



Universidad Nacional Autónoma de México
PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS

BÚSQUEDA DE BLANCOS TERAPÉUTICOS ASOCIADOS A
FENOTIPOS DE CÁNCER DE MAMA

TESIS

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS BIOMÉDICAS

PRESENTA:

Q.F.B. Guillermo de Anda Jáuregui

TUTOR PRINCIPAL: Dr. Enrique Hernández Lemus, Instituto de
Ecología

COMITÉ TUTOR:

Dr. Alfredo Hidalgo Miranda, Facultad de Medicina
Dra. Mariana Benítez Keinrad, Instituto de Ecología

Ciudad de México. Julio, 2016



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS
GENÓMICA COMPUTACIONAL

BÚSQUEDA DE BLANCOS
TERAPÉUTICOS ASOCIADOS A
FENOTIPOS DE CÁNCER DE MAMA

T E S I S

QUE PARA OBTENER EL GRADO DE:
DOCTOR EN CIENCIAS BIOMÉDICAS

PRESENTA:

Q.F.B. GUILLERMO DE ANDA JÁUREGUI

DIRECTOR DE TESIS:

DR. ENRIQUE HERNÁNDEZ LEMUS



CIUDAD DE MÉXICO

JULIO, 2016

Agradecimientos

Agradezco al CONACYT por la beca de doctorado 324432 // 262974 asignada a mi CVU 420970.

Agradezco al Programa de Doctorado en Ciencias Biomédicas de la Universidad Nacional Autónoma de México por todas las facilidades para la realización de este proyecto.

Agradezco al Instituto Nacional de Medicina Genómica (INMEGEN), a la Dirección de Investigación, a la Dirección de Enseñanza y Divulgación y a la Subdirección de Genómica Poblacional.

Agradecimientos, dedicatorias...

Índice general

| | | |
|----------|---|-----------|
| 1 | Introducción: Planteamiento del problema | 1 |
| 2 | Marco teórico | 4 |
| 2.1 | Cáncer de mama | 4 |
| 2.2 | Heterogeneidad en el cáncer de mama | 6 |
| 2.2.1 | Clasificaciones | 6 |
| 2.2.2 | Tratamiento | 12 |
| 2.3 | Biología de sistemas | 13 |
| 2.4 | Redes biológicas | 15 |
| 2.4.1 | <i>Pathways</i> , vías o rutas biológicas | 17 |
| 2.4.2 | Entrecruzamiento de vías | 18 |
| 3 | Metodologías | 23 |
| 3.1 | Base de datos de microarreglos de expresión de cáncer de mama | 23 |
| 3.1.1 | Composición de la base de datos | 23 |
| 3.1.2 | Análisis de datos de microarreglos | 24 |
| 3.2 | Metodologías de clasificación en subtipos moleculares | 25 |
| 3.3 | Expresión diferencial y enriquecimiento de vías | 28 |
| 3.3.1 | Análisis de expresión diferencial | 28 |
| 3.3.2 | Análisis de enriquecimiento de vías | 29 |
| 3.4 | Construcción de redes | 30 |
| 3.4.1 | Redes transcripcionales | 30 |

| | | |
|-------|--|-----------|
| 3.4.2 | Redes de entrecruzamiento de vías | 32 |
| 3.4.3 | Análisis de redes | 34 |
| 3.5 | Exploración de blancos terapéuticos para cáncer de mama | 34 |
| 3.5.1 | Generación del espacio de blancos terapéuticos moleculares de cáncer de mama | 34 |
| 3.5.2 | Minería de bases de datos farmacológicas | 35 |
| 3.5.3 | Clasificación de los blancos farmacológicos | 36 |
| 3.6 | Flujo de trabajo para la exploración de vías biológicas asociadas a activi- dad farmacológica | 37 |
| | Resultados | 39 |
| 4 | Clasificación en subtipos moleculares | 41 |
| 5 | Expresión diferencial y enriquecimiento por subtipos | 44 |
| 5.1 | Análisis de expresión diferencial | 44 |
| 5.2 | Enriquecimiento de vías | 46 |
| 6 | Arquitectura de redes transcripcionales de subtipos moleculares de cáncer de mama | 50 |
| 6.1 | Redes transcripcionales de subtipos moleculares | 51 |
| 6.1.1 | Composición de las redes | 53 |
| 6.1.2 | Parámetros topológicos globales de las redes | 54 |
| 6.2 | Redes transcripcionales filtradas por anotación funcional | 57 |
| 6.2.1 | La estructura de las redes transcripcionales se preserva al filtrarse por anotación funcional | 57 |
| 6.2.2 | Parámetros topológicos de las redes filtradas | 60 |
| 7 | Redes de entrecruzamiento | 61 |

| | | |
|-----------|---|------------|
| 7.1 | Las redes de entrecruzamiento permiten integrar la información de una colección de vías biológicas | 61 |
| 7.2 | Las redes de entrecruzamiento reconstruyen las interacciones sistémicas descritas por diferentes colecciones de rutas biológicas | 63 |
| 7.3 | Las redes de entrecruzamiento muestran una fuerte intercomunicación entre procesos | 63 |
| 7.4 | Las distribuciones de grado y <i>betweenness centrality</i> de las redes de entrecruzamiento muestran diferencias entre las colecciones de vías | 65 |
| 7.5 | Consideraciones | 68 |
| 8 | Espacio de blancos terapéuticos moleculares para subtipos de cáncer de mama | 69 |
| 8.1 | Espacio de blancos terapéuticos moleculares con fármacos asociados a ellos | 70 |
| 8.1.1 | Blancos moleculares con la misma dirección de desregulación en todos los subtipos | 72 |
| 8.1.2 | Blancos moleculares desregulados únicamente en un subtipo molecular | 73 |
| 8.1.3 | Blancos moleculares para tres subtipos. | 75 |
| 8.1.4 | Blancos moleculares para dos subtipos | 76 |
| 8.2 | Espacio de blancos moleculares actualmente sin tratamientos asociados . | 79 |
| 9 | Caso de estudio: entrecruzamiento de la vía de estrógeno | 81 |
| 9.1 | Flujo de trabajo para el estudio de la desregulación en vías entrecruzadas con la vía de señalización de estrógeno | 84 |
| 9.2 | Desregulación de la vía de estrógeno | 85 |
| 9.2.1 | Red de entrecruzamiento de estrógeno | 90 |
| 9.2.2 | Desregulación de vías entrecruzadas | 92 |
| 10 | Conclusiones | 106 |

| | |
|--|------------|
| 11 Perspectivas | 118 |
| 11.1 Otras formas de construcción de redes transcripcionales | 118 |
| 11.1.1 Más información de redes transcripcionales | 119 |
| 11.2 Modelado de las redes de entrecruzamiento | 120 |
| 11.3 Otras estrategias de exploración del espacio de blancos terapéuticos | 121 |
| 11.4 Experimentos funcionales para cuantificar la influencia de los procesos entrecruzados sobre efectos farmacológicos | 121 |
| 11.5 Exploración del fenómeno de entrecruzamiento sobre otros blancos te- rapéuticos | 122 |
| 12 Consideraciones finales | 124 |
| A Análisis de datos de microarreglos | 144 |
| B Subtipificación molecular | 149 |
| C Expresión Diferencial | 152 |
| D Enriquecimiento de vías | 154 |
| E Redes de Entrecruzamiento | 157 |
| F Minería de DGIdb | 160 |

Índice de figuras

| | | |
|-----|--|----|
| 2-1 | <i>Hallmarks</i> del cáncer | 5 |
| 2-2 | Clasificación histológica del cáncer de mama | 9 |
| 2-3 | Blancos de acción de fármacos para cáncer de mama | 13 |
| 2-4 | Algunas características de los sistemas biológicos como fenómenos complejos | 14 |
| 2-5 | Diferentes espacios moleculares de fenómenos biológicos | 15 |
| 2-6 | Representación gráfica simplificada de un <i>pathway</i> | 18 |
| 2-7 | Entrecruzamiento de vías | 19 |
| 2-8 | Representación gráfica de resistencia a un efecto farmacológico por efecto del entrecruzamiento de vías | 21 |
| 3-1 | Flujo de trabajo para el estudio de desregulación de vías entrecruzadas. . | 38 |
| 4-1 | Gráfico de componentes principales de muestras clasificadas por CITBCMST | 42 |
| 4-2 | Gráfico de componentes principales de muestras clasificadas por PAM50 . | 42 |
| 4-3 | Gráfico tridimensional de componentes principales de muestras subtipificadas | 43 |
| 5-1 | Diagrama de Venn que muestra transcritos sobreexpresados en subtipos contra el fenotipo sano | 45 |
| 5-2 | Diagrama de Venn que muestra transcritos subexpresados en subtipos contra el fenotipo sano. | 45 |
| 5-3 | Enriquecimiento de vías de KEGG en GAGE. Las vías sobreexpresadas están en rojo. Las vías subexpresadas están en verde. Las vías en negro no están significativamente enriquecidas. La intensidad del color es proporcional a la significancia estadística (valor de q). | 47 |

| | | |
|-----|--|----|
| 5-4 | Enriquecimiento de vías de Reactome en GAGE. Las vías sobreexpresadas están en rojo. Las vías subexpresadas están en verde. Las vías en negro no están significativamente enriquecidas. La intensidad del color es proporcional a la significancia estadística (valor de q). | 49 |
| 6-1 | Redes transcripcionales para subtipos moleculares de cáncer de mama | 52 |
| 6-2 | Diagrama de Venn de los nodos que componen las redes de subtipos moleculares de cáncer de mama | 53 |
| 6-3 | Distribución de grado de las redes transcripcionales. Los paneles A-D muestran la distribución de grado, en escala log-log, de las redes para los subtipos luminal A, luminal B, basal y HER2-enriquecido. El panel E muestra la distribución de grado, en escala log-log, para la red de tejido mamario sano. | 56 |
| 6-4 | Redes transcripcionales para subtipos moleculares de cáncer de mama filtradas por anotación funcional | 58 |
| 6-5 | Diagrama de Venn de los nodos que componen las redes transcripcionales de subtipos moleculares de cáncer de mama, filtradas por factores transcripcionales y moléculas contenidas en KEGG | 59 |
| 7-1 | Redes de entrecruzamiento | 62 |
| 7-2 | Distribuciones de longitud de caminos más cortos | 65 |
| 7-3 | Distribuciones de grado | 66 |
| 7-4 | Distribuciones de <i>Betweenness Centrality</i> | 67 |
| 8-1 | Espacio de blancos terapéuticos moleculares | 71 |
| 8-2 | Blancos moleculares con la misma dirección de desregulación en todos los subtipos | 72 |
| 8-3 | Desregulados únicamente en un subtipo. Up = Sobreexpresados en el subtipo, Down = Subexpresados en el subtipo | 74 |
| 8-4 | Desregulados en tres subtipos, sin cambios en el otro | 75 |

| | | |
|------|---|----|
| 8-5 | Subexpresados en dos, sin cambios en los otros | 77 |
| 8-6 | Sobreexpresados en dos, sin cambios en los otros | 77 |
| 8-7 | Espacio de blancos farmacológicos sin compuestos con actividad farmacológica asociados. | 80 |
| 9-1 | 17 β -Estradiol, Principal hormona sexual femenina. | 82 |
| 9-2 | Vía de Señalización de Estrógeno. KEGG (Kanehisa <i>et al.</i> , 2014) | 83 |
| 9-3 | Identificación de las vías que entrecruzan con la vía de señalización de estrógeno. | 85 |
| 9-4 | Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo luminal A | 86 |
| 9-5 | Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo luminal B | 87 |
| 9-6 | Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo basal | 88 |
| 9-7 | Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo HER2-enriquecido | 89 |
| 9-8 | Identificación de las vías que entrecruzan con la vía de señalización de estrógeno. | 90 |
| 9-9 | Identificación de las vías que entrecruzan con la vía de señalización de estrógeno. | 91 |
| 9-10 | Patrones de desregulación de vías entrecruzadas con la vía de señalización de estrógeno. | 93 |
| 9-11 | Desregulación de expresión génica en la vía de apoptosis en el subtipo luminal A | 95 |
| 9-12 | Desregulación de expresión génica en la vía de apoptosis en el subtipo luminal B | 96 |
| 9-13 | Desregulación de expresión génica en la vía de apoptosis en el subtipo basal | 97 |

| | |
|--|-----|
| 9-14 Desregulación de expresión génica en la vía de apoptosis en el subtipo HER2-enriquecido | 98 |
| 9-15 Desregulación de expresión génica en la vía de señalización de calcio en el subtipo luminal A. | 100 |
| 9-16 Desregulación de expresión génica en la vía de señalización de calcio en el subtipo luminal B | 101 |
| 9-17 Desregulación de expresión génica en la vía de señalización de calcio en el subtipo basal | 102 |
| 9-18 Desregulación de expresión génica en la vía de señalización de calcio en el subtipo HER2-enriquecido | 103 |
| 9-19 Las vías de inmunidad que entrecruzan con estrógeno lo hacen a través de vías generales de señalización | 104 |

Resumen

El presente trabajo propone un marco conceptual para el estudio de la desregulación genómica en diversos subtipos moleculares que capturan la heterogeneidad del cáncer de mama, y las posibles implicaciones terapéuticas de dicha heterogeneidad en el contexto de biología de sistemas.

Se desarrollaron e implementaron estrategias para la subtipificación molecular por medio de microarreglos de expresión de cáncer de mama. Se implementaron metodologías de análisis de datos genómicos, tales como análisis de expresión diferencial, análisis de enriquecimiento de vías, e inferencia y análisis de redes transcripcionales. Se desarrollaron flujos de trabajo para la integración de estos datos, y la exploración de sus consecuencias en el contexto de blancos terapéuticos.

A partir de una base de datos de microarreglos de expresión (Affymetrix[®] HGU133A) con 819 muestras de tumores primarios de mama y 61 muestras de tejido mamario sano, se obtuvo un conjunto de muestras clasificadas en cuatro subtipos moleculares: luminal A, luminal B, basal y HER2-enriquecido.

Se identificaron más de 5800 genes diferencialmente expresados entre las muestras tumorales y el tejido mamario sano, con 3555 de estos desregulados en todos los subtipos moleculares. Se identificaron patrones de desregulación de vías específicos para cada subtipo molecular.

Se construyeron redes transcripcionales para cada subtipo molecular, las cuales mostraron notables diferencias en los transcritos que las componen y en la arquitectura general de las redes.

Se desarrolló la estrategia de *redes de entrecruzamiento de vías* que permite conectar las vías biológicas comprendidas en diferentes bases de datos. Se utilizó esta herramienta para explorar resistencias a tratamiento antiestrogénico debidas al entrecruzamiento en subtipos moleculares de cáncer de mama.

A partir de datos de expresión diferencial y el número de vías en las que participan, 1198 moléculas susceptibles de modulación farmacológica fueron ubicadas en un espacio de blancos terapéuticos.

A partir de estas exploraciones se encontraron diferencias notables entre el tejido mamario sano y el tejido canceroso, tanto en desregulación génica, desregulación de vías y arquitectura transcripcional. Se encontraron diferencias en estos mismos niveles entre los diferentes subtipos moleculares, aunque también se encontraron muchas características

comunes. En el presente trabajo se muestra la utilidad del uso de las diferentes herramientas desarrolladas e implementadas tanto para estudiar fenómenos como la resistencia de fármacos conocidos como para explorar posibles nuevas alternativas terapéuticas.

Capítulo 1

Introducción: Planteamiento del problema

El cáncer es una enfermedad compleja (Knox, 2010), en la que se engloban un conjunto amplio de enfermedades que involucran la desregulación de procesos biológicos que tiene como consecuencia una división descontrolada de las células.

En particular, el cáncer de mama es una enfermedad de importancia para la salud pública. Ocupa el primer lugar en número de casos de cáncer diagnosticados en mujeres a nivel mundial (Ferlay *et al.*, 2015) y es causa del 14% de muertes relacionadas con cáncer en México. Desde el año 2006, esta enfermedad es la primera causa de muerte en mujeres mayores a 25 años en nuestro país(Chávarri-Guerra *et al.*, 2012).

El cáncer de mama es una enfermedad muy heterogénea (Polyak, 2011). La alta diversidad biológica que existe entre diferentes tumores resulta en diferencias sustanciales en el riesgo de progresión de la enfermedad, el pronóstico y la respuesta a tratamientos. Esta diversidad ha llevado a la implementación de diferentes estrategias de clasificación (Viale, 2012).

Comprender la heterogeneidad en el cáncer de mama es fundamental para entender las diferencias en eficacia terapéutica (Rivenbark *et al.*, 2013). Existen algunos tratamientos farmacológicos que aprovechan características biológicas específicas de algunos tipos de cáncer de mama, pero en su mayoría, los fármacos aprobados para el tratamiento de cáncer de mama actúa a través de mecanismos quimiotóxicos, con fuertes efectos adversos, que de ninguna forma garantizan el éxito del tratamiento. Es por esto que la importancia de entender la biología del cáncer de mama, sus blancos farmacológicos y sus interacciones sigue siendo una línea de investigación de relevancia biomédica.

Una *enfermedad compleja* es aquella en la que el fenotipo patológico es el resultado de la contribución de muchos elementos. En el caso del cáncer, hay una contribución tanto de factores ambientales como de factores biológicos de tipo heredable. Al abordar el estudio de la enfermedad nos enfrentamos a alteraciones biológicas en varios niveles como el genómico, transcriptómico, metabólico. Retomando la heterogeneidad del cáncer de mama, al tratar de estudiarlo nos enfrentamos en realidad a un conjunto de condiciones patológicas afectadas por muchos factores; el problema claramente no es trivial.

El avance de las tecnologías *ómicas* y de la información permite tener acceso a los elementos necesarios para enfrentar el problema. Existen técnicas de alta resolución que generan mediciones a gran escala de manera automatizada, con una velocidad y precisión inigualable con técnicas anteriores; bases de datos que compendian el conocimiento biológico y médico de la humanidad de manera pública. Herramientas matemáticas y los sistemas informáticos con poder de cómputo suficiente para implementarlos. El reto es tener la capacidad de *integrar* estas herramientas. La *Biología de Sistemas* es una disciplina emergente que plantea el modelado de los sistemas biológicos desde una perspectiva interdisciplinaria y no reduccionista.

El **objetivo** del presente trabajo es establecer un *marco teórico* para la *integración* de datos genómicos en el contexto de vías biológicas, tomando en cuenta las interconexiones entre estas, para la *generación de hipótesis* de posible relevancia farmacológica para la

diversidad de fenotipos cancerosos de mama. Para ello, se implementaron metodologías de análisis de datos genómicos de cáncer de mama: subtipificación molecular, análisis de expresión diferencial, análisis de enriquecimiento de vías, inferencia y análisis de redes transcripcionales.

Capítulo 2

Marco teórico

2.1. Cáncer de mama

El cáncer de mama es cáncer que ocurre en el tejido del seno; particularmente, aquel derivado del desarrollo desregulado de células epiteliales en el tejido mamario.

En general, las distintas enfermedades identificadas como cáncer involucran la rápida producción de células anormales (en morfología y función), lo que conlleva a un desarrollo anormal de tejido, con un crecimiento descontrolado y agresivo; en muchos casos, con la capacidad de invadir sitios distantes dentro del organismo.

Los procesos asociados a estas condiciones patológicas son complejos desde el punto de vista fisiológico y molecular. Hanahan y Weinberg (2000) describieron un conjunto de características comunes al cáncer, los llamados *hallmarks* del cáncer. Un trabajo posterior (Hanahan y Weinberg, 2011) elaboró en cuáles son estas características. La figura 2-1, adaptada del antes citado artículo, muestra estas características.



Figura 2-1: *Hallmarks* del cáncer.

El cáncer de mama es un problema de salud pública mundial. (IARC, 2013). Los datos epidemiológicos más recientes indican que cerca del 12% de los casos de cáncer diagnosticados se tratan de cáncer de mama, siendo este el cáncer más diagnosticado en mujeres (Ferlay *et al.*, 2015).

Las cifras en México indican un aumento en el riesgo de cáncer de mama, llegando a un 5% en el año 2010. El cáncer de mama se asocia a un 14% de las muertes relacionadas con cáncer en México (Chávarri-Guerra *et al.*, 2012). Datos del Instituto Nacional de

Estadística, Geografía e Informática (INEGI) indican que se trata de la primera causa de muerte en mujeres de entre 35 y 44 años de edad. El cáncer de mama tiene un fuerte impacto en muchos aspectos: además del daño físico, representa un gasto económico, tanto para los pacientes como para los sistemas de salud (Campbell y Ramsey, 2009; Knaul *et al.*, 2009; Andersen *et al.*, 2015) y conlleva un fuerte impacto emocional (Saeedi-Saedi *et al.*, 2015).

2.2. Heterogeneidad en el cáncer de mama

Una de las características más relevantes del cáncer de mama es su alta heterogeneidad (Polyak, 2011). Esta heterogeneidad se refleja en implicaciones clínicas (Bedard *et al.*, 2013): diversidad en el pronóstico, riesgo de recurrencia, progresión de la enfermedad y sensibilidad al tratamiento farmacológico (Zardavas *et al.*, 2015). El impacto de la heterogeneidad de la enfermedad hace necesaria la implementación de *clasificaciones* para su estudio y la elección de las mejores opciones de tratamiento para cada paciente, dependiendo de la clasificación que haya recibido cada caso.

2.2.1. Clasificaciones

Existen diferentes estrategias de clasificación para el cáncer de mama. Dependiendo del criterio empleado para la clasificación, estas pueden dividirse en:

- Escalas: grado, etapa
- Clasificación histopatológica
- Estatus de expresión de receptores
- Clasificación molecular

A continuación describiremos brevemente las clasificaciones, haciendo énfasis en la clasificación molecular, ya que es en esta en la que se centra el presente trabajo.

Escalas: grado, estadíos

Diferentes escalas categorizan los tumores de mama dependiendo de características clínicas observables.

La *escala de Nottingham* o *Elston-Ellis* (Elston *et al.*, 1991) cuantifica la diferenciación de las células del tumor a partir de la observación de características en el microscopio. Cada parámetro es evaluado en una escala de 1 a 3 puntos.

- Formación de túbulos: Mide qué porcentaje del tumor forma estructuras ductales normales. 1: Formación de túbulos mayor a 75 % del tumor. 2: Formación de túbulos entre 10 % y 75 % del tumor. 3: Formación de túbulos en menos del 10 % del tumor.
- Pleomorfismo nuclear: Hace referencia a la variabilidad que existe en la forma y tamaño de los núcleos. 1: Variación mínima. 2: Variación moderada. 3: Variación alta.
- Cuenta mitótica: Cuantifica el número de células mitóticas. 1: 0 a 9 cuentas en 10 campos, a una amplificación de 25X. 2: 10 a 19 cuentas en 10 campos, a una amplificación de 25X. 3: 20 o más cuentas en 10 campos, a una amplificación de 25X.

A partir de estos puntajes, se define un valor de *grado del tumor*:

- Grado I: 3 a 5 puntos. Bien diferenciado, mejor pronóstico.
- Grado II: 6 o 7 puntos. Moderadamente diferenciado. Pronóstico medio.
- Grado III: 8 o 9 puntos. Pobremente diferenciado. Peor pronóstico.

La clasificación de los tumores en *estadíos* cuantifica la extensión de la propagación del cáncer. El *sistema TNM*¹ (Egner, 2010) es un sistema que contempla tres parámetros principales:

¹La reproducción íntegra de la clasificación está limitada por derechos de autor

- Tamaño de la masa tumoral.
- Ubicación anatómica de nodos linfáticos infiltrados.
- Presencia de metástasis.

A partir de estos parámetros, se puede determinar el estadio de la progresión del tumor en:

- Estadio 0
- Estadio 1
- Estadio 2
- Estadio 3
- Estadio 4

Clasificación histológica

La clasificación histopatológica se basa en características estructurales y de patrones de crecimiento observables a través de microscopía en muestras de biopsia de tumores. Esta clasificación es extensivamente descriptiva de las características observables del tejido (Tavassoli y Devilee, 2003). Primordialmente se separan los tumores en aquellos que proliferan en el tejido epitelial sin invadir tejido circundante (carcinoma *in situ*) y aquellos que lo hacen (carcinoma invasivo).

En la figura 2-2, tomada de Malhotra *et al.* (2010) podemos observar dicha clasificación (la imagen también refiere al grado previamente descrito).

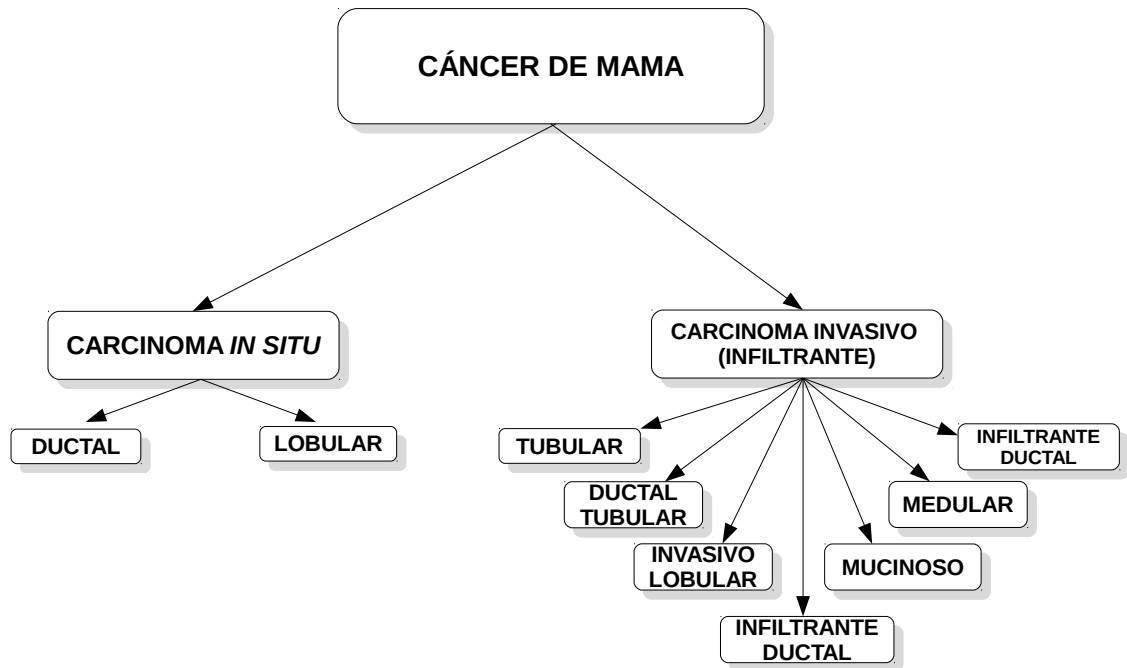


Figura 2-2: Clasificación histológica del cáncer de mama, adaptada de Malhotra *et al.* (2010).

Estatus de receptores

El *estatus de receptores*, cuantificables mediante técnicas inmunohistoquímicas, permite una clasificación de relevancia clínica.

Cuando se habla de estatus de receptores, se está haciendo referencia a las siguientes moléculas:

- *Receptores de estrógeno (ER)*: Receptores nucleares, sensibles a señales de hormonas estrogénicas, que actúan como factores transcripcionales relacionados a procesos de división celular. En el capítulo *Caso de estudio: entrecruzamiento de la vía de estrógeno* se aborda un poco más este receptor.

- *Receptores de progesterona (PR)*: Receptores nucleares, sensibles a señales de progesterona, también actúan como factores transcripcionales que llevan a la expresión de dimorfismo sexual.
- *Receptor HER2*: También conocido como *ERBB2*, es un receptor membranal sensible a señales del factor de crecimiento epidermal, asociado a procesos de proliferación y antiapoptóticos.

El estatus de estos receptores permite hacer consideraciones clínicas (Dunnwald *et al.*, 2007; Parise y Caggiano, 2014), teniéndose un mejor pronóstico para aquellos tumores que expresan receptores, y un mal pronóstico para los llamados *triples negativos*.

Uno de los factores que llevan a estas diferencias en pronóstico es que el estatus de receptores se relaciona con las opciones terapéuticas disponibles (Zaha, 2014). Más adelante se profundizará en esto.

Subtipos moleculares

Los *subtipos moleculares intrínsecos* son clasificaciones basadas en niveles de biomarcadores asociados a características fenotípicas determinadas. A diferencia de las estrategias antes descritas, en las que la clasificación se hace a partir de características clínico-patológicas, los subtipos moleculares son producto de un paradigma genómico (Kittaneh *et al.*, 2013).

La primera propuesta de subtipos moleculares de aceptación general son los subtipos descritos por Perou *et al.* (2000). En este y otros trabajos (Sørli *et al.*, 2001, 2003) se identificaron subtipos a través de estrategias de agrupamiento no supervisado de muestras a partir de niveles de expresión génica, a partir de las cuales se pudo implementar un algoritmo clasificador basado en niveles de expresión en microarreglos (*Prediction Analysis of Microarrays*), PAM50 (Parker *et al.*, 2009), basado en una *firma molecular* de 50 genes. Cuatro son los subtipos moleculares más ampliamente descritos y asociados a características biológicas particulares. Estos subtipos son: luminal A, luminal B, basal, y

HER2-enriquecido (HER2+). Existe controversia en la interpretación de algunos subtipos moleculares identificados, como el *normal-like* (Badve *et al.*, 2011); estudios posteriores han identificado nuevos subtipos moleculares (Prat y Perou, 2011). El presente trabajo se enfoca en los subtipos luminal A, luminal B, basal, y HER2-enriquecido (HER2+).

Luminal A Este subtipo tiene un perfil de expresión similar al del tejido epitelial mamario luminal. Se trata del subtipo más común, ya que cerca del 40 % de los cánceres de mama corresponden a este subtipo. Característicamente, este subtipo presenta sobreexpresión del receptor de estrógenos y genes regulados por este. Regularmente no presenta sobreexpresión del receptor HER2, ni de genes relacionados con la proliferación. Se trata del subtipo con el mejor pronóstico (Hu *et al.*, 2006).

Luminal B Este subtipo se caracteriza por una mayor variabilidad en la expresión de receptores de estrógeno y genes relacionados al receptor HER2 que el subtipo luminal A, y una mayor expresión de genes proliferativos. Se han identificado mutaciones en *TP53* e inestabilidad genética asociada a este subtipo, así como un mayor riesgo de recaídas. Cerca del 20 % de los tumores de mama pertenecen a este subtipo (Sørlie *et al.*, 2001).

HER2-enriquecido (HER2+) Este subtipo está definido por la sobreexpresión del receptor ERBB2, generalmente asociada a una amplificación a nivel cromosómico (Burstein, 2005). Por lo general, estos tumores suelen ser negativos a receptores de estrógeno y progesterona, y tienen un pronóstico menos alentador que los subtipos luminales.

Basal El subtipo basal tiene patrones de expresión similares a los del epitelio mamario basal. Este subtipo presenta subexpresión de los receptores a estrógeno, progesterona y HER2; la mayor parte de los tumores triple negativos pertenecen a este subtipo. A este subtipo se asocia inestabilidad genómica, alta expresión de genes de proliferación, y grados histológicos más altos (Bayraktar y Glück, 2013).

2.2.2. Tratamiento

El tratamiento del cáncer de mama consta de tres componentes principales: tratamiento quirúrgico, radioterapia y tratamiento farmacológico, ya sea en un contexto neoadyuvante (previo a cirugía) o adyuvante (posterior a la cirugía). (Matsen y Neumayer, 2013).

Como se mencionó anteriormente, las opciones de tratamiento farmacológico están íntimamente ligadas al estatus de receptores (y por consiguiente, a los subtipos moleculares) (Zaha, 2014). Existen tres categorías principales de fármacos actualmente disponibles para el tratamiento de cáncer de mama:

- *Tratamiento antihormonal*: Este tratamiento bloquea la transducción de señales hormonales de crecimiento celular. Tradicionalmente, se recomienda para aquellos tumores en los que se sobreexpresan receptores hormonales. La categoría contiene fármacos como los inhibidores competitivos y agonistas parciales de receptor de estrógeno (tamoxifen) e inhibidores de aromatasa (anastrozol), que inhiben la producción de hormonas esteroides.
- *Terapia de anticuerpos*: Esta terapia consiste en el uso de anticuerpos monoclonales dirigidos a moléculas específicas de las células tumorales. En este caso, el tratamiento basado en esta estrategia más difundido es el uso de *trastuzumab* contra el receptor HER2, para aquellos tumores que presentan sobreexpresión de dicho receptor.
- *Quimioterapia citotóxica*: Dentro de esta categoría se encuentran todas las terapias que emplean algún mecanismo de daño a las células en replicación, como la alquilación del ADN (ciclofosfamida), la intercalación del ADN (antraciclinas), inhibición de la síntesis de ADN (metrotexato), o la estabilización de microtúbulos (taxanos). Las terapias citotóxicas son las que presentan la mayor cantidad de efectos adversos, siendo las más inespecíficas. Estas terapias no dependen del estatus de receptores,

por lo que se tratan de las únicas opciones actuales para tumores evaluados como triple negativos.

En la figura 2-3, adaptada de Baselga (2011), podemos ver una representación abstracta de los sitios de acción de fármacos para cáncer de mama.

