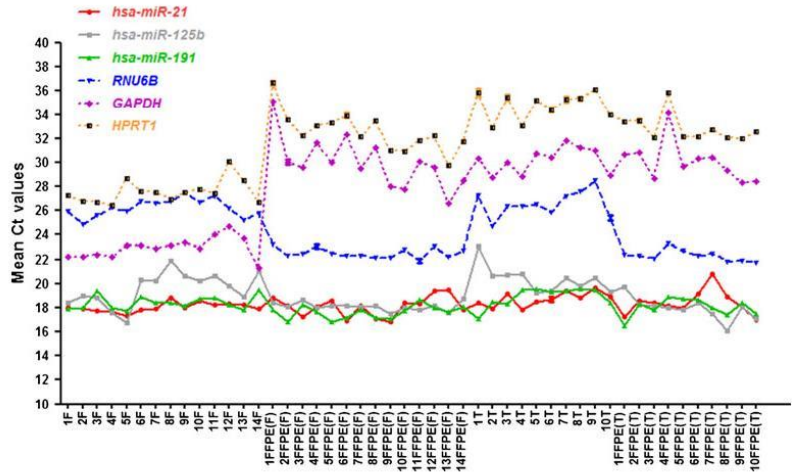
Well-preserved, freshly frozen tissue samples constitute an optimal material for carrying out molecular studies and for using high-throughput technologies, but this type of samples are not always easily available in large amounts. Thus, it is increasingly common to use other tissue sources like FFPE samples, which preserve tissue morphology optimally and are available in the archives of Pathology and Histology

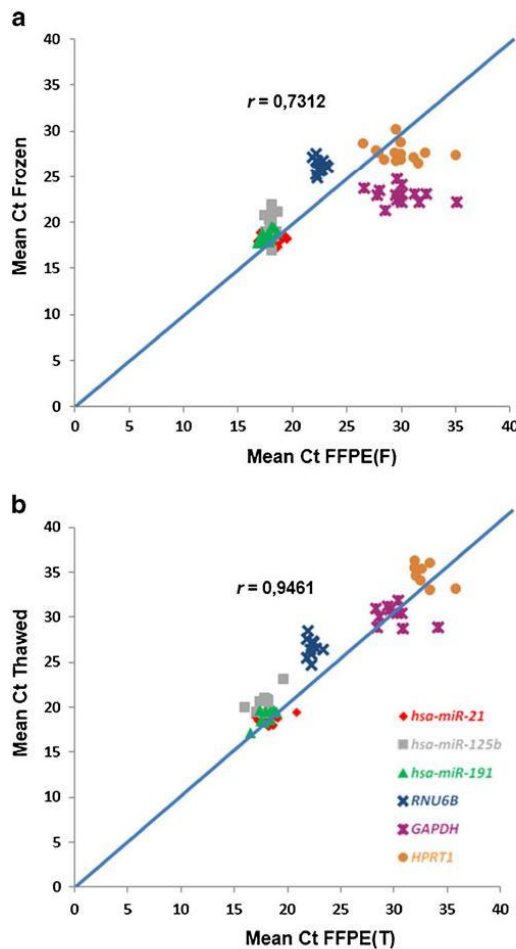
departments worldwide; such sources constitute a real treasure for carrying out biomedical studies [22].

In addition, the preservation of frozen tissue samples is expensive and laborious and it is not exempt of disastrous incidents that could compromise the integrity of stored material, like an uncontrolled tissue thawing process at room temperature. The histologic alterations we have described in our thawed samples are a good indicator of tissue degradation and illustrate the underlying molecular damage that can occur (Fig. 1). Although modern biobanks address this and other issues within a comprehensive policy of quality assurance in an effort to minimize the frequency of such disastrous events, their occurrence cannot be completely ruled out.

The main objective of this study was to analyze the integrity of total and the small fraction of RNA and the recovery of

**Fig. 4** Analysis of specific RNA species by qRT-PCR. Three miRNAs (*hsa-miR-21*, *hsa-miR125b*, and *hsa-miR-191*; solid lines), one snoARN (*RNU6B*; dashed line), and two mARNs (*GAPDH* and *HPRT1*; dotted lines) were evaluated. Each sample was analyzed by qRT-PCR in triplicate using identical amounts of total RNA and TaqMan® MicroRNA Assays or TaqMan® Gene Expression Assays. The mean  $C_t$  values and the standard deviation of each sample for each specific RNA are shown





**Fig. 5** Correlation of  $C_t$  values between matched samples for the specific RNAs analyzed. **a** Pearson's correlation of mean  $C_t$  values between optimally frozen tissue samples ( $y$ -axis) and matched FFPE samples ( $x$ -axis). **b** Pearson's correlation of mean  $C_t$  values between thawed tissue samples ( $y$ -axis) and matched FFPE samples ( $x$ -axis). Pearson's correlation coefficient ( $r$ ) was determined, with  $p < 0.0001$  and 95 % confidence interval (two-tailed test)

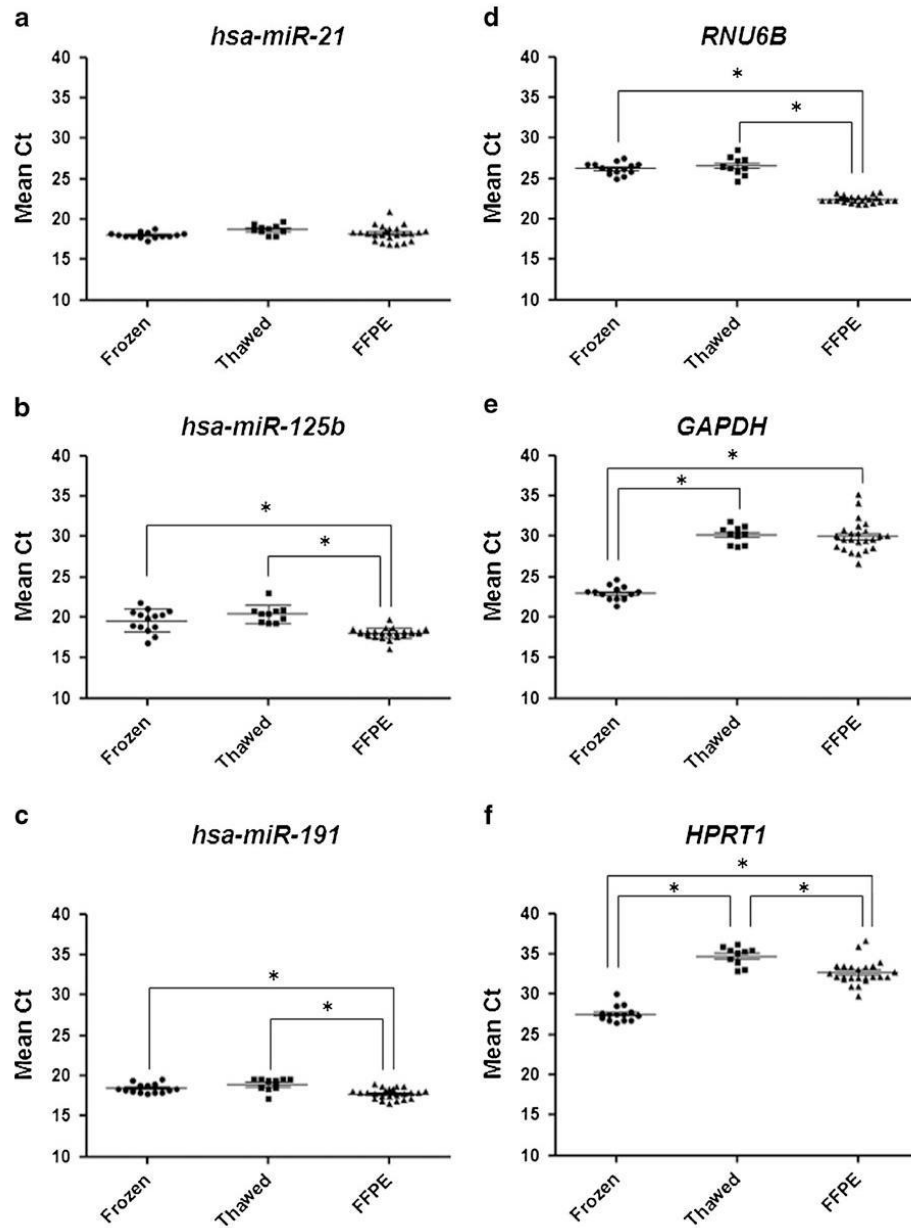
specific RNA species (mRNAs, snoRNAs, and miRNAs) from different tissue sample sources: optimally frozen tissue samples from a Biobank, FFPE samples from an archive of a pathology department, and frozen tissue samples that underwent an uncontrolled thawing process.

The analysis of total RNA showed that thawed and FFPE samples were highly degraded (the mean RIN value was nearly 2 in both cases; Table 1). However, differences in the electrophoretic profiles were observed: FFPE samples showed an accumulation of RNA fragments smaller than 200 nt, while

thawed samples showed a distribution of RNA fragments that reached sizes of nearly 1,000 nt (Fig. 2). This could be mainly explained by the different degradation process that occurred in each sample type. Storage at  $-80^\circ\text{C}$  suppresses intracellular metabolism, but most cell structures are disrupted. When tissue is thawed, intracellular and extracellular RNases can be activated again, mediating a high level of RNA degradation [23]. On the other hand, it has previously been reported that the formalin fixation and paraffin embedding process results in a marked reduction of detectable mRNA molecules [24–26]. This process causes enzymatic degradation and chemical modification of RNA giving rise to cross-links with proteins and making RNA extraction difficult [25, 27, 28]. Thus, a digestion step with proteinase K is required to eliminate cross-links and facilitate RNA extraction from FFPE samples. The longer an RNA molecule is, the more likely cross-links will remain after proteinase K digestion, hence small RNA molecules will be easier to extract from this type of samples and fragments larger than 200 nt will be harder to recover from FFPE samples. Overall, these findings suggest that the amount of total RNA that can be extracted from an FFPE tissue sample is only a fraction of what can be extracted from its corresponding frozen tissue [13, 26].

The analysis of small RNA using 10 ng of total RNA with the Agilent Small RNA Assay (Agilent Technologies) clearly showed distinguishable profiles between the three groups of samples. Optimally frozen sample profiles showed three characteristic peaks corresponding to tRNAs, rRNAs, and snoRNAs, thawed sample profiles were jagged due to the presence of degraded RNA fragments, and FFPE sample profiles were quite homogeneous with a characteristic predominant peak around 60–80 nt (Fig. 3). This result suggests that the visualisation of the small RNA electrophoretic profiles may allow differentiation between these three groups of samples. Currently, the Agilent Small RNA Assay is one of the few possibilities for evaluating small RNA integrity and quantity (Table 2), but this method presents some limitations. One of these limitations is that fragments with a size range of 10–40 nt are considered as miRNAs (Fig. 3), although the literature describes miRNAs as being 17–25 nt [29–31]. Another important limitation is related to RNA degradation, which gives rise to small RNA fragments that can reach the analytical range evaluated by this assay, leading to an overestimation of the amount of small RNAs [29]. This could explain the high quantity of small RNAs and miRNAs in degraded samples (thawed and FFPE) compared with optimally cryopreserved samples (frozen) (Table 2). The high percentage of miRNAs recovered in thawed samples ( $45\% \pm 4$ , Table 2) was surprising and this could be a consequence of the reactivation of RNases that occurs in these samples. Overall, in our opinion and according to Dijkstra et al. [31], currently there is no optimal technique available that can determine yield, purity, and integrity for miRNAs.