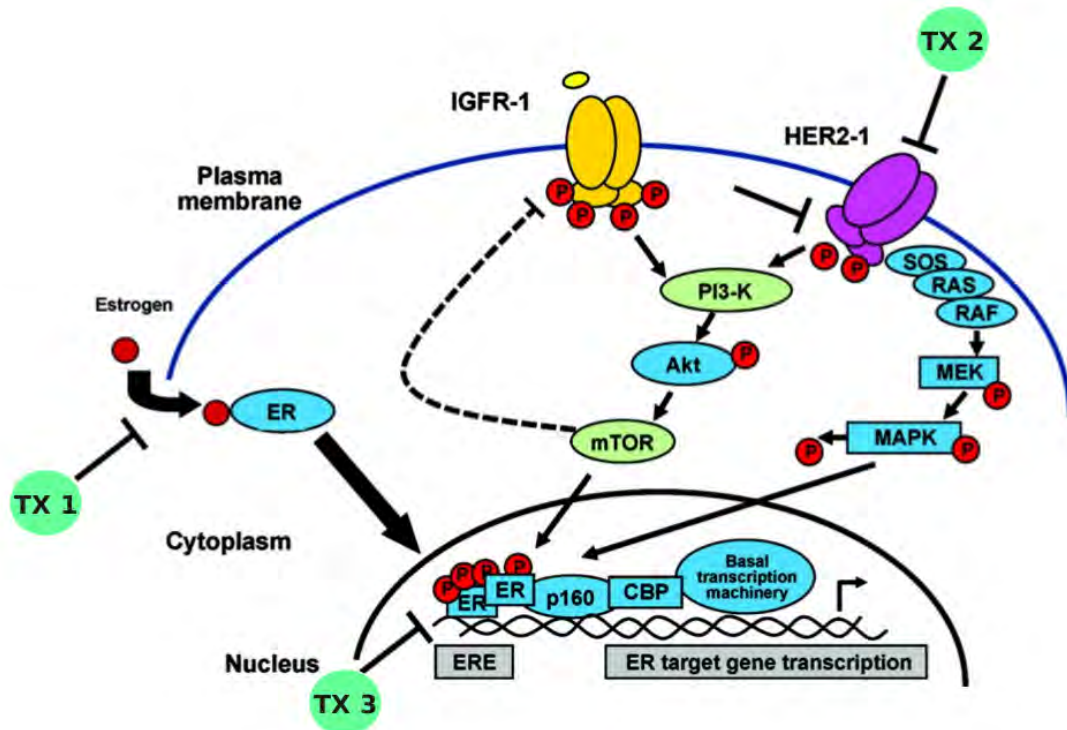


Figura 2-3: Blancos de acción de fármacos para cáncer de mama. Cuando se presenta una sobreexpresión de receptores hormonales, se puede optar por el bloqueo hormonal (TX1). Cuando se sobreexpresa un receptor membranal, como HER2, es posible utilizar anticuerpos contra dicho receptor (TX2). De no existir estos blancos, es necesario utilizar terapia citotóxica, que ataque las células durante su replicación (TX3).

Como se puede apreciar, las opciones de tratamiento farmacológico actuales para cáncer de mama son limitadas.

2.3. Biología de sistemas

La *biología de sistemas* es una disciplina de reciente creación que estudia las interacciones complejas dentro de los elementos que componen los sistemas biológicos. La biología

de sistemas busca conjuntar los datos obtenidos con tecnologías masivas (*ómicas*), el extenso cuerpo de información documental disponible a través de tecnologías de la información, y las herramientas físico-matemáticas necesarias para su análisis (Hernández-Lemus *et al.*, 2015).

Vistos desde esta óptica, los sistemas biológicos se comportan como fenómenos complejos. En la figura 2-4, tomada de Hernández-Lemus (2014), podemos observar algunas de las características de estos sistemas.

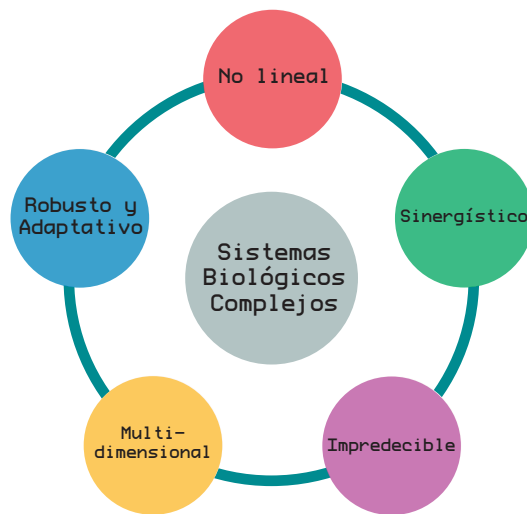


Figura 2-4: Algunas características de los sistemas biológicos como fenómenos complejos

Estas características pueden ser observadas en el cáncer, y en particular en el cáncer de mama. En la enfermedad están involucrados desregulación hormonal, desregulación metabólica, inestabilidad genómica, desregulación de procesos de inflamación y respuesta inmune, múltiples procesos de señalización, entre otros. La figura 2-5, tomada de Brazhnik *et al.* (2002) ilustra cómo estos procesos participan en diferentes espacios moleculares.

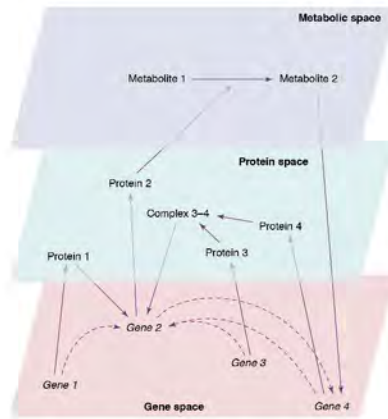


Figura 2-5: Diferentes espacios moleculares de fenómenos biológicos

Importancia de la biología de sistemas para el estudio del cáncer

El estudio del cáncer desde la perspectiva de biología de sistemas es un área de mucho interés en la actualidad. Como se mencionó anteriormente, en el cáncer intervienen muchos factores diferentes. En el trabajo fundamental de Hanahan y Weinberg (2000) se menciona que el cáncer es una *enfermedad de vías*, por lo que tratar de ubicar dichos factores dentro de su contexto biológico es deseable.

La subtipificación molecular es un ejemplo de éxito de la integración de tecnologías *ómicas* empleando herramientas matemáticas, con relevancia clínica. Con esto en mente, se puede pensar en la implementación de otras metodologías computacionales que contemplen las relaciones entre los factores moleculares involucrados en cáncer para identificar nuevas líneas de acción, con beneficios terapéuticos.

2.4. Redes biológicas

Una herramienta muy útil para describir los sistemas biológicos de manera integrativa es el uso de *teoría de grafos*. De esta manera, es posible modelar los fenómenos biológicos como *redes*. Formulamos la siguiente definición de red biológica (de Anda-Jáuregui *et al.*, 2015):

Definición. Una **red biológica** se define formalmente como un grafo $\mathcal{G}(V, E)$ sobre un dúplex formado por dos conjuntos, un conjunto V de nodos o vértices ($v_i \in V$) conformado por **biomoléculas**, y un conjunto E de aristas o enlaces que conectan dichos vértices ($e_i \in E$) representando interacciones físicas o químicas entre dichas biomoléculas. La regla de conectividad puede ser representada por una matriz de adyacencia $\mathcal{A} = A_{i,j}$, donde $A_{i,j} \neq 0$ implica una interacción no nula entre las biomoléculas v_i y v_j .

Se puede intuir de esta definición que existen muchos tipos de redes biológicas distintas. Algunas de estas (Pavlopoulos *et al.*, 2011) son:

- **Redes de interacción proteína - proteína:** Compuestas de nodos que representan proteínas, y aristas que representan las interacciones fisicoquímicas entre estas que conllevan a procesos biológicos celulares.
- **Redes de señalización:** Estas redes integran tanto proteínas como otras especies químicas involucradas en la transducción de señales hacia, dentro y desde las células.
- **Redes bioquímicas - metabólicas:** En estas redes los nodos representan reactivos y productos de reacciones catalíticas representadas por los enlaces de la red.
- **Redes transcripcionales:** En estas los nodos representan genes, y los enlaces representan una relación de coexpresión entre ellos. Un caso particular de estas son las *redes de regulación génica*, en las cuales se captura la información de control en la expresión.

2.4.1. *Pathways*, vías o rutas biológicas

De manera general, se conoce como *pathway*, *vía biológica* o *ruta biológica*² a un conjunto de moléculas biológicas cuyas interacciones dan lugar a un proceso biológico. Utilizando la anterior definición de red biológica, se propone la siguiente definición:

Definición. *Dada una red biológica, pueden definirse trayectorias sobre dicha red – secuencias de interacciones concatenadas $\mathcal{P}_{m \rightarrow n} = v_m \rightarrow v_{m+1} \rightarrow v_{m+2} \cdots \rightarrow v_{n-1} \rightarrow v_n$ conectando un conjunto de moléculas $v_m, v_{m+1}, v_{m+2}, \dots, v_{n-1}, v_n$ – que representan procesos biológicos definidos. Defínase como **pathway, vía biológica** o **ruta biológica** al subgrafo de la red biológica compuesto por dichas trayectorias asociadas a un proceso biológico de interés.*

Las rutas biológicas tienen sentido como herramientas conceptuales para describir los fenómenos biológicos. De esta manera, diferentes esfuerzos han dado lugar a diferentes *colecciones de rutas biológicas*, bases de datos que contienen descripciones de rutas basadas en el conocimiento biológico actual. Algunas de estas bases de datos son la *Encyclopedia of Genes and Genomes* (Kanehisa *et al.*, 2014), *Biocarta* (Nishimura, 2001), *Pathway Interaction Database* (Schaefer *et al.*, 2009), y *Reactome* (Joshi-Tope *et al.*, 2005). Pensar en rutas biológicas permite tener un nivel conceptual intermedio para pasar de las observaciones funcionales al plano molecular y de regreso, localizando dichas moléculas en un contexto funcional.

²El término *pathway* se traduce al español como *vía* o *ruta*. Dependiendo del contexto y área de estudio, se prefiere una traducción u otra: por ejemplo, *ruta metabólica* o *vía de señalización*. En el presente trabajo se ha decidido, para poder aplicarse de manera general, utilizar ambas traducciones de manera sinónima.

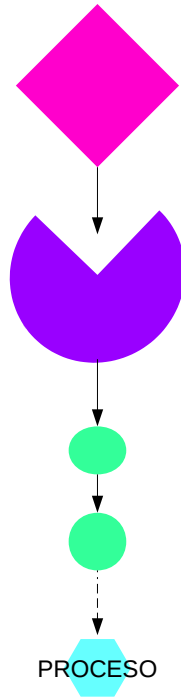


Figura 2-6: Representación gráfica simplificada de un *pathway*. Una molécula inicial (rombo rosa) interactúa con otra molécula receptora (semicírculo morado). Esta molécula inicia una secuencia de interacciones con moléculas subsecuentes (verdes) hasta llegar a una molécula que efectúa algún proceso biológico (hexágono azul).

2.4.2. Entrecruzamiento de vías

Al pensar en las vías biológicas como subgrafos (es decir, subconjuntos de nodos y enlaces) de una red biológica global, es evidente que existe la posibilidad de que dos vías biológicas compartan moléculas. Llamamos a este fenómeno *entrecruzamiento de vías* o *crosstalk*.

Definición. Dada una red biológica con dos vías \mathcal{P}_A y \mathcal{P}_B , éstas **entrecruzan** si existe un conjunto $v^{A,B}$ (que contiene al menos una molécula $v_i^{A,B}$) de tal forma que $\forall v_i^{A,B} \in v^{A,B}$ se sostiene que $v_i^{A,B} \in \mathcal{P}_A$ y $v_i^{A,B} \in \mathcal{P}_B$. Si A es el conjunto completo de moléculas en \mathcal{P}_A y B es el conjunto completo de moléculas en \mathcal{P}_B entonces

$v^{A,B} = A \cap B$. Cada molécula en $v_i^{A,B} \in v^{A,B}$ participa en un **evento de entrecruzamiento** entre las vías \mathcal{P}_A y \mathcal{P}_B .

En la figura 2-7 vemos una representación elemental de un fenómeno de entrecruzamiento: dos vías, ABC (Azul) y DBE (dorada), entrecruzan a través de la molécula B.

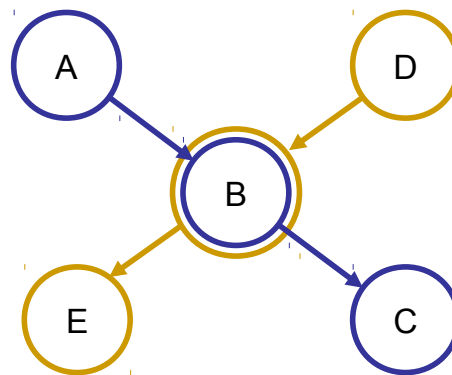
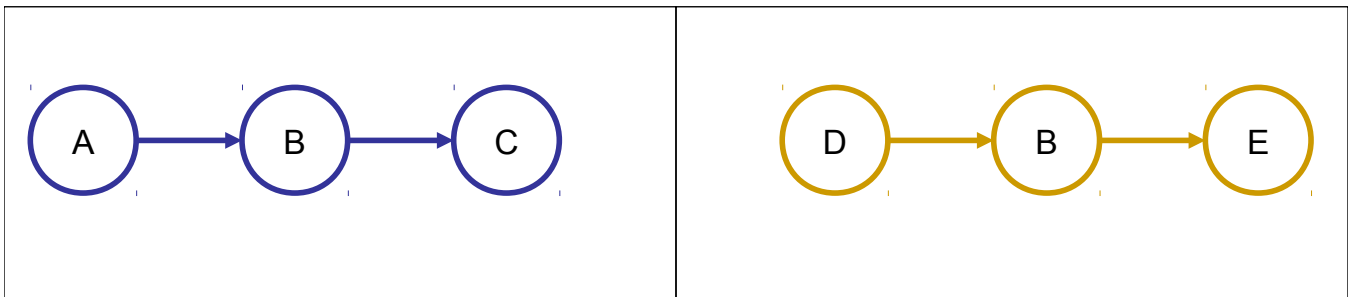


Figura 2-7: Entrecruzamiento de vías

Los términos *entrecruzamiento* o *crossstalk* son utilizados en otros contextos biológicos, particularmente para expresar comunicación entre entidades en diferentes niveles de

descripción (Taylor *et al.*, 2004). En el presente trabajo se estudiará el fenómeno descrito en la definición antes propuesta, predominantemente en el contexto de interconexión de vías asociadas a procesos biológicos. Se utilizarán de manera indistinta los términos en español y en inglés.

El fenómeno de entrecruzamiento de vías es de mucho interés para el estudio de la patología desde el punto de vista de biología de sistemas (Kreeger y Lauffenburger, 2010; Alberghina *et al.*, 2014). Retomando la idea del cáncer como una enfermedad de vías, el concepto de entrecruzamiento permite identificar la *comunicación* entre vías. Se ha observado que el fenómeno de entrecruzamiento de vías está implicado en procesos patológicos, como el cáncer de mama (Wang *et al.*, 2016). En particular, el fenómeno de entrecruzamiento se ha asociado a posibles mecanismos de desarrollo de resistencias farmacológicas (Viedma-Rodríguez *et al.*, 2014; Azad *et al.*, 2015).

Resistencia farmacológica por efecto de entrecruzamiento de vías

El entrecruzamiento de vías puede afectar los efectos de la modulación (farmacológica) de ciertas vías. La figura 2-8 considera un *pathway* asociado a un proceso patológico (piénsese, por ejemplo, en proliferación acelerada por señalización de estrógeno). Un fármaco F (por ejemplo, tamoxifen) puede bloquear la vía y por ende el proceso. Pero un *pathway* (moléculas naranjas) puede entrecruzar a través de una molécula (X) y continuar la señalización río abajo, saltando el bloqueo farmacológico. Llamemos a este fenómeno *resistencia farmacológica por efecto de entrecruzamiento de vías*.

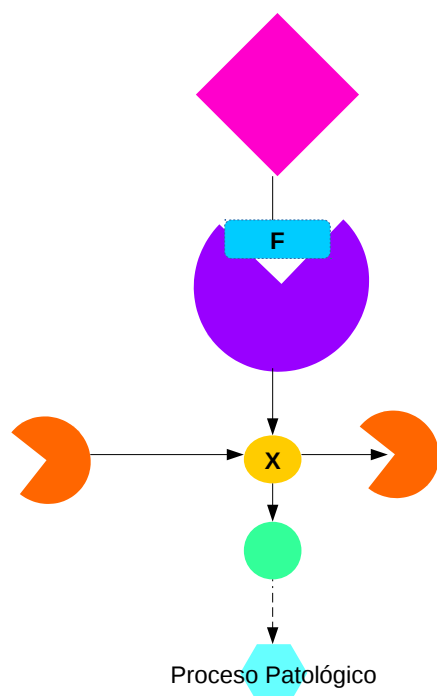


Figura 2-8: Representación gráfica de resistencia a un efecto farmacológico por efecto del entrecruzamiento de vías. Una vía que lleva a un proceso patológico es un blanco terapéutico; para modularla, se interviene farmacológicamente el receptor en la vía (semicírculo morado) interfiriendo con la molécula señalizadora (rombo rosa); sin embargo, la vía de interés *entrecruza*, a través de la molécula X, con otra vía (semicírculos naranjas); el resultado es que, a pesar de la intervención farmacológica, el proceso patológico puede continuar por la señalización de la segunda vía.

Existen muchos ejemplos de vías biológicas entrecruzadas que se asocian al desarrollo de resistencias a tratamientos farmacológicos (Baker *et al.*, 2014b; Cui *et al.*, 2012; Fiorio *et al.*, 2008). Estrategias de diseño racional de fármacos modernas consideran los efectos que el fenómeno de entrecruzamiento puede tener sobre la eficacia terapéutica (Nahta, 2012). De la misma forma, considerar el entrecruzamiento de vías es importante si se piensa en estrategias de *reposicionamiento de fármacos*, es decir, el uso de fármacos conocidos para el tratamiento de otras enfermedades (Azmi y Mohammad, 2014), o la *polifarmacia* (Jia *et al.*, 2009).

El entrecruzamiento de vías como una herramienta para la búsqueda de blancos terapéuticos para cáncer de mama

El entrecruzamiento de vías puede ser útil como una medida muy directa y evidente de la influencia sistémica de una molécula: una molécula que participa en muchas vías biológicas tiene una probabilidad mayor de tener más efectos en el sistema biológico. Si dicha molécula se perturba, podríamos esperar una mayor perturbación del sistema global.

Dos implicaciones de importancia para el estudio farmacológico se derivan de lo anterior:

- Por un lado, la modulación farmacológica de una molécula de mucha influencia puede afectar varios blancos terapéuticos río abajo simultáneamente, modulando varios *pathways* involucrados con la enfermedad.
- Por otro lado, la modulación farmacológica de una molécula de mucha influencia tiene asociada una probabilidad de *eventos adversos* mayor.

Como ya se mencionó también, al pensar en el sistema biológico como una red conformada por distintos caminos, los puntos de entrecruzamiento son precisamente los cruces de caminos a través de los que se comunican dichas vías. Conocer estos puntos de cruce permite identificar propiedades del sistema en forma global, sabiendo cuales son los caminos que confluyen hacia el proceso patológico, y como difieren del proceso sano. Con esto en mente, se puede también identificar el papel de una vía particular en relación con otras; pensando en las vías susceptibles de modulación farmacológica, este conocimiento puede ayudar para evitar los fenómenos de resistencia ya descritos. Con estas consideraciones en mente, el presente trabajo utiliza el concepto de entrecruzamiento de vías en las diferentes exploraciones realizadas, con el fin de integrar las observaciones encontradas en un contexto sistémico que describa los distintos fenotipos de cáncer de mama estudiados.

Capítulo 3

Metodologías

3.1. Base de datos de microarreglos de expresión de cáncer de mama

El presente trabajo tiene como un antecedente directo trabajo previo de nuestro laboratorio computacional para la generación de un compendio extenso de datos de microarreglos de expresión de cáncer de mama (Baca-López *et al.*, 2012). Esta base de datos (denominada *KBDB* en honor a su principal curadora) contiene un conjunto de 880 microarreglos de expresión de mRNA, recopilados de diez experimentos diferentes cuyos resultados se encuentran públicamente en el *Gene Expression Omnibus* (Edgar *et al.*, 2002).

3.1.1. Composición de la base de datos

Todos estos experimentos se realizaron de acuerdo al protocolo GPL96 , con extracciones de mRNA total en la plataforma de microarreglos Affymetrix[®] HGU133A que tiene una cobertura de 12500 genes. 61 de las muestras contenidas corresponden a tejido mamario sano, el cual fue utilizado como grupo de control para los contrastes realizados en este trabajo. Las 819 muestras restantes corresponden a muestras de tejido de

cáncer de mama primario, sin recibir tratamiento farmacológico. El cuadro 3-1 describe la contribución de cada experimento a la base de datos.

Cuadro 3-1: Origen de las muestras que componen la *KBDB*

| ID serie GEO | No. Tumores | No. Controles | Referencia |
|--------------|-------------|---------------|-------------------------------|
| GSE1456 | 159 | | Pawitan <i>et al.</i> (2005) |
| GSE1561 | 49 | | Farmer <i>et al.</i> (2005) |
| GSE2603 | 99 | | Minn <i>et al.</i> (2005) |
| GSE2990 | 61 | | Sotiriou <i>et al.</i> (2006) |
| GSE3494 | 4 | | Miller <i>et al.</i> (2005) |
| GSE4922 | 249 | | Ivshina <i>et al.</i> (2006) |
| GSE7390 | 198 | | Desmedt <i>et al.</i> (2007) |
| GSE6883 | | 3 | Liu <i>et al.</i> (2007) |
| GSE9574 | | 15 | Tripathi <i>et al.</i> (2008) |
| GSE15852 | | 43 | Ni <i>et al.</i> (2010) |
| Total | 819 | 61 | |

3.1.2. Análisis de datos de microarreglos

Los datos crudos de microarreglos fueron procesados siguiendo el algoritmo *Robust Multi-array Average* (Irizarry *et al.*, 2003). Este algoritmo ha sido implementado en el grupo de trabajo Baca-López *et al.* (2009), permitiendo la normalización en conjunto de microarreglos de lotes experimentales distintos Tovar *et al.* (2015).

Para un conjunto de n microarreglos con un número p de sondas:

1. Corrección de fondo de las señales de cada microarreglo, utilizando una transformación B.
2. Normalización de los microarreglos por cuantiles, ajustando de las distribuciones de señales entre muestras.
3. Agregar las señales de cada sonda ajustándolas a un modelo lineal.

El resultado de este tratamiento es una *matriz de expresión* con niveles de expresión comparables para cada sonda y para cada muestra. En el apéndice A se muestra código detallado para esta sección.

3.2. Metodologías de clasificación en subtipos moleculares

El presente trabajo parte de considerar la importancia en la heterogeneidad del cáncer de mama. Esta heterogeneidad se refleja en muchas variables clínicas y fisiológicas. Constando con datos experimentales de microarreglos de expresión, la mejor estrategia de clasificación a la que se puede acceder es la subtipificación molecular, ya que esta depende únicamente de niveles de expresión de los genes identificados como clasificadores. Los subtipos moleculares se asocian a fenotipos, entre los cuales existen diferencias bien determinadas.

Los algoritmos clasificadores de subtipos moleculares son conocidos como *single sample predictors*. Estos algoritmos parten de un conjunto de muestras de entrenamiento en el cual se han identificado los diferentes subtipos. Para cada subtipo, se establece un *centroide* a partir de los niveles de expresión promedio de genes clasificadores. Una vez identificados estos centroides, se pueden asignar subtipos a muestras sin clasificar. Para cada muestra problema, se calcula el centroide de los mismos genes, y se asigna el subtipo más cercano.

La clasificación en subtipos moleculares más extendida es la propuesta por Sørлие *et al.* (2001). El algoritmo utilizado como estándar de oro para esta clasificación es PAM50 (Parker *et al.*, 2009), basado en los niveles de expresión de 50 genes.

El siguiente pseudocódigo describe la asignación de subtipos moleculares

```
1   
2 Sea E un conjunto de entrenamiento de microarreglos de expresion
```



```

3
4 Sea P un conjunto de microarreglos de expresion que se busca
   subtipificar
5
6 Sea G un conjunto de genes de clasificacion
7     medidos en cada microarreglo de P
8
9 Centrar los valores de expresion de P
10     Para cada gen en G
11         Calcular mediana de expresion
12             de las medidas de expresion en cada
13                 microarreglo de P
14     Para cada muestra en P
15         Para cada gen en G
16             Restar mediana
17
18 Calcular centroides de cada subtipo en E
19
20 Para cada muestra en P
21     Calcular distancia a centroides
22     Asignar al subtipo del centroide mas cercano

```

En el presente trabajo se utilizaron dos clasificadores moleculares: el antes citado PAM50, y CITBCMST (Guedj *et al.*, 2012), un algoritmo basado en una propuesta de refinamiento de los subtipos moleculares, basado en la expresión de 241 conjuntos de sondas de la plataforma Affymetrix HGU133A. Este segundo algoritmo fue escogido por ser complementario a PAM50.

En el presente trabajo se buscaba encontrar conjuntos de subtipos moleculares muy bien

caracterizados, con alta heterogeneidad entre subtipos y homogeneidad dentro de los subtipos. Para ello, se buscó eliminar muestras cuya asignación a un subtipo pudiera ser dudosa. Para esto utilizamos la *estrategia de clasificadores en tándem*, en la cual se preservaron aquellas muestras que fueron clasificadas de manera congruente por ambos clasificadores.

El siguiente pseudocódigo describe la estrategia:

```
1 Dado un conjunto de microarreglos control L
2   para cada microarreglo l en L
3     CITBCMST(l)
4     q ← max(distancia al centroide para todos los l en L)
5     q ← max. valor aceptado de dist. centroide
6
7 Dado un conjunto de microarreglos control C
8   para cada microarreglo c en C
9     CITBCMST(c)
10    Si dist.centroide(c) < q
11      asignar subtipo a c
12      c ∈ D
13      D ← conjunto de microarreglos clasificados por
14          CITBCMST
15
16   para cada microarreglo c en D
17     PAM50(c)
18     asignar subtipo a c
19
20 Si subtipo CITBCMST de c = subtipo PAM50 de c
21     c ∈ E
22     E ← conjunto de muestras clasificadas por
23         subtipo molecular
24     E = {luminal A, luminal B, basal, Her2+ }
```

En el apéndice B se muestra código para utilizar los algoritmos PAM50 y CITBCMST sobre un conjunto de muestras de cáncer de mama.

3.3. Expresión diferencial y enriquecimiento de vías

El estudio de los distintos fenotipos de cáncer de mama requiere la identificación de las diferencias que existen entre estos y el tejido mamario sano, por ejemplo a partir de datos de expresión de genes. Estas comparaciones fueron realizadas a dos niveles: análisis de expresión diferencial, y enriquecimiento de vías.

3.3.1. Análisis de expresión diferencial

Para el análisis de expresión diferencial se utilizaron métodos de ajuste de modelos lineales, contenidos en el paquete de R - Bioconductor `limma`, *linear models for microarrays* (Smyth, 2005).

En este caso, se realizaron cuatro comparaciones entre los cuatro subtipos tumorales: luminal A, luminal B, basal y HER2+, y el fenotipo de tejido mamario sano. Para ello, se utilizaron los datos de microarreglos de expresión previamente subtipificados, haciéndose un ajuste a modelos lineales utilizando regresión robusta a nivel de conjuntos de sondas. Para nuestros fines, se consideraron diferencialmente expresados aquellos conjuntos de sondas que:

- presentan un *log fold change* mayor a $|1|$. El *log fold change* es la proporción de cambio en el nivel de expresión entre condiciones experimentales, expresado en logaritmo base 2.
- presentan un estadístico B mayor a 5. El estadístico B es un indicador de consistencia estadística en los contrastes entre grupos de niveles de expresión.

En el apéndice C se muestra código para implementar este análisis.

3.3.2. Análisis de enriquecimiento de vías

Teniendo en mente comparaciones a nivel de desregulación de vías, se buscó implementar una estrategia de análisis de enriquecimiento de vías. En este caso, se busca identificar cuales son las vías biológicas que, en conjunto, presentan una desregulación a nivel transcripcional en los subtipos moleculares observados, en comparación con el tejido mamario sano.

Existen muchas estrategias para el análisis de enriquecimiento de vías (García-Campos *et al.*, 2015). Con el fin de incorporar la información transcripcional de los datos experimentales sin sesgos, se optó por la implementación de un algoritmo de tipo *Functional Class Scoring* (Khatri *et al.*, 2012) En este tipo de análisis, dada una colección de conjuntos de genes (por ejemplo, *pathways*), y conjuntos de datos de expresión (por ejemplo, microarreglos) de diferentes fenotipos, se realiza una comparación estadística entre los niveles de expresión de los elementos de cada conjunto de genes entre los fenotipos, y se asigna un valor de significancia estadística a la diferencia en la expresión de cada conjunto de genes.

El algoritmo implementado en este trabajo fue **GAGE**, *Generally Applicable Geneset Enrichment* (Luo *et al.*, 2009). **GAGE** está basado en un procedimiento parametrizado de aleatorización de genes sobre la distribución de niveles de expresión de genes individuales.

Para cada conjunto de microarreglos por fenotipo canceroso, se realizaron comparaciones uno-a-uno entre todas las muestras del fenotipo y las muestras del fenotipo sanos. Se escogieron dos colecciones de *pathways* para este trabajo: KEGG y Reactome, para verificar que los resultados no dependen de la colección de *pathways* utilizada. En el apéndice D se muestra código para implementar este análisis.

3.4. Construcción de redes

El uso de *redes* permite la representación de las interacciones entre los elementos que participan en fenómenos biológicos. En el presente trabajo, se implementaron dos tipos de redes diferentes: redes transcripcionales, y redes de entrecruzamiento. A continuación se describen las metodologías para la generación de ambas.

3.4.1. Redes transcripcionales

Se infirieron redes transcripcionales a partir de datos de microarreglos de expresión. Para esto, se utilizaron los datos de microarreglos previamente clasificados utilizando las metodologías de clasificación antes descritas. Para la construcción de redes transcripcionales, se utilizó una implementación del algoritmo ARACNE (Margolin *et al.*, 2006) paralelizado (Tovar *et al.*, 2015) para el cálculo de *información mutua* (MI), una medida de la dependencia estadística de variables: en nuestro caso, la expresión de genes. La implementación de ARACNE paralelizado se encuentra disponible en <https://github.com/CSB-IG/parallel-aracne>

Para generar las redes por subtipo se siguió el algoritmo:

Para cada conjunto de microarreglos subtipificados:

- 1) Se calculó información mutua a nivel de sondas, para todo par de sondas contenidas en la plataforma de microarreglos Affymetrix HGU133A (22283 sondas).
- 2) Se conservaron las 10000 interacciones de mayor valor de información mutua.
- 3) Se descartaron aquellas interacciones que contienen sondas que no mapean a un transcrito nombrado por el *HUGO Genome Nomenclature Committee*.

A partir de estas redes crudas, se siguieron dos estrategias de *pulido* distintas.

Islas mayores a 25 nodos

En este caso retiramos aquellas islas menores a 25 nodos. Esto pensando en que estas interacciones no aportan mucha información a la interconexión de los nodos dentro de la red.

Filtrado por anotación funcional

En este caso, optamos por hacer un filtrado de los nodos en función de la anotación funcional. De esta manera, las interacciones encontradas a nivel transcripcional en nuestras redes pueden ser contextualizadas en relación a las funciones biológicas ya descritas. Para ello, se utilizaron dos fuentes. Por un lado, la descripción funcional de las redes en *KEGG*. Por otro lado, identificamos aquellos transcritos que han sido descritos como factores transcripcionales, siguiendo el trabajo de Vaquerizas *et al.* (2009).

3.4.2. Redes de entrecruzamiento de vías

El fenómeno de entrecruzamiento de vías establece intercomunicación entre diferentes rutas biológicas. A través de estas interconexiones se pueden tener contribuciones de procesos al desarrollo de un fenotipo. Es por esta razón que dicho fenómeno ha sido ampliamente estudiado y se ha encontrado que es importante para el estudio del cáncer de mama (Azad *et al.*, 2015; Baker *et al.*, 2014a; Guo *et al.*, 2011; Hollmén *et al.*, 2015; Karamouzis *et al.*, 2015; Wang *et al.*, 2016).

Una red biológica compuesta por los transcritos codificantes en el genoma tiene un número de nodos del orden de 10^5 . Al integrarse los elementos no codificantes del genoma, la complejidad de la red puede aumentar notablemente. Haciendo uso del fenómeno de entrecruzamiento y la anotación funcional de rutas biológicas, en el presente trabajo se presenta la propuesta de generación de *redes de entrecruzamiento*, las cuales permiten la exploración de la comunicación entre vías biológicas en un espacio más acotado.

Consideremos la definición de *pathway*, ruta o vía biológica proporcionada en el marco teórico del presente trabajo, y considerando una *colección de pathways* como una base de datos que contiene un conjunto de rutas biológicas curadas por expertos a partir de datos experimentales o literatura científica.

Para la generación de las redes de entrecruzamiento se ha seguido el siguiente algoritmo:

```

1
2 Dada una coleccion de pathways W
3
4 Poblar una red donde cada pathway pw sea un nodo
5
6 Para cada par de pathways  $pw_i$  y  $pw_j$  en W
7 Calcular el indice de Jaccard de las moléculas contenidas en  $pw_i$ 
   y  $pw_j$ 
8
9  $J_{pw_i,pw_j} = \frac{pw_i \cap pw_j}{pw_i \cup pw_j}$ 
10
11 Si
12
13  $J_{pw_i,pw_j} > 0$ 
14
15 Establecer un enlace entre  $pw_i$  y  $pw_j$ 

```

Una implementación ejecutable de este algoritmo se encuentra en el apéndice [E](#).

El resultado es una red ponderada no dirigida. En ella, los nodos son vías biológicas, y las aristas son el índice de Jaccard como una medida del entrecruzamiento, donde un mayor valor de J representa un mayor número de moléculas compartidas entre vías. Aquellas vías que no comparten moléculas tienen una intersección vacía y por lo tanto, su índice de Jaccard es de 0, de tal forma que no existe una conexión entre ellas. Sí dos vías son completamente idénticas, su índice de Jaccard será de 1. Valores de J cercanos a 1 representarían vías que comparten muchos elementos y difieren en un número pequeño de moléculas. Al considerar únicamente las moléculas que componen las vías, somos capaces de recuperar las interacciones biológicas que pueden estar omitidas en la anotación de una de las vías entrecruzadas, pero no de la otra. De esta manera, se obtiene un contexto

completo del sistema biológico, con una complejidad menor a la de una red biológica compuesta por moléculas específicas. En el presente trabajo, se construyeron redes de entrecruzamiento para cuatro colecciones de vías biológicas: KEGG (Kanehisa *et al.*, 2014), Biocarta (Nishimura, 2001), Pathway Interaction Database (Schaefer *et al.*, 2009), y Reactome (Joshi-Tope *et al.*, 2005).

3.4.3. Análisis de redes

Para la descripción de las redes, se obtuvieron las siguientes medidas: distribución de grado (número de nodos a los que se conecta un nodo), número de componentes conectados, diámetro de la red y coeficiente de agrupamiento. Los cálculos de parámetros topológicos estructurales de redes se realizaron utilizando las implementaciones de los algoritmos relevantes contenidas en la herramienta NetworkAnalyzer de Cytoscape 3 (Shannon *et al.*, 2003).

3.5. Exploración de blancos terapéuticos para cáncer de mama

Para la exploración sistemática de blancos terapéuticos moleculares para cáncer de mama, se desarrolló la estrategia de la generación de un *espacio de blancos terapéuticos moleculares de cáncer de mama*. En el capítulo *Espacio de blancos terapéuticos moleculares para subtipos de cáncer de mama* se explora dicho espacio.

3.5.1. Generación del espacio de blancos terapéuticos moleculares de cáncer de mama

Para nuestro estudio, nos enfocaremos en moléculas codificadas en el genoma.

Definición 1. *Una molécula codificada en el genoma es aquella biomolécula que puede ser generada a partir de la transcripción de información contenida en una secuencia genética.*

Toda molécula codificada en el genoma puede ser descrita en función de sus niveles de expresión. Dado que los niveles de expresión de cada molécula codificada en el genoma pueden variar entre distintos fenotipos, se puede definir cada molécula como un vector, donde cada componente es la magnitud de sus niveles de expresión. Adicionalmente, se puede asociar a cada molécula una medida de su influencia sistémica. Una medida de esta influencia puede ser el número de vías biológicas a las cuales se ha asociado dicha molécula. De esta manera, se puede representar cada molécula como un vector multidimensional, con componentes de niveles de expresión asociados a fenotipos, y una componente de influencia sistémica. La colección de estos objetos vectoriales compone un espacio vectorial.

Para la generación de nuestro espacio, consideramos cuatro fenotipos: los cuatro subtipos moleculares de cáncer de mama. Como una medida de la expresión, decidimos utilizar los niveles de *log-fold change* entre cada subtipo con el fenotipo sano (obtenidos del Análisis de Expresión Diferencial antes descrito). Se descartaron todas aquellas moléculas que no se consideraron diferencialmente expresadas (es decir, se descartan aquellas que no cumplen con *log-fold change* mayor a ± 1 y estadístico B mayor a 5). Como medida de la influencia, se examinaron 1330 vías descritas contenidas en la *Molecular Signature Database* del Instituto Broad, en la sección C2/CP (*curated gene sets/ canonical pathways* (Subramanian *et al.*, 2005)).

3.5.2. Minería de bases de datos farmacológicas

Se buscaron asociaciones de agentes farmacológicos para cada molécula contenida dentro del espacio. Para este propósito, se desarrolló una herramienta programática para

la minería de la meta-base de datos *Drug-Gene Interaction database* (Griffith *et al.*, 2013; Wagner *et al.*, 2015). La implementación de la herramienta programática de minería se puede encontrar en el apéndice **F**.

Aquellas moléculas identificadas con al menos una asociación a un agente farmacológico se preservaron en el espacio de blancos terapéuticos moleculares de cáncer de mama propiamente. Aquellas moléculas para las cuales no se identificaron agentes farmacológicos se conformaron en un espacio alterno, el *espacio de blancos moleculares actualmente sin tratamientos asociados*.

3.5.3. Clasificación de los blancos farmacológicos

Al expresar los blancos como variables vectoriales, se tiene la posibilidad de utilizar distintos abordajes para su categorización. Como un primer acercamiento, se utilizó una estrategia semicuantitativa en dos niveles. En el primer nivel, se identificó la direccionalidad de la expresión diferencial (es decir, sobreexpresados, subexpresados, o sin cambio) entre subtipos; con base en esto, se agruparon las moléculas sobreexpresadas en todos los subtipos, las subexpresadas en todos los subtipos, las sobreexpresadas únicamente en un subtipo, etc... Para cada molécula, se identificó el número de vías biológicas en las que participa como una medida de su potencial influencia sistémica. Los blancos moleculares identificados en una única vía biológica podrán tener efectos sistémicos más acotados. Por su parte, las moléculas asociadas a más de un vía biológica podrán tener efectos más extensos, aunque también aumenta la probabilidad de potenciales efectos adversos. Finalmente, algunas de las moléculas no se encuentran asociadas a vías biológicas; consideramos que eso es indicativo de una falta de conocimiento actual del contexto biológico en el que se desenvuelven dichas moléculas, y por lo tanto no es posible actualmente predecir su nivel de influencia sistémica.

3.6. Flujo de trabajo para la exploración de vías biológicas asociadas a actividad farmacológica

La información generada con las diferentes metodologías desarrolladas y herramientas implementadas requiere de estrategias de integración en el contexto de vías biológicas. La figura 3-1 describe un flujo de trabajo propuesto para la exploración de vías biológicas. Se parte de una o un conjunto de vías biológicas que resulten de interés. A partir de las *redes de entrecruzamiento* se localizan aquellas vías con las que pueden interaccionar. Este flujo se alimenta con datos de expresión por fenotipo. Con estos datos es posible utilizar *enriquecimiento de vías* para la subred de rutas biológicas que entrecruzan con la vía de interés original. Los *perfiles de expresión* resultantes dan una noción general de las posibles influencias sistémicas en las alteraciones fenotípicas observadas en las rutas de interés.

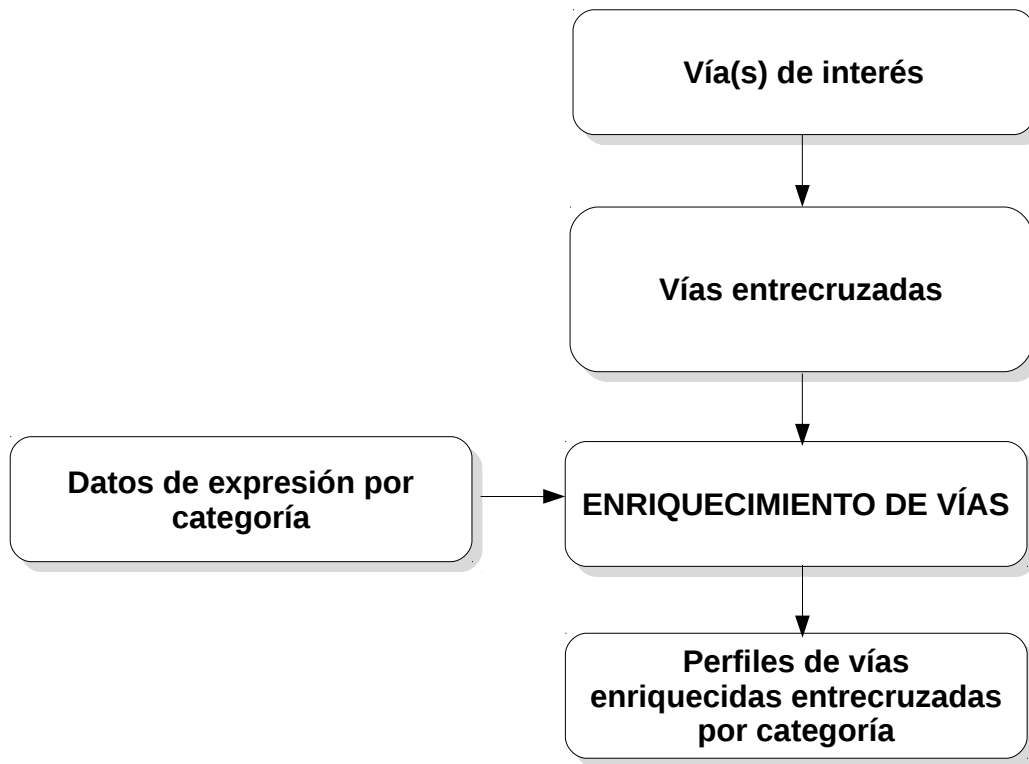


Figura 3-1: Flujo de trabajo para el estudio de desregulación de vías entrecruzadas.

En el capítulo titulado *Caso de estudio: entrecruzamiento de la vía de estrógeno* se muestra un ejemplo de esta estrategia para explorar los posibles efectos del entrecruzamiento de vías en la eficacia terapéutica de tratamientos actuales para cáncer de mama.

Resultados

En los siguientes capítulos se presentan los resultados de la aplicación de las metodologías antes descritas. De manera general, los resultados obtenidos son los siguientes:

- La estrategia de clasificación molecular implementada permitió la subtipificación del conjunto original de microarreglos de muestras de expresión en subconjuntos de los subtipos moleculares luminal A, luminal B, basal y HER2-enriquecido. La estrategia permitió obtener conjuntos subtipificados de alta homogeneidad intra-grupo y alta heterogeneidad intergrupo, apropiados para los análisis subsecuentes.
- El análisis de expresión diferencial de genes mostró que existen muchas diferencias a nivel transcripcional entre el fenotipo sano y los subtipos moleculares de cáncer de mama. Encontramos también que existen muchas moléculas que se encuentran diferencialmente expresadas en todos los subtipos moleculares con respecto a los sanos, mientras que existen pocas moléculas desreguladas exclusivamente en un subtipo específico.
- El análisis de enriquecimiento de vías mostró que existe una desregulación muy extensa de procesos en los fenotipos cancerosos con respecto del fenotipo sano. A pesar de que todos los subtipos estudiados mostraron una gran cantidad de procesos desregulados, encontramos que todos los procesos presentan perfiles de enriquecimiento únicos.

- La arquitectura de redes transcripcionales de los fenotipos cancerosos es muy distinta a la arquitectura transcripcional del tejido mamario sano, teniendo este último una red dominada por un gran componente conectado, mientras que las redes de cáncer están conformadas por varios componentes. Además, se observa que ninguna de las redes, tanto del fenotipo sano como de subtipos cancerosos, está compuesta de exactamente los mismos elementos.
- Las redes de entrecruzamiento permitieron la integración de la información de procesos. Estas redes mostraron que la mayoría de los procesos biológicos descritos no son independientes, en tanto que existen muchos elementos comunes entre ellos.
- Se generó un espacio de blancos terapéuticos que permitió integrar la información de expresión diferencial y participación en vías biológicas para identificar potenciales blancos terapéuticos para cáncer de mama, tanto generales como subtipo-específicos.
- Las metodologías implementadas fueron utilizadas para estudiar el fenómeno de resistencia a tratamientos antiestrogénicos en cáncer de mama. Se identificó cómo la vía de señalización de estrógeno se encuentra desregulada en todos los subtipos moleculares, aún sin la sobreexpresión del receptor de estrógeno, y cómo algunos procesos que entrecruzan con esta vía (por ejemplo, apoptóticos y de respuesta inmune) pueden estar involucrados en el desarrollo de resistencia antiestrogénica.

Capítulo 4

Clasificación en subtipos moleculares

Partiendo del conjunto original de microarreglos de expresión de cáncer de mama primario, con 819 muestras, nuestra estrategia de clasificación *en tandem* permitió asignar un subtipo molecular a 493 muestras de cáncer de mama. Las figuras 4-1 y 4-2 muestran una representación gráfica de los resultados de los algoritmos clasificadores CITBCMST y PAM50, respectivamente, en forma de un gráfico de componentes principales.

Es importante mencionar que con esta estrategia de clasificación buscábamos tener un conjunto de muestras clasificadas que fueran homogéneas dentro de cada subtipo, y heterogéneas entre subtipos, para capturar de mejor manera las diferencias entre subtipos. La estrategia presentada es adecuada para nuestros fines de investigación, pero hay que resaltar que no se trata de una estrategia apropiada para el contexto clínico, en tanto que las muestras que no cumplen con nuestros estrictos estándares de clasificación son descartadas.

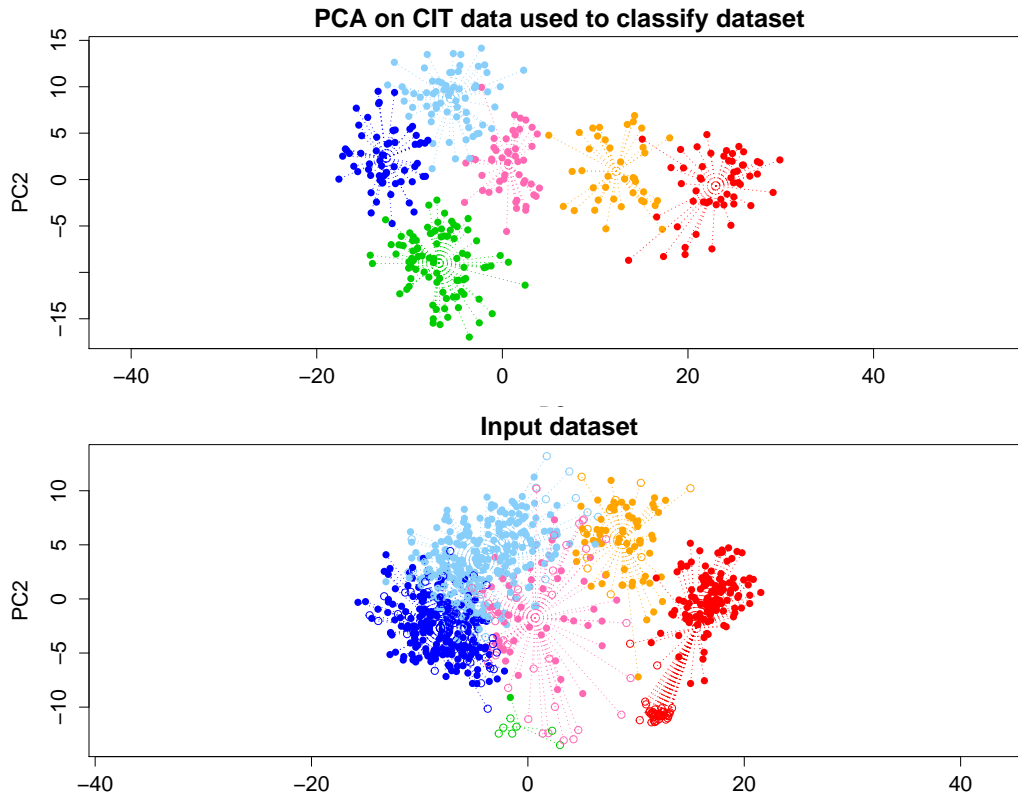


Figura 4-1: Gráfico de componentes principales de muestras clasificadas por CITBCMST (Panel superior, conjunto de entrenamiento, panel inferior, conjunto de prueba)

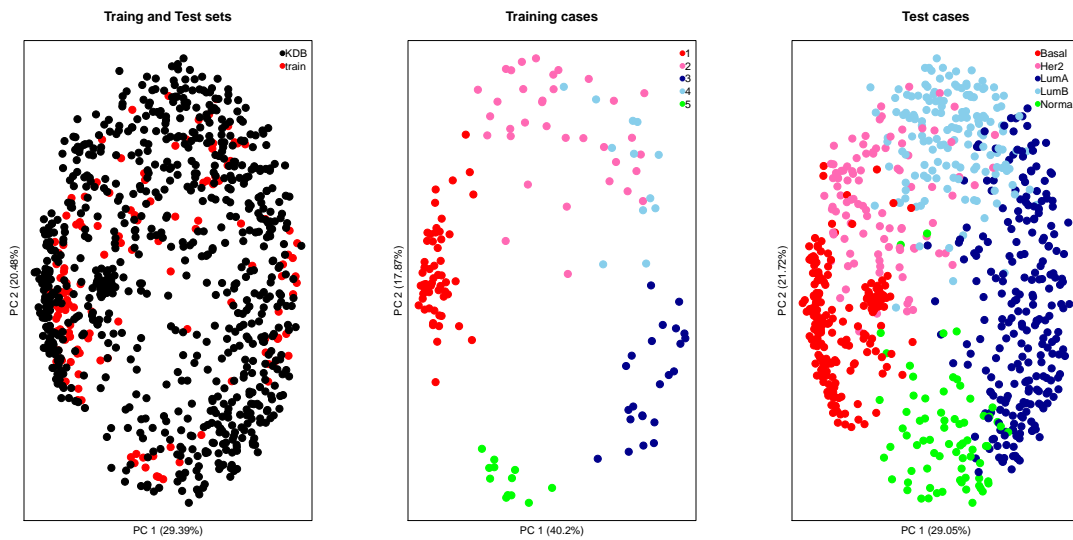


Figura 4-2: Gráfico de componentes principales de muestras clasificadas por PAM50 (Panel izquierdo, unión de los conjuntos de entrenamiento y prueba, panel central, conjunto de entrenamiento, panel derecho, conjunto de prueba)

La figura 4-3 muestra una representación tridimensional de las muestras clasificadas de manera congruente por ambos algoritmos. En esta imagen se puede notar que los conjuntos de subtipos moleculares recuperados cumplen con las condiciones buscadas de tener alta heterogeneidad entre grupos y contar con homogeneidad intragrupos.

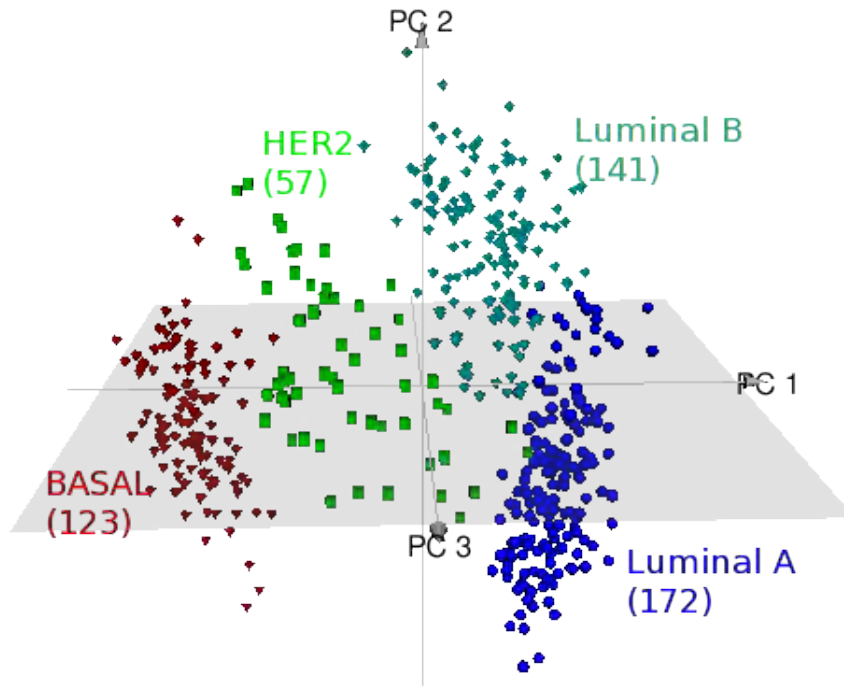


Figura 4-3: Gráfico tridimensional de componentes principales de muestras subtipificadas, con concordancia en la clasificación por CITBCMST y PAM50

Los conjuntos moleculares obtenidos sirvieron como la base que permitió continuar con las diferentes exploraciones realizadas en el presente trabajo.

Capítulo 5

Expresión diferencial y enriquecimiento por subtipos

Un punto central para estudiar fenotipos patológicos es identificar diferencias con el fenotipo sano. En el presente trabajo, se utilizó información transcriptómica de diferentes subtipos moleculares de cáncer de mama. Para analizar dicha información, se trabajó a dos niveles distintos: por un lado, la expresión diferencial de genes; por el otro, el enriquecimiento de vías.

5.1. Análisis de expresión diferencial

Las figuras 5-1 y 5-2 muestran la desregulación de transcritos apropiadamente identificados con un *Gene Symbol*, separados en sobreexpresados y subexpresados.

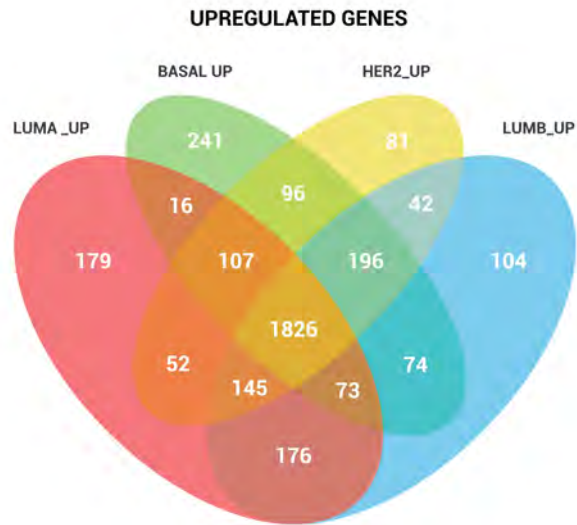


Figura 5-1: Diagrama de Venn que muestra transcritos sobreexpresados en subtipos contra el fenotipo sano. Nótese que la intersección de los cuatro subtipos es mucho mayor que los genes alterados únicamente en algún subtipo molecular

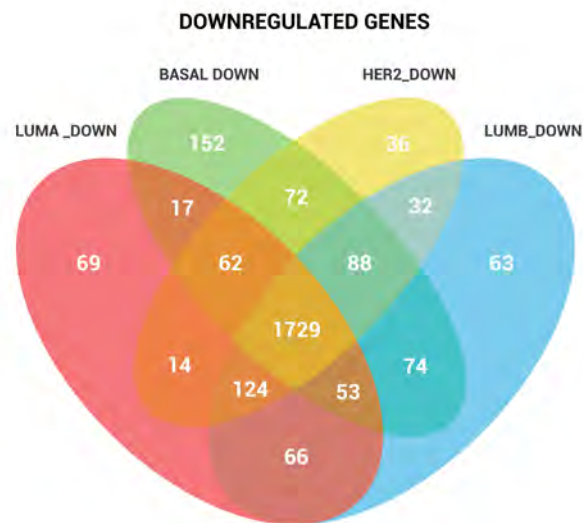


Figura 5-2: Diagrama de Venn que muestra transcritos subexpresados en subtipos contra el fenotipo sano. Nuevamente nótese que la intersección de los cuatro subtipos es mucho mayor que los genes alterados únicamente en algún subtipo molecular

En estas figuras podemos observar que los fenotipos cancerosos muestran marcadas diferencias a nivel transcripcional con el fenotipo sano. Se aprecia que la mayor parte de los transcritos desregulados se encuentran compartidos entre los diferentes subtipos moleculares: la desregulación de estas moléculas parece estar asociada a procesos generales del cáncer de mama.

Por su parte, podemos observar que existen moléculas desreguladas que no se asocian a un subtipo, presentándose todas las combinaciones hasta tener moléculas subexpresadas o sobreexpresadas únicamente en un subtipo molecular. Estas moléculas estarán asociadas a las diferencias existentes entre subtipos moleculares, hasta explicar características específicas de cada uno de los subtipos moleculares.

5.2. Enriquecimiento de vías

Las diferencias a nivel de procesos biológicos pueden resultar más reveladoras para estudiar diferencias entre fenotipos, especialmente si lo que se busca es eventualmente llegar a restaurar las condiciones fisiológicas saludables (por ejemplo, mediante modulación farmacológica). Por esto, consideramos el uso de análisis de enriquecimiento de vías.

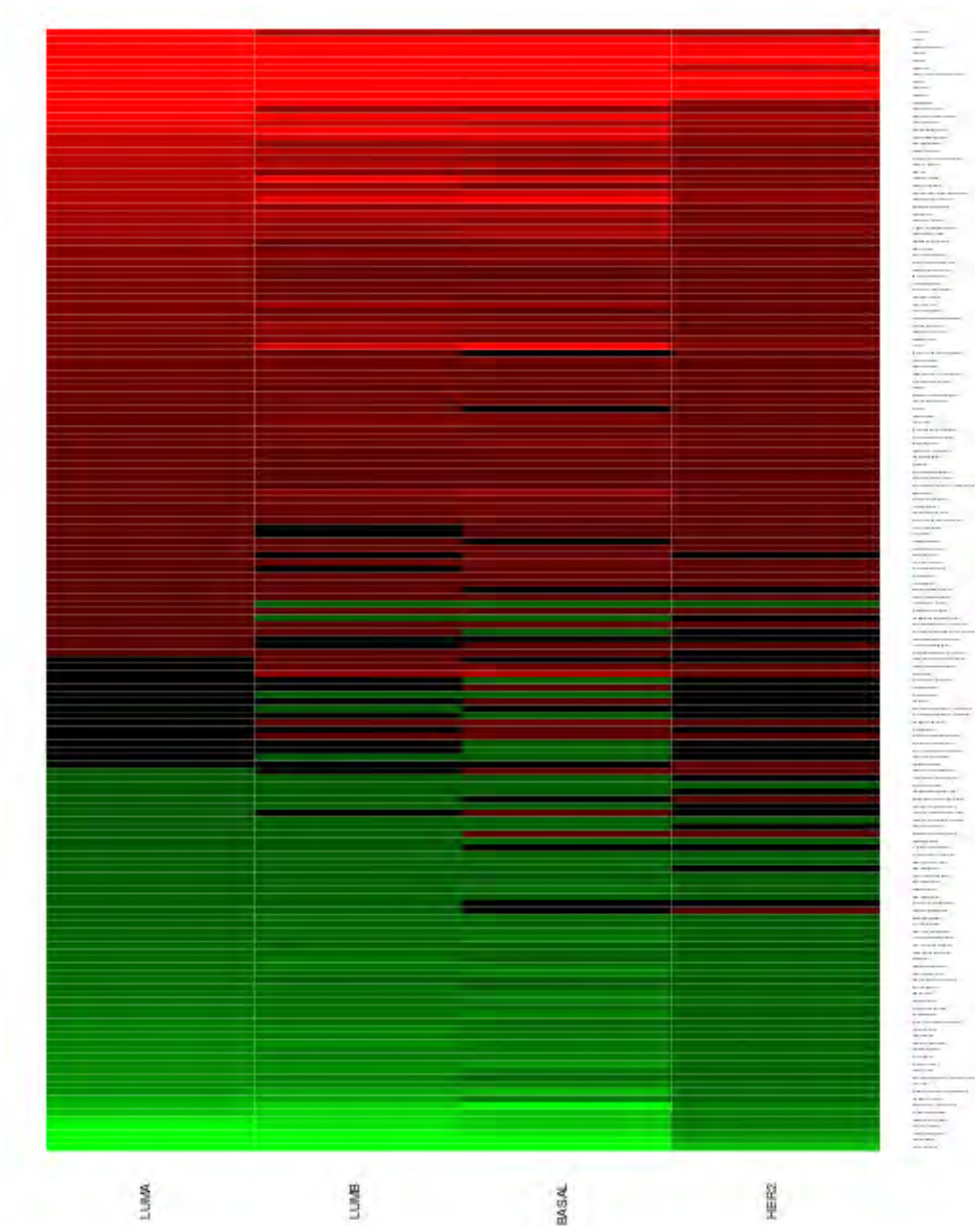


Figura 5-3: Enriquecimiento de vías de KEGG en GAGE. Las vías sobreexpresadas están en rojo. Las vías subexpresadas están en verde. Las vías en negro no están significativamente enriquecidas. La intensidad del color es proporcional a la significancia estadística (valor de q).

En la figura 5-3 se muestra un mapa de calor que muestra el enriquecimiento de todas las vías contenidas en KEGG para los cuatro subtipos de cáncer de mama, comparados con tejido mamario sano, a partir de los datos de microarreglos de expresión. Se puede apreciar que existe una desregulación extendida de una gran cantidad de procesos en to-

dos los subtipos moleculares, esperable al tratarse de muestras cancerosas, y congruente con los resultados de expresión diferencial a nivel de genes individuales. Llama la atención que la mayoría de los procesos desregulados se encuentran desregulados *en la misma dirección* (sobre o subexpresados) en todos los subtipos moleculares, salvo algunos casos puntuales. Estos resultados nuevamente nos llevan a pensar en procesos de desregulación comunes a los procesos generales de cáncer. Observemos, sin embargo, que ninguno de los subtipos moleculares cuenta con el mismo perfil de desregulación, por lo que podemos pensar que estas diferencias puntuales a nivel transcriptómico se asocian a las diferencias observadas entre fenotipos.

En la figura 5-4 se presentan datos de enriquecimiento de vías sobre los mismos conjuntos de subtipos moleculares, pero utilizando vías descritas en Reactome, que tiene un número mayor de vías que KEGG. Podemos observar que nuevamente existe una amplia desregulación de procesos en todos los subtipos moleculares, y que de la misma forma ningún subtipo molecular muestra un mismo perfil de desregulación.

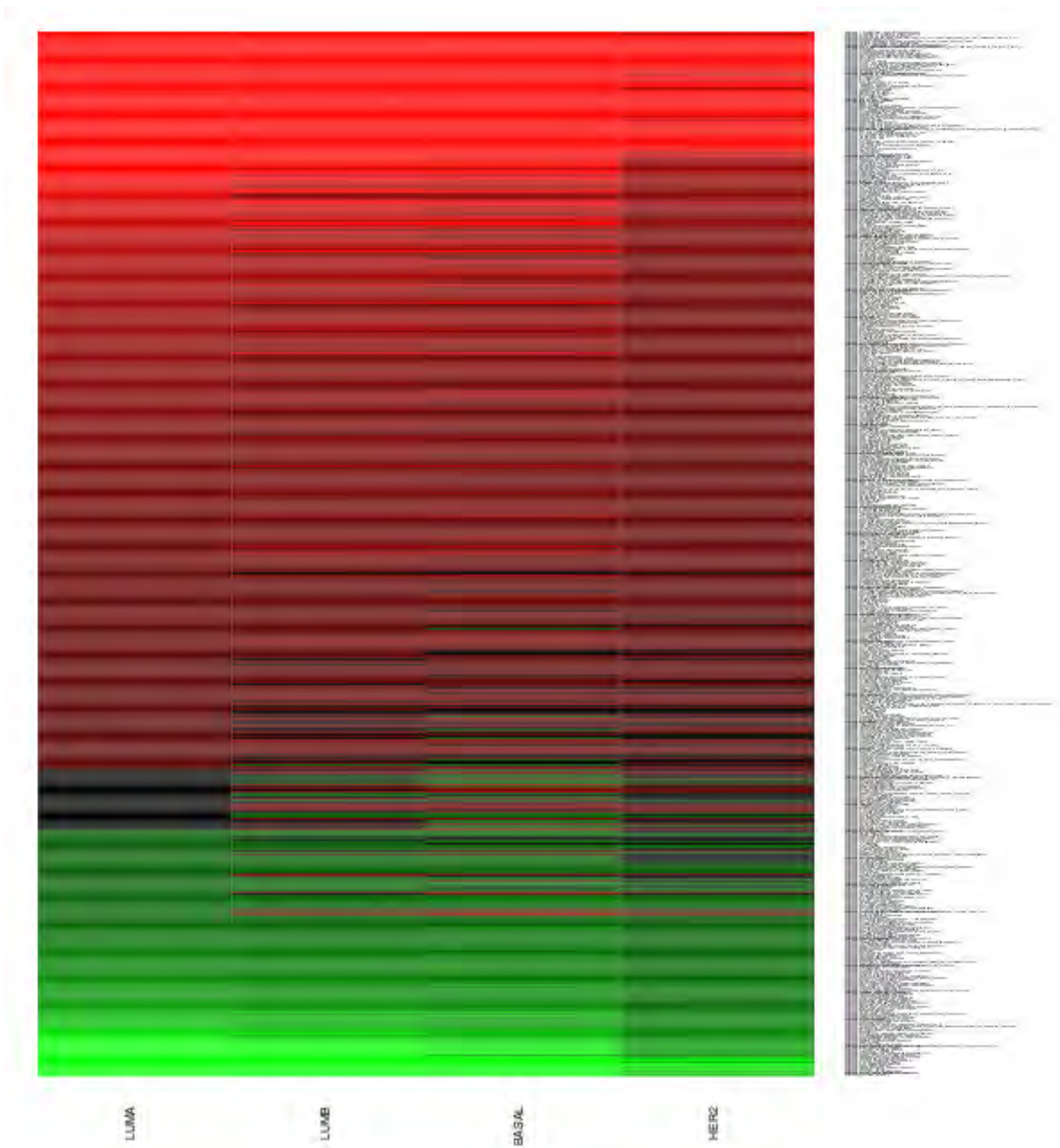


Figura 5-4: Enriquecimiento de vías de Reactome en GAGE. Las vías sobreexpresadas están en rojo. Las vías subexpresadas están en verde. Las vías en negro no están significativamente enriquecidas. La intensidad del color es proporcional a la significancia estadística (valor de q).