**Fig. 6** Comparative analysis of the three sample groups. The ANOVA analysis of variance of the three groups of samples (frozen, thawed, and FFPE) for each of the specific RNA species was performed: *hsa-miR-21* (a), *hsa-miR-125b* (b), *hsa-miR-191* (c), *RNU6B* (d), *GAPDH* (e), and *HPRT1* (f). Mean  $C_t$  values of triplicates for each of the samples, grouped

in sample types were determined. Mean values and standard deviation for each group are shown. Statistical significance was evaluated with the Newman–Keuls multiple comparison test ( $p < 0.01$ ) and significant differences between groups are indicated (\*)

In this study we also analyzed the recovery of three miRNAs (*miR-21*, *miR-125b*, and *miR-191*), one snoRNA

(*RNU6B*) and two mRNAs (*GAPDH* and *HPRT1*) in all three groups of samples (optimally frozen, thawed, and FFPE

samples) by qRT-PCR using identical input amounts of total RNA (Fig. 4). Our results show a high degree of correlation for miRNA detection when comparing matched tissue samples (frozen vs. FFPE [Fig. 5a] and thawed vs. FFPE [Fig. 5b]). It is interesting to note that small RNAs were detected at lower  $C_t$  values in FFPE samples than in frozen/thawed samples (Fig. 6a–d). This may be a consequence of the different extraction method used to isolate RNA from FFPE samples than that used for frozen/thawed samples. As mentioned above, residual RNA–protein cross-links that are not removed in FFPE samples following the digestion by proteinase K prevent some longer RNA molecules from being extracted. Therefore, the smaller a molecule of RNA is, the easier it is to recover [13], which resulting in a higher representation of small RNA molecules in FFPE tissue samples compared to that in frozen/thawed tissue samples. Previous studies support the high degree of correlation in miRNA detection between FFPE tissue samples and their matched frozen tissues [13, 14, 16, 32]. In addition, a study performed in serum/plasma samples also demonstrated the high stability of miRNAs, even after incubation at room temperature for up to 24 h and after multiple freeze/thaw cycles [33]. Furthermore, the usefulness of thawed tissue samples for carrying out gene and protein expression analysis has recently been studied [34]. However, to our knowledge, this is the first study in which the stability of small RNA molecules (snoRNA and miRNAs) is analyzed in thawed tissue samples, and then compared to that of frozen and FFPE tissue samples. We demonstrate that specific miRNA molecules are similarly recovered in different tissue sample sources (frozen, thawed, and FFPE), which supports their high degree of stability.

Regarding mRNAs, *GAPDH* and *HPRT1* were detected at quite low  $C_t$  values in optimally frozen tissue samples, while we found similar recovery of these mRNA in degraded samples (thawed and FFPE) (Fig. 6e–f). It has previously been described that differences in  $C_t$  values between FFPE and frozen samples are higher when large amplicons are analyzed [13], so FFPE tissue samples present limitations for RNA-based molecular studies due to RNA degradation caused by the formalin fixation and the embedding process. Here we show that thawed tissue samples behave similarly to FFPE samples when mRNA molecules are analyzed.

We conclude that miRNA detection is robust irrespective of the quality of the tissue sample. Thus, these tiny molecules constitute very useful analytes to be studied from compromised quality tissue samples due to their high stability and their resistance to degradation [22]. Considering the diverse roles of miRNAs in a broad spectrum of biological processes that have been described in recent years and their enhanced stability, the study of these small molecules may lead to the discovery of new biomarkers and also the performance of retrospective molecular analysis based on miRNAs that will facilitate the development of new diagnostic and therapeutic applications [12].

Our results provide information to biobank managers and also to the scientific community regarding possible uses of compromised quality biospecimens, so we recommend the evaluation of these compromised biological samples before discarding them. In this regard, a new field of research in biobanking is emerging, not only focused on how to collect, store, and provide biological material and data [35], but also on how to take full advantage of the differently preserved biospecimens stored in biobanks to be used for biomedical research.

**Acknowledgments** This study has been supported by Fundación LeCado-Proyecto Flor de Vida. LPC is funded by the National Health Institute Carlos III FEDER (RD09/0076/00132). MPC is funded by the Generalitat Valenciana Conselleria d'Educació VALi+d program (ACIF/2011/207). GR is funded by the Spanish Ministry of Health (CP08\_00069). We want to particularly acknowledge the patients for their participation and INCLIVA Biobank, integrated in the Spanish Hospital Biobanks Network (ReTBioH) and supported by the National Health Institute Carlos III/FEDER and Ministerio de Economía y Competitividad (grant number: RD09/0076/00132). We also thank the Unit for Multigenic Analysis from the Central Unit for Medical Research (UCIM/INCLIVA) for total RNA analysis and the Genomic Unit from Central Service for the Support to Experimental Research (SCSIE) for small RNA analysis.

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

1. Cambon-Thomsen A, Ducournau P, Gourraud PA, Pontille D (2003) Biobanks for genomics and genomics for biobanks. *Comp Funct Genomics* 4(6):628–634. doi:10.1002/cfg.333
2. Bresters D, Schipper ME, Reesink HW, Boeser-Nunnink BD, Cuypers HT (1994) The duration of fixation influences the yield of HCV cDNA-PCR products from formalin-fixed, paraffin-embedded liver tissue. *J Virol Methods* 48(2–3):267–272
3. Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, Jordan RC (2002) Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. *Mod Pathol* 15(9):979–987. doi:10.1097/01.MP.0000026054.62220.FC
4. Cronin M, Pho M, Dutta D, Stephans JC, Shak S, Kiefer MC, Esteban JM, Baker JB (2004) Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 164(1):35–42. doi:10.1016/S0002-9440(10)63093-3
5. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5(7):522–531. doi:10.1038/nrg1379nrg1379
6. Hui AB, Shi W, Boutros PC, Miller N, Pintilie M, Fyles T, McCreedy D, Wong D, Gerster K, Waldron L, Jurisica I, Penn LZ, Liu FF (2009) Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Lab Invest* 89(5):597–606. doi:10.1038/labinvest.2009.12
7. Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* 1(2):155–161. doi:10.1038/nmeth717

8. Esquela-Kerscher A, Slack FJ (2006) Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* 6(4):259–269. doi:10.1038/nrc1840
9. O'Day E, Lal A (2010) MicroRNAs and their target gene networks in breast cancer. *Breast Cancer Res* 12(2):201. doi:10.1186/bcr2484
10. Garofalo M, Quintavalle C, Di Leva G, Zanca C, Romano G, Taccioli C, Liu CG, Croce CM, Condorelli G (2008) MicroRNA signatures of TRAIL resistance in human non-small cell lung cancer. *Oncogene* 27(27):3845–3855. doi:10.1038/onc.2008.6
11. Lawrie CH (2008) MicroRNA expression in lymphoid malignancies: new hope for diagnosis and therapy? *J Cell Mol Med* 12(5A):1432–1444. doi:10.1111/j.1582-4934.2008.00399.x
12. Sandoval J, Peiro-Chova L, Pallardo FV, Garcia-Gimenez JL (2013) Epigenetic biomarkers in laboratory diagnostics: emerging approaches and opportunities. *Expert Rev Mol Diagn* 13(5):457–471. doi:10.1586/erm.13.37
13. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, Guenther SM, O'Leary JJ, Sheils O (2007) Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol* 7:36. doi:10.1186/1472-6750-7-36
14. Doleshal M, Magotra AA, Choudhury B, Cannon BD, Labourier E, Szafranska AE (2008) Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* 10(3):203–211. doi:10.2353/jmoldx.2008.070153
15. Zhang X, Chen J, Radcliffe T, Lebrun DP, Tron VA, Feilottter H (2008) An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. *J Mol Diagn* 10(6):513–519. doi:10.2353/jmoldx.2008.080077
16. Lovendorf MB, Zibert JR, Hagedorn PH, Glue C, Odum N, Ropke MA, Skov L (2012) Comparison of microRNA expression using different preservation methods of matched psoriatic skin samples. *Exp Dermatol* 21(4):299–301. doi:10.1111/j.1600-0625.2012.01445.x
17. Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK, Fedele V, Ginzinger D, Getts R, Haqq C (2006) Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer* 5:24. doi:10.1186/1476-4598-5-24
18. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* 103(7):2257–2261. doi:10.1073/pnas.0510565103
19. Si ML, Zhu S, Wu H, Lu Z, Wu F, Mo YY (2007) miR-21-mediated tumor growth. *Oncogene* 26(19):2799–2803. doi:10.1038/sj.onc.1210083
20. Gee HE, Buffa FM, Camps C, Ramachandran A, Leek R, Taylor M, Patil M, Sheldon H, Bets G, Homer J, West C, Ragoussis J, Harris AL (2011) The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer* 104(7):1168–1177. doi:10.1038/sj.bjc.6606076
21. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN (2005) Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* 85(1):154–159. doi:10.1038/labinvest.3700208
22. Klopffleisch R, Weiss AT, Gruber AD (2011) Excavation of a buried treasure—DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histol Histopathol* 26(6):797–810
23. Ma Y, Dai H, Kong X (2012) Impact of warm ischemia on gene expression analysis in surgically removed biosamples. *Anal Biochem* 423(2):229–235. doi:10.1016/j.ab.2012.02.003
24. von Smolinski D, Leverkoehne I, von Samson-Himmelstjerna G, Gruber AD (2005) Impact of formalin-fixation and paraffin-embedding on the ratio between mRNA copy numbers of differently expressed genes. *Histochem Cell Biol* 124(2):177–188. doi:10.1007/s00418-005-0013-0
25. Srinivasan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 161(6):1961–1971. doi:10.1016/S0002-9440(10)64472-0
26. Abrahamsen HN, Steiniche T, Nexø E, Hamilton-Dutoit SJ, Sorensen BS (2003) Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. *J Mol Diagn* 5(1):34–41. doi:10.1016/S1525-1578(10)60449-7
27. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K (1999) Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* 27(22):4436–4443
28. Korga A, Wilkolaska K, Korobowicz E (2007) Difficulties in using archival paraffin-embedded tissues for RNA expression analysis. *Postepy Hig Med Dosw (Online)* 61:151–155
29. Becker C, Hammerle-Fickinger A, Riedmaier I, Pfaffl MW (2010) mRNA and microRNA quality control for RT-qPCR analysis. *Methods* 50(4):237–243. doi:10.1016/j.ymeth.2010.01.010
30. Ambros V (2001) microRNAs: tiny regulators with great potential. *Cell* 107(7):823–826
31. Dijkstra JR, Mekenkamp LJ, Teerenstra S, De Krijger I, Nagtegaal ID (2012) MicroRNA expression in formalin-fixed paraffin embedded tissue using real time quantitative PCR: the strengths and pitfalls. *J Cell Mol Med* 16(4):683–690. doi:10.1111/j.1582-4934.2011.01467.x
32. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, Ju J (2007) Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* 13(10):1668–1674. doi:10.1261/ma.642907
33. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogossova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105(30):10513–10518. doi:10.1073/pnas.0804549105
34. Oliveira IB, Ramos DR, Lopes KL, Souza RM, Heimann JC, Funakawa LN (2012) Isolated total RNA and protein are preserved after thawing for more than twenty-four hours. *Clinics (Sao Paulo)* 67(3):255–259
35. Thasler WE, Thasler RM, Schelcher C, Jauch KW (2013) Biobanking for research in surgery: are surgeons in charge for advancing translational research or mere assistants in biomaterial and data preservation? *Langenbecks Arch Surg* 398(4):487–499. doi:10.1007/s00423-013-1060-y



## RESEARCH ARTICLE

## Open Access

# MicroRNA profile in very young women with breast cancer

María Peña-Chilet<sup>1</sup>, María T Martínez<sup>1</sup>, Jose A Pérez-Fidalgo<sup>1</sup>, Lorena Peiró-Chova<sup>2</sup>, Sara S Oltra<sup>1</sup>, Eduardo Tormo<sup>1</sup>, Elisa Alonso-Yuste<sup>3</sup>, Beatriz Martínez-Delgado<sup>4</sup>, Pilar Eroles<sup>1</sup>, Joan Climent<sup>1</sup>, Octavio Burgués<sup>3</sup>, Jaime Ferrer-Lozano<sup>2,3</sup>, Ana Bosch<sup>1,5</sup>, Ana Lluch<sup>1</sup> and Gloria Ribas<sup>1\*</sup>

## Abstract

**Background:** Breast cancer is rarely diagnosed in very young women (35 years old or younger), and it often presents with distinct clinical-pathological features related to a more aggressive phenotype and worse prognosis when diagnosed at this early age. A pending question is whether breast cancer in very young women arises from the deregulation of different underlying mechanisms, something that will make this disease an entity differentiated from breast cancer diagnosed in older patients.

**Methods:** We performed a comprehensive study of miRNA expression using miRNA Affymetrix2.0 array on paraffin-embedded tumour tissue of 42 breast cancer patients 35 years old or younger, 17 patients between 45 and 65 years old and 29 older than 65 years. Data were statistically analyzed by t-test and a hierarchical clustering via average linkage method was conducted. Results were validated by qRT-PCR. Putative targeted pathways were obtained using DIANA miRPath online software.

**Results:** The results show a differential and unique miRNA expression profile of 121 miRNAs (p-value <0.05), 96 of those with a FDR-value <0.05. Hierarchical clustering grouped the samples according to their age, but not by subtype nor by tumour characteristics. We were able to validate by qRT-PCR differences in the expression of 6 miRNAs: miR-1228\*, miR-3196, miR-1275, miR-92b, miR-139 and miR-1207. Moreover, all of the miRNAs maintained the expression trend. The validated miRNAs pointed out pathways related to cell motility, invasion and proliferation.

**Conclusions:** The study suggests that breast cancer in very young women appears as a distinct molecular signature. To our knowledge, this is the first time that a validated microRNA profile, distinctive to breast cancer in very young women, has been presented. The miRNA signature may be relevant to open an important field of research in order to elucidate the underlying mechanism in this particular disease, which in a more clinical setting, could potentially help to identify therapeutic targets in this particular set of patients.

**Keywords:** Breast cancer, Young women, miRNA, Molecular profile

## Background

Breast cancer (BC) is the most common cancer in women worldwide. It is estimated that breast cancer will account for up to 29% of all new cases of cancer in women in the USA in 2013 [1]. Although the median age at onset is 61 years, approximately one in forty women diagnosed with early breast cancer is very young, constituting 5 to 7% of all cancer deaths in these women [2,3].

Breast cancer in very young women is typically more aggressive than in their older counterparts, in part owing to the over-representation of high-grade, triple-negative tumours in the former patients. Indeed, very young patients diagnosed with breast cancer usually present larger and poorly differentiated tumours, lymph node invasion, HER2 overexpression and an absence of hormone receptor (HR) expression [4,5]. It is still a matter of controversy, however, as to whether breast cancer in very young patients (BCVY) is a biologically unique entity that should be considered separately or if its behaviour and outcome is solely due to the higher frequency

\* Correspondence: gribasdespuig@gmail.com

<sup>1</sup>Medical Oncology and Hematology Unit, INCLIVA Biomedical Research Institute, Av. Blasco Ibañez, 17, Valencia 46010, Spain  
Full list of author information is available at the end of the article



© 2014 Peña-Chilet et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

of more aggressive subtypes. There are contradictory studies on the identification of young age as an independent prognostic factor [5-7].

MicroRNAs (miRNAs) are post-transcriptional regulators that bind to complementary sequences on target messenger RNA transcripts (mRNAs), usually resulting in translational repression or target degradation and gene silencing [8-10]. Deregulation of many of the miRNA's expression has been linked to various types of disease [11-13]; (<http://www.mir2disease.org/>). The expression of several miRNAs has been found to be deregulated in some types of cancer [14,15]. High levels of some miRNA have been linked to stem cell promotion [16-18], while others exhibit a reduced expression of many, promoting loss of differentiation [19,20]. Both are common traits in tumour development, but many other unknown underlying mechanisms could be affected [21]. Moreover, miRNA expression profiling has been previously shown to be a useful tool for classifying different cancer risk stratification, outcome prediction and classification of histological subtypes [22,23]. Regarding breast cancer, several studies have identified molecular markers, such as miR-21, miR-9, let-7, miR-205, miR-200 family, miR-126 and miR-335, which correlate to tumours of high metastatic and proliferative capacity, larger size and thus poorer prognosis [18,24-29].

As previously stated, very little is known about the biology of BCVY, and published studies have focused mainly on hereditary tumours and genomic traits [30,31]. Nevertheless, the data from two previous published studies point to the possibility that BCVY could have a distinct molecular identity [32,33].

We performed a study with the aim to identify whether BCVY had a different miRNA profile compared with older age. A secondary objective was to identify those miRNA that are typically up- or downregulated in the tumours from very young patients, which could reveal the ongoing mechanisms of the development of the tumour and of its aggressiveness at an early age.

## Methods

### Sample selection

Tissue samples were obtained from patients undergoing surgery for breast cancer at the Hospital Clínico Universitario of Valencia from among the Spanish population. As shown in Figure 1, we selected two groups of women according to their age, one group of women 35 years old or younger (BCVY), and the other of those older than 65 years of age. After excluding those samples with known BRCA1/2 mutation, we got a total of 148 women 35 years old or younger and 3140 patients over 65 years of age with an available biopsy sample before any treatment. We screened for BRCA1/2 status according to the ESMO Guidelines for Hereditary

Breast Cancer that follow standard criteria for high-risk patients and from those, we excluded the positive cases from our study as well as those with known or suspected family history of breast cancer. All the participants, with one exception, were Caucasian women of European origin and gave written, informed consent by signing an approved Biobank document allowing research use of the remaining diagnostic tumour material. The study was approved by the Institutional Health Incliva-Hospital Clínico Ethics Committee.

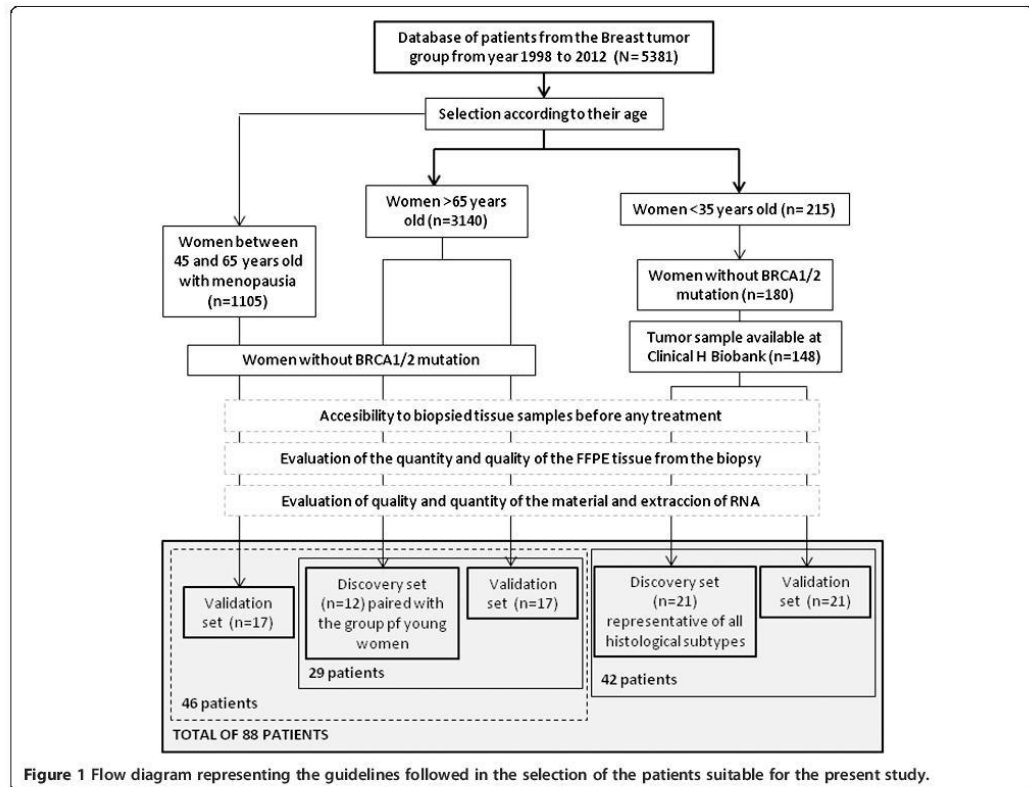
After evaluating the quantity and quality of the material, a total of 33 breast cancer patients were included in this study: 21 women diagnosed before the age of 35 categorized as the BCVY group, and 12 breast cancer patients older than 65 years old and matched by tumour subtype (Table 1). We also collected tissue samples from three young cancer-free women undergoing mammary reduction surgery, with ages 28, 34 and 40 years old, and BMI within normal values (between 20 and 24).

For the validation set, we included a total of 55 new independent samples from HCUV, selected with the same aforementioned criteria, 21 of which were BCVY patients, 17 were older than 65 years, and 17 individuals were post-menopausal women between 45 and 65 years old. We included this new age group in order to better establish the cut-off age of the miRNA profile. Overall, the present study was performed on a total of 88 breast cancer patients.

All of the tissue samples were biopsied from primary tumour and locked in FFPE (Formalin-fixed paraffin-embedded). Three- $\mu$ m thick paraffin sections were stained with haematoxylin and eosin (H&E) in order to obtain a histological control of all samples. All tissue samples contained >30% tumour material. Tumour grade was assessed based on Bloom-Richardson scoring system. Images of the sample's tissues can be seen in Additional file 1.

To obtain the molecular characteristics of the tumour, presence of molecular markers ER, PR, HER2 and Ki67 was evaluated following the ACO/CAP guidelines. ER and PR status were obtained by IHC staining and HER2 Immunohistochemical staining on TMA sections was performed by the EnVision method with a heat-induced antigen retrieval step. Staining results were assessed by a pathologist. ER and PR were scored based on two-stage scoring system: positive (1) for >10% of ER/PR positive cells and negative (0) for less than 10%, as described previously. Proliferation was assessed measuring percentage of Ki-67 expression. HER2 was called positive either by detection of ERBB2 gene amplification by FISH analysis and/or 3+ staining by DAKO system on HercepTest™. Where duplicate cores gave discordant results, the higher score was used. Breast cancer tumours were classified into four subtypes based on IHC-model (Tang P. *et al.*, 2009) as: luminal A (ER+ and/or PR+, HER2-, Ki67 < 14%);





luminal B (ER + and/or PR+, HER2-, Ki67 > 14%); Triple Negative (ER-, PR-, HER2-) and HER2 overexpressed/amplified (ER-, PR-, HER2+), plus an additional group Luminal/HER2 (ER + and/or PR+, HER2+).

#### RNA isolation

Total RNA was isolated using RecoverAll Total Nucleic Acid Isolation Kit (Applied Biosystems by Life Technologies, Carlsbad, California, USA) following the manufacturer's protocol. RNA concentration was measured using a NanoDrop ND 2000 UV-vis Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington DE, USA).

RNA integrity determined by RIN (RNA integrity number) value was assessed with the Agilent 2100 Bioanalyzer using the RNA 6000 Nano Assay (Agilent Technologies Inc., Santa Clara, CA, USA).

#### miRNA microarray

Microarray expression profiling was performed using GeneChip® miRNA 2.0 Array (Affymetrix, Santa Clara, CA, USA), containing a total of 15,644 probes, in 11 replicates, including 1100 human mature miRNA, their

precursors, 2334 human snoRNA (small nucleolar RNA) and scaRNA (small Cajal body-specific RNA) annotated in the miRBase v.15 database. Hybridisation and scanning were performed according to the Affymetrix standard protocol, using Affymetrix Expression Console software. The microarray dataset is publicly available at GEO database (GSE48088) <http://www.ncbi.nlm.nih.gov/geo/info/linking.html>.

#### Array data processing and analysis

Data passed quality controls implemented in Expression Console software, build 1.2.0.20, and QC Tool (Affymetrix, Santa Clara, CA, USA). All data were normalised by the DABG-RMA, detected above background (DABG) Robust Multichip Average (RMA) method. We selected 1100 human miRNAs, and set as the threshold for low expression the highest common intensity value of miRNA from vegetable organisms (not supposed to be expressed in humans) included in the microarray. After filtering the data, in order to determine the differences in expression pattern between tumours from BCVY patients and from older ones, we carried out differential expression analyses