Capítulo 6

Arquitectura de redes transcripcionales de subtipos moleculares de cáncer de mama

La estructura de las redes transcripcionales está relacionada con las características de un fenotipo determinado (MacNeil y Walhout, 2011; Carter *et al.*, 2013). La estructura de las redes transcripcionales recuperada a partir de información experimental es dependiente de la estrategia empleada para su reconstrucción. En este sentido, el uso de información mutua como una medida de interacción transcripcional resulta óptimo, en tanto que permite establecer relaciones no lineales entre transcritos; dicho de otra forma, permite identificar si existe dependencia estadística en los niveles de expresión de dos genes, más allá de la simple coexpresión. Para el presente estudio se utilizaron dos estrategias para la reconstrucción de las redes: en una, el único criterio de filtrado fue el valor de información mutua de las interacciones; en el otro, se consideró también la anotación de los nodos, sea como factores transcripcionales (Vaquerizas *et al.*, 2009) o bien, moléculas asociadas a procesos descritos en *KEGG* (Kanehisa *et al.*, 2014).

6.1. Redes transcripcionales de subtipos moleculares

Para la construcción de estas redes, el principal criterio fue obtener redes transcripcionales con valores altos de información mutua para sus interacciones. De esta manera, las interacciones reportadas en la red tienen una mayor significancia estadística y son más representativas del fenotipo. Para este estudio es muy importante la comparabilidad de las redes. Las redes inferidas estadísticamente son sensibles a los criterios empleados para su construcción. En ese sentido, nuestras redes son comparables entre sí al ser construidas siguiendo el algoritmo descrito en la sección metodológica. Es relevante remarcar esto, ya que la comparación de nuestras redes con redes transcripcionales inferidas con otras estrategias no es directa, y dependiendo de la estrategia, puede no ser posible.

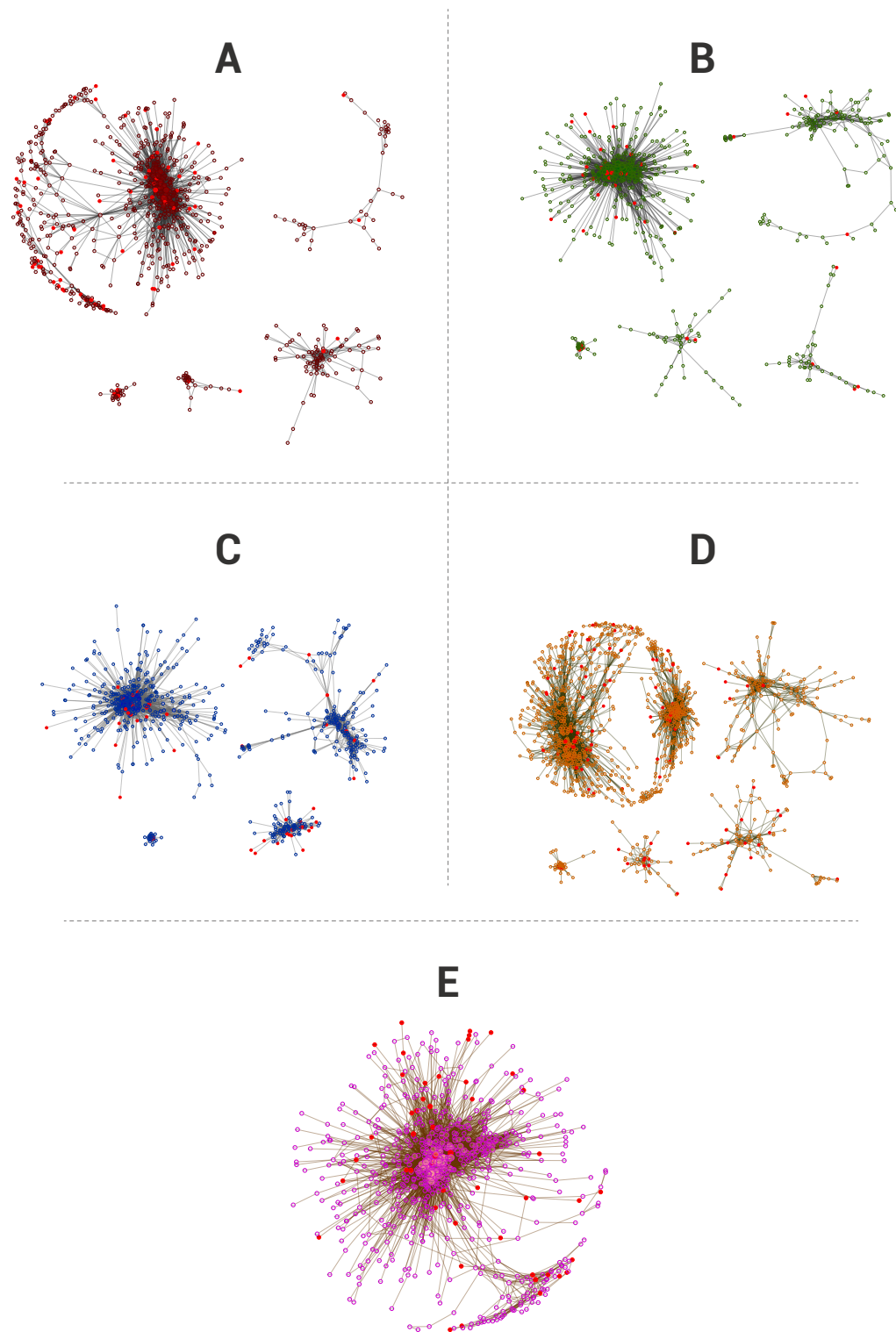


Figura 6-1: Redes transcripcionales para subtipos moleculares de cáncer de mama. Los paneles A-D muestran redes para los subtipos luminal A, luminal B, basal y HER2-enriquecido respectivamente. El panel E muestra una red transcripcional para tejido mamario sano. Los nodos identificados como factores transcripcionales se encuentran coloreados en rojo

6.1.1. Composición de las redes

Los nodos que componen cada una de las redes varían dependiendo de los fenotipos. En la figura 6-2 se puede apreciar que existe un número reducido de nodos compartidos por todos los subtipos moleculares, y un número importante de nodos exclusivos para cada subtipo molecular. Esto es indicativo de que en cada subtipo existen dependencias estadísticas distintas entre transcritos, lo que puede asociarse a las diferencias fenotípicas observadas.

El cuadro 6-1 muestra una comparación similar entre cada subtipo molecular y la red de tejido mamario sano. Utilizando el índice de Jaccard como una medida de similitud, encontramos que los subtipos difieren en al menos la mitad de los nodos que componen sus redes transcripcionales, siendo la del subtipo basal la más distinta de la de tejido mamario sano.

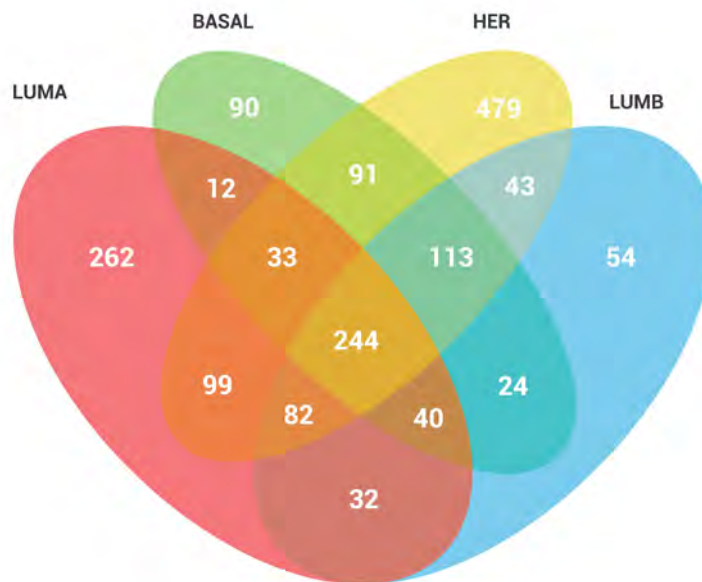


Figura 6-2: Diagrama de Venn de los nodos que componen las redes de subtipos moleculares de cáncer de mama

Cuadro 6-1: Índice de Jaccard (J) de nodos de redes transcripcionales de subtipos moleculares vs nodos de la red de tejido mamario sano

| Subtipo | J (vs Sanos) |
|---------|--------------|
| LumA | 0.0313 |
| LumB | 0.0113 |
| Basal | 0.0111 |
| HER2+ | 0.0161 |

6.1.2. Parámetros topológicos globales de las redes

Cuadro 6-2: Parámetros de redes transcripcionales

| Parámetro | LumA | LumB | Basal | HER2+ | Tejido Mamario Sano |
|-------------------------------|-------|-------|-------|-------|---------------------|
| Nodos | 805 | 632 | 646 | 1184 | 709 |
| Aristas | 4567 | 4772 | 4494 | 5124 | 6220 |
| Componentes Conectados | 5 | 5 | 4 | 5 | 1 |
| Diámetro de la red | 16 | 16 | 10 | 20 | 10 |
| <i>Clustering Coefficient</i> | 0.436 | 0.527 | 0.560 | 0.330 | 0.374 |

El cuadro 6-2 muestra parámetros topológicos globales de las redes. Se puede observar que tanto el número de nodos como de aristas se encuentra en el mismo orden de magnitud tanto para los subtipos como para el tejido mamario sano, lo que sostiene la comparabilidad de las redes inferidas. El *clustering coefficient* (coeficiente de agrupación) global es una medida de la agrupación de nodos en la red (Barabasi y Oltvai, 2004), en el que valores más altos indican una mayor densidad de conexiones dentro de la red. Se puede observar que existen diferencias en la agrupación entre los distintos fenotipos. La existencia de agrupaciones dentro de los fenotipos puede relacionarse además con una estructura modular (Ravasz *et al.*, 2002). En ese sentido, existe la posibilidad de encontrar estructuras de *comunidades* funcionales dentro de las redes transcripcionales (Holme *et al.*, 2003). Una medida relacionada es el número de componentes conectados en la red. Podemos observar que en las redes construidas, los diferentes fenotipos muestran diferencias en el número de componentes conectados que presentan. De esta manera, llama la atención que el tejido mamario sano presenta un menor número de componentes co-

nectados que los fenotipos cancerosos. Hay evidencia que muestra que la desconexión de redes puede hacer más eficiente el flujo dentro de las mismas (Braess, 2005): La paradoja de Braess describe el caso de como, cuando se cierra una autopista, el flujo vehicular en las vías alternas eventualmente se estabiliza, y por el contrario, al abrir una nueva autopista, el flujo vehicular se hace más pesado. Tomando esta idea, podemos pensar que la desconexión a nivel transcripcional puede estar dando lugar a una *hiperfuncionalización celular*, en la que el flujo de información (en este caso transcripcional) se vuelve más ágil en cada componente desconexo, eficientando la expresión de características observables como *hallmarks* del cáncer.

Distribución de grado

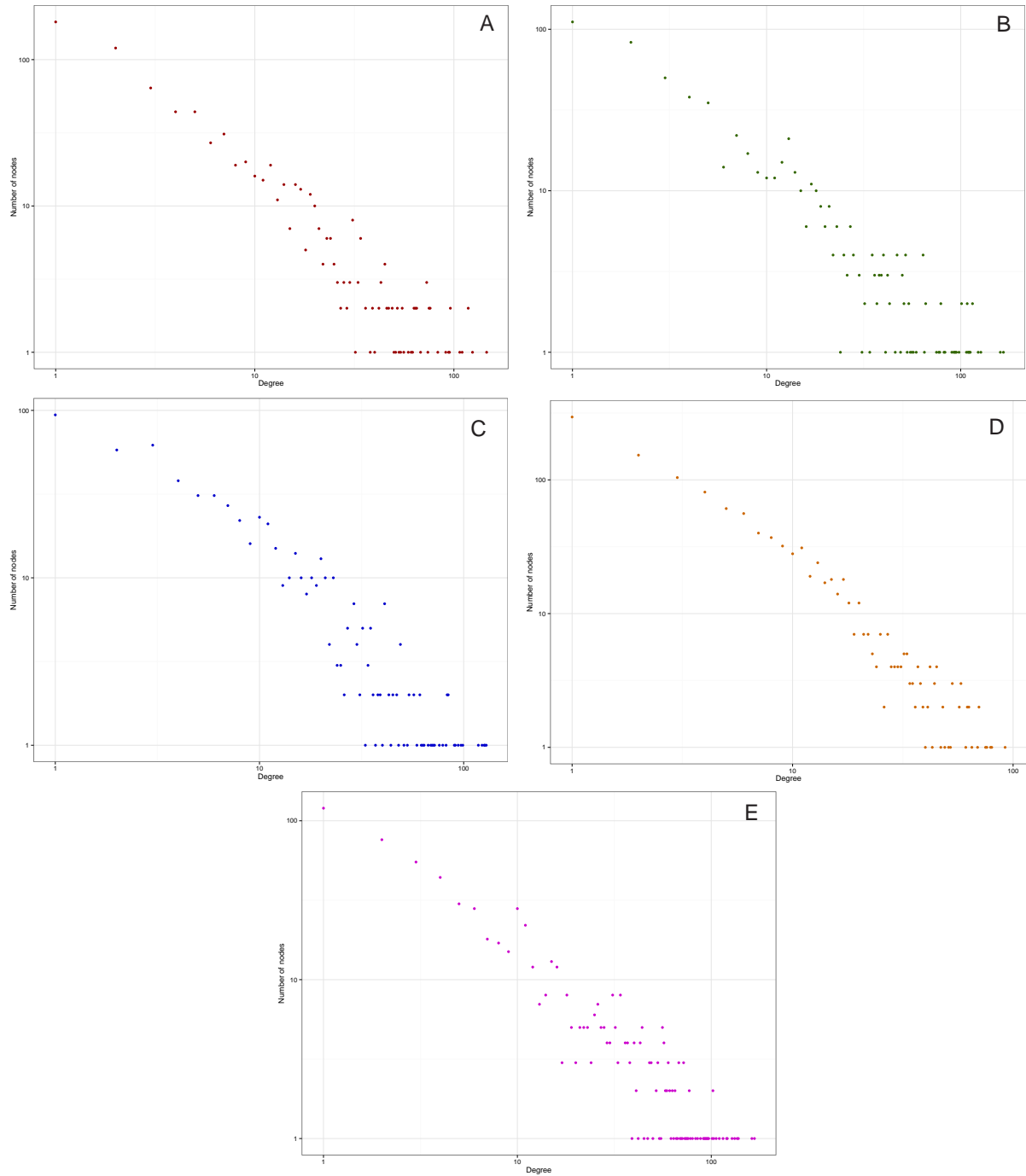


Figura 6-3: Distribución de grado de las redes transcripcionales. Los paneles A-D muestran la distribución de grado, en escala log-log, de las redes para los subtipos luminal A, luminal B, basal y HER2-enriquecido. El panel E muestra la distribución de grado, en escala log-log, para la red de tejido mamario sano.

El valor de grado de cada nodo es el número de nodos con los que el primero se conecta. Observamos que todos los fenotipos tienen una distribución de grado de *cola larga*: un número alto de nodos se conectan con un número bajo de nodos, mientras que un número muy pequeño de nodos se conectan con una gran cantidad de nodos. El grado de cada nodo puede ser utilizado como una medida de la *influencia* de dicho nodo en la red: Un transcrito con un valor de grado alto tendrá muchas conexiones a nivel transcripcional, y por lo tanto su estado influye más en la red. En ese sentido, la modulación farmacológica de un nodo de alto grado puede tener un mayor efecto en la modificación del fenotipo.

6.2. Redes transcripcionales filtradas por anotación funcional

Hacer un filtrado de las redes a partir de un criterio funcional permite hacer análisis de mayor significado biológico sobre las redes. En particular, identificar aquellos transcritos que tienen una función de factores transcripcionales resulta importante ya que las redes inferidas con la metodología empleada no son direccionadas (Tovar *et al.*, 2015), por lo que esta es una forma de asignar direccionalidad a las interacciones.

6.2.1. La estructura de las redes transcripcionales se preserva al filtrarse por anotación funcional

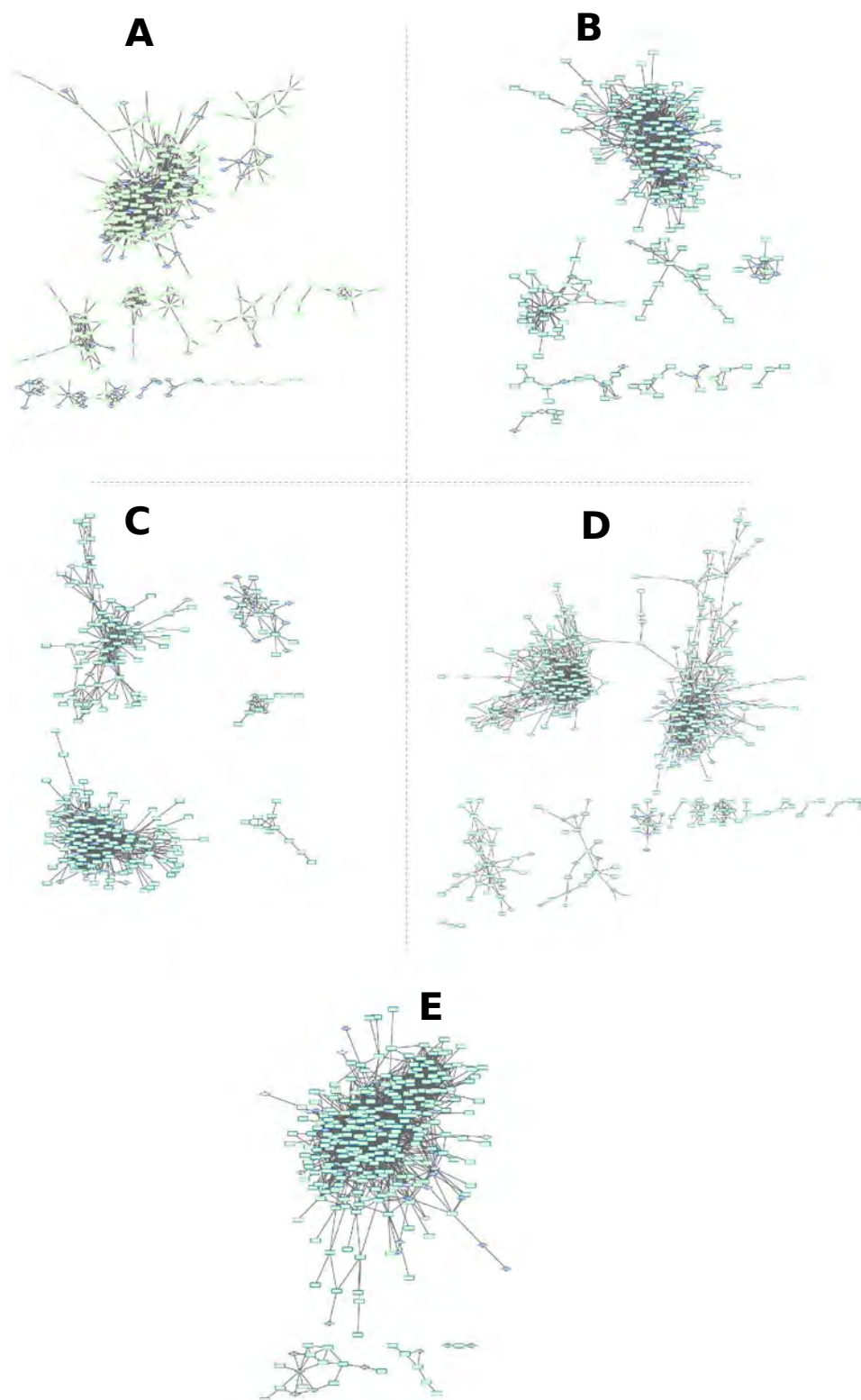


Figura 6-4: Redes transcripcionales para subtipos moleculares de cáncer de mama filtradas por anotación funcional. Los paneles A-D muestran redes para los subtipos luminal A, luminal B, basal y HER2-enriquecido respectivamente. El panel E muestra una red transcripcional para tejido mamario sano. Los nodos coloreados en verde son aquellos nodos que están identificados en algún *pathway* de *KEGG*. Los nodos coloreados en azul son factores transcripcionales anotados en (Vaquerizas *et al.*, 2009). Los nodos coloreados en rosa pertenecen a ambas categorías.

El filtrado de las redes transcripcionales considerando un criterio de anotación funcional, preserva la arquitectura de las redes, a pesar de tener un número menor de nodos y aristas. De esta manera, esta estrategia de filtrado nos permite ganar poder de interpretación biológica sin perder la información obtenida de la inferencia estadística de la red.

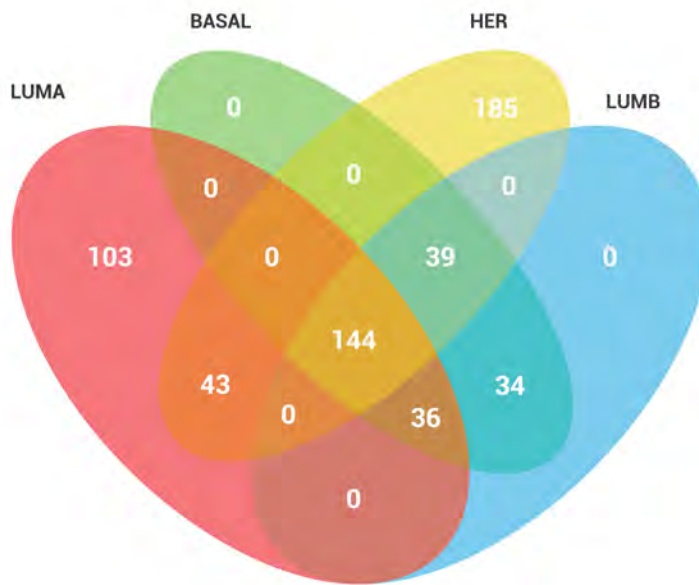


Figura 6-5: Diagrama de Venn de los nodos que componen las redes transcripcionales de subtipos moleculares de cáncer de mama, filtradas por factores transcripcionales y moléculas contenidas en KEGG

En la figura 6-5 observamos muchas de las intersecciones vacías. Llama la atención también observar que no existen transcritos asociados únicamente al subtipo luminal B o al subtipo Basal. Esto puede ser debido principalmente a una falta de anotación y conocimiento de la función biológica de las moléculas asociadas exclusivamente a estos fenotipos.

6.2.2. Parámetros topológicos de las redes filtradas

El cuadro 6-3 muestra diferencias en relación al cuadro 6-2. Observamos una diferencia en el número de nodos y aristas, resultado esperado al realizar un filtrado distinto de la red. El mismo fenómeno sucede con el diámetro de las redes. Por su parte, el *clustering coefficient* de las redes transcripcionales de los diferentes fenotipos se ve muy ligeramente modificado ante el filtrado, por lo que los indicios de modularidad de las redes se mantienen.

Cuadro 6-3: Parámetros de Redes Transcripcionales (nodos anotados en *KEGG* o como factores transcripcionales)

| Parámetro | LumA | LumB | Basal | HER2+ | Tejido Mamario Sano |
|-------------------------------|--------|--------|-------|--------|---------------------|
| Nodos | 329 | 253 | 268 | 414 | 268 |
| Aristas | 1009 | 1042 | 1154 | 1486 | 1069 |
| <i>Clustering Coefficient</i> | 0.4121 | 0.4721 | 0.52 | 0.3838 | 0.41 |
| Componentes conectados | 18 | 12 | 6 | 13 | 4 |
| Diámetro | 10 | 7 | 7 | 14 | 8 |

Capítulo 7

Redes de entrecruzamiento

Las redes de entrecruzamiento describen la comunicación que existe entre conjuntos de rutas asociadas a procesos biológicos descritos. Estas redes de entrecruzamiento resultan útiles para contextualizar procesos biológicos individuales, identificando otros procesos que pueden tener una influencia sobre uno, o bien ser influenciados por el mismo. A continuación se presentan redes de entrecruzamiento construidas para cuatro colecciones de rutas biológicas: KEGG, Biocarta, Pathway Interaction Database y Reactome.

7.1. Las redes de entrecruzamiento permiten integrar la información de una colección de vías biológicas

En la figura 7-1 podemos observar la representación gráfica de las redes de entrecruzamiento generadas para las cuatro colecciones de vías biológicas.

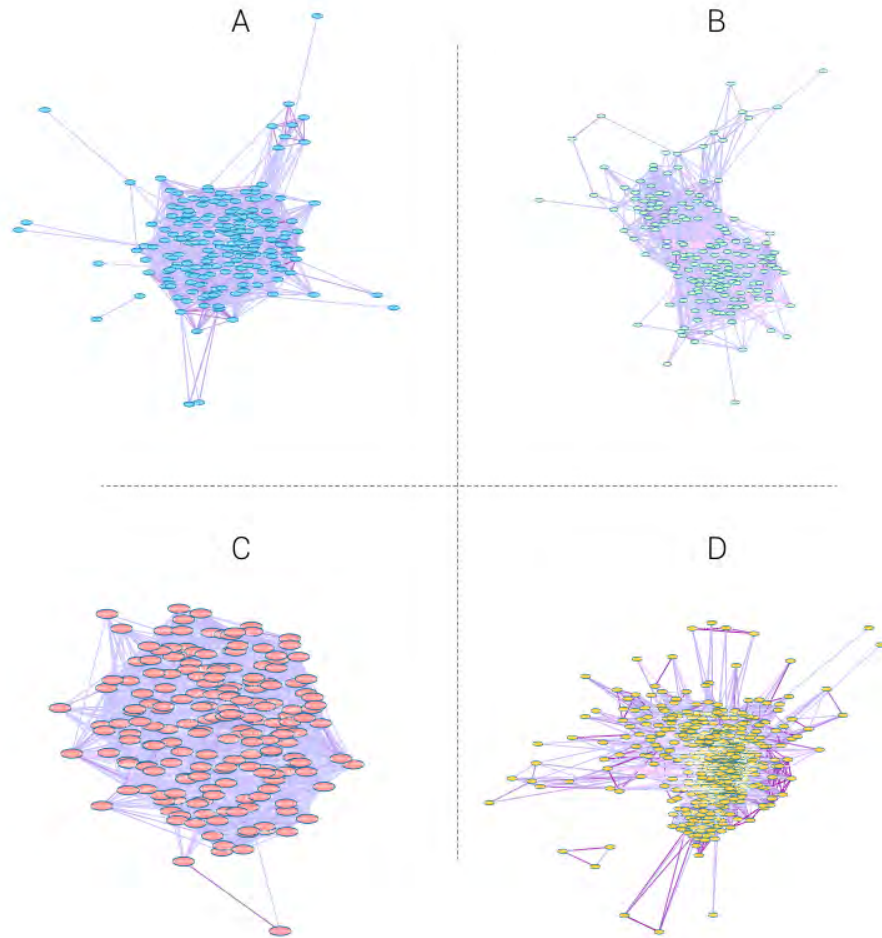


Figura 7-1: Redes de entrecruzamiento. Los paneles A-D muestran redes de entrecruzamiento para las bases de datos Biocarta, KEGG, PID y Reactome

Como se puede observar, en los cuatro casos tenemos redes densamente conectadas, mostrando comunicación entre los diferentes procesos descritos como rutas individuales. De esta manera podemos identificar que el fenómeno de entrecruzamiento se trata de un fenómeno ubicuo y recuperado en las diferentes bases de datos.

Resulta importante resaltar que diferentes bases de datos presentan diferentes descripciones de procesos distintos. En ese sentido consideramos que, para evitar tener nodos virtualmente duplicados (que tendrían enlaces con valores de J cercanos a 1) resulta

mejor construir redes de entrecruzamiento para cada colección de *pathways* individual.

7.2. Las redes de entrecruzamiento reconstruyen las interacciones sistémicas descritas por diferentes colecciones de rutas biológicas

En el cuadro 7-1 observamos algunos parámetros de las redes, que muestran diferencias en la composición de las redes.

Observamos que el número de nodos es distinto para cada colección; en este sentido, KEGG, PID y Biocarta son de tamaños comparables, mientras que Reactome es una colección notablemente más grande.

Llama la atención observar que el número de aristas es distinto para las diferentes redes. Observamos en KEGG el menor número de entrecruzamientos, seguido por Biocarta. Reactome, teniendo un número mucho mayor de nodos, tiene un número de entrecruzamientos mucho mayor. Interesantemente, PID, a pesar de tener un número de *pathways* comparable con KEGG y Biocarta, tiene una mucha mayor proporción de *pathways* entrecruzados. Estas diferencias muestran que las diferentes colecciones capturan el fenómeno de entrecruzamiento de vías, si bien con diferentes niveles de éxito. En ese sentido, las diferentes colecciones recuperan un contexto sistémico global.

7.3. Las redes de entrecruzamiento muestran una fuerte intercomunicación entre procesos

Encontramos que las redes para KEGG y PID cuentan con un único componente conectado, mientras que las redes de Reactome y Biocarta presentan dos componentes conectados; sin embargo, el segundo componente cuenta con tres y dos nodos respectiva-

mente.

Observamos que el diámetro de las redes va de un valor de 3 (para PID), pasando por 5 (Reactome y Biocarta) hasta un máximo de 6 (KEGG). Este resultado muestra una distancia entre *pathways* muy corta. Explorando la distribución de caminos más cortos en la figura 7-2, podemos observar que para todas las colecciones, la longitud de caminos más cortos promedio más frecuente es cercana a dos: las diferentes rutas son cercanas entre sí.

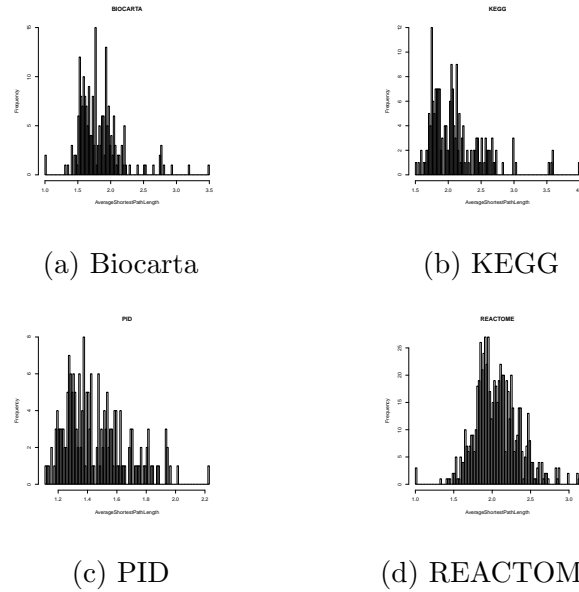


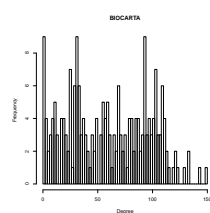
Figura 7-2: Distribuciones de longitud de caminos más cortos

7.4. Las distribuciones de grado y *betweenness centrality* de las redes de entrecruzamiento muestran diferencias entre las colecciones de vías

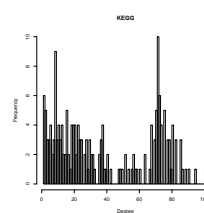
En la figura 7-3 observamos que existen diferencias en la distribución de grado de los diferentes *pathways* que integran cada una de las colecciones. Biocarta muestra una distribución de grado por lo general homogénea, en la que existe similar número de vías baja, mediana y altamente conectadas; PID muestra mayoritariamente vías altamente conectadas; KEGG muestra una cierta bimodalidad, entre vías escasamente conectadas y vías altamente conectadas; finalmente Reactome tiene muchas más vías bajamente conectadas, y escasas vías altamente conectadas. Por su parte, la figura 7-4, muestra que en todas las colecciones existen pocas vías de alta centralidad, lo que es de esperarse dada la alta conectividad de las redes.

Cuadro 7-1: Parámetros de Redes de Entrecruzamiento

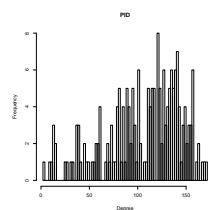
| Parámetro | Biocarta | KEGG | PID | Reactome |
|-------------------------------|----------|-------|-------|----------|
| Nodos | 216 | 188 | 195 | 674 |
| Aristas | 6519 | 3745 | 10337 | 29194 |
| <i>Clustering Coefficient</i> | 0.626 | 0.608 | 0.745 | 0.696 |
| Componentes conectados | 2 | 1 | 1 | 2 |
| Diámetro | 5 | 6 | 3 | 5 |



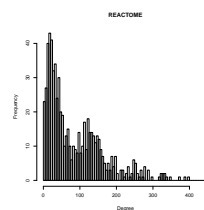
(a) Biocarta



(b) KEGG

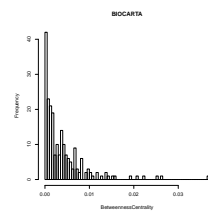


(c) PID

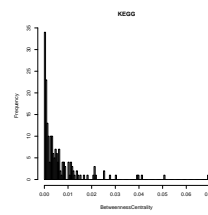


(d) Reactome

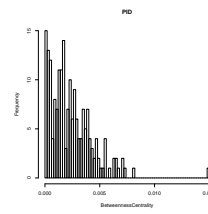
Figura 7-3: Distribuciones de grado



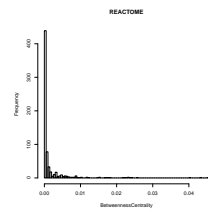
(a) Biocarta



(b) KEGG



(c) PID



(d) Reactome

Figura 7-4: Distribuciones de *Betweenness Centrality*

7.5. Consideraciones

Las redes presentadas en esta sección reconstruyen los contextos sistémicos de las vías biológicas. Existen consideraciones que se deben tomar en cuenta al trabajar con este abordaje. Una de las ventajas de este abordaje es que se reduce notablemente el número de elementos en las redes. Se pasa de un espacio del orden de 22000 elementos (pensando en genoma completo) a espacios de unos cientos de nodos, reduciendo la complejidad del sistema. Esto es una ventaja para la exploración de *grano grueso* del contexto biológico. Existe una pérdida de definición asociada, notable en la pérdida de la direccionalidad de las interacciones; a este nivel, no es posible conocer la dirección en la que una vía puede estar influyendo a otra, o si existe una influencia mutua. Sin embargo, dado que se conocen los elementos de cada vía, recuperar esta información es trivial. La estrategia de redes de entrecruzamiento permite identificar cuáles son posibles procesos biológicos que pueden estar influyendo en otros procesos de interés, y limitar o ampliar la exploración biológica tanto como sea requerido.