**Table 1 Clinical tumour characteristics of the sample groups used in the present study**

Age group	Discovery set N = 34 (%)		Validation set N = 55 (%)		
	<35 (n = 22)*	>65 (n = 12)	<35 (n = 21)	45-65 (n = 17)	>65 (n = 17)
Age mean (SD)	31.41 (4.04)	73.3 (10.18)	31.4 (2.87)	56.94 (5.58)	69.4 (4.48)
BMI mean (SD)	21.98 (3.53)	29.44 (6.66)	23.95 (5.50)	30.15 (8.04)	30.23 (6.55)
Histological grade					
I	1 (4.55)	3 (25.00)	3 (14.29)	8 (47.06)	3 (17.65)
II	8 (36.36)	3 (25.00)	7 (33.33)	5 (29.41)	12 (70.59)
III	13 (59.09)	6 (50.00)	11 (52.38)	4 (23.53)	2 (11.76)
Histological Type					
Ductal	19 (86.36)	11 (91.67)	19 (90.48)	13 (76.47)	12 (70.59)
Lobular	0 (0.00)	0 (0.00)	1 (4.76)	0 (0.00)	2 (11.76)
Others	3 (13.64)	1 (8.33)	1 (4.76)	4 (23.53)	3 (17.65)
Tumour size					
< 2 cm	3 (13.64)	9 (75.00)	3 (14.29)	11 (4.71)	14 (82.35)
2-5 cm	15 (68.18)	2 (16.67)	15 (71.43)	6 (35.29)	3 (17.65)
> 5 cm	4 (18.18)	1 (8.33)	3 (14.29)	0 (0.00)	0 (0.00)
Nodal status					
Positive	10 (45.45)	4 (33.33)	7 (33.33)	6 (35.29)	3 (17.65)
Negative	12 (54.55)	8 (66.67)	14 (66.67)	11 (64.71)	14 (82.35)
Receptors					
ER+	17 (77.27)	9 (75.00)	14 (66.67)	14 (82.35)	14 (82.35)
ER -	5 (22.73)	3 (25.00)	7 (33.33)	3 (17.65)	3 (17.65)
PR+	15 (68.18)	7 (58.33)	14 (66.67)	13 (76.47)	13 (76.47)
PR -	7 (31.82)	5 (41.67)	7 (33.33)	4 (23.53)	4 (23.53)
HER2+	10 (45.45)	3 (25.00)	5 (23.81)	2 (11.76)	2 (11.76)
HER2 -	12 (54.55)	9 (75.00)	16 (76.19)	15 (88.24)	15 (88.24)
Ki67					
1-14%	5 (22.73)	3 (25.00)	5 (23.81)	8 (47.06)	5 (29.41)
14-30%	8(36.36)	5 (41.67)	11 (52.38)	7 (41.18)	10 (58.82)
>30%	9 (40.91)	4 (33.33)	5 (23.81)	2 (11.76)	2 (11.76)
Histological subtype					
Luminal A	5 (22.73)	3 (25.00)	3 (14.29)	8 (47.06)	5 (29.41)
Luminal B	6 (27.27)	4 (33.33)	10 (47.62)	4 (23.53)	9 (52.94)
TN	2 (9.09)	2 (16.67)	3 (14.29)	3 (17.65)	2 (11.76)
Luminal/HER2	5 (22.73)	2 (16.67)	2 (9.52)	0 (0.00)	1 (5.88)
HER2	4 (18.18)	1 (8.33)	3 (14.29)	2 (11.76)	0 (0.00)

BMI stands for Body Mass Index and is expressed in terms of kg/m<sup>2</sup>. ER: Estrogen receptor; PR: Progesterone receptor. HER2: ErbB2 receptor; +/-: presence (+) and absence (-) of receptor overexpression. HER2 is considered positive (+) when immunohistochemical analyses show +++/+++ or +++/+++ (and FISH shows HER2 amplification). Subtypes have been categorized according to Hormonal Receptors, HER2 expression and Ki67 value. \*One of the initial 22 samples younger than 35 years old, was removed from the study due to methodology QC thresholds.

with the POMELO II tool (<http://asterias.bioinfo.cnio.es>). We performed a *t*-test by permutation testing, and *p*-values were adjusted for multiple comparisons by Benjamini & Hochberg False Discovery Rate (FDR). Those miRNAs with a FDR *p*-value <0.05 were considered statistically significant. Average linkage hierarchical clustering was performed to obtain clusters of data sets, using

Gene Cluster and Treeview software (<http://www.eisenlab.org/eisen/>).

In order to evaluate whether the miRNA profile detected was a result of age differences or if it was influenced by confounding factors such as tumour size, histological grade and other features characteristic of more aggressive tumours, we performed a *t*-test analysis on the filtered

miRNA expression data after correcting by the influence of nodal status, histological grade, percentage of ki67 expression and tumour size. Data were adjusted obtaining the correlation between miRNAs' expression and tumour characteristics via linear regression with R Studio (<http://www.rstudio.com/>), using the remaining residuals for the correction.

#### Validation by qRT-PCR

Quantitative real time-PCR (qRT-PCR) of selected miRNAs was performed on breast tumour tissue samples in independent series of women younger than 35 years old, women between 45 and 65 years old, and women older than 65 years using TaqMan microRNA Assays (Applied Biosystems by Life Technologies, Carlsbad, California, USA). Normalisation was done with RNU43 snoRNA. The data were managed using the Applied Biosystems software RQ Manager v1.2.1. Relative expression was calculated by using the comparative Ct method and obtaining the fold-change value ( $\Delta\Delta Ct$ ). Data analyses were performed via GraphPad Prism v6.00. The ANOVA test was performed, and Tukey's test was used to correct for multiple comparisons (p-value threshold of 0.05). Brown-Forsythe test was used to assess the homogeneity of the variances in the different samples groups. Finally, the association of the validated miRNAs was adjusted by breast cancer subtypes using RStudio (<http://www.rstudio.com/>).

#### Pathway enrichment analysis and candidate gene searching

DIANA miRPath pathway enrichment analysis was used to gain insight into global molecular networks and canonical pathways related to differentially expressed miRNAs (<http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index>). The software performs an enrichment analysis of multiple miRNA target genes comparing each set of miRNA targets to all known KEGG pathways. Those pathways showing a FDR p-value <0.05 were considered significantly enriched between classes under comparison. We also searched for candidate genes using Target-scan online software (<http://www.targetscan.org/>) and previously published data.

## Results

### miRNA expression profiling in primary breast tumours from discovery set

After initial pre-processing of 1100 miRNA expression data, we discarded miRNAs with uniformly low expression, reducing the number of miRNA to a total of 251 hsa-miRNAs. Finally we detected 119 miRNAs differentially expressed between the two groups (p-value < 0.05) and after adjusting for FDR, 96 miRNAs remained significant (Additional file 2). In addition, all of these miRNAs' expression remained significantly different even after

correcting by possible confounding factors (tumour size, nodal status, ki67% and histological grade), suggesting that young age is an independent factor (Additional file 3).

Mammary tissue from three cancer-free women was included as healthy control group. When unsupervised hierarchical clustering was performed two clearly different groups of samples (Figure 2). One group included all except one of the BC tissue samples from very young women. The other group contained all the samples from both older women and normal tissue. We were able to identify two major nodes of differentially regulated miRNA in the BCVY group. As shown in Figure 2, the 52 miRNAs overexpressed in the BCVY group can be seen clustered in the top node, 115, (correlation 0.383) while the lower node, 119, shows the 67 miRNAs found to be repressed in this BCVY set of tumours (correlation 0.102).

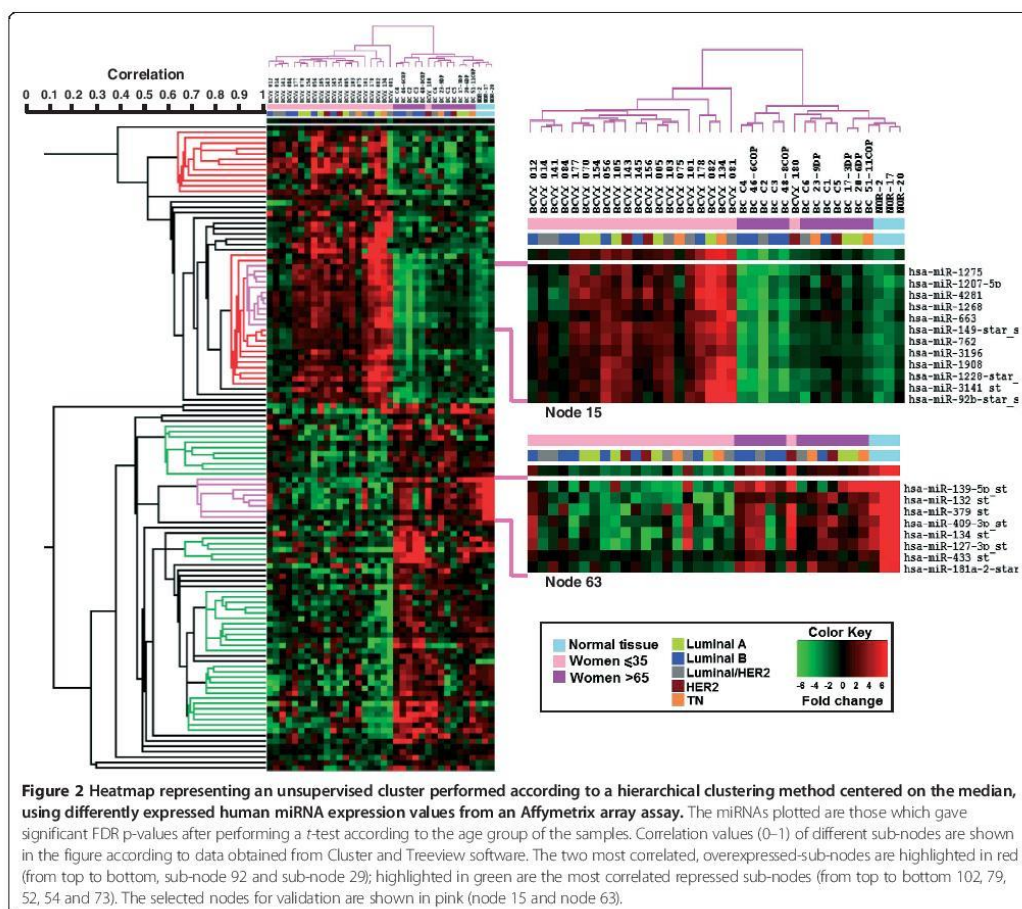
From the overexpressed node (node 115) we were able to identify two differentiated sub-nodes: node 92 (correlation value 0.63) and node 84 (correlation value 0.66), that appears to be more homogenous inside each group of patients. Within node 84 we selected sub-node 15 (correlation 0.923) including miRNAs with the highest significant expression difference between Young vs Old women-Normal. When scanning the repressed node 119, we identified six interesting sub-nodes, covering most of the node: sub-node 102 (correlation 0.582), sub-node 79 (correlation 0.667), sub-node 76 (correlation 0.678), sub-node 63 (correlation 0.72), sub-node 54 (correlation 0.75) and sub-node 52 (correlation 0.75). While sub-nodes 54 and 52 reached the highest correlation value, we selected sub-node 63 for the further study, since the expression of the whole cluster in cancer-free tissue samples was remarkably high, and also because the BC patients showed levels uniformly higher than the BCVY ones. MicroRNAs from the selected sub-nodes are listed in Figure 2, and their expression fold change and p-values are shown in Additional file 4.

Regarding the hsa-miRNA-star (hsa-miR-92\*, hsa-miR-149\* and hsa-miR-1228\*), when the main form of both miRNA species was also differentially expressed (hsa-miR-92\* and hsa-miR-149\*), we selected the main form of each miRNA (hsa-miR-92-3p and hsa-miR-149-5p) in order to get more reliable results.

### Pathway enrichment and selection of candidate miRNAs for validation step

With the 20 selected miRNAs, composing the two most striking sub-nodes, we performed in silico analyses and database searches to determine which ones might be more biologically relevant in the breast cancer context, and eligible for validation. Considering that the miRNAs target from small to a very large number of mRNA





transcripts, common deregulation of a set of miRNA could have a potential impact on various biological pathways. To assess which pathways could be affected by the deregulation, we used DIANA miRPath v2.0. When studying miRNAs from the selected sub-nodes, KEGG pathway enrichment analysis revealed several pathways overrepresented with a FDR p-value <0.05, including adhesion and mobility related pathways (extracellular matrix-receptor interactions, glycosaminoglycan biosynthesis, adherens junction, focal adhesion, cell adhesion molecules, regulation of actin cytoskeleton), biological proliferation-or differentiation-related pathways relevant in cancer development or progression (MAPK signalling, endocytosis, axon guidance), circadian rhythm and several cardiomyopathies. Within the genes involved in the pathways related with the selected miRNAs, we found many integrins, laminins, collagen, MAPkinases, semaforins, interleukins,

claudins and calmodulins, among others. More information and FDR p-values associated with each pathway are shown in Additional file 5.

We have finally selected 12 miRNAs (7 from the sub-node 15 and 5 from sub-node 63) as putative candidate targets to perform further studies according to their significance, previous published studies, and implications in biological pathways: hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-149\*, hsa-miR-762, hsa-miR-3196, hsa-miR-1228\*, hsa-miR-92\*, hsa-miR-139-5p, hsa-miR-132, hsa-miR-379, hsa-miR-409-3p, and hsa-miR-433. Supplementary information is available in Additional file 6.

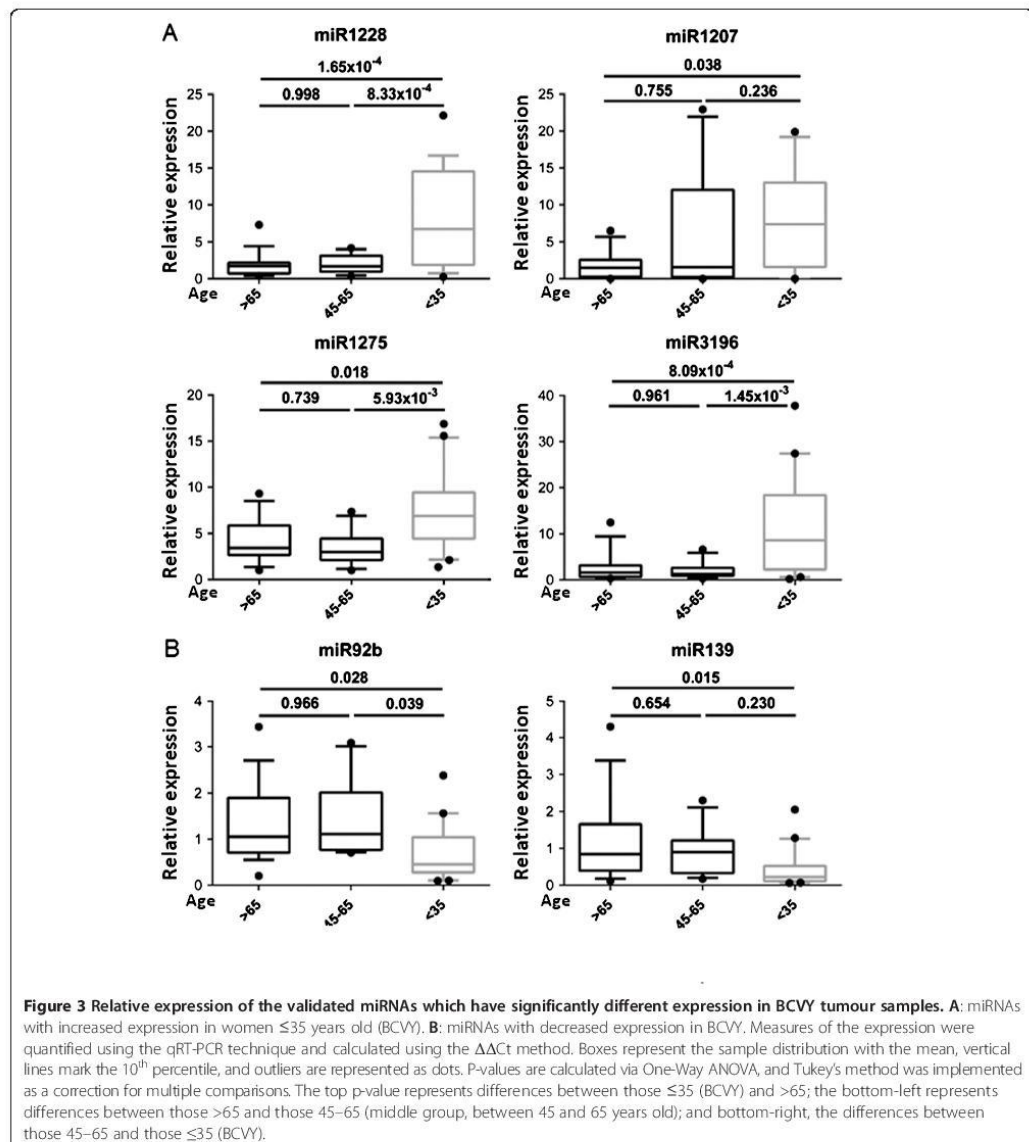
#### qRT-PCR validation in an independent set of samples

We performed a validation of the 12 selected miRNAs using qRT-PCR, on an independent set of samples with similar characteristics as those used in the Discovery

phase. However, in the validation set, we classified the 55 samples into three groups (21 tumours from women younger than 35 years, 17 tumours from women aged between 45 and 65 and finally 17 tumours from women older than 65) (see the clinical characteristics for the entire sample set in Table 1).

We did not find evidence suggesting that women older than 65 had any differential characteristics with

the 45 to 65 year age group in terms of miRNA expression (Figure 3). Six of the twelve selected miRNAs were confirmed to have a differential expression across groups: miR-1228\* (p-value  $4.73 \times 10^{-5}$ ), miR-3196 (p-value  $1.98 \times 10^{-4}$ ), miR-1275 (p-value  $2.86 \times 10^{-3}$ ), miR-92b (p-value 0.012), miR-139-5p (p-value 0.018), and miR-1207 (p-value 0.047). After Tukey's correction for multiple comparisons, we obtained p-values between



each group, confirming that the differences were higher between BCVY samples and the other two older groups (Figure 3). The first four values agree with an upregulation in BCVY, while the last two, miR-92b and miR-139-5p, were indeed downregulated in young women group (see Figure 3 and Additional file 2). Afterward, a *t*-test analysis was used to compare the BCVY group with the two older patient groups, treating the last two as only one group. The results obtained were the following: miR-1228 (p-value  $7.13 \times 10^{-6}$ ), miR-3196 (p-value  $3.23 \times 10^{-5}$ ), miR-1275 (p-value  $7.77 \times 10^{-4}$ ), miR-92b (p-value  $2.94 \times 10^{-3}$ ), miR-139-5p (p-value  $6.57 \times 10^{-3}$ ), and miR-1207 (p-value 0.523). After adjusting for tumour subtype, four of them remained statistically significant: miR-1228 (p-value 0.0032), miR-1275 (p-value 0.0010), miR-3196 (p-value 0.0095), miR-139 (p-value 0.0172). It is worth mentioning that all of the other miRNAs tested did indeed lose expression, as was expected; however, no significant value could be obtained due to a lack of detection sensitivity.

#### Putative functional implication of the validated miRNAs

In order to establish a possible underlying biological implication with the validated deregulated miRNAs (hsa-miR-1228\*, hsa-miR-3196, hsa-miR-1275, hsa-miR-1207-5p, hsa-miR-92b, hsa-miR-139-5p), we used online software and miRNA databases to perform an exhaustive search for enrichment of candidate pathways and genes.

Moreover, we text-mined all published studies related to the six validated miRNAs with the terms cancer and/or putatively target genes. We found several enriched pathways related to cell motility and invasion: adherens junction (FDR p-value  $1.03 \times 10^{-4}$ ), glycosaminoglycan biosynthesis (FDR p-value  $2.23 \times 10^{-3}$ ), cell adhesion molecules (CAMs) (FDR p-value 0.035) and several related genes such as integrins, claudins and cadherins. We also found enriched pathways involved in proliferation and apoptosis: calcium signalling (FDR p-value 0.035) and Wnt signalling (FDR p-value 0.041) (Table 2).

#### Discussion

Our data lead us to think of a differentiated miRNA signature in the group of patients diagnosed at  $\leq 35$  years. We have detected a wide range of differentially regulated miRNAs in BCVY when compared to tumours diagnosed in older women and healthy breast tissue. The fact that the older women's miRNA profile was highly similar to the healthy breast tissue profile (obtained from women under 40 years old) highlights the distinctiveness of the miRNA profile being characteristic of BCVY patients. These differences were maintained after validation, in an independent series of samples using a different technique, in which controls between 45 and 65 years were included, suggesting that our findings are consistent and not a result of the false positives characteristic of high throughput technologies. Although from a limited sample size, the in

**Table 2 Significantly enriched signalling pathways associated to the validated differentially expressed miRNAs**

KEGG Pathway	FDR	Number of miRNAs	Putative target genes
Adherens junction	$1.03 \times 10^{-4}$	4	ACTB, WASL, RAC2, IGF1R, PTPRF, TJP1, CDH1, CSNK2B, FARP2, PTPRJ, SSX2IP, PTPRB, PVRL1
Steroid biosynthesis	$1.49 \times 10^{-4}$	2	SC5DL, DHCR24, CYP27B1
Glycosaminoglycan biosynthesis - chondroitin sulfate	$2.23 \times 10^{-3}$	2	CHPF
Dilated cardiomyopathy	$2.23 \times 10^{-3}$	4	ACTB, ADCY1, ITGB8, ITGB3, CACNG7, TTN, ADCY3, ITGA5, ITGB5, DMD, PRKACA, CACNA1C, CACNA2D2, IGF1, CACNA2D1, CACNA1D, PRKACB
Lysine degradation	0.014	3	PLOD3, ASH1L, SETD8, MLL2, MLL, SETD1A
Taste transduction	0.014	4	GNG13, SCNN1B, GNB1, PRKACA, SCNN1A, PRKACB
Cell adhesion molecules (CAMs)	0.035	3	NEGR1, ITGB8, MPZ, SDC3, CLDN19, CLDN11, PTPRF, CDH1, CD276, NRXN2, CD86, ICOSLG, NFASC, PVRL1, NLGN3
Calcium signaling pathway	0.035	4	PRKCA, ADCY1, ATP2B2, BDKRB2, ADCY3, ITPKB, SLC8A2, PDE1B, CACNA11, CAMK2A, NOS1, PRKACA, NTSR1, CACNA1C, PHKA1, MYLK3, CACNA1D, GRIN2A, PRKACB, CHRM1
Fc gamma R-mediated phagocytosis	0.036	2	PRKCA, WASL, RAC2, PLA2G4F, ASAP3, GSN, FCGR1A, MARCKS, PIP5K1A, PLD2, AKT3, VASP, MYO10, ARF6
Hypertrophic cardiomyopathy (HCM)	0.036	4	ACTB, ITGB8, ITGB3, CACNG7, TTN, ITGA5, ITGB5, DMD, CACNA1C, CACNA2D2, IGF1, CACNA2D1, CACNA1D
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	0.039	3	ACTB, DSC2, ITGB8, ITGB3, CACNG7, ITGA5, ITGB5, DMD, CACNA1C, CACNA2D2, CACNA2D1, CACNA1D
Wnt signaling pathway	0.041	4	PRKCA, DVL3, WNT16, ROCK1, RAC2, VANGL1, PPP2CA, PPP2R5C, CAMK2A, WNT6, PRICKLE1, PPP2R5B, JUN, APC2, PPP2R1A, PRKACA, VANGL2, CSNK2B, WNT9A, PRKACB

FDR refers to p-values adjusted by Benjamini & Hochberg's False Discovery Rate. Number of miRNAs is the count of miRNAs involved in the pathway and targeting one or more genes of the pathway.



silico searches done, pointed out to pathways closely related to high cell proliferation, motility and invasion, which support the fact that tumours from young woman have an aggressive behaviour and are prone to metastasis. Other studies in breast cancer showing miRNA profiles characterized by bad prognosis, revealed similar signalling deregulated pathways [34].

Recently, miRNA profiling has arisen as a major approach study technique, that aims to gain more insight into tumour biology, and widespread deregulated miRNA has been demonstrated in various tumour types [35]. It has been shown that specific miRNA, such as miR-195 and let-7a, may play a role as a potential diagnostic tool and a new biomarker for both detecting non-invasive, early breast cancer as well as monitoring treatment effectiveness [36,37]. Furthermore, other studies have revealed miRNA expression to be specific for each breast tissue type, but it seems not to exist a specific miRNA profile distinctive of BC subtypes, which may point to both differences in the biology of tumours, as well as to patterns of response to treatment [38-44]. As suggested in previous studies, our results reinforce the applicability of the study of miRNA expression in FFPE samples, since the small size of the miRNA preserves them from degradation [45]. In addition, the use of Taq-Man-assay for microRNA expression profiling and validation in FFPE breast tissue samples is reported to offer excellent inter-experimental reproducibility and biological accuracy [41,46].

Several studies confirm that BC in very young patients is a more aggressive disease given the high frequency of adverse prognosis factors or more aggressive subtypes at this age [4]. In most studies published the cut-off age for defining young women is 45 years. However, results from a large Korean breast cancer study with 9885 premenopausal patients detects a sharp increase in risk of death from breast cancer in women younger than 35 years old, establishing this age as the more reasonable cut-off defining young age-onset breast cancer [47]. Following this recommendation, we have determined 35 years old as the threshold for the young group, since we considered that, despite this cut-off makes harder patient's recruitment within the young age group, it identifies a more homogeneous population.

Whether breast cancer in the young is a distinct biological entity or just a reflection of a higher percentage of cases with aggressive phenotypes (that are otherwise indistinctive from aggressive tumours in older patients) remains a matter of controversy and an important outstanding question with possible therapeutic implications [6].

In order to minimize clinical-pathological putative confounding effects, we have performed several comparisons in the expression of these miRNA in terms of tumour size, axillary node involvement, histological grade and

ki67, with the results showing no significant differences. This absence of significant differences, having considered the most relevant adverse prognostic factors, suggests that young age could be indeed a factor independent of subtype, size, nodal involvement, or proliferative characteristics that merits further study.