Es necesario recordar que las redes de entrecruzamiento se tratan de construcciones meramente teóricas, y que son dependientes de la calidad y la cobertura de la anotación biológica; mientras esta aumente, las redes de entrecruzamiento describirán mejor los contextos biológicos.

Finalmente, la construcción de estas redes está pensada para poder integrar información experimental y situarla en un contexto biológico. En el capítulo *Caso de estudio: entrecruzamiento de la vía de estrógeno* se muestra un ejemplo de la aplicación de una de estas redes para explorar diferencias fenotípicas y sus posibles implicaciones farmacológicas.

Capítulo 8

Espacio de blancos terapéuticos moleculares para subtipos de cáncer de mama

En este capítulo discutimos el espacio de blancos terapéuticos moleculares para los distintos subtipos de cáncer de mama. Con este abordaje podemos identificar aquellas moléculas que, al encontrarse desreguladas en alguna de las manifestaciones estudiadas de cáncer de mama, resultan potenciales blancos para tratamiento farmacológico. Estas moléculas además están categorizadas a partir de los subtipos en los que las encontramos desreguladas, y del número de vías en las que participan como una medida de su papel sistémico.

8.1. Espacio de blancos terapéuticos moleculares con fármacos asociados a ellos

El espacio de blancos terapéuticos moleculares con fármacos asociados a ellos consiste en un espacio vectorial pentadimensional, el cual está poblado por moléculas codificadas en el genoma que cumplen las siguientes dos condiciones:

- a) Presentan expresión diferencial en al menos uno de los subtipos moleculares de cáncer de mama: luminal A, luminal B, basal, HER2+
- b) Existe al menos un compuesto con actividad farmacológica reportada sobre la molécula de interés

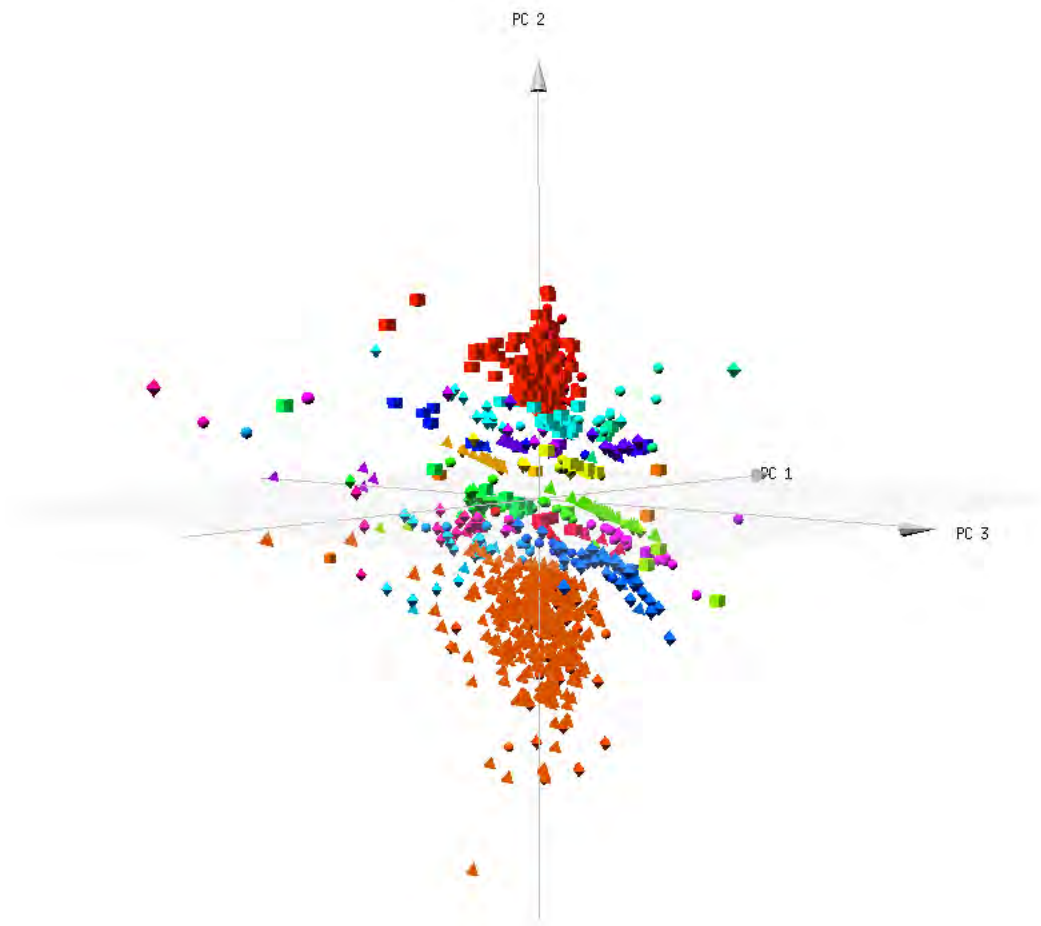


Figura 8-1: Espacio de blancos terapéuticos moleculares susceptibles a modulación farmacológica (con al menos un compuesto con actividad farmacológica asociado a ellos). Representación tridimensional de los elementos (genes) en un espacio de tres componentes principales. Los diferentes colores y formas de los elementos indican pertenencia a alguna de las categorías descritas más adelante en el presente capítulo

Este espacio está poblado por 1198 moléculas que cumplen con ambas condiciones.

La exploración de este espacio permite identificar potenciales sitios de intervención farmacológica, localizando oportunidades de *reposicionamiento de fármacos*. El reposicionamiento de fármacos es una línea de investigación importante en tanto que gran parte de los costos de desarrollo del fármaco, en particular aquellos asociados a demostrar la seguridad de los mismos, han sido ya cubiertos (Strittmatter, 2014), de tal forma que la investigación puede centrarse a identificar la aplicabilidad del tratamiento para una

condición de interés. En nuestro caso, el cáncer de mama.

En la imagen 8-1 los colores de cada molécula representan una de estas categorías. A continuación, se presenta la exploración de las distintas categorías.

8.1.1. Blancos moleculares con la misma dirección de desregulación en todos los subtipos

En esta categoría se ubican aquellas moléculas que se encuentran desreguladas en la misma dirección en todos los subtipos de cáncer de mama explorados. En este caso, los fármacos que pudieran repositionarse ofrecerían opciones terapéuticas generales, atractivas como tratamientos de amplio espectro subtipo-independientes.

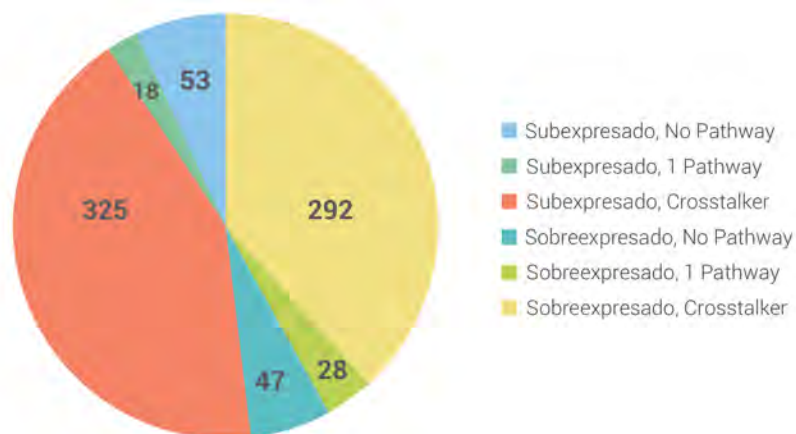


Figura 8-2: Blancos moleculares con la misma dirección de desregulación en todos los subtipos

En la figura 8-2 observamos que existe una proporción similar de blancos terapéuticos sobreexpresados y subexpresados.

Cuadro 8-1: Algunos blancos farmacológicos compartidos por todos los subtipos

| Blanco Molecular | Dirección de desregulación | No. <i>Pathways</i> | Fármaco |
|------------------|----------------------------|---------------------|--------------|
| ALOX5 | Subexpresado | 1 | Diclofenaco |
| ADCY2 | Subexpresado | 1 | Forscolina |
| EGFR | Subexpresado | 71 | HEGF |
| MTOR | Subexpresado | 59 | MHY1485 |
| OPRM1 | Subexpresado | 11 | Morfina |
| TOP1 | Sobreexpresado | 1 | Irinotecan |
| TSPO | Sobreexpresado | 1 | PK-11195 |
| AKT1 | Sobreexpresado | 193 | Perifosina |
| RAF1 | Sobreexpresado | 157 | Sorafenib |
| MAP2K2 | Sobreexpresado | 110 | Trametinib |
| COL1A1 | Sobreexpresado | 22 | Halofuginona |
| A2M | Sobreexpresado | 12 | Bacitracina |

Interesantemente, blancos de algunos tratamientos quimioterapéuticos actuales, como la topoisomerasa, se encuentran dentro de esta categoría, lo que concuerda con la idea de blancos de amplio espectro en este espacio.

8.1.2. Blancos moleculares desregulados únicamente en un sub-tipo molecular

En esta categoría se encuentran blancos terapéuticos cuya utilidad está limitada a un subtipo único.

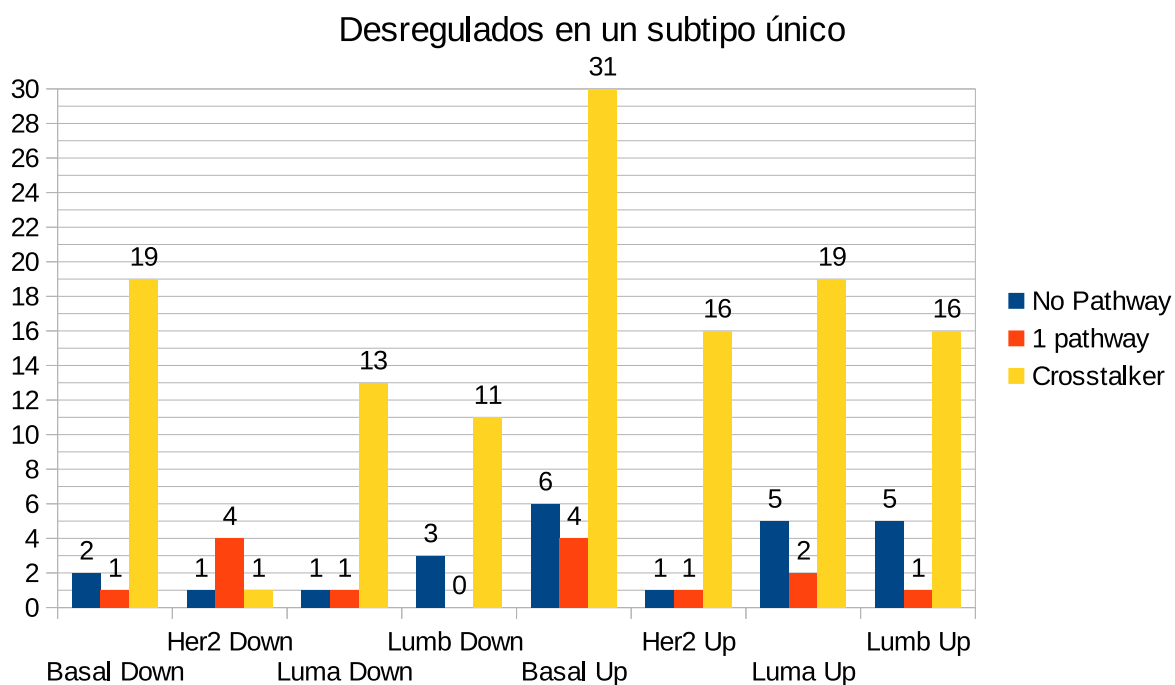


Figura 8-3: Desregulados únicamente en un subtipo. Up = Sobreexpresados en el subtipo, Down = Subexpresados en el subtipo

En la figura 8-3 podemos observar que se tienen potenciales blancos farmacológicos para todos los subtipos. En ese sentido, el subtipo basal es el que cuenta con una mayor cantidad de potenciales blancos propios.

Cuadro 8-2: Algunos blancos farmacológicos subtipo-específicos

| Blanco Molecular | Subtipo / Dirección de desregulación | No. <i>Pathways</i> | Fármaco |
|------------------|--------------------------------------|---------------------|-------------------------|
| AKR1C2 | LumB Subexpresado | 2 | Ác. Ursodesoxicólico |
| CCL2 | Basal Sobreexpresado | 17 | Danazol |
| IMPA2 | Basal Sobreexpresado | 2 | Litio |
| PTEN | Her2+ Sobreexpresado | 62 | <i>CAS 1372540-25-4</i> |
| ABL1 | Luma Sobreexpresado | 28 | Imatinib |

8.1.3. Blancos moleculares para tres subtipos.

Los blancos que pertenecen a esta categoría pueden tener una utilidad extendida, pero pueden presentar una ineficacia para algún subtipo particular.

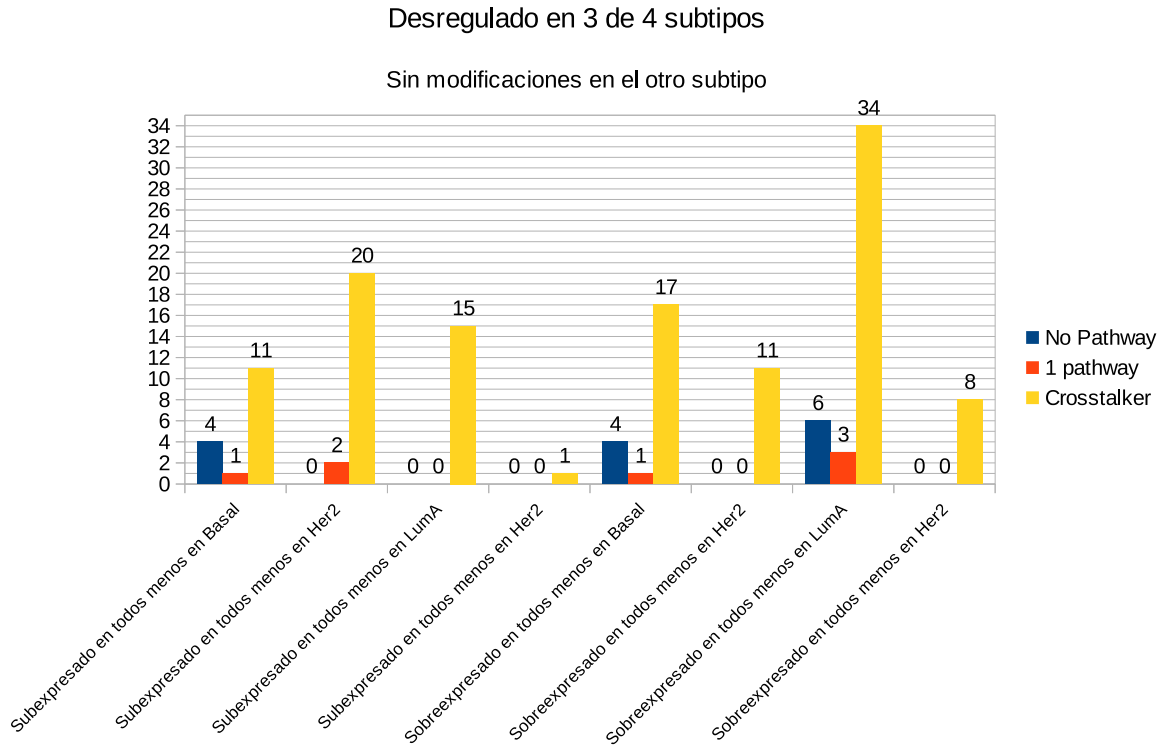


Figura 8-4: Desregulados en tres subtipos, sin cambios en el otro

En la figura 8-4 se puede observar que el subtipo más afectado por este fenómeno es el luminal A.

Cuadro 8-3: Algunos blancos farmacológicos con posible ineficacia subtipo-específica

| Blanco Molecular | Dirección de desregulación... Excepto en el subtipo... | No. <i>Pathways</i> | Fármaco |
|------------------|--|---------------------|-----------------|
| GABBR | Subexpresado... Basal | 15 | Barclofen |
| GHR | Subexpresado... Her2 | 16 | EP-51216 |
| OXTR | Subexpresado... Luma | 9 | Oxitocina |
| KEAP1 | Sobreexpresado... Her2 | 6 | Dimetilfumarato |
| CASP3 | Sobreexpresado... Luma | 50 | Minociclina |
| MMP14 | Sobreexpresado... Lumb | 5 | Marimastat |

8.1.4. Blancos moleculares para dos subtipos

En esta categoría se encuentran moléculas que presentan desregulación en dos subtipos únicamente. El ejemplo canónico de esta categoría sería el receptor de estrógeno (ESR1), blanco terapéutico de tamoxifen y otros fármacos antiestrogénicos. En el capítulo titulado *Caso de estudio: entrecruzamiento de la vía de estrógeno* se discutirá a profundidad este blanco terapéutico.

Subexpresado en dos subtipos

Sin cambios en los otros dos

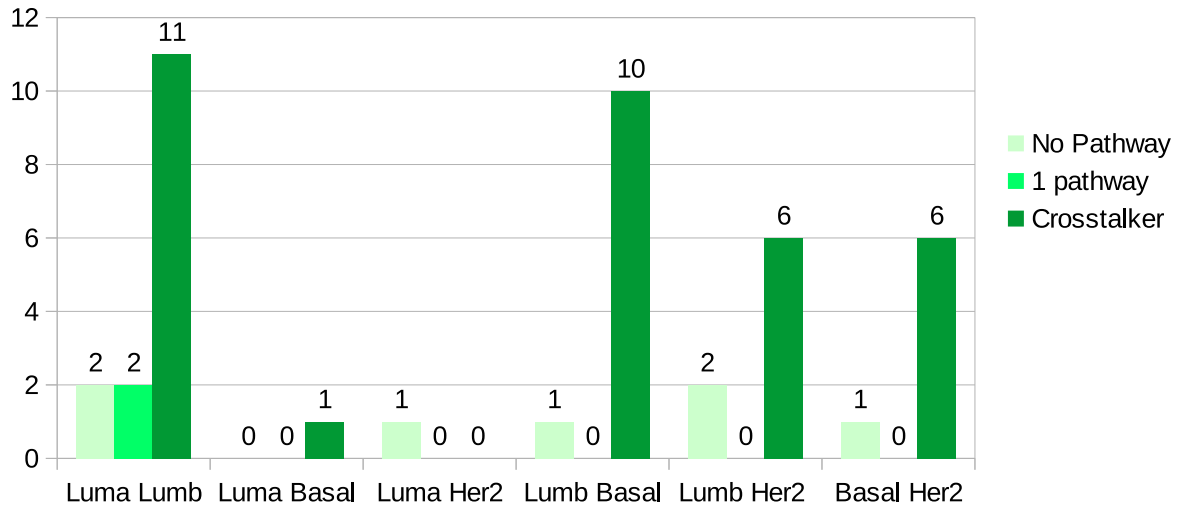


Figura 8-5: Subexpresados en dos, sin cambios en los otros

Sobreexpresado en dos subtipos

sin cambios en los otros

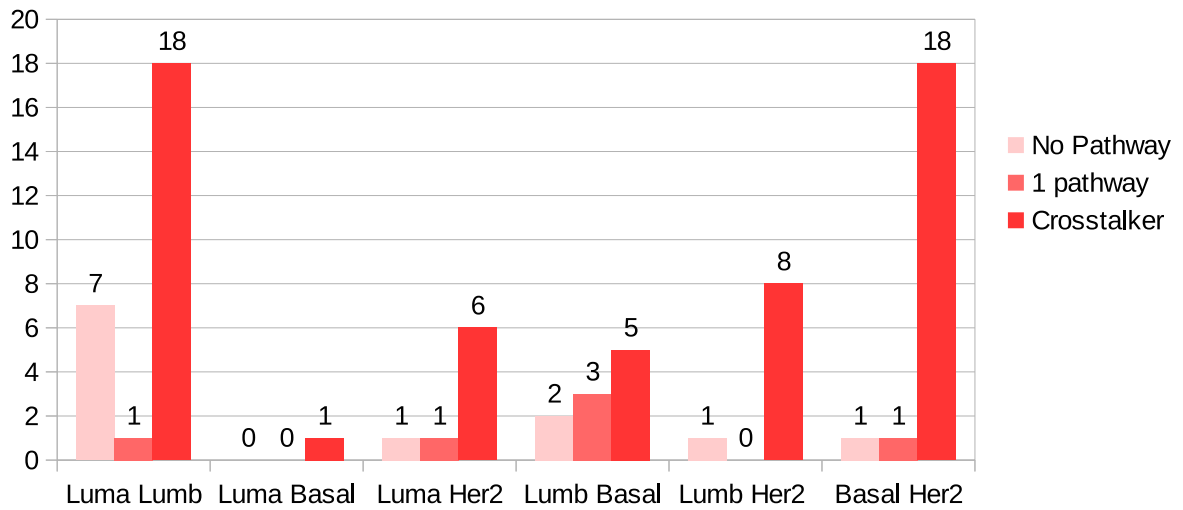


Figura 8-6: Sobreexpresados en dos, sin cambios en los otros

Cuadro 8-4: Algunos blancos farmacológicos subtipo-específicos

| Blanco Molecular | Subtipo / Dirección de desregulación | No. <i>Pathways</i> | Fármaco |
|------------------|--------------------------------------|---------------------|----------------------|
| ADORA2B | Luma/Lumb Subexpresado | 10 | Adenosina |
| ADORA3 | Luma/Basal Subexpresado | 7 | Adenosina |
| GRIN2B | Lumb/Basal Sobreexpresado | 18 | Acetilcisteína |
| MMP1 | Basal/Her2 Sobreexpresado | 15 | Marimastat |
| MMP3 | Luma/Her2 Sobreexpresado | 7 | Galardina |
| BIRC2 | Luma/Basal Sobreexpresado | 25 | GDC0917 |
| CCND1 | Luma/Lumb Sobreexpresado | 57 | Trióxido de arsénico |
| CDK4 | Lumb/Basal Sobreexpresado | 35 | Palbociclib |

8.2. Espacio de blancos moleculares actualmente sin tratamientos asociados

El espacio de blancos moleculares actualmente sin tratamientos asociados es un espacio complementario al espacio de blancos terapéuticos moleculares. Este espacio está poblado por moléculas codificadas en el genoma que cumplen las siguientes dos condiciones:

- a) Presentan expresión diferencial en al menos uno de los subtipos moleculares de cáncer de mama: luminal A, luminal B, basal, HER2+
- b) No existe al menos un compuesto con actividad farmacológica reportada sobre la molécula de interés

Como se puede apreciar, se trata de moléculas desreguladas para las cuales actualmente no se cuenta con un tratamiento farmacológico capaz de modularlas.

Como se hizo para el Espacio de Blancos Terapéuticos Moleculares, es posible categorizar las moléculas que lo componen a partir de sus niveles de expresión en cada subtipo y su nivel de influencia. En la figura 8-7 los colores de cada molécula representan la categoría a la que pertenecen.

Este espacio de blancos moleculares puede ser clasificado para su exploración; en la figura 8-7 se muestran las moléculas clasificadas con una estrategia igual a la seguida para el espacio de blancos terapéuticos moleculares con fármacos asociados; sin embargo, podemos observar que en este caso los blancos se distribuyen en un número más extenso de categorías.

No se realizó una exploración profunda de este espacio, dado que la propuesta de potenciales fármacos para estos requiere de estrategias de *diseño racional de fármacos* que van más allá de los alcances del trabajo realizado. Se presenta este resultado como un posible punto de partida para investigaciones relacionadas con el diseño quimioinformático de fármacos para cáncer de mama.

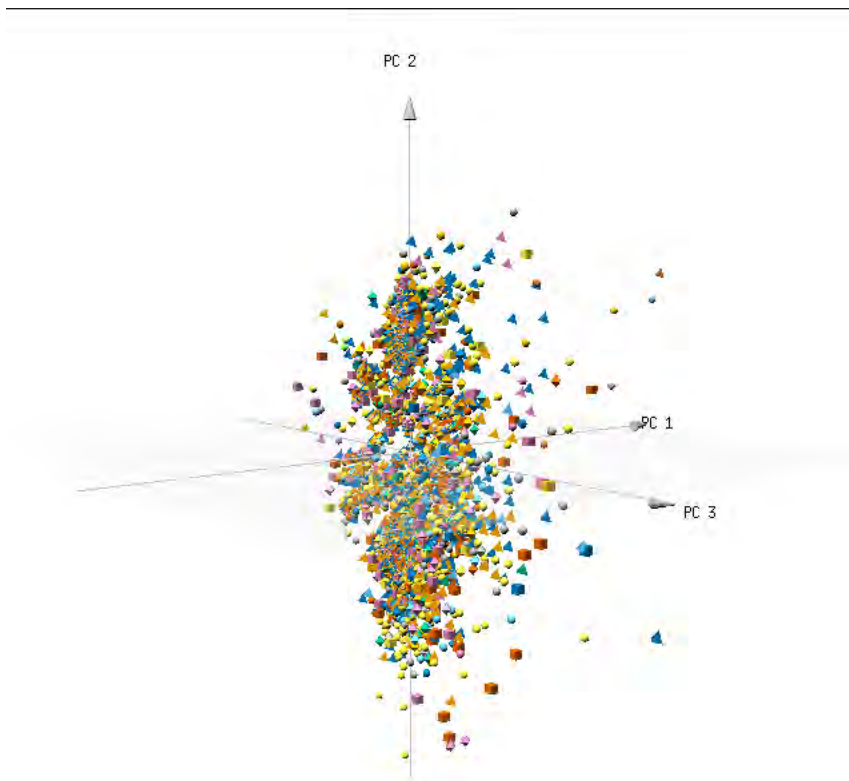


Figura 8-7: Espacio de blancos farmacológicos sin compuestos con actividad farmacológica asociados. Representación tridimensional de los elementos (genes) en un espacio de tres componentes principales. Los diferentes colores y formas de los elementos indican pertenencia a alguna categoría definida por los niveles de expresión en los cuatro subtipos y el número de *pathways* en que participan

Capítulo 9

Caso de estudio: entrecruzamiento de la vía de estrógeno

Consideramos que el marco conceptual desarrollado da pie a muchas líneas de investigación relacionadas a la farmacología del cáncer de mama. Como prueba de concepto, y para probar la aplicabilidad de nuestras metodologías, se decidió explorar la influencia del fenómeno de entrecruzamiento en distintos subtipos moleculares sobre un blanco terapéutico probado: el receptor de estrógeno. Los resultados de este trabajo, fueron recientemente publicados (de Anda-Jáuregui *et al.*, 2015).

En el presente capítulo se discutirá un caso de estudio en el cual se aplicaron los flujos de trabajo desarrollados. Se exploró el fenómeno de entrecruzamiento en la vía de señalización de estrógeno, como componente en la resistencia a tratamientos antiestrogénicos (Viedma-Rodríguez *et al.*, 2014).

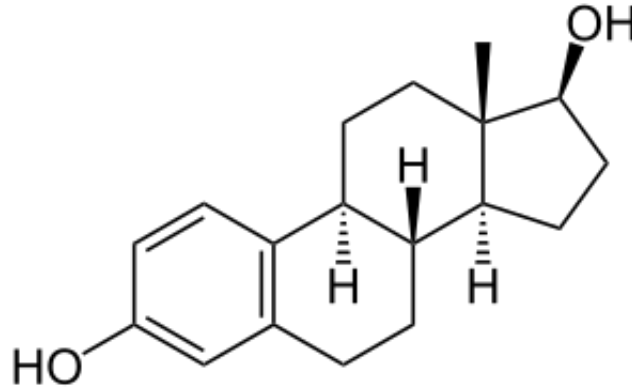


Figura 9-1: 17β -Estradiol, Principal hormona sexual femenina.

La vía de señalización de estrógeno está asociada al crecimiento, desarrollo y diferenciación celular; es fundamental para el desarrollo del sistema reproductor femenino (Hastings *et al.*, 2013). Las células mamarias son susceptibles de estimulación por estrógenos, los cuales actúan como señales de proliferación y supervivencia Levin (2005). Los estrógenos se unen a receptores de estrógeno nucleares, los cuales pueden actuar como factores transcripcionales. Adicionalmente, existen receptores de estrógeno membranales que pueden generar reacciones fisiológicas a través de proteínas que actúan como segundos mensajeros. Existe retroalimentación a través de estos segundos mensajeros. (Zwart *et al.*, 2011)

En muchos cánceres de mama existe desregulación en la señalización de estrógenos. (Liang y Shang, 2013). Como ya se ha discutido, por esa razón se trata de un blanco farmacológico de interés para el cáncer de mama. Como un caso de estudio para la integración de datos, estudiamos la desregulación de la vía de señalización de estrógeno y el posible impacto del entrecruzamiento de vías sobre ella.

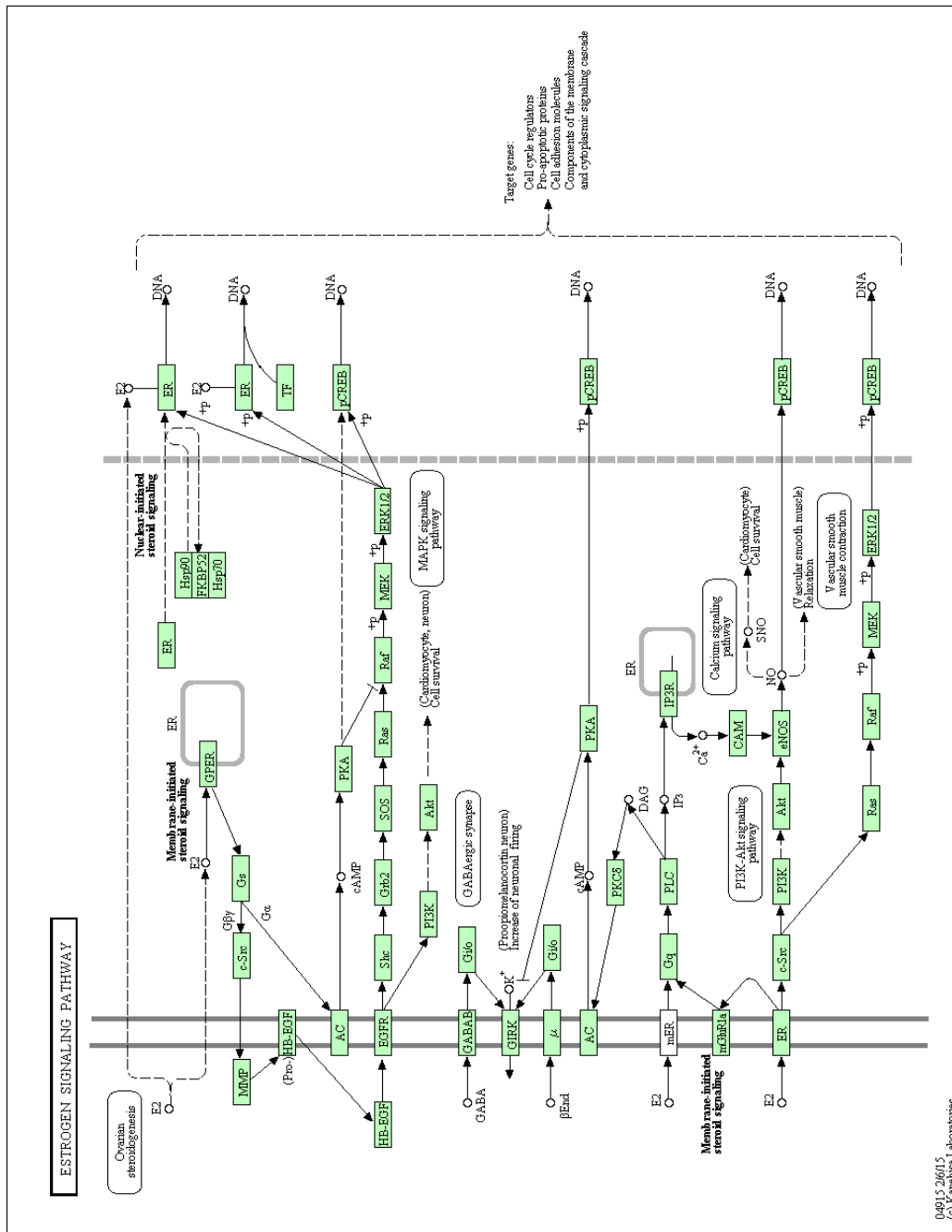


Figura 9-2: Vía de Señalización de Estrógeno. KEGG (Kanehisa *et al.*, 2014)

9.1. Flujo de trabajo para el estudio de la desregulación en vías entrecruzadas con la vía de señalización de estrógeno

El flujo de trabajo previamente descrito fue aplicado para el estudio de la desregulación de la vía de señalización de estrógeno y las vías que entrecruzan con ella. En este caso, se partió de la información contenida en la base de datos KEGG (Kanehisa *et al.*, 2014) y los datos experimentales previamente subtipificados (descritos antes en este trabajo). Las vías de interés para este estudio fueron seleccionadas como aquellas que entrecruzan con la vía de señalización de estrógeno. La desregulación a nivel de vías fue identificada utilizando *GAGE*.