Moreover, no distinct therapy is considered for young or very young patients, but in many cases these patients are over-treated as a consequence of their young age. Thus identification of new prognostic or diagnostic platforms and identification of specific targets is needed for this subgroup [48].

Data about genomic profile in young breast cancer patients are scarce in the literature; however, Colak and cols. have recently published an age-specific gene expression signature in breast cancer [33]. Although, the study focused on women younger than 45, it also included 6 samples from those younger than 35 (BCVY). The study pointed to genes barely implicated in the pathways found in our study. That notwithstanding, the authors proposed that there is a different molecular profile characterizing breast cancer in very young women.

A recent retrospective study analysed data from 20 data sets and assessed the role of these genetic signatures in predicting the prognosis in breast cancer in young women. This study confirmed that BC patients aged 40 or less were diagnosed more frequently with oestrogen receptors and HER2 negative tumours. Furthermore, although proliferation-related gene signatures were not associated with age in this study, in two independent cohorts BC in younger patients was associated with immature epithelial cell and growth factor signalling pathways, leading to the conclusion that BC in young women seems to be a distinct entity beyond the intrinsic breast cancer subtype classification [32].

The data obtained with our miRNA profiling supports the evidence proposed by Yau and cols that could be detected an age-dependent signature in BC with a number of samples from young women very similar to ours. Although the conclusions are raised on gene expression analyses the signaling pathways detected including cell cycle, mammary gland development and extracellular matrix (ECM) are also highlighted in our study [49].

In our study, the most significantly set of deregulated miRNAs in BCVY pointed out to pathways related to apoptosis, cell motility, proliferation, mitotic regulatory processes and the PI3K and IGF1R transduction that confer tumours high metastatic capacity, increasing progression and invasion (Table 3). Furthermore, many of the pathways involved include integrins and laminins as target genes, described as proliferation genes portraits of breast tumours [50].

If we take into account the set of validated miRNAs (upregulated miR-1228\*, miR-3196, miR-1275, miR-1207

**Table 3 Published information about validated microRNA target genes and implication in cancer**

miRNAs	Gene	Target/Function	Published papers
miR-1228	<i>MOAP1</i>	<i>Bcl</i> homologous. When repressed, allows cells to escape from apoptosis.	Yan et al. <i>Apoptosis</i> . 2012 [69]
miR-3196	<i>PAX2</i> , <i>THTPA</i> , <i>PIK3R2</i> , <i>BBC3</i>	<b>PAX2</b> . Implicated in suppression of translation (through WT1). Associated to low breast cancer risk. Repression of <i>PAX2</i> would promote a more aggressive breast cancer. <b>THTPA</b> . Metastasis tumour suppressor. <b>PIK3R2</b> . Proliferation pathway. Anti-apoptotic. <b>BBC3</b> . Pro-apoptotic gene and associated to tumour size.	Beauchemin et al. <i>Molecular Cancer</i> . 2011 [53]. Kovacevic et al. <i>Biochem Biophys Acta</i> . 2008 [58]. Tajnik et al. <i>Cancer Biomarkers</i> . 2012 [66]. Wong et al. <i>PlosOne</i> . 2012 [68]
miR-1275	<i>IGF1</i> , <i>NFIX</i> , <i>Claudin11</i>	<b>IGF1</b> related to tumour proliferation. <b>NFIX</b> hypermethylated in breast cancer lines. <b>Claudin 11</b> , cellular adhesion molecule, associated to invasion and capacity of metastasis when downregulated.	Castañó et al. <i>Cancer Discovery</i> . 2013 [54]. Awsare, et al. <i>Oncology reports</i> . 2011 [51]. Webb et al. <i>BMC Cell Biol</i> . 2013 [67]. Özata et al. <i>Endocr Relat Cancer</i> . 2011 [62]. Lian et al. <i>Int J Oncol</i> . 2012 [60]. Katsushima et al. <i>J Biol. Chem.</i> 2012 [57]
miR-1207	<i>DHCR24</i> , <i>Claudins</i>	<b>DHCR24</b> expression decreased in metastatic prostate cancer. <b>Claudins</b> . Associated with cell motility and tumour invasion and spread of tumour cells and metastasis.	Romanuk et al. <i>Am J Pathol</i> . 2009 [64]. Webb et al. <i>BMC Cell Biol</i> . 2013 [67]
miR-139	<i>RAP1B</i> , <i>c-FOS</i> , <i>IGF1R</i> , <i>TOP2A</i> <i>CXCR4</i>	<b>RAP1B</b> , family RAS (oncogene). It has a pseudogene. <b>IGF1R</b> Low expression of miR in colo-rectal cancer is associated to more advanced tumours and less survival. <b>FOS</b> proteins are implicated in cell cycle, differentiation and cellular transformation. Downregulation induces increase in apoptosis and more cellular differentiation. <b>TOP2A</b> , topoisomerase. Upregulation is associated to tumour proliferation. <b>CXCR4</b> , associated with progression and metastasis in CRC. Regulated by HER2-CD44 via miR139.	Guo et al. <i>Cell Biology</i> . 2012 [56]. Milde-Langosch et al. <i>Breast Cancer Res Treat</i> . 2013 [61]. Shen et al. <i>Biochemical Pharmacology</i> . 2012 [65]. Fan et al. <i>Cell Biochem Funct</i> . 2012 [55]. Bao et al. <i>Gastroenterology</i> . 2011 [52]
miR-92b	<i>PSMD10</i> , <i>FOX2</i>	<b>PSMD10</b> . Inhibition of the protein will slow down tumour progression in hepatocarcinoma. High increased expression related to worse prognostic in glioma. <b>FOX2</b> . RNA-binding protein regulates alternative splicing.	Leidner et al. <i>PlosOne</i> . 2013 [59]. He et al. <i>Nature</i> . 2005 [15]. Qian et al. <i>Gastroenterology</i> . 2012 [63]

Genes targeted by miRNAs were obtained neither via *in silico* predictions with Targetscan or searching for previously experimental published works. Most relevant genes are included in the table and most interesting cancer-related pathways or mechanisms has been added to *function* column.

and downregulated miR-139-5p and miR-92b) we were able to detect in the literature their implications in escaping of apoptosis and therefore, contributing to migration and invasion [15,51-69]. Downregulation of one particular miRNA, miR-139, might promote gastric tumour progression and metastasis via upregulating CXCR4, Bao and cols suggest that miR-139 might be suppressed by upstream HER2 signaling, a strongly deregulated pathway in breast cancer [52].

We cannot fully reject an influence of demographics or epidemiological factors inherent to young women in the tumour profile obtained. Bearing this in mind, we believe that the inclusion of healthy young breast tissue in the analysis might have minimized these putative confounding epidemiologic aspects.

Oestrogen levels are an age-dependent factor increasing the risk of breast cancer, which can themselves modulate the expression of several miRNAs. Moreover, miRNAs can as well regulate oestrogen receptor levels [70,71]. None of the miRNAs described in the literature related to oestrogen appears to be deregulated in the BCVY profile presented. The different profile obtained between BCVY and normal breast samples from young women (with similar oestrogen levels) supports the idea that the intrinsic levels of oestrogen are not responsible of the miRNA deregulation.

Early pregnancies and lactation are considered protector factors in BC and they could bias the comparison between young and older woman. In our series, older women in general have early age at first birth and still have developed breast cancer, while in the young women we have representation of both circumstances and all of them fall in the same miRNA profile, highlighting young age as the main leading factor of the profiling.

Mammary density is the strongest risk factor for non-familial breast cancer among women, apart from older age. Young women present denser breast tissue in mammographies, which makes more difficult the diagnosis. In addition, differences in normal breast microenvironment seem to exert an influence on the behavior of breast cancer cells in premenopausal women [72]. A recent molecular profiling study performed on adjacent tissue of breast tumours, classifies extratumoral stromal microenvironments into two primary gene expression phenotypes (Active and Inactive). The Active subtype has high expression of genes involved in activation of fibrosis, cellular movement, increased TWIST expression, and positive expression of TGF- $\beta$  signatures. Inactive phenotype is over-represented by samples from young women and according to Perez and cols, means higher levels of cell adhesion and cell-cell contact genes [73,74]. The miRNA profile presented in our study pointed out to pathways also related



with cell motility and adhesion, however the upregulation of implicated miRNAs indicates a downregulation of the targeted genes and therefore, stroma that correlates with young and denser breast tissue is not the driver in our deregulated miRNAs signature.

Finally, BMI (Body Mass Index) has been recently related to BC risk and survival. In older women higher BMI results in elevated risk of developing BC, while in young women is a protector factor [75]. In our study the mean value of BMI in BCVY is within normal ranges, which makes difficult any speculation about its role in BCVY.

In our study all patients but one are confirmed to be of European Caucasian ethnicity. This implies that our miRNA profile might not be representative of other genetic backgrounds such as African descent, where exists a higher proportion of breast cancer in young women. On the other side, the study has been performed in a very homogeneous population which reinforces the conclusions obtained and we encourage the replication of our findings in other ethnicity breast cancer collections.

## Conclusions

Our results suggest that breast cancer in young patients appears to be a different biological entity. Previously published works and in silico enrichment analyses of the miRNAs deregulated in young women suggest that targeted genes involved in proliferation pathways, cell adhesion, apoptosis, extracellular matrix and cell motility, deregulation of these pathways may lead to more aggressive, proliferative and metastatic tumour phenotypes. A more detailed analysis of those miRNA significantly up- or downregulated could guide both to the establishment of a different hypothesis about the potential molecular mechanisms involved in the carcinogenesis and to identify more specific therapeutic targets in this particular set of patients.

## Additional files

**Additional file 1: Haematoxylin and eosin staining images from tissue samples used in the present study.** Sections from FFPE tissue blocks stained with haematoxylin and eosin. **A, B**: correspond to samples from normal healthy mammary tissue. **C, D**: represent the mammary tumour from young women. **E, F**: show breast tumour tissue obtained from older women.

**Additional file 2: List and p-values of miRNAs significantly differences in expression between BCVY and BC.** Significantly different expressed miRNAs among two sample groups: young women in breast cancer (BCVY, younger than 35 years old) and older than 65 years women with breast cancer (BC). P-values and FDR p-values (corrected for Benjamini and Hochberg's False Discovery Rate for multiple comparisons) were obtained performing a *t*-test with 200000 permutations. Nomenclature of miRNAs belong to miRBase v.15.

**Additional file 3: List and p-values of miRNAs significantly differences in expression between BCVY and BC.** Significantly different expressed miRNAs among two sample groups: young women in

breast cancer (BCVY, younger than 35 years old) and older than 65 years women with breast cancer (BC), headed by Age FDR (from Additional file 1: Table S1), the association of miRNA was further analyzed correcting for tumour characteristics (grade, tumour size, Ki67 and nodal status) to assess whether there are any confounding factor. FDR p-values are corrected for Benjamini and Hochberg's False Discovery Rate for multiple comparisons and were obtained performing a *t*-test with 200000 permutations. Nomenclature of miRNAs belong to miRBase v.15.

**Additional file 4: MicroRNA selected for this study from miRNAs significantly associated with young women's breast cancer in discovery set and its validation.** Clusters were constructed using Average hierarchical clustering method. Statistics of the discovery set were calculated via *t*-test. FDR stands for False Discovery rate multitest corrected p-value. Only miRNAs remarked in bold were validated on an independent set of patients. Fold change of the validation set were calculated using  $\Delta\Delta Ct$  method and p-values were estimated via ANOVA test comparing the age groups and adjusted for multiple testing using Tukey's method. Bold denotes significant associated results in validation set.

**Additional file 5: Putative pathways enriched with the miRNAs included in the sub-node 15 and 63.** Code between brackets refers to the pathway code on KEGG, hsa means *homo sapiens*. FDR refers to p-value adjusted by False Discovery Rate. 20 miRNAs were included in a pathway enrichment analysis and obtained the genes union results.

**Additional file 6: Information about miRNAs selected for validation.** Information of the miRNAs selected for validation, including 3p or 5p form (3' or 5'), MIMAT ID, family and sequence of the mature miRNA. \*\*\* refers to the minor form of the miRNA, according to miRBase v.15 nomenclature.

## Abbreviations

ANOVA: Analysis of variances; BC: Breast cancer; BCVY: Breast cancer in very young women; BMI: Body mass index; BRCA1/2: Breast cancer gene 1/2; cDNA: Complementary DNA; Ct: Mean cycle above threshold; DABG-RMA: Detected above background-robust multichip average; FDR: False discovery rate; FFPE: Formalin fixed-paraffin embedded; H&E: Hematoxylin and eosin; HCUV: Valencian Clinic-University Hospital; HER2: Human epidermal growth factor receptor 2; HR: Hormone receptor; Hsa: Homo sapiens; IGF1R: Insulin-like growth factor receptor; MAPK: Mitogen-activated protein kinase; miR: microRNA; miRNA: microRNA; mRNA: messenger RNA; PCR: Polymerase chain reaction; PI3K: Phosphoinositol-3-kinase; qRT-PCR: Quantitative real time-PCR; RIN: RNA integrity number; RNA: Ribonucleic acid; RNU43: Ribonucleoprotein; scaRNA: Small cajal-body RNA; snoRNA: Small nucleolar RNA; Wnt: Wingless-type MMTV integration site family.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MPC processed the material and extracted RNA, prepare samples for Affymetrix array, performed qRT-PCR, acquired and analyzed all the data and performed statistical analyses, carried out databases and literature searching and contributed in the drafting of the manuscript; MTM performed the selection of patients suitable for the study, helped processing the samples, gave intellectual support, performed literature searching and contributed writing the manuscript; JAPF conceived the study, participated in the patients selection, contributed in the interpretation of the results and in the manuscript writing; LPC processed samples, optimized the used technique and acquired data, also provide useful information and intellectual content to the manuscript drafting; ET, PE and JC helped in the sample processing, interpretation of the results and revised final manuscript; BMD helped in the data analyses, interpretation of the results and in drafting the manuscript; EA, OB and JFL performed the pathological analyses, characterized the tumour tissue samples and performed immunohistochemical techniques, also provide the FFPE material and provide intellectual content and information for the manuscript drafting; AB has been involved in revising the manuscript critically for important intellectual content; AL performed the conceptual design of the study, supervision of the whole study, revised the manuscript and provide intellectual content, also participated in the acquisition of

funding; GR conceived and supervised the whole study, contributed in the interpretation of the data, gave intellectual support, contributed in the manuscript writing and helped in the acquisition of funding. All authors have given final approval of the version to be published.

#### Acknowledgements

MPC is funded by the Generalitat Valenciana VALI + d, AClF/2011/270. MTM is funded by "Rio Hortega Project" (CM12/00264). GR is a FIS "Miquel Servet" Researcher. AB holds a Translational Research Grant awarded by the Spanish Society of Medical Oncology (SEOM). This project was carried out thanks to Fundación LeCádiz – proyecto Flor de Vida and co-funded by FIS project P113/00606 and FEDER. We would like to give thanks to all the patients and volunteers for their participation and also to the INCLIVA Biobank, integrated into the Spanish Hospital Biobanks Network (ReTBioH) and supported by the Instituto de Salud Carlos III/FEDER (grant number: RD09/0076/00132). We also wish to thank several private Breast cancer associations that funded this study and the Unit for Multigenic Analysis from the Central Unit for Medical Research (UCIM/INCLIVA) for the performance of the Affymetrix microRNA profiles.

#### Author details

<sup>1</sup>Medical Oncology and Hematology Unit, INCLIVA Biomedical Research Institute, Av. Blasco Ibañez, 17, Valencia 46010, Spain. <sup>2</sup>Biobank, INCLIVA Biomedical Research Institute, Av. Blasco Ibañez, 17, Valencia, Spain. <sup>3</sup>Pathology Unit, INCLIVA Biomedical Research Institute, Av. Blasco Ibañez 17, Valencia, Spain. <sup>4</sup>Laboratory rare diseases, ISCIII, Sinesio Delgado, 4, Madrid, Spain. <sup>5</sup>Memorial Sloan-Kettering Cancer Center, 417 E. 68th St, New York, NY 10065, USA.

Received: 10 December 2013 Accepted: 15 July 2014

Published: 21 July 2014

#### References

1. Siegel R, Naishadham D, Jemal A: **Cancer statistics, 2013.** *CA Cancer J Clin* 2013, **63**(1):11–30.
2. Cardoso F, Lobl S, Paganì O, Graziottin A, Panizza P, Martinich L, Gentilini O, Peccatori F, Fourquet A, Delalage S, Marotti L, Penault-Llorca F, Kotti-Kitromilidou AM, Rodger A, Harbeck N: **The European Society of Breast Cancer Specialists' recommendations for the management of young women with breast cancer.** *Eur J Cancer* 2012, **48**(18):3355–3377.
3. Kothari AS, Fentiman IS: **Breast cancer in young women.** *Int J Clin Pract* 2002, **56**(3):184–187.
4. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, Perou CM: **Breast Carcinomas Arising at a Young Age: Unique Biology or a Surrogate for Aggressive Intrinsic Subtypes?** *J Clin Oncol* 2011, **29**(1):e18–e20.
5. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, Wang Y, Marcom PK, Marks JR, Febbo PG, Nevins JR, Potti A, Blackwell KL: **Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression.** *J Clin Oncol* 2008, **26**(20):3324–3330.
6. Colleani M, Anders CK: **Debate: the biology of breast cancer in young women is unique.** *Oncologist* 2013, **18**(4):e13–e15.
7. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JA, Zhang Y, Wang Y, Marcom K, Marks JR, Mukherjee S, Nevins JR, Blackwell KL, Potti A: **Age-specific differences in oncogenic pathway deregulation seen in human breast tumors.** *PLoS One* 2008, **3**(1):e1373.
8. Bartels CL, Tsongalis GJ: **MicroRNAs: novel biomarkers for human cancer.** *Clin Chem* 2009, **55**(4):623–631.
9. Chen K, Rajewsky N: **The evolution of gene regulation by transcription factors and microRNAs.** *Nat Rev Genet* 2007, **8**(2):93–103.
10. Kusenda B, Miraz M, Mayer J, Pospisilova S: **MicroRNA biogenesis, functionality and cancer relevance.** *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2006, **150**(2):205–215.
11. Hammond SM: **RNAi, microRNAs, and human disease.** *Cancer Chemother Pharmacol* 2007, **58**(Suppl 1):s63–s68.
12. Care A, Catalucci D, Felicetti F, Bonci D, Addario A, Gallo P, Bang ML, Segnalinì P, Gu Y, Dalton ND, Elia L, Latronico MV, Hoydal M, Autore C, Russo MA, Dorn GW 2nd, Ellingsen O, Ruiz-Lozano P, Peterson KL, Croce CM, Peschle C, Condorelli G: **MicroRNA-133 controls cardiac hypertrophy.** *Nat Med* 2007, **13**(5):613–618.
13. Beveridge NJ, Gardiner E, Carroll AP, Tooney PA, Cairns MJ: **Schizophrenia is associated with an increase in cortical microRNA biogenesis.** *Mol Psychiatry* 2009, **15**(12):1176–1189.
14. Hammond SM: **MicroRNAs as oncogenes.** *Curr Opin Genet Dev* 2006, **16**(1):4–9.
15. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM: **A microRNA polycistron as a potential human oncogene.** *Nature* 2005, **435**(7043):828–833.
16. Gunaratne PH: **Embryonic stem cell microRNAs: defining factors in induced pluripotent (iPS) and cancer (CSC) stem cells?** *Curr Stem Cell Res Ther* 2009, **4**(3):168–177.
17. Nimmo RA, Slack FJ: **An elegant miRror: microRNAs in stem cells, developmental timing and cancer.** *Chromosoma* 2009, **118**(4):405–418.
18. Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, Diehn M, Liu H, Panula SP, Chiao E, Dirbas FM, Somlo G, Pera RA, Lao K, Clarke MF: **Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells.** *Cell* 2009, **138**(3):592–603.
19. Gurtan AM, Ravi A, Rahl PB, Bosson AD, Jnabpatis CK, Bhutkar A, Whittaker CA, Young RA, Sharp PA: **Let-7 represses Nr61 and a mid-gestation developmental program in adult fibroblasts.** *Genes Dev* 2013, **27**(8):941–954.
20. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ: **RAS is regulated by the let-7 microRNA family.** *Cell* 2005, **120**(5):635–647.
21. Gurtan AM, Sharp PA: **The Role of miRNAs in Regulating Gene Expression Networks.** *J Mol Biol* 2013, **425**(19):3582–3600.
22. Di Leva G, Croce CM: **miRNA profiling of cancer.** *Curr Opin Genet Dev* 2013, **23**(1):3–11.
23. Jay C, Nemanitis J, Chen P, Fulgham P, Tong AW: **miRNA profiling for diagnosis and prognosis of human cancer.** *DNA Cell Biol* 2007, **26**(5):293–300.
24. Gravgard KH, Lyng MB, Laenkhölm AV, Sokilde R, Nielsen BS, Litman T, Ditzel HJ: **The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer.** *Breast Cancer Res Treat* 2012, **134**(1):207–217.
25. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ: **The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.** *Nat Cell Biol* 2008, **10**(5):593–601.
26. Iorio MV, Ferracin M, Liu CG, Veronesi A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Menard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin GA, Querzoli P, Negrini M, Croce CM: **MicroRNA gene expression deregulation in human breast cancer.** *Cancer Res* 2005, **65**(16):7065–7070.
27. Lee YM, Lee JY, Ho CC, Hong QS, Yu SL, Tzeng CR, Yang PC, Chen HW: **miRNA-34b as a tumor suppressor in estrogen-dependent growth of breast cancer cells.** *Breast Cancer Res* 2011, **13**(6):R116.
28. Sun X, Qin S, Fan C, Xu C, Du N, Ren H: **Let-7: a regulator of the ERalpha signaling pathway in human breast tumors and breast cancer stem cells.** *Oncol Rep* 2013, **29**(5):2079–2087.
29. Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang Q, Bos PD, Gerald WL, Massague J: **Endogenous human microRNAs that suppress breast cancer metastasis.** *Nature* 2008, **451**(7175):147–152.
30. Pastrello C, Polesel J, Della Puppa L, Viel A, Maestro R: **Association between hsa-mir-146a genotype and tumor age-of-onset in BRCA1/BRCA2-negative familial breast and ovarian cancer patients.** *Carcinogenesis* 2010, **31**(12):2124–2126.
31. Tanic M, Yanowsky K, Rodriguez-Antona C, Andres R, Marquez-Rodas I, Osorio A, Benitez J, Martinez-Delgado B: **Deregulated miRNAs in hereditary breast cancer revealed a role for miR-30c in regulating KRAS oncogene.** *PLoS One* 2012, **7**(6):e38847.
32. Azim HA Jr, Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, Halbe-Kains B, Piccart MJ, Sotiriou C, Loi S: **Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling.** *Clin Cancer Res* 2012, **18**(5):1341–1351.
33. Colak D, Nofal A, Albakheet A, Nirmal M, Jepre H, Eldali A, Al-Tweigeri T, Tulbah A, Ajarim D, Malik OA, Inan MS, Kaya N, Park BH, Bin Amer SM: **Age-specific gene expression signatures for breast tumors and cross-species conserved potential cancer progression markers in young women.** *PLoS One* 2013, **8**(5):e63204.