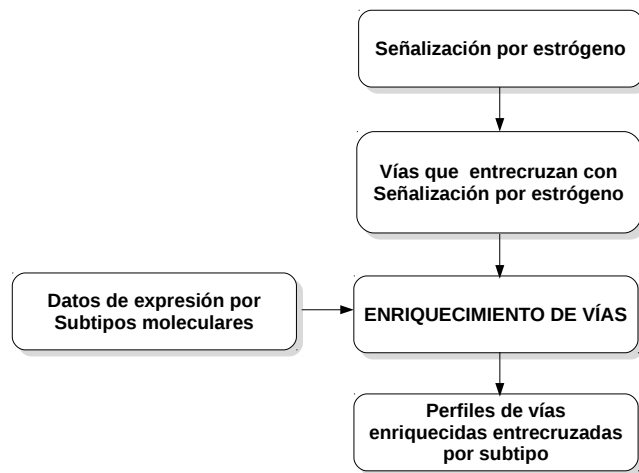


Figura 9-3: Identificación de las vías que entrecruzan con la vía de señalización de estrógeno.

9.2. Desregulación de la vía de estrógeno

De los resultados de desregulación de vías previamente presentados se observa que la vía de señalización de estrógeno se encuentra desregulada en todos los subtipos moleculares. Esto resulta relevante, dado que tradicionalmente la actividad de estrógeno se asocia a los subtipos luminales. Al explorarse a nivel genético, observamos que el receptor de estrógeno está diferencialmente expresado a la alta en subtipos luminales y a la baja en subtipos no luminales. Sin embargo, es muy importante notar que moléculas *río abajo* en la vía se encuentran activas en todos los subtipos moleculares. La implicación es que existe una actividad en la vía asociada a la señalización de estrógeno independiente de los niveles de expresión del receptor de estrógeno. Por esta razón, esta vía parece ser un componente de importancia para todos los distintos subtipos de cáncer de mama estudiados.

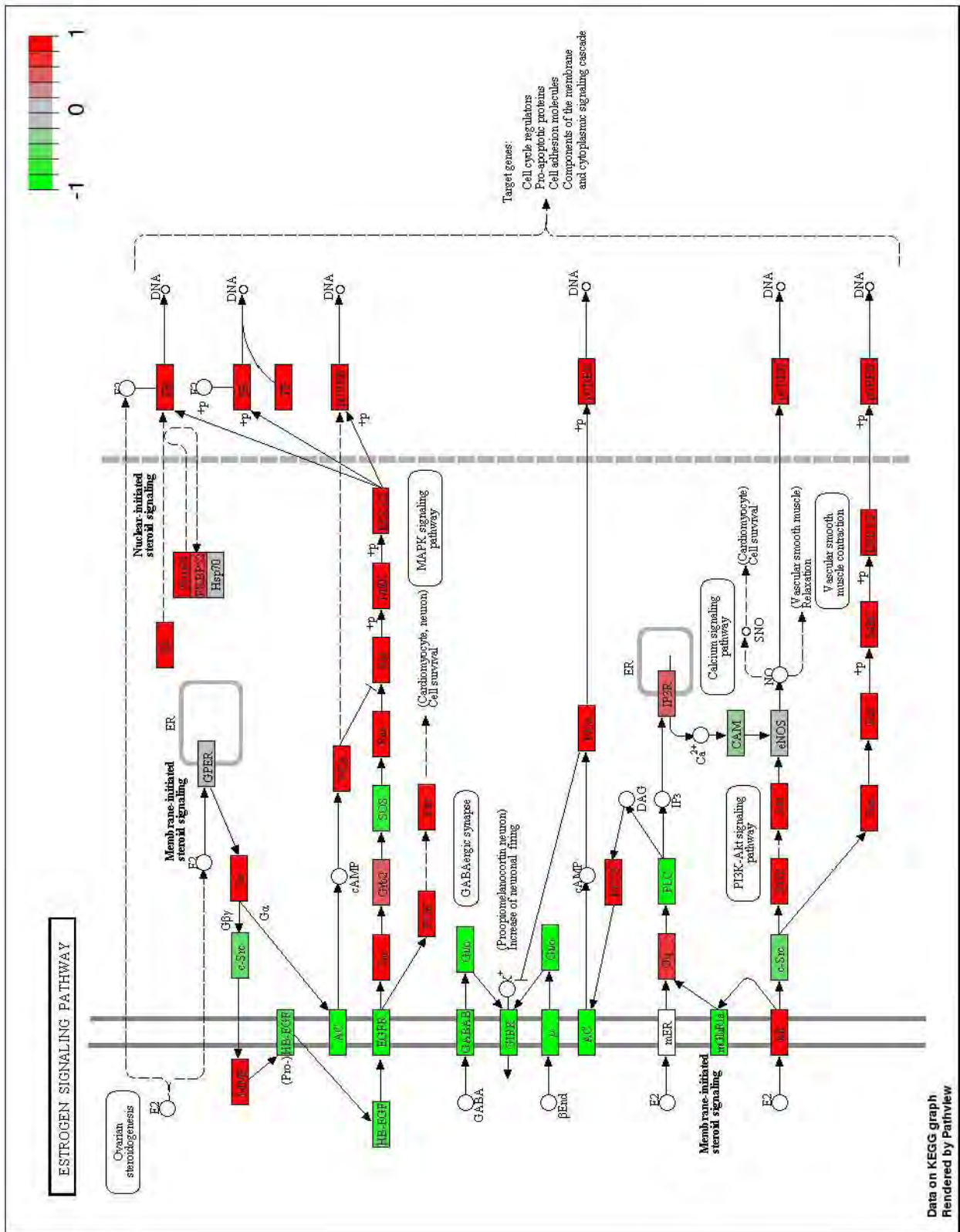


Figura 9-4: Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo luminal A. Verde es subexpresión, rojo es sobreexpresión con respecto a tejido mamario sano.

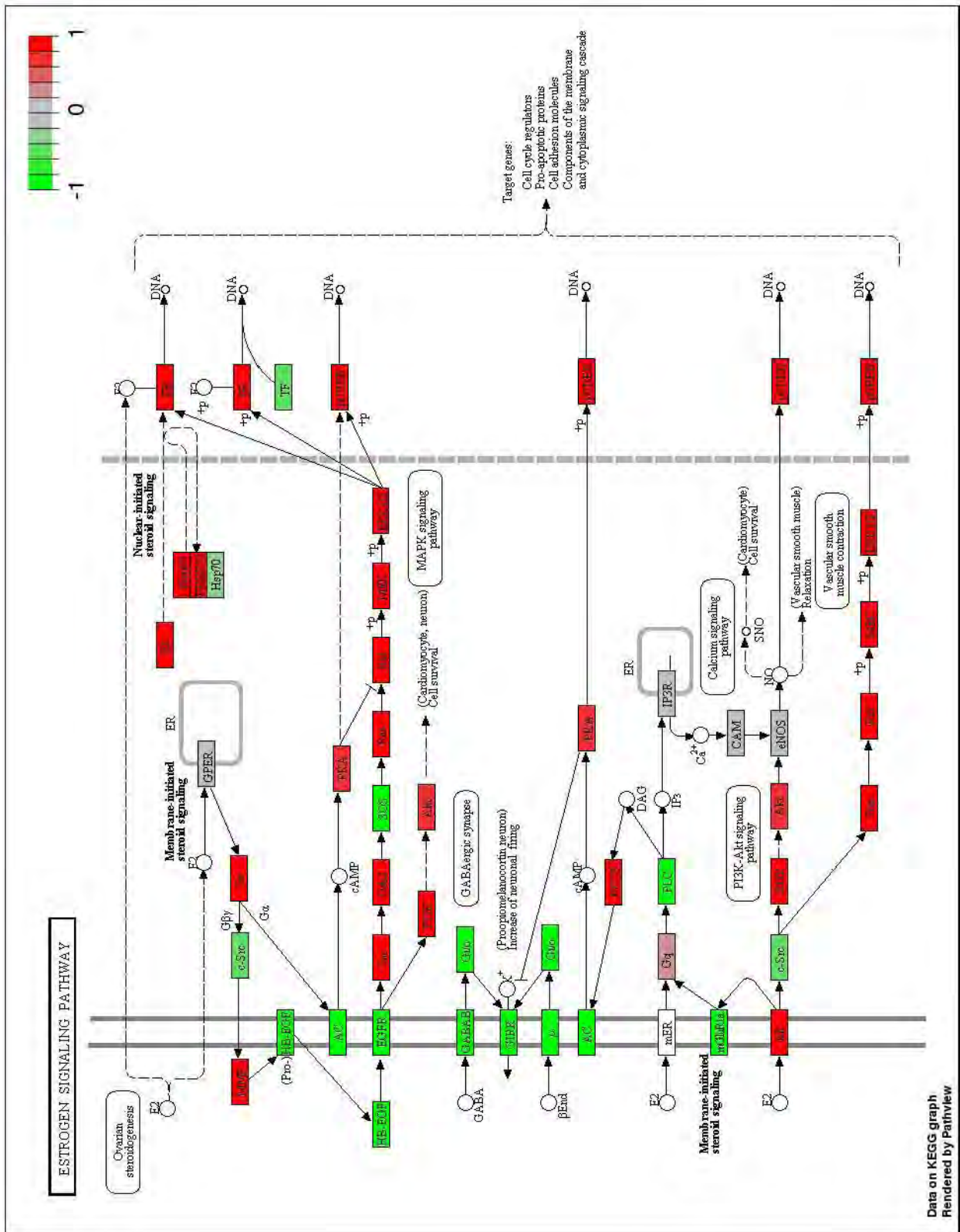


Figura 9-5: Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo luminal B. Verde es subexpresión, rojo es sobreexpresión con respecto a tejido mamario sano.

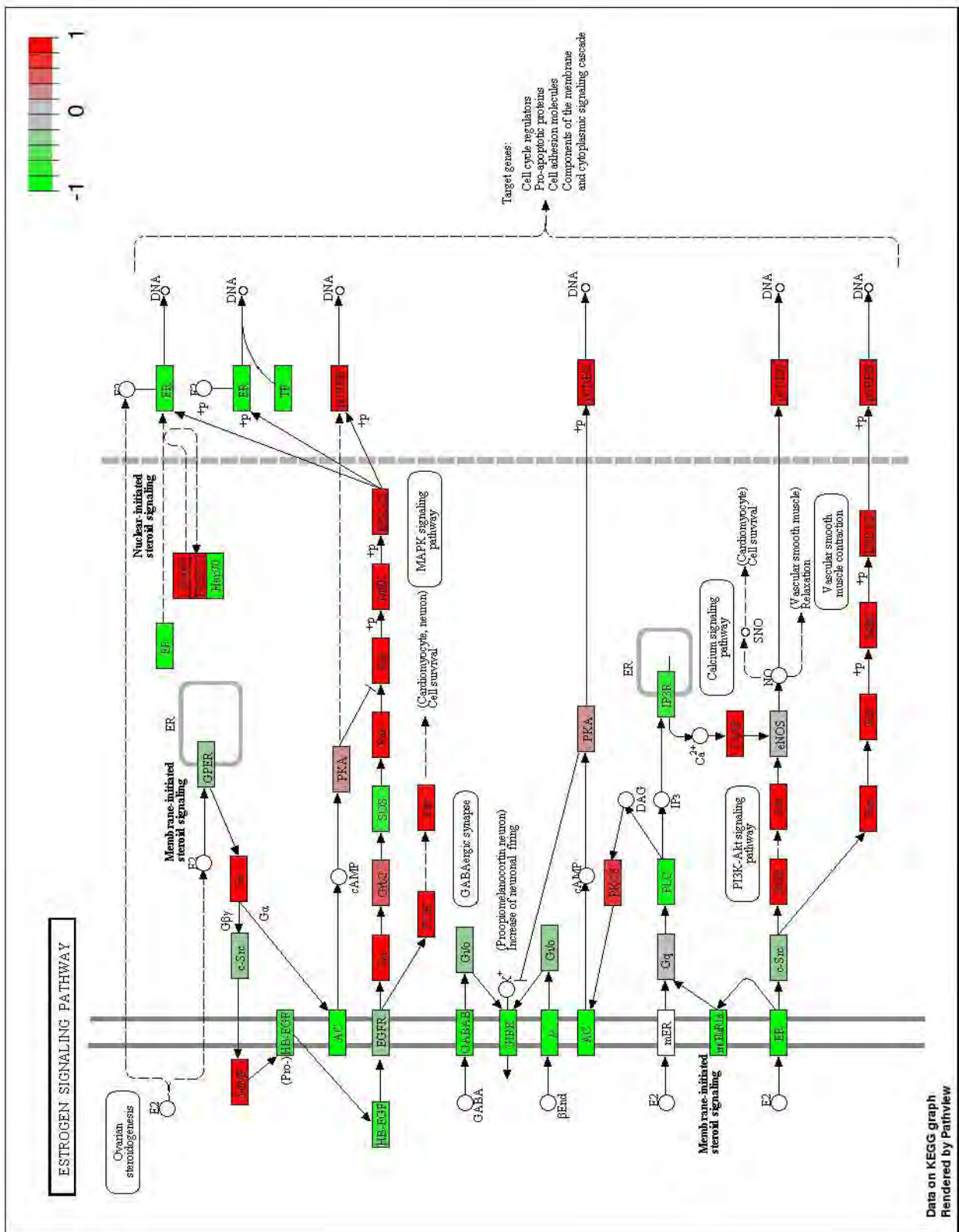


Figura 9-6: Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo basal. Verde es subexpresión, rojo es sobreexpresión con respecto a tejido mamario sano.

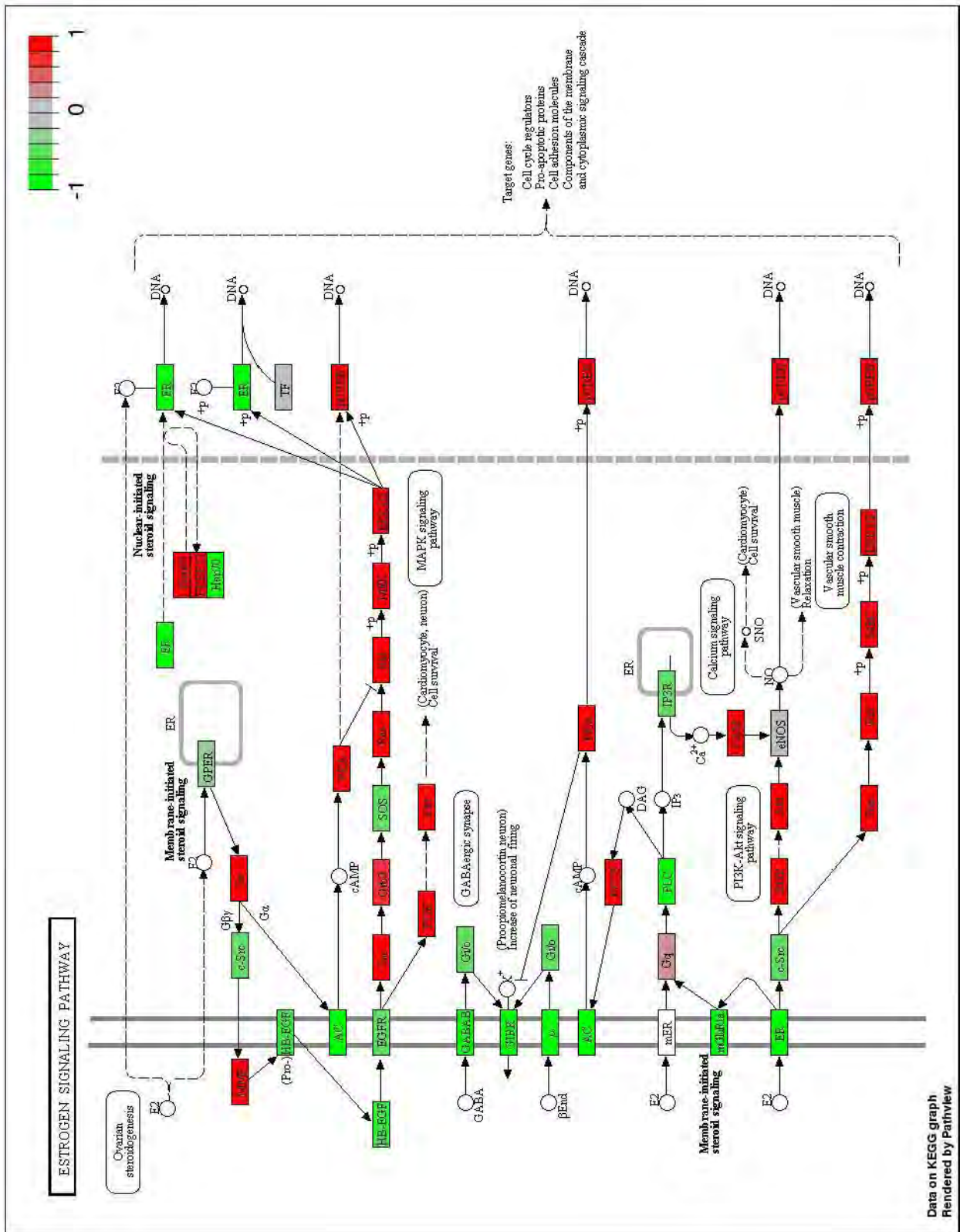


Figura 9-7: Desregulación de expresión génica en la vía de señalización de estrógeno en el subtipo HER2-enriquecido. Verde es subexpresión, rojo es sobreexpresión con respecto a tejido mamario sano.

9.2.1. Red de entrecruzamiento de estrógeno

Se identificaron las vías que entrecruzan con la vía de señalización de estrógeno, y que por esta comunicación pueden tener una influencia sobre la vía de señalización de estrógeno (siendo una posible fuente de resistencia a su bloqueo) o bien, ser afectadas por la actividad estrogénica. En la figura 9-8 podemos observar una representación gráfica de dicho entrecruzamiento de vías.

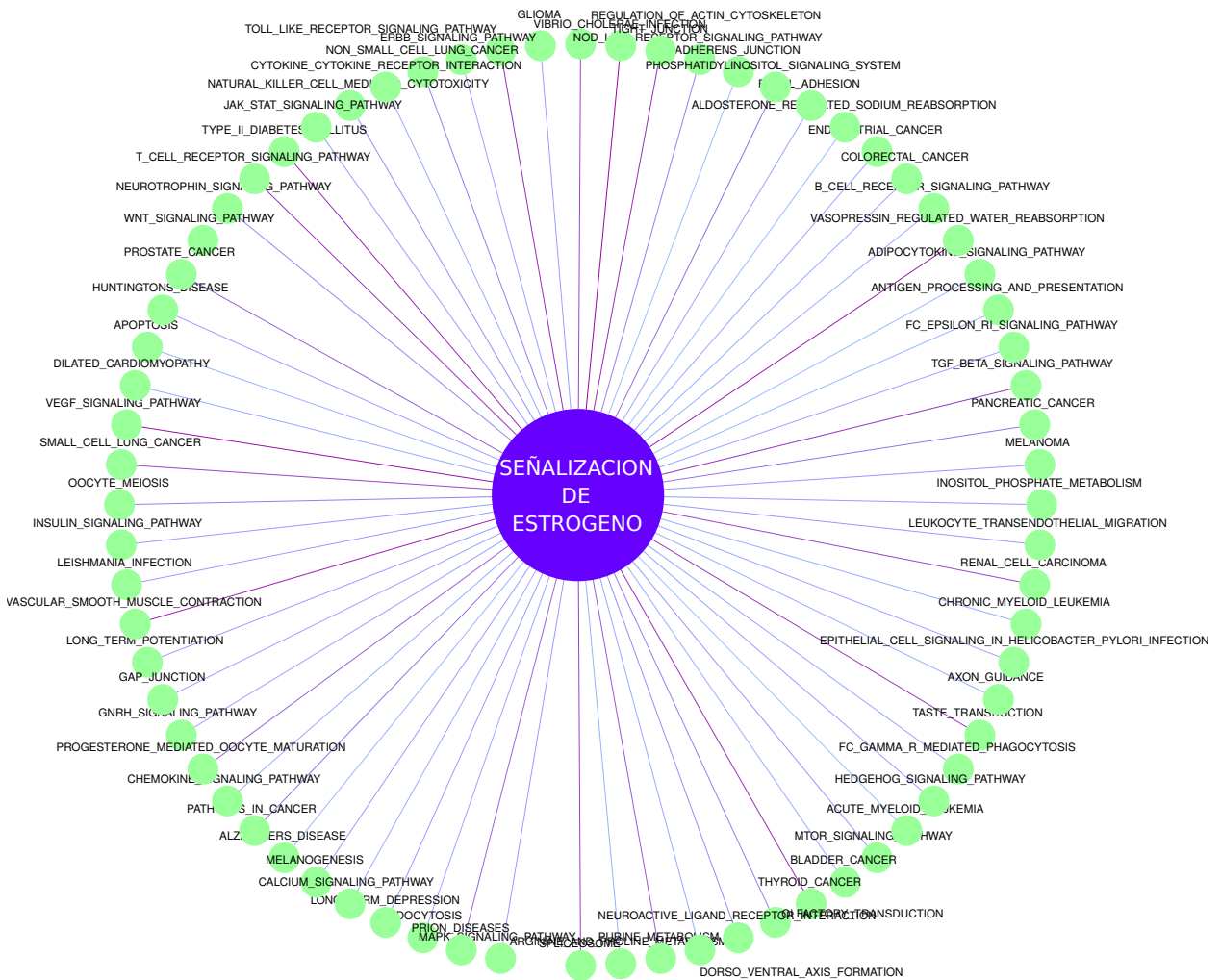


Figura 9-8: Identificación de las vías que entrecruzan con la vía de señalización de estrógeno.

Por su parte, la figura 9-9 muestra la comunicación que se presenta entre las vías que entrecruzan con la vía de señalización de estrógeno. Se puede apreciar que existe una alta conectividad entre dichas vías. Dicha alta conectividad resulta en un sistema redundante, lo que confiere robustez al sistema. Dicho de otra forma, aún si removieramos alguna de las vías en la red, la red continuaría conectada, y seguirían existiendo caminos que permiten llegar de una vía determinada a otra.

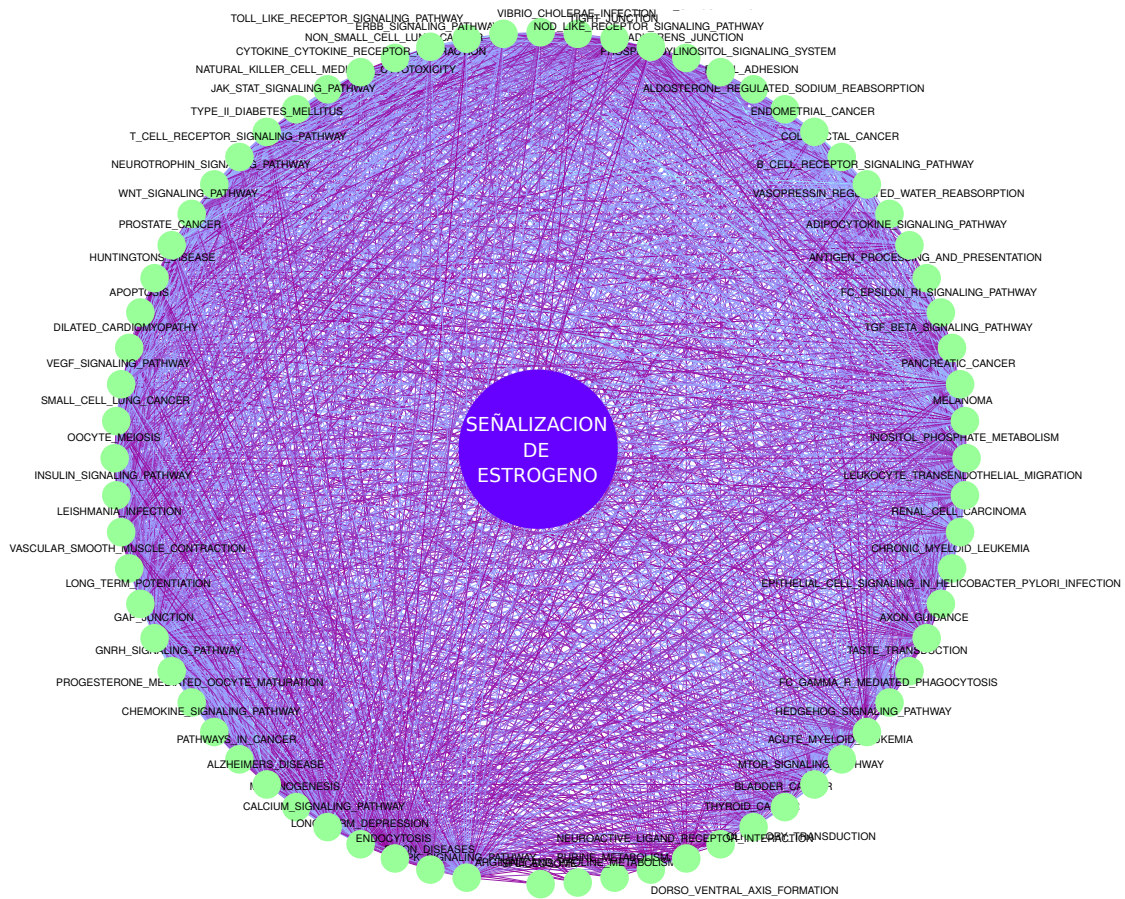


Figura 9-9: Identificación de las vías que entrecruzan con la vía de señalización de estrógeno.

9.2.2. Desregulación de vías entrecruzadas

Al estudiarse las vías que entrecruzan con la señalización de estrógeno, podemos observar que existe una desregulación extendida en gran cantidad de estas vías, en todos los subtipos moleculares. De hecho, todas las vías que entrecruzan con la vía de señalización de estrógeno están desreguladas en al menos uno de los subtipos moleculares de cáncer de mama estudiados. Esta extensa desregulación puede ser una explicación a la observación previa de la desregulación de la vía de señalización de estrógeno: la activación de las moléculas de la vía se genera gracias a la desregulación de las otras vías entrecruzadas. Esta observación, en conjunto con la estructura altamente conectada de la red, hace posible pensar en el fenómeno de *resistencias farmacológicas por compensación de vías entrecruzadas* que puede estar actuando en el caso del tratamiento antiestrogénico, al presentarse activaciones de la vía de señalización de estrógeno independientes al ligando (VanHook, 2010), (Carascossa *et al.*, 2010).

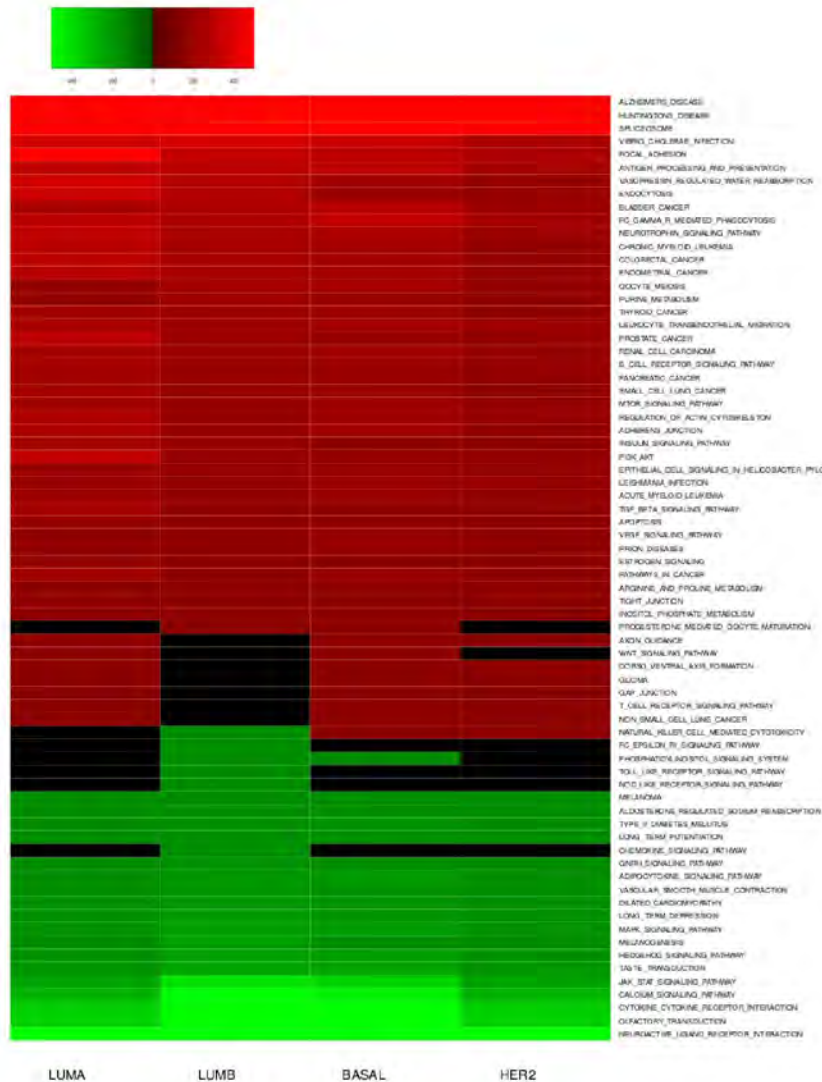


Figura 9-10: Patrones de desregulación de vías entrecruzadas con la vía de señalización de estrógeno.

A continuación se discuten algunas de las implicaciones del entrecruzamiento de la vía de señalización de estrógeno con vías implicadas en actividades farmacológicas anticancerígenas: apoptosis, y respuesta inmune.

Apoptosis

Se observó una desregulación de la vía de apoptosis en todos los subtipos moleculares de cáncer de mama. Esto era de esperarse, ya que la evasión de la apoptosis es uno de

los *hallmarks* del cáncer (Hanahan y Weinberg, 2000).

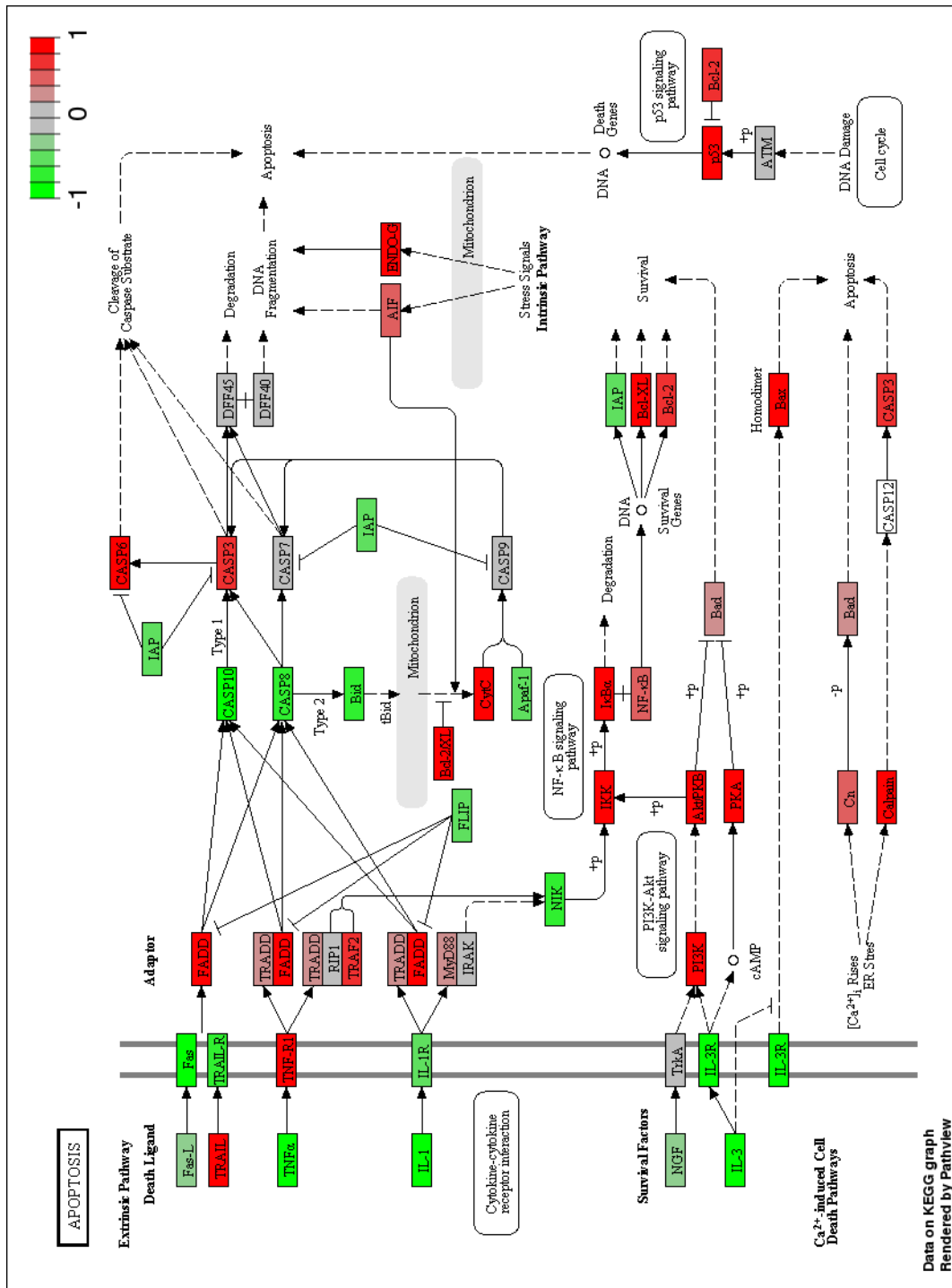


Figura 9-11: Desregulación de expresión génica en la vía de apoptosis en el subtipo luminal A. Genes en rojo están sobreexpresados, genes en verde están subexpresados

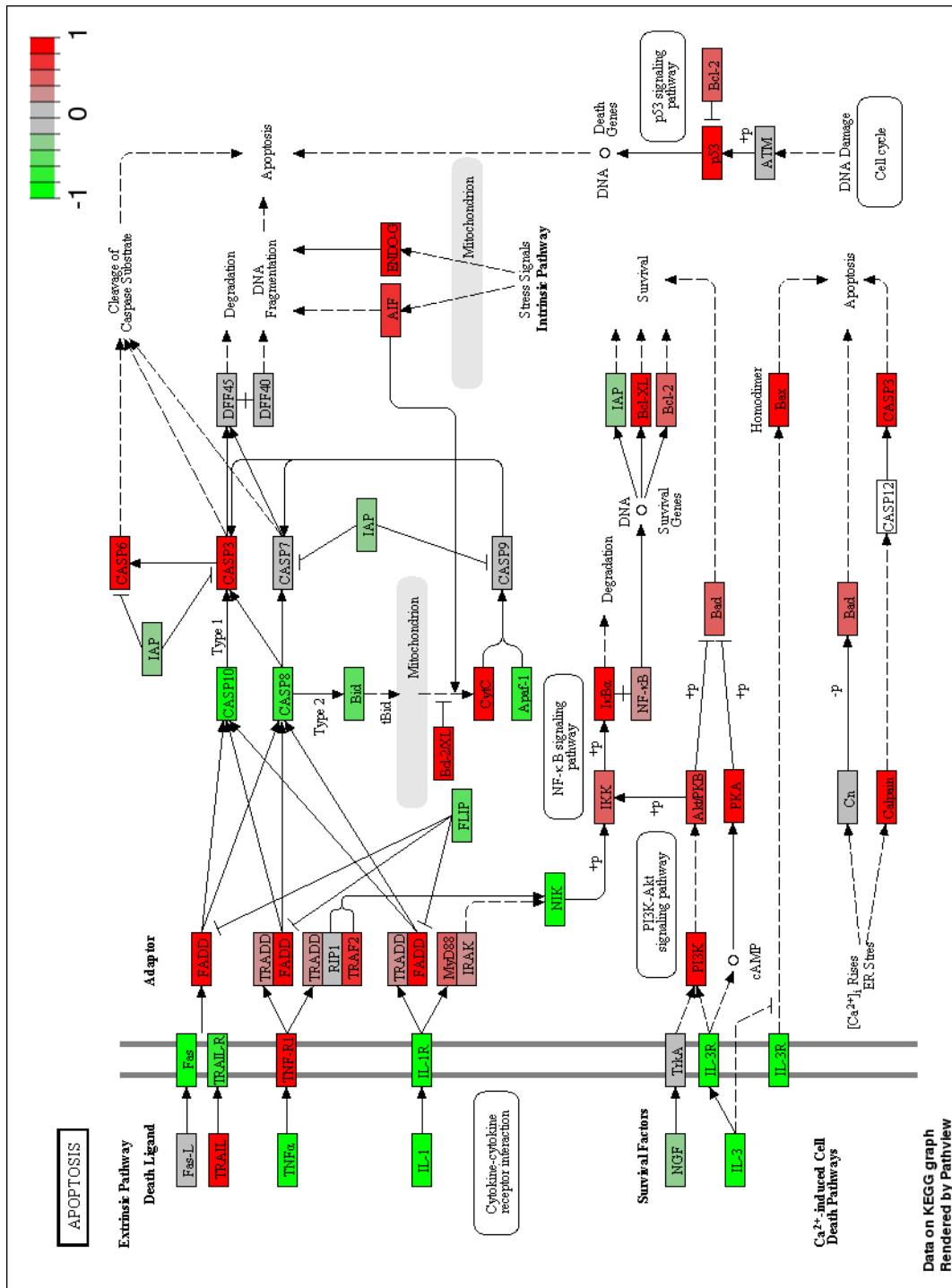


Figura 9-12: Desregulación de expresión génica en la vía de apoptosis en el subtipo luminal B. Genes en rojo están sobreexpresados, genes en verde están subexpresados

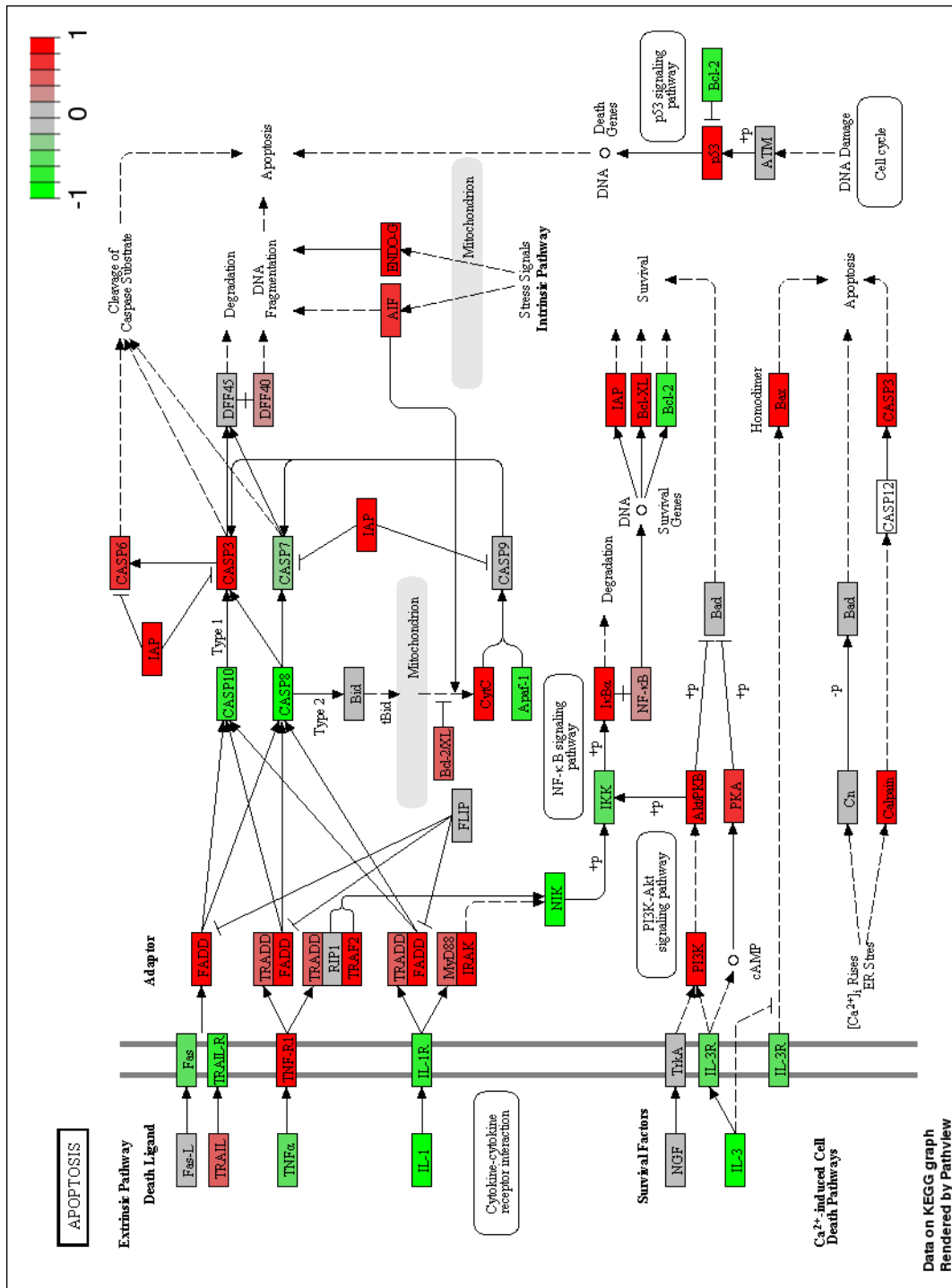


Figura 9-13: Desregulación de expresión génica en la vía de apoptosis en el subtipo basal. Genes en rojo están sobreexpresados, genes en verde están subexpresados

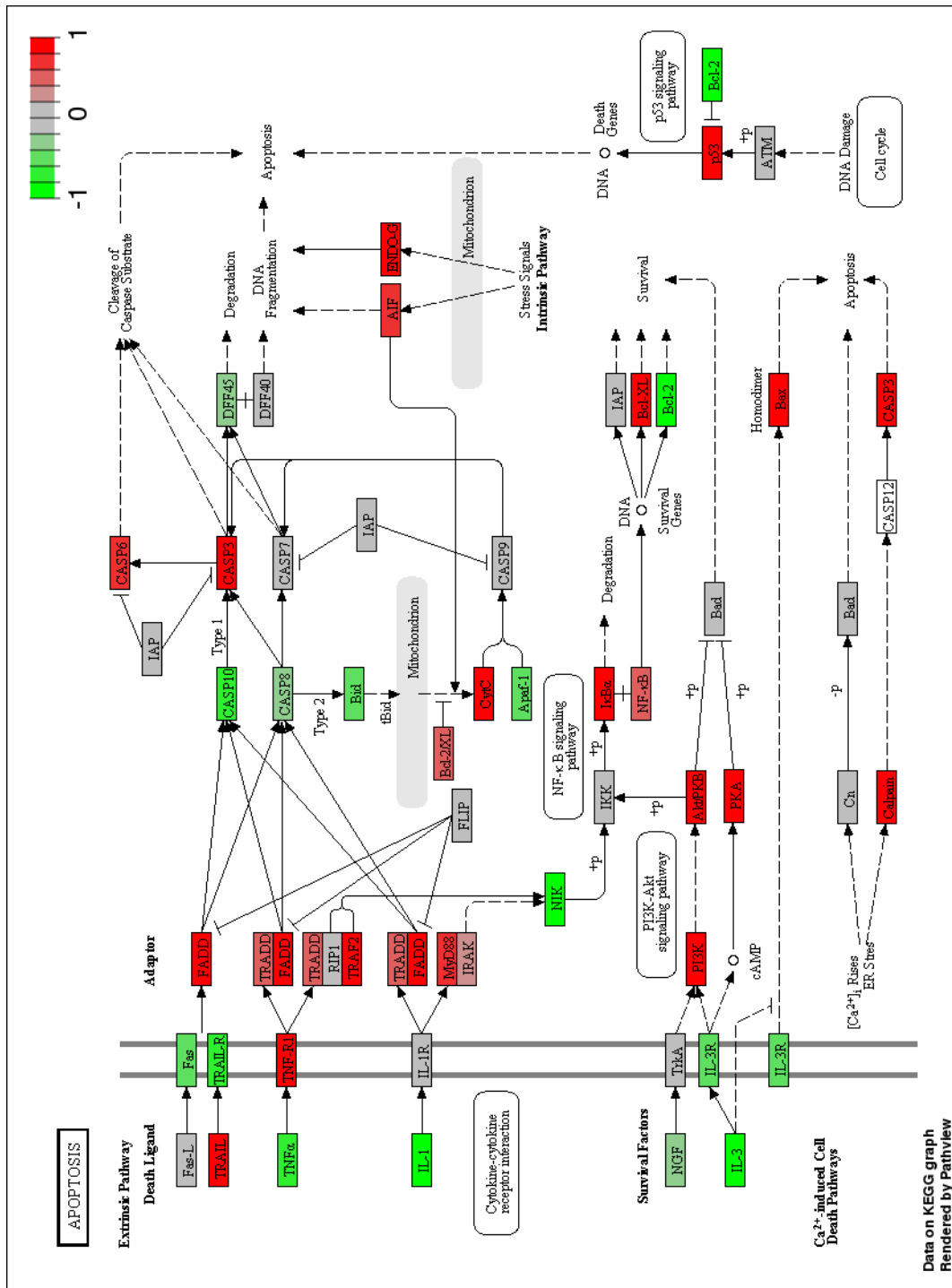


Figura 9-14: Desregulación de expresión génica en la vía de apoptosis en el subtipo HER2-enriquecido. Genes en rojo están sobreexpresados, genes en verde están subexpresados

El bloqueo de la señal estrogénica permite limitar los efectos de supervivencia y proliferación de ésta. De igual forma, se ha reportado que algunos antiestrógenos (como tamoxifen) pueden actuar como inductores de apoptosis (Mandlekar y Kong, 2001; Liu *et al.*, 2014). De esta manera, encontramos que el entrecruzamiento de vías entre señalización de estrógeno y apoptosis tiene relevancia terapéutica.

La expresión de dos moléculas de la familia Bcl: Bcl-2, una antiapoptótica, y Bcl-XL, una molécula de supervivencia, en los distintos subtipos moleculares, llamó la atención. Bcl-XL se encuentra sobreexpresada en todos los subtipos moleculares, mientras que Bcl-2 se encuentra sobreexpresada en los fenotipos luminales, y subexpresada en los no luminales. Se ha discutido la importancia de una ruta asociada a Bcl-2 (Clarke *et al.*, 2012) que conecta los procesos de autofagia y apoptosis. Existe evidencia experimental de que la inhibición farmacológica de Bcl-2 permite recuperar la sensibilidad a tratamiento antiestrogénico en células que han desarrollado resistencia.

Una de las señales de regulación de apoptosis es la concentración de calcio intracelular. Uno de los puntos de entrecruzamiento entre las vías de apoptosis y señalización de estrógeno es precisamente la señalización por calcio (Swami, 2010; Chen *et al.*, 2013). Nuestro análisis de enriquecimiento de vías muestra que la vía de señalización de calcio se encuentra subexpresada en todos los subtipos moleculares. Identificamos la modulación de la señalización de calcio como un potencial blanco farmacológico para evitar la resistencia debida al entrecruzamiento. Esta es una línea que muy recientemente se ha explorado (Azimi *et al.*, 2014).

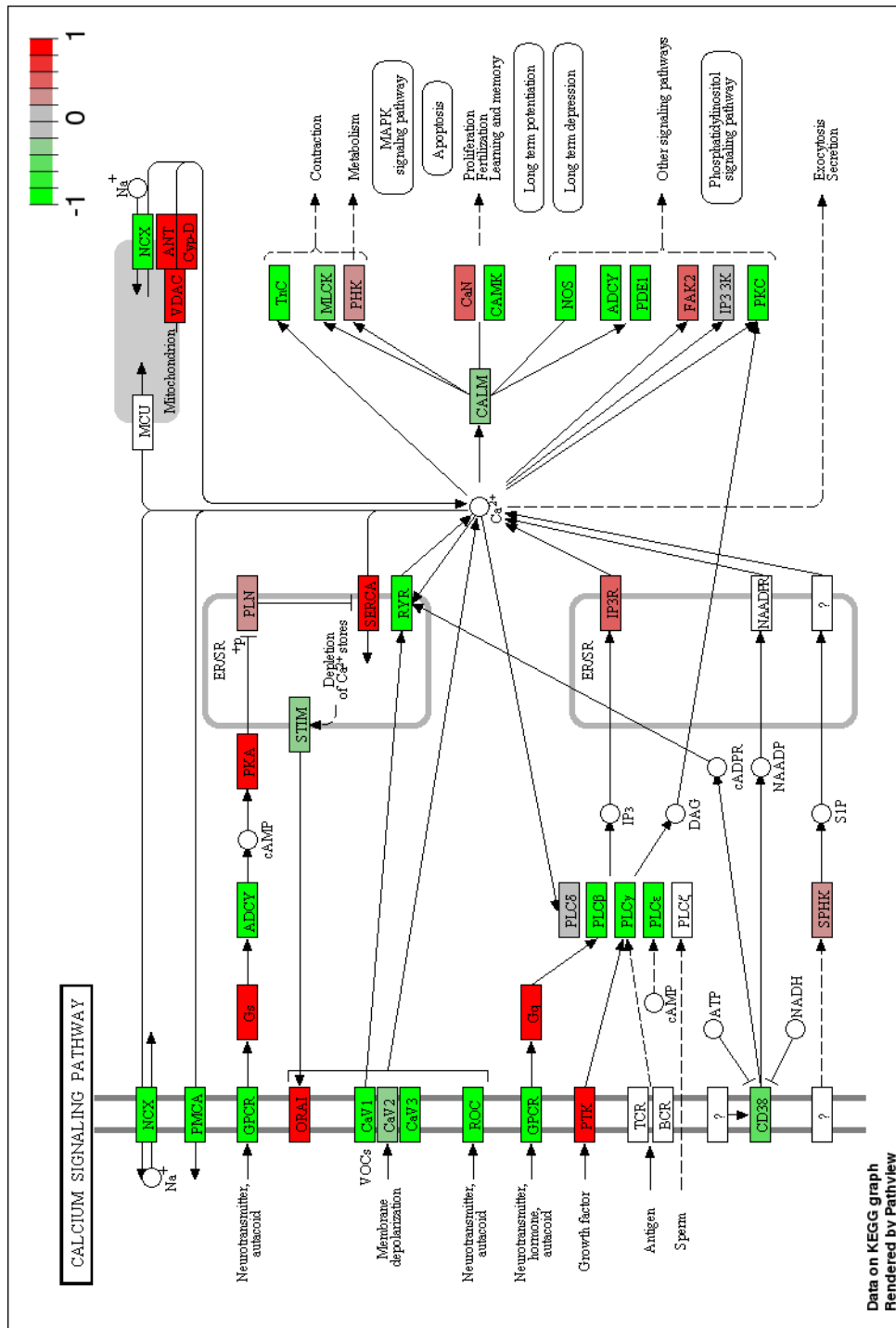


Figura 9-15: Desregulación de expresión génica en la vía de señalización de calcio en el subtipo luminal A. Genes en rojo están sobreexpresados, genes en verde están subexpresados

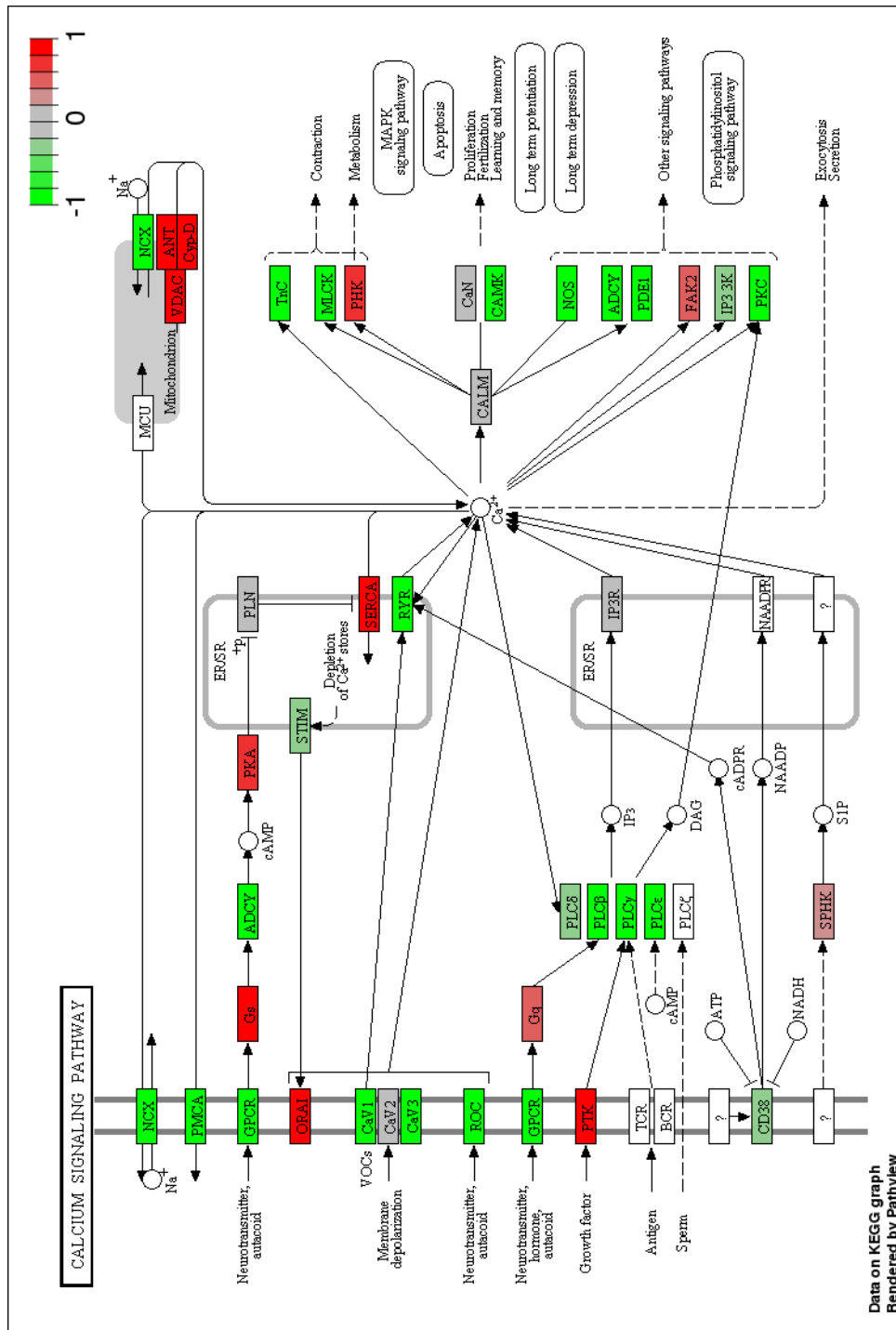


Figura 9-16: Desregulación de expresión génica en la vía de señalización de calcio en el subtipo luminal B. Genes en rojo están sobreexpresados, genes en verde están subexpresados

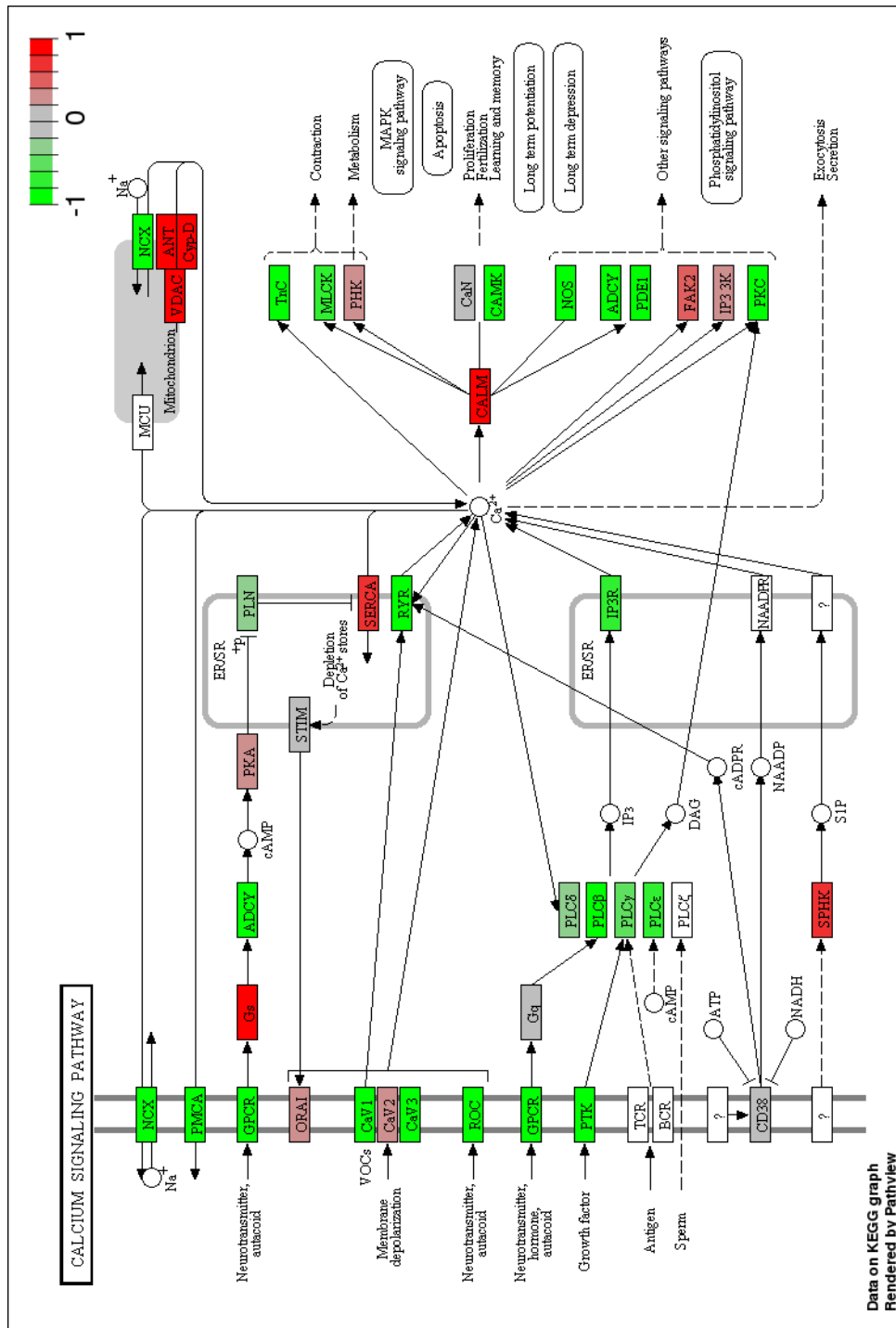


Figura 9-17: Desregulación de expresión génica en la vía de señalización de calcio en el subtipo basal. Genes en rojo están sobreexpresados, genes en verde están subexpresados

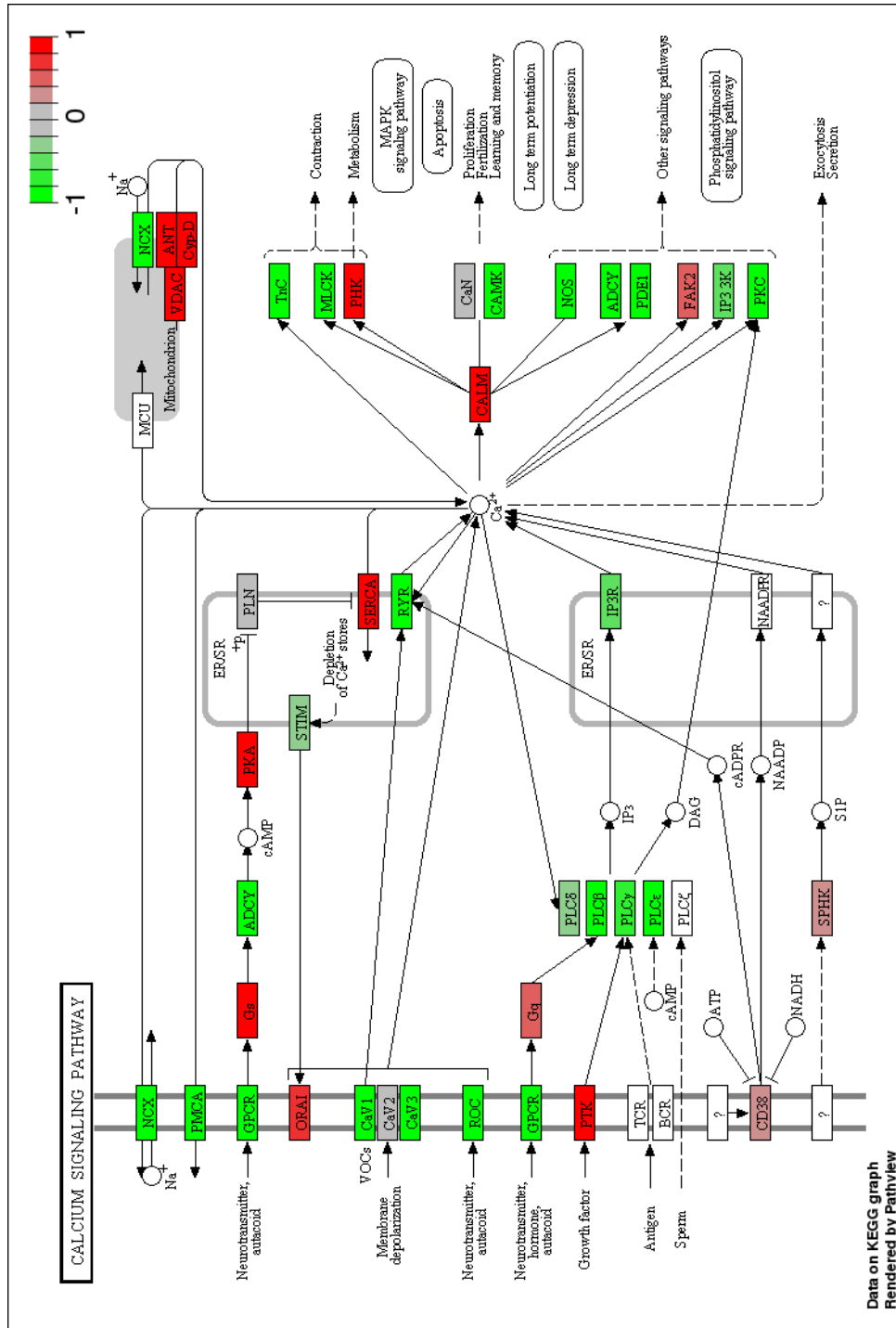


Figura 9-18: Desregulación de expresión génica en la vía de señalización de calcio en el subtipo HER2-enriquecido. Genes en rojo están sobreexpresados, genes en verde están subexpresados

Vías de inmunidad:

La evasión de los mecanismos de inmunidad ha sido integrada a los *hallmarks* del cáncer (Hanahan y Weinberg, 2011). Por esta razón, las implicaciones de su desregulación son relevantes en el contexto del tratamiento antiestrogénico. El entrecruzamiento de la vía de señalización de estrógeno con las vías relacionadas a mecanismos inmunes se da a través de moléculas generales de transducción de señales, como son la señalización por MAPK, PI3K, JAK-STAT y calcio, etc. La figura 9-19 ilustra este fenómeno.

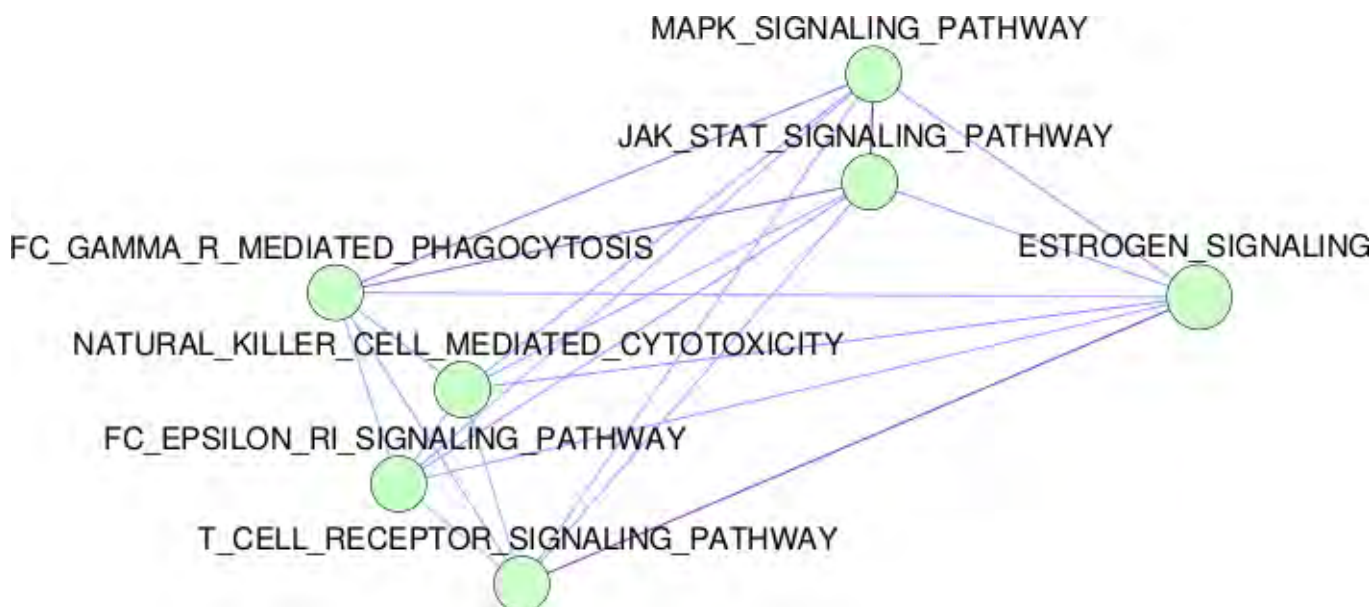


Figura 9-19: Las vías de inmunidad que entrecruzan con estrógeno lo hacen a través de vías generales de señalización

Llama la atención que existe una desregulación de vías de inmunidad en todos los subtipos, pero cada subtipo presentó un perfil de enriquecimiento distinto. El perfil más distinto fue el de el subtipo luminal B, siendo el único en el que tanto la vía de citotoxicidad mediada por células NK como la vía de señalización de quimiocinas se encuentra subexpresada, y el único subtipo en el que la vía de señalización del receptor de células T no está desregulada. Existe evidencia de efectos inmunomoduladores de los tratamientos antiestrogénicos (Behjati y Frank, 2009). Los resultados aquí presentados muestran una posible contribución del fenómeno de entrecruzamiento a la inmunomodulación, y la po-

sibilidad de diferencias en estos efectos entre subtipos moleculares. El entrecruzamiento a través de transductores compartidos abre la posibilidad a un fenómeno de retroalimentación entre las vías de inmunidad y la señalización de estrógeno. Datos recientes (Cook *et al.*, 2014) concuerdan con esta idea. En dicho trabajo, el tratamiento antiestrogénico induce un flujo autofágico, que a su vez parece inducir la resistencia al tratamiento antiestrogénico. A su vez, la acción estrogénica sobre células inmunes (macrófagos y monocitos) da lugar a modificaciones en los procesos inmunes. Esta evidencia experimental es congruente con la importancia que se propone en el entrecruzamiento de señalización de estrógeno e inmunidad.

Con base en estos resultados se puede observar que la vía de señalización de estrógeno no está aislada, sino que con ella entrecruza un número importante de vías formando una red altamente conectada, involucrados en múltiples procesos biológicos y por lo tanto posiblemente involucrados en la naturaleza de la patología.

Capítulo 10

Conclusiones

El presente trabajo muestra un panorama general de las diferencias y similitudes entre distintos subtipos moleculares de mama, así como las diferencias y similitudes que existen en relación al tejido mamario sano.

Este trabajo propone la implementación de metodologías que pueden ser útiles para la *generación de hipótesis* que deberán ser demostradas de manera experimental. Las exploraciones aquí presentadas sirven para dar una nueva perspectiva sobre el cáncer de mama, su heterogeneidad y las implicaciones sobre el tratamiento farmacológico. En ese sentido, a continuación se discuten algunas observaciones generales, y se presentan, a manera de *prueba de concepto*, algunos ejemplos particulares que, a partir de los resultados obtenidos y evidencia documental, pueden dar lugar a líneas de investigación de interés.

Existen diferencias entre el cáncer y el tejido mamario sano

En los múltiples análisis realizados, encontramos que existen fuertes diferencias entre el tejido mamario sano, y el tejido mamario canceroso.

Las diferencias en niveles de expresión génica entre tejido sano y tejido enfermo han sido exploradas desde hace tiempo, particularmente en el contexto de cáncer (Zhang *et al.*,

1997), identificándose como uno de los orígenes de la enfermedad (Wang *et al.*, 2015). En este trabajo se identificaron no menos de 5900 genes diferencialmente expresados entre el tejido mamario sano y el tejido canceroso, confirmando que la desregulación en cáncer de mama no se limita a pocos genes, sino que se trata de un fenómeno extenso.

La desregulación de la expresión de genes se puede asociar a una desregulación en diferentes procesos biológicos. Como se ha mencionado antes, se han asociado muchos procesos desregulados al cáncer (Giancotti, 2014; Hanahan y Weinberg, 2011, 2000). En este trabajo se identificó que la desregulación de procesos en cáncer de mama, a nivel transcripcional, no se encuentra limitada a un número de procesos reducido, y que la mayoría de los procesos biológicos conocidos se encuentran alterados en el fenotipo patológico.

La arquitectura de redes transcripcionales presenta otro nivel de diferencia entre el tejido mamario sano y el tejido mamario canceroso. La red de tejido mamario sano está compuesta comparte muy pocos genes con las redes de tejido canceroso. Más llamativo aún es el hecho de que la red de tejido mamario sano se compone predominantemente de un único componente de gran tamaño, altamente conectado, mientras que las redes transcripcionales de tejido canceroso se componen de múltiples componentes. Aparentemente en el fenotipo sano, existe una regulación tal que da lugar a redes altamente conectadas, fenómeno que se pierde ante un proceso patológico. Una de las posibles implicaciones de esta observación es la independización de procesos: la desconexión de las redes podría ser el origen de algunos de los *hallmarks* del cáncer, como la evasión de las señales de regulación tanto positivas como negativas.

Podemos observar que las diferencias entre el tejido mamario sano y el tejido canceroso se dan a diferentes niveles y de una manera muy extendida. En el presente trabajo utilizamos información transcripcional primordialmente, pero se puede extender lo aquí presentado a otros tipos de datos, como información genotípica o proteómica, de contarse con la tecnología. Se esperaría que el nivel de complejidad aumentara. Por un lado, esto puede parecer un problema, pues existen muchos factores que contribuyen a la enferme-

dad. Pero por otro lado, esta misma diversidad de factores puede ser una ventaja desde el punto de vista de investigación biomédica: existen muchos factores que pueden ser objetos de estudio para la búsqueda de alternativas en la terapia.

Los subtipos moleculares de cáncer de mama presentan similitudes y diferencias

El presente trabajo mostró que la heterogeneidad del cáncer de mama puede ser identificada a diferentes niveles.

Una primera observación es que fue posible implementar estrategias de clasificación de muestras de cáncer de mama, a partir de datos de expresión, útiles para comparaciones entre fenotipos. Es importante observar que las estrategias de clasificación actuales están pensadas con un enfoque clínico: toda muestra analizada, será asociada a una categoría. La estrategia de clasificación propuesta e implementada en este trabajo está orientada a la comparación de fenotipos. En ese sentido, al utilizar clasificadores independientes y conservar únicamente aquellas muestras donde los clasificadores coinciden, se obtuvieron conjuntos bien diferenciados pero homogéneos dentro de los grupos.

Dado que la clasificación en subtipos moleculares está basada en diferencias a nivel de expresión de genes, era de esperarse que encontraríamos diferencias transcripcionales. Uno de los resultados más llamativos del estudio fue encontrar no solo una amplia desregulación de genes entre subtipos cancerosos y tejido sano, sino que la mayoría de los genes desregulados se encuentran desregulados de manera similar en *todos* los subtipos moleculares. La posible implicación es que estos genes participen en procesos centrales cuya desregulación da lugar al fenómeno canceroso. Algo interesante es que si bien estos genes pueden estar sobre o subexpresados en todos los subtipos, no necesariamente lo están en la misma magnitud. Un ejemplo en este trabajo es el gen ERBB2, responsable del receptor HER2. Tradicionalmente (y de manera natural) se asocia al receptor HER2 con el subtipo molecular HER2-enriquecido. Aquí encontramos que ERBB2 se encuentra so-

breexpresado en los cuatro subtipos moleculares; Sin embargo, la expresión de ERBB2 en el subtipo HER2-enriquecido es mucho mayor que en cualquiera de los otros subtipos; es decir, la sobreexpresión de ERBB2 es *más característica* del subtipo HER2-enriquecido. Existen algunos genes que se encuentran desregulados en diferentes subtipos de manera distinta. Un ejemplo es el receptor de estrógeno ESR1, sobreexpresado en los subtipos luminales pero subexpresado en los subtipos no-basales.

Se identificaron también genes que están expresados diferencialmente sólo en algún subtipo. Por ejemplo, se identificó la sobreexpresión de CCL2 en el subtipo basal. La secreción de CCL2 por parte del tejido mamario da lugar al reclutamiento de monocitos inflamatorios (Bonapace *et al.*, 2014). Hay reportes que identifican desregulación de CCL2 en el subtipo luminal B (Fang *et al.*, 2015), y se ha propuesto su modulación para el tratamiento de tumores de mama dependientes de estrógeno (Svensson *et al.*, 2015). Los resultados de este trabajo apuntan a que el tratamiento con moduladores de CCL2 sería más apropiado para tumores basales. Por ello, se propone evaluar dicho efecto en un modelo experimental.

El análisis de desregulación a nivel de vías muestra que la desregulación de una gran cantidad de vías no es exclusiva de un subtipo específico; por el contrario, la mayor parte de las vías analizadas se encuentra desregulada en cada subtipo.

La mayoría de las vías desreguladas identificadas se encuentran desreguladas de la misma forma en todos los subtipos moleculares. Es interesante observar que muchas de estas vías se asocian a procesos relacionados con crecimiento celular, señalización, supervivencia, y respuesta inmune: procesos que hacen referencia a los ya citados *hallmarks* del cáncer. El enriquecimiento de vías remarca que existe una serie de desregulaciones que son comunes a distintas manifestaciones cancerosas, como pueden ser distintos subtipos moleculares de cáncer de mama.

Por otro lado, es interesante notar que cada uno de los subtipos cuenta con un *perfil de enriquecimiento* característico; es decir, es posible recuperar las diferencias fenotípicas asociadas a los subtipos moleculares, más allá del nivel de desregulación transcripcional,

en el nivel de procesos. Se han desarrollado algoritmos clasificadores basados en *pathways*, por ejemplo *Pathifier* (Drier *et al.*, 2013). En un trabajo centrado en cáncer de mama (Livshits *et al.*, 2015), se identificaron con este algoritmo nuevos subtipos basados en vías, aunque su relevancia clínica no ha sido validada.

La construcción de redes transcripcionales permite un acercamiento distinto al estudio de las diferencias entre subtipos. Es interesante observar como las redes, construidas con una estrategia que hace factible su comparabilidad, están compuestas por transcritos distintos. Esta observación muestra que existen diferencias en el rol de diferentes genes a nivel de la arquitectura transcripcional asociada a los fenotipos.

La medida de *relevancia* de los genes dentro de la red más inmediata es el *grado*, el número de vecinos de un nodo en la red; en este caso, el número de genes a los que se asocia un gen particular en función de la dependencia estadística en su expresión, medida como información mutua.

Encontramos que algunos de los genes en las redes transcripcionales de fenotipos cancerosos tienen un valor de grado alto en todos los subtipos moleculares. Un ejemplo es CNR2. CNR2 codifica el receptor a canabinoides tipo 2 (CB2), asociado a efectos inmunomoduladores de endocannabinoides (Munro *et al.*, 1993). Se han identificado alteraciones en CB2 en distintos cánceres (Pisanti *et al.*, 2013; Jha *et al.*, 2012; Guida *et al.*, 2010), incluyendo mama (Pérez-Gómez *et al.*, 2015; Nasser *et al.*, 2011; Sophocleous *et al.*, 2015). Los resultados del presente trabajo identificaron a CNR2 entre los 30 nodos de mayor grado en los cuatro subtipos cancerosos. Esto contribuye a señalar a CNR2 como un gen de importancia en los diversos subtipos de cáncer de mama. Resulta interesante además, observar que sí bien CNR2 tiene un número alto de vecinos en todos los subtipos, no se trata de los mismos vecinos; de esta manera, se puede pensar que aunque CNR2 juega un papel importante en las redes transcripcionales de cáncer de mama, lo hace interactuando con diferentes moléculas, dando lugar así a diferencias fenotípicas entre los subtipos.

Existen algunos genes encontrados en las redes transcripcionales que difieren en grado, lo que lleva a pensar que tienen un papel más preponderante en la arquitectura transcrip-

cional de algunos subtipos que de otros. Un ejemplo de gen en esta categoría es LUZP4. LUZP4 codifica una proteína de tipo *leucin zipper*. Se ha identificado que esta proteína está involucrada en la exportación de mRNA (Viphakone *et al.*, 2015). Los resultados del análisis de las redes transcripcionales muestran que este transcrito tiene un valor de grado alto en los subtipos luminales y en el subtipo basal, con más de cien vecinos en cada uno de estos subtipos, mientras que tiene un número reducido en el subtipo HER2-enriquecido. Dada su función, se podría considerar que LUZP4 está involucrado en la exportación de mRNAs importantes para los subtipos luminales y el basal, pero resulta menos relevante en el fenotipo HER2-enriquecido.

También se identificaron genes que solo participan en la red de un subtipo particular. Ejemplo de este caso es LCK. LCK codifica la proteína tirosina cinasa linfocito-específica. Esta proteína está involucrada en transducción de señales. Se encontró LCK participando únicamente en la red transcripcional del subtipo HER2-enriquecido. Existen reportes de la utilidad pronóstica de una firma molecular asociada a LCK en tumores de mama HER2-enriquecidos (Rody *et al.*, 2009). La red transcripcional en el presente trabajo parece dar sustento a esta observación, ya que identifica un papel importante de LCK en la arquitectura de red.

Se pueden explotar las diferencias entre subtipos para buscar opciones terapéuticas

Como se ha expresado previamente, la diversidad de elementos involucrados en el cáncer de mama en el presente trabajo podría llevar a pensar que el diseño de estrategias para su tratamiento es un problema poco resoluble. Sin embargo, el tratar de *integrar* y *contextualizar* dichos elementos permite otro tipo de acercamientos a dicho problema. En ese sentido, las exploraciones realizadas en el presente trabajo permiten identificar

diferentes potenciales blancos terapéuticos con características diversas.

La diversidad del cáncer de mama da lugar a la búsqueda de opciones terapéuticas que pueden ir desde las de *amplio espectro* hasta opciones muy específicas para fenotipos particulares. Diferentes abordajes para la exploración *sistemática* de dichas opciones terapéuticas pueden llevar a identificar potenciales blancos relevantes a diferentes niveles. El nivel de exploración de la expresión diferencial es posiblemente el más natural para iniciar una búsqueda; la comparación de diferencias en niveles de expresión entre casos y controles es un diseño experimental muy evidente, y una estrategia clásica de búsqueda de fármacos. Los tratamientos dirigidos actuales, como el trastuzumab para los tumores de mama que sobreexpresan HER2 o el tratamiento antiestrogénico para los estrógeno-positivos surgen de esta estrategia.

Los resultados del presente trabajo respaldan la lógica detrás de estas terapias: los tumores luminales fueron identificados como estrógeno-positivos, por lo que son sensibles al bloqueo de la señal de estrógeno. Por su parte, identificamos que el receptor HER2 es mucho más característico del subtipo HER2-enriquecido, por lo que atacarlo resultará útil en este subtipo; los resultados obtenidos apuntan, además, a un posible uso de la terapia anti-HER2 en otros subtipos moleculares.

A partir de los resultados de este trabajo se pueden identificar potenciales sitios de acción farmacológica, con esta misma estrategia. Pero recordar que el presente trabajo ha partido de la idea de una exploración *integrativa*, de *biología de sistemas* del fenómeno. Por esa razón, el estudio de las moléculas desreguladas se beneficia de tener medidas cuantitativas de las interacciones de las moléculas. Una medida muy accesible fue identificar en qué *pathways* participan las moléculas.

Existen moléculas que se han asociado a una sola vía, y por lo tanto se han asociado a una función biológica particular. Por otro lado, existen muchas moléculas por las que *entrecruzan* varias vías. La modulación farmacológica de una molécula que se encuentra en muchas vías puede ser deseable, en tanto que se pueden tener efectos terapéuticos con un alcance más amplio dentro del sistema biológico (tanto a nivel celular como en el

organismo). Pero al mismo tiempo, la probabilidad de alterar otras funciones, es decir, la aparición de *efectos adversos* aumenta. Es importante resaltar que no se propone una *preferencia* entre una clase y la otra: más bien, se propone contar con esta información *a priori* para la toma de decisiones en el reposicionamiento de fármacos (Novac, 2013), la polifarmacia (Jia *et al.*, 2009) y el diseño racional de medicamentos (Nikolsky *et al.*, 2005), con el fin de reducir costos y conseguir tratamientos seguros y eficaces (Benson y van der Graaf, 2014).

Con estas consideraciones en mente, algunas de las moléculas que identificamos pueden resultar interesantes como potenciales blancos terapéuticos en cáncer de mama. COL1A1, que codifica la proteína colágeno tipo I alfa 1, presenta sobreexpresión en todos los subtipos analizados. Se encontró esta molécula participando en 22 de los *pathways* analizados. La desregulación de esta molécula se ha identificado en cáncer de mama (Helleman *et al.*, 2008). Además, se ha identificado un posible papel de la sobreexpresión de esta molécula en la resistencia a quimioterapia en otros tipos de cáncer (Yu *et al.*, 2014). Hay reportes de agentes farmacológicos capaces de inhibir la transcripción de esta molécula (Louneva *et al.*, 2003), por lo que la evaluación del efecto terapéutico de su modulación para el cáncer de mama puede resultar atractiva.

Otra molécula identificada cuya modulación farmacológica puede contribuir al tratamiento de diferentes subtipos moleculares de cáncer de mama es el receptor de andrógenos. Este receptor, como el receptor de estrógenos, tiene varias funciones, tanto como factor transcripcional como en señalización (Heinlein y Chang, 2002). Se encontró a esta molécula participando en 13 de las rutas biológicas analizadas. Llama la atención que el receptor de andrógenos se encuentra sobreexpresado en los subtipos luminales y el subtipo HER2-enriquecido, pero se encuentra subexpresado en el subtipo basal. Se ha estudiado la importancia del receptor de andrógenos y su potencial uso como blanco terapéutico en cáncer de mama (Chia *et al.*, 2015; Cochrane *et al.*, 2014; Mishra *et al.*, 2012). Se ha identificado que la pérdida de la expresión del receptor se asocia a recurrencia en tumores basales (Thike *et al.*, 2014), y su potencial uso como marcador y blanco terapéutico

en tumores triple negativos (Zakaria *et al.*, 2015; Safarpour y Tavassoli, 2014; Gasparini *et al.*, 2014; Shah *et al.*, 2013). También se ha estudiado el entrecruzamiento de la señalización por el receptor de andrógenos con la señalización de estrógeno (Karamouzis *et al.*, 2015) y con la señalización por el receptor HER2 (Naderi y Hughes-Davies, 2008). Existen moléculas capaces de actuar como agonistas y como antagonistas del receptor. En este caso, podríamos pensar en la posibilidad de modular una misma molécula en formas distintas dependiendo del subtipo molecular. Los resultados de este trabajo apoyan la relevancia del estudio del receptor de andrógenos como blanco terapéutico.

El acercamiento al cáncer de mama a través de redes transcripcionales proporciona un nivel de entendimiento del fenómeno diferente al obtenido a través de los análisis antes descritos. Con este abordaje, se asigna *importancia* a las moléculas basados en las características que presentan en el contexto de las redes transcripcionales, y no necesariamente en función de los niveles cuantificados experimentalmente. Como ya se ha discutido, las redes transcripcionales fueron capaces de identificar diferencias entre subtipos que no aparecen al nivel de la expresión de genes. Estas diferencias permiten identificar potenciales blancos terapéuticos de otras maneras.

Previamente se discutió la importancia del gen CNR2 en el contexto de redes transcripcionales, y se mencionó que se trata de un potencial blanco terapéutico de amplio espectro, subtipo-independiente, para cáncer de mama. El uso de agentes cannabinoides para el tratamiento del cáncer ha sido explorado (Velasco *et al.*, 2016; Chakravarti *et al.*, 2014), incluido el uso de agonistas en cáncer de mama (Qamri *et al.*, 2009), y en particular para tumores triple negativos (Morales *et al.*, 2015). Las redes transcripcionales apuntan a que la modulación farmacológica de esta molécula puede tener un efecto más general en la diversidad de cáncer de mama, que deberá ser validado experimentalmente. En el caso de potenciales blancos subtipo-específicos, se discutió el gen LCK, de alto grado en la red transcripcional de cáncer de mama HER2-enriquecido, y ausente del resto de las redes. Se ha identificado que la proteína codificada por LCK es un blanco farmacológico modulable a través de fármacos como dasatinib e imatinib (Harr *et al.*, 2010;

Lee *et al.*, 2010). En este caso, la evaluación del efecto del uso de dichos fármacos para el tratamiento de tumores HER2-enriquecidos, ya sea en combinación con el tratamiento tradicional o de manera individual, puede ser interesante.

Se han presentado varias estrategias que permiten la exploración de nuevos blancos terapéuticos para cáncer de mama, y se han mostrado ejemplos de como las estrategias pueden identificar moléculas de interés, de manera que puedan dirigir el diseño de experimentos funcionales que validen su utilidad. Resulta importante observar como las estrategias no son mutuamente excluyentes, y bien pueden ser complementadas. Los espacios de blancos moleculares presentados en este trabajo fueron construidos considerando únicamente una medida de influencia sistémica, basada en la anotación funcional de las moléculas. Sin embargo, se ha presentado también la estrategia de identificar el papel sistémico de las moléculas en el contexto de cada subtipo molecular, utilizando redes transcripcionales inferidas: Las estrategias pueden ser complementarias. Conforme las tecnologías masivas avancen, se puede incluso integrar información de otros tipos de experimentos: regulación a través de miRNAs, efectos de RNAs largos no codificantes, datos proteómicos. Las estrategias aquí presentadas son generalizables.

Las diferencias entre subtipos pueden dar lugar a diferencias en la respuesta a tratamiento farmacológico

Se ha expuesto ampliamente como la heterogeneidad existente en el cáncer de mama da lugar a una amplia variabilidad en la respuesta a tratamientos. Las metodologías aquí presentadas permiten explorar, además de potenciales nuevos blancos terapéuticos, las interacciones que pueden tener los *blancos terapéuticos conocidos* que den lugar a ineficacias terapéuticas. Esto es importante por dos razones: primero, porque puede permitir identificar blancos terapéuticos adicionales que actúen *en conjunto* con los tratamientos tradicionales, y segundo, porque como ya se mencionó, considerar los potenciales nuevos blancos terapéuticos en un contexto sistémico *a priori* puede hacer más eficiente su desa-

rollo.

Se escogió el caso de estudio de la insensibilidad a la modulación de la *señalización de estrógeno* por tratarse de un problema que hace uso de las estrategias implementadas:

- Se sabe que no existe una causa única de resistencia (Viedma-Rodríguez *et al.*, 2014).
- Los efectos del tratamiento farmacológico se presentan tanto a nivel de expresión *genes* como de *vías*
- Se consideran diferencias en la susceptibilidad de tratamiento antiestrogénico entre subtipos
- Moléculas relacionadas con la sensibilidad al tratamiento antiestrogénico distintas a receptores de estrógeno pueden ser potenciales blancos farmacológicos.

Como ya se ha discutido, los tratamientos antiestrogénicos se basan en la sobreexpresión del receptor de estrógeno. Dado que el receptor se encuentra sobreexpresado únicamente en los subtipos luminales, el tratamiento es efectivo únicamente en dichos subtipos.

Sin embargo, cuando pasamos a un nivel de *vías*, encontramos que la vía de señalización de estrógeno está alterada en todos los subtipos moleculares, independientemente del estado del receptor de estrógeno. El fenómeno puede explicarse cuando se contextualiza la señalización de estrógeno en una *red de entrecruzamiento*; la vía conecta con otro gran número de vías, compartiendo moléculas que pueden estar alteradas por otros procesos *independientes del receptor de estrógeno*. La implicación es que las contribuciones terapéuticas del bloqueo antiestrogénico relacionadas con la vía de señalización pueden perderse, porque el bloqueo es *compensado* por las vías que entrecruzan.

La exploración permitió identificar vías entrecruzadas con la vía de señalización de estrógeno; se identificó en la literatura evidencia experimental de la relación de dichas vías con la sensibilidad al tratamiento antiestrogénico, y potenciales moléculas en dichas vías susceptibles de modulación farmacológica. Es importante mencionar que tomar estas observaciones como un indicio de una eficacia del tratamiento antiestrogénico en los

subtipos no luminales es una interpretación que puede ser errónea. Una interpretación más cautelosa sería considerar que las vías, en este caso de señalización, comparten muchas moléculas comunes (es decir, *entrecruzan*). De la misma manera, muchas de estas moléculas comunes se encuentran desreguladas en los diferentes subtipos moleculares: a pesar de las diferencias en expresión de genes, existen similitudes entre subtipos al nivel de expresión de vías.

El flujo de trabajo utilizado para esta exploración es muy versátil. Este caso de estudio sirve como una *prueba de concepto* de su utilidad para la exploración de fenómenos de relevancia farmacológica. La estrategia puede ser utilizada para la exploración de otros fenómenos, tanto biológicos como farmacológicos, de interés.

Capítulo 11

Perspectivas

En la presente tesis se expone un marco de trabajo para la exploración de blancos farmacológicos en el contexto de la biología de sistemas de subtipos moleculares de cáncer de mama.

Siendo este un primer acercamiento, se dio mayor relevancia a la integración de las herramientas, sacrificando un nivel de profundización mayor sobre los resultados obtenidos que pueden servir como puntos de partida para siguientes nuevas líneas de investigación. Las siguientes son algunas de estas líneas que se desprenden de los resultados obtenidos en la implementación de las técnicas implementadas en este trabajo.

11.1. Otras formas de construcción de redes transcripcionales

La estrategia empleada para la construcción de redes transcripcionales es cardinal para poder realizar interpretaciones biológicas. En el presente trabajo se empleó un criterio para la reconstrucción de las redes basado primordialmente en el peso de las interacciones (en función de la información mutua); adicionalmente integramos un segundo criterio biológico de anotación funcional. Esta misma estrategia puede ser modificada, tanto por la parte cuantitativa (empleando otros valores de Información Mutua), como por la parte

biológica (utilizando otras bases de datos).

Pueden construirse redes transcripcionales para un conjunto de moléculas de interés. De la misma forma que filtramos aquellas moléculas que están descritas en KEGG, podemos hacer filtrados para moléculas específicas: por ejemplo, moléculas de procesos concretos o aquellas identificadas en experimentos funcionales.

Otra consideración importante es la direccionalidad de las redes. Las redes transcripcionales presentadas en el capítulo correspondiente son no dirigidas. En ellas identificamos aquellos nodos que han sido descritos como factores transcripcionales, y mencionamos que estos pueden ser empleados para establecer direccionalidad a partir de un criterio biológico. Sin embargo, no nos aventuramos a establecer dicha direccionalidad ya que no contamos con suficiente información biológica para describir todos los nodos (¿cómo interpretar las interacciones entre dos factores transcripcionales? ¿O las interacciones entre dos moléculas que *no* son factores transcripcionales?). La construcción de redes de *regulación* transcripcional requiere la integración de información validada experimentalmente sobre las interacciones transcripcionales en la red. Las redes transcripcionales que aquí se presentan pueden orientar en el diseño de nuevos experimentos.

11.1.1. Más información de redes transcripcionales

Las redes transcripcionales que presentamos capturan diferencias entre los diferentes subtipos cancerosos y con el fenotipo sano, mostrando que la arquitectura transcripcional es diversa y puede estar asociada a diferencias fenotípicas.

Una observación interesante es que las redes mostraron una tendencia al *agrupamiento*. La estructura de nuestras redes indica la posibilidad de la existencia de estructuras de *comunidades* transcripcionales: conjuntos de moléculas fuertemente asociadas (Moreira-Filho *et al.*, 2014). Dadas las diferencias en la composición de las redes transcripcionales, las comunidades en ellas tendrían que ser distintas. Más interesante sería identificar si las comunidades se asocian a *funciones biológicas* concretas. La implementación de metodologías de esta naturaleza ha sido iniciada ya en nuestro grupo de trabajo, y pronto

será aplicada a las redes aquí presentadas.

Otra línea de investigación que se desprende de las redes transcripcionales es la identificación de *reguladores transcripcionales maestros*. Un regulador transcripcional maestro es un factor de transcripción expresado al comienzo del desarrollo de un fenotipo y que participa en la definición de dicho fenotipo (Chan y Kyba, 2013). Dadas las diferencias en la composición de las redes por subtipos moleculares, podemos inferir que existirán diferencias en los reguladores transcripcionales maestros para cada uno de los subtipos moleculares de cáncer de mama. Más aún, existe la posibilidad de que la modulación farmacológica de estos permita revertir el fenotipo, haciéndolos atractivos blancos terapéuticos. Algunas estrategias para la identificación de reguladores transcripcionales han sido implementadas en el grupo de trabajo (Tovar *et al.*, 2015).

11.2. Modelado de las redes de entrecruzamiento

El fenómeno de entrecruzamiento ha sido descrito previamente (Donato *et al.*, 2013). Sin embargo, el enfoque presentado, empleando dicho fenómeno para la construcción de redes de *procesos* interconectados no había sido desarrollado antes.

Nuevamente se debe resaltar que las redes de entrecruzamiento reducen notablemente el número de nodos, comparados con una red de moléculas individuales (por ejemplo, una red metabólica). Esto puede ser una ventaja para la implementación de algoritmos que permitan el modelado de la *dinámica* de las redes, cuantificando la contribución en la desregulación de procesos individuales hacia el desarrollo de un fenotipo. Algunos esfuerzos se están desarrollando en el grupo de trabajo para poder implementar estos modelos partiendo de los datos disponibles.

11.3. Otras estrategias de exploración del espacio de blancos terapéuticos

El espacio de blancos terapéuticos presentado en el capítulo correspondiente se propone como una herramienta novedosa para la exploración de alternativas terapéuticas en enfermedades heterogéneas, como es el caso del cáncer de mama. La estrategia puede ser implementada para el estudio de otras enfermedades heterogéneas en las que diferencias a nivel de expresión sean relevantes (Espinal-Enríquez *et al.*, 2015), y es susceptible a ser expandida tanto a nivel de datos experimentales como datos de interconexión sistémica (sean estos de la literatura, como en el caso presentado, o inferidos a través de otras estrategias, por ejemplo, redes transcripcionales).

En este trabajo presentamos una estrategia de exploración basada en un sistema de clasificación muy elemental, con particiones en función de las magnitudes descritas para cada molécula. Una ventaja de la representación de moléculas desreguladas presentada es que esta no es la única estrategia de categorización posible. Existe la posibilidad de implementar diferentes *algoritmos de agrupamiento* sobre las moléculas desreguladas que permitan identificar patrones en las moléculas desreguladas. Una línea de investigación particular en este sentido es identificar si estos conjuntos teóricos de blancos terapéuticos son susceptibles de una modulación conjunta: es decir, si se pueden inferir potenciales combinaciones terapéuticas.

11.4. Experimentos funcionales para cuantificar la influencia de los procesos entrecruzados sobre efectos farmacológicos

La exploración del fenómeno de entrecruzamiento en la vía de señalización de estrógeno mostró que la vía de estrógeno es una vía altamente conectada en el contexto celular, y

a cuya alteración patológica pueden contribuir muchas vías distintas.

Una limitación que encontramos al momento de implementar nuestras metodologías fue que si bien contamos con datos suficientes para la descripción del fenómeno en los subtipos moleculares, existen pocos datos experimentales de efectos farmacológicos en la diversidad de subtipos moleculares.

Nuestros resultados muestran que la desregulación de vías es un fenómeno extendido, y que incluso vías que tradicionalmente no se asocian a algunos fenotipos pueden estar desreguladas en ellas (como la vía de señalización de estrógeno en el subtipo basal). Estos resultados llevan a repensar en la necesidad de experimentos biológicos sobre estos subtipos. Todavía más importantes a futuro serían experimentos multifactoriales en los que se profundice en la influencia que puede tener la modulación farmacológica de otras vías sobre el efecto tanto de fármacos ya conocidos como para moléculas nuevas.

En el caso de estudio de la vía de estrógenos, profundizamos un poco en los efectos ya descritos del entrecruzamiento con la vía de apoptosis y mencionamos evidencias respecto a la posible contribución de vías de inmunidad; experimentos sobre estas vías, así como otras de las identificadas serían de interés tanto en la búsqueda de nuevas opciones terapéuticas.

11.5. Exploración del fenómeno de entrecruzamiento sobre otros blancos terapéuticos

El caso de estudio de la vía de estrógeno es una prueba de concepto de la aplicabilidad de nuestras estrategias para la exploración farmacológica de subtipos moleculares de cáncer de mama. La misma estrategia puede ser planteada para estudiar otros blancos terapéuticos identificados.

Dentro del espacio de blancos terapéuticos encontramos a los blancos terapéuticos tradicionales para cáncer de mama. Un estudio similar al realizado para señalización de estrógeno sobre las vías asociadas a ERBB2 o las vías apoptóticas atacadas por agentes

quimioterapéuticos pueden revelar las influencias de otras vías en su eficacia. De manera similar, contextualizar algunos blancos terapéuticos menos ortodoxos, como el receptor de andrógenos (Chia *et al.*, 2015). Todavía más interesante es la posibilidad de explorar las vías biológicas en las que están involucrados otros blancos, que no han sido explorados, que identificamos en nuestro espacio de blancos terapéuticos.

Capítulo 12

Consideraciones finales

En este trabajo se presenta un marco conceptual de biología de sistemas para el estudio de subtipos moleculares de cáncer de mama y sus implicaciones terapéuticas.

Encontramos que existen desregulaciones comunes dentro de la heterogeneidad del cáncer de mama — *el cáncer de mama es cáncer de mama*, y tiene características comunes.

También encontramos diferencias a nivel transcripcional entre distintos subtipos de cáncer de mama, que pueden ayudar a explicar la heterogeneidad fenotípica observada en el cáncer de mama.

Las desregulaciones, tanto comunes como específicas, de los diferentes subtipos de cáncer de mama pueden ser situadas en su contexto sistémico. Implementamos diferentes estrategias complementarias, tanto previamente establecidas como nuevas propuestas, para hacer esto.

Finalmente, establecemos algunas estrategias de como alimentar el conocimiento farmacológico con los modelos propuestos.

Los resultados aquí reportados tienen la intención de ser *generadores de hipótesis* que guíen futuras líneas de investigación experimental. Con este trabajo se pretende establecer bases teóricas que permitan la transición a una *farmacología de sistemas* del cáncer de mama.

Bibliografía

- ALBERGHINA, L., GAGLIO, D., MORESCO, R.M., GILARDI, M.C., MESSA, C., Y VANONI, M. A systems biology road map for the discovery of drugs targeting cancer cell metabolism. *Current