34. de Rinaldis E, Gazinska P, Mera A, Modrusan Z, Fedorowicz GM, Burford B, Gillett C, Marra P, Grigoriadis A, Dornan D, Holmberg L, Pinder S, Tutt A: **Integrated genomic analysis of triple-negative breast cancers reveals novel microRNAs associated with clinical and molecular phenotypes and sheds light on the pathways they control.** *BMC Genomics* 2013, **14**(1):643.
35. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR: **MicroRNA expression profiles classify human cancers.** *Nature* 2005, **435**(7043):834–838.
36. Heneghan HM, Miller N, Kelly R, Newell J, Kerin MJ: **Systemic miRNA-195 differentiates breast cancer from other malignancies and is a potential biomarker for detecting noninvasive and early stage disease.** *Oncologist* 2010, **15**(7):673–682.
37. Heneghan HM, Miller N, Kerin MJ: **Circulating miRNA signatures: promising prognostic tools for cancer.** *J Clin Oncol* 2010, **28**(29):e573–e574. author reply e575–576.
38. Enerly E, Steinfling I, Kleivi K, Leivonen SK, Aure MR, Russnes HG, Ronneberg JA, Johnsen H, Navon R, Rodland E, Makela R, Naume B, Perala M, Kallioniemi O, Kristensen VN, Yakhini Z, Borresen-Dale AL: **miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors.** *PLoS One* 2011, **6**(2):e16915.
39. Farazi TA, Horlings HM, Ten Hoeve JJ, Mihalovic A, Halfwerk H, Morozov P, Brown M, Hafner M, Reyaf F, van Kouwenhove M, Kreike B, Sie D, Hovestadt V, Wessels LF, van de Vijver MJ, Tuschl T: **MicroRNA sequence and expression analysis in breast tumors by deep sequencing.** *Cancer Res* 2011, **71**(13):4443–4453.
40. Rothe F, Ignatiadis M, Chaboteaux C, Halbe-Kains B, Kheddoumi N, Majaj S, Badran B, Fayyad-Kazan H, Desmedt C, Harris AL, Piccart M, Sotiriou C: **Global microRNA expression profiling identifies miR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer.** *PLoS One* 2011, **6**(6):e20980.
41. Bockmeyer CL, Christgen M, Muller M, Fischer S, Ahrens P, Langer F, Kreipe H, Lehmann U: **MicroRNA profiles of healthy basal and luminal mammary epithelial cells are distinct and reflected in different breast cancer subtypes.** *Breast Cancer Res Treat* 2011, **130**(3):735–745.
42. Cascione L, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, Alder H, He G, Vecchione A, Croce CM, Shapiro CL, Huebner K: **Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer.** *PLoS One* 2013, **8**(2):e55910.
43. Dvinge H, Git A, Graf S, Salmon-Divon M, Curtis C, Sottoriva A, Zhao Y, Hirst M, Arminen J, Miska EA, Chin SF, Provenzano E, Turashvili G, Green A, Ellis I, Aparicio S, Caldas C: **The shaping and functional consequences of the microRNA landscape in breast cancer.** *Nature* 2013, **497**(7449):378–382.
44. Majumder S, Jacob ST: **Emerging role of microRNAs in drug-resistant breast cancer.** *Gene Expr* 2012, **15**(3):141–151.
45. Liu A, Xu X: **MicroRNA isolation from formalin-fixed, paraffin-embedded tissues.** *Methods Mol Biol* 2011, **724**:259–267.
46. Hui AB, Shi W, Boutros PC, Miller N, Pintilie M, Fyles T, McCreedy D, Wong D, Gerster K, Waldron L, Jurisica I, Penn LZ, Liu FF: **Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues.** *Lab Invest* 2009, **89**(5):597–606.
47. Han W, Kang SY: **Relationship between age at diagnosis and outcome of premenopausal breast cancer: age less than 35 years is a reasonable cut-off for defining young age-onset breast cancer.** *Breast Cancer Res Treat* 2010, **119**(1):193–200.
48. Goldhirsch A, Gelber RD, Yothers G, Gray RJ, Green S, Bryant J, Gelber S, Castiglione-Gertsch M, Coates AS: **Adjuvant therapy for very young women with breast cancer: need for tailored treatments.** *J Natl Cancer Inst Monogr* 2001, **30**:44–51.
49. Yau C, Fedele V, Roydasgupta R, Fridlyand J, Hubbard A, Gray JW, Chew K, Dairkee SH, Moore DH, Schittulli F, Tommasi S, Paradiso A, Albertson DG, Benz CC: **Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers.** *Breast Cancer Res* 2007, **9**(5):R59.
50. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**(6797):747–752.
51. Awware NS, Martin TA, Haynes MD, Matthews PN, Jiang WG: **Claudin-11 decreases the invasiveness of bladder cancer cells.** *Oncol Rep* 2011, **25**(6):1503–1509.
52. Bao W, Fu HJ, Xie QS, Wang L, Zhang R, Guo ZY, Zhao J, Meng YL, Ren XL, Wang T, Li Q, Jin BQ, Yao LB, Wang RA, Fan DM, Chen SY, Jia LT, Yang AG: **HER2 interacts with CD44 to up-regulate CXCR4 via epigenetic silencing of microRNA-139 in gastric cancer cells.** *Gastroenterology* 2011, **141**(6):2076–2087. e2076.
53. Beauchemin D, Lacombe C, Van Themsche C: **PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype.** *Mol Cancer* 2011, **10**:148.
54. Gastano Z, Marsh T, Tadipatri R, Kuznetsov HS, Al-Shahrour F, Paktinat M, Greene-Colozzi A, Nilsson B, Richardson AL, McAllister SS: **Stromal EGF and IGF1 Together Modulate Plasticity of Disseminated Triple Negative Breast Tumors.** *Cancer Discov* 2013, **3**(8):922–935.
55. Fan Q, He M, Deng X, Wu WK, Zhao L, Tang J, Wen G, Sun X, Liu Y: **Derepression of c-Fos caused by MicroRNA-139 down-regulation contributes to the metastasis of human hepatocellular carcinoma.** *Cell Biochem Funct* 2012, **31**(4):319–324.
56. Guo H, Hu X, Ge S, Qian G, Zhang J: **Regulation of RAP1B by miR-139 suppresses human colorectal carcinoma cell proliferation.** *Int J Biochem Cell Biol* 2012, **44**(9):1465–1472.
57. Katsushima K, Shirjo K, Natsume A, Ohka F, Fujii M, Osada H, Sekido Y, Kondo Y: **Contribution of microRNA-1275 to Claudin11 protein suppression via a polycomb-mediated silencing mechanism in human glioma stem-like cells.** *J Biol Chem* 2012, **287**(33):27396–27406.
58. Kovacevic Z, Fu D, Richardson DR: **The iron-regulated metastasis suppressor, Ndrp-1: identification of novel molecular targets.** *Biochim Biophys Acta* 2008, **1783**(10):1981–1992.
59. Leidner RS, Li L, Thompson CL: **Dampening enthusiasm for circulating microRNA in breast cancer.** *PLoS One* 2013, **8**(3):e57841.
60. Lian ZQ, Wang Q, Li WP, Zhang AQ, Wu L: **Screening of significantly hypermethylated genes in breast cancer using microarray-based methylated-CpG island recovery assay and identification of their expression levels.** *Int J Oncol* 2012, **41**(2):629–638.
61. Milde-Langosch K, Kam T, Muller V, Witzel I, Rody A, Schmidt M, Wirtz RM: **Validity of the proliferation markers Ki67, TOP2A, and RacGAP1 in molecular subgroups of breast cancer.** *Breast Cancer Res Treat* 2013, **137**(1):57–67.
62. Ozata DM, Caramuta S, Velazquez-Fernandez D, Akcakaya P, Xie H, Hoog A, Zedenius J, Backdahl M, Larsson C, Lui WO: **The role of microRNA deregulation in the pathogenesis of adrenocortical carcinoma.** *Endocr Relat Cancer* 2011, **18**(6):643–655.
63. Qian YW, Chen Y, Yang W, Fu J, Gao J, Ren YB, Zhu JJ, Su B, Luo T, Zhao XF, Dai RY, Li JJ, Sun W, Wu MC, Feng GS, Wang HY: **p28(GANK) prevents degradation of Oct4 and promotes expansion of tumor-initiating cells in hepatocarcinogenesis.** *Gastroenterology* 2012, **142**(7):1547–1558. e1514.
64. Romanuk TL, Ueda T, Le N, Haile S, Yong TM, Thomson T, Vessella RL, Sadar MD: **Novel biomarkers for prostate cancer including noncoding transcripts.** *Am J Pathol* 2009, **175**(6):2264–2276.
65. Shen K, Liang Q, Xu K, Cui D, Jiang L, Yin P, Lu Y, Li Q, Liu J: **miR-139 inhibits invasion and metastasis of colorectal cancer by targeting the type I insulin-like growth factor receptor.** *Biochem Pharmacol* 2012, **84**(3):320–330.
66. Tajnik M, Strazisar M, Volavsek M, Bostjancic E, Glavac D: **BBC3 is down-regulated with increased tumor size independently of p53 expression in head and neck cancer.** *Cancer Biomark* 2012, **11**(5):197–208.
67. Webb PG, Spillman MA, Baumgartner HK: **Claudins play a role in normal and tumor cell motility.** *BMC Cell Biol* 2013, **14**:19.
68. Wong FY, Liem N, Xie C, Yan FL, Wong WC, Wang L, Yong WP: **Combination therapy with gossypol reveals synergism against gemcitabine resistance in cancer cells with high BCL-2 expression.** *PLoS One* 2012, **7**(12):e50786.
69. Yan B, Zhao JL: **miR-1228 prevents cellular apoptosis through targeting of MOAP1 protein.** *Apoptosis* 2012, **17**(7):717–724.
70. Leivonen SK, Makela R, Ostling P, Kohonen P, Haapa-Paananen S, Kleivi K, Enerly E, Aakula A, Hellstrom K, Sahlberg N, Kristensen VN, Borresen-Dale AL, Saviranta P, Perala M, Kallioniemi O: **Protein lysate microarray analysis to identify microRNAs regulating estrogen receptor signaling in breast cancer cell lines.** *Oncogene* 2009, **28**(44):3926–3936.
71. Queiros AM, Eschen C, Flegner D, Karargis G, Dworatzek E, Westphal C, Sanchez Ruderisch H, Regitz-Zagrosek V: **Sex- and estrogen-dependent regulation of a miRNA network in the healthy and hypertrophied heart.** *Int J Cardiol* 2013, **169**(5):331–338.
72. Fleming JM, Miller TC, Quinones M, Xiao Z, Xu X, Meyer MJ, Ginsburg E, Veenstra TD, Vonderhaar BK: **The normal breast microenvironment of**

- premenopausal women differentially influences the behavior of breast cancer cells in vitro and in vivo. *BMC Med* 2010, **8**:27.
73. Roman-Perez E, Casbas-Hernandez P, Pirone JR, Rein J, Carey LA, Lubet RA, Mani SA, Amos KD, Troester MA: **Gene expression in extratumoral micro-environment predicts clinical outcome in breast cancer patients.** *Breast Cancer Res* 2012, **14**(2):R51.
74. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, Wesolowska E, Boyd NF, Johnson NB, Figueroa JD, Sherman ME, Troester MA: **Relationship of mammographic density and gene expression: analysis of normal breast tissue surrounding breast cancer.** *Clin Cancer Res* 2013, **19**(18):4972–4982.
75. Anderson WF, Jatoi I, Sherman ME: **Qualitative age interactions in breast cancer studies: mind the gap.** *J Clin Oncol* 2009, **27**(32):5308–5311.

doi:10.1186/1471-2407-14-529

**Cite this article as:** Peña-Chilet et al.: MicroRNA profile in very young women with breast cancer. *BMC Cancer* 2014 **14**:529.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

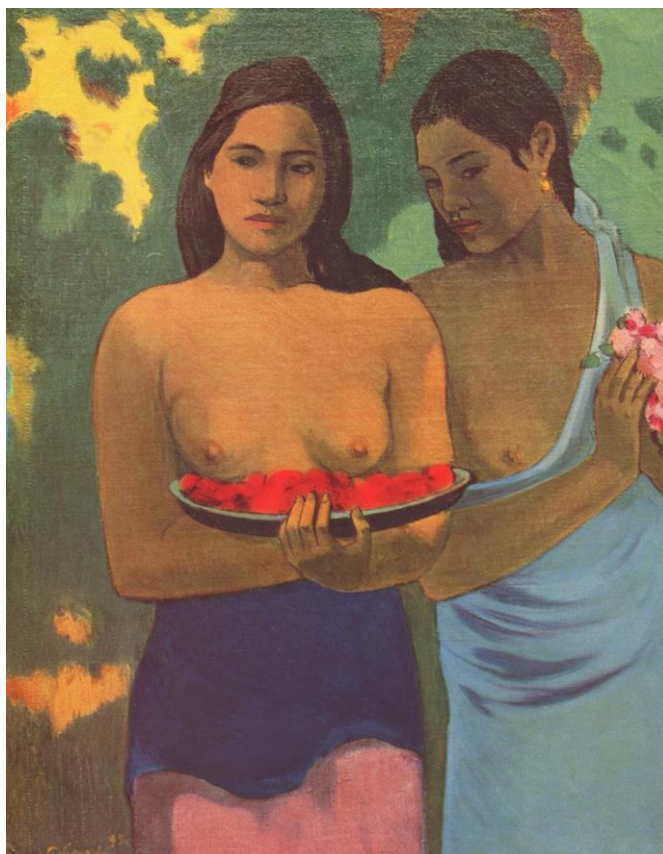
- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)









*Pechos con flores rojas / Dos tahitianas (con flores de mango)*. Paul Gauguin, 1899. Óleo sobre lienzo. Museo de Arte Metropolitano, Nueva York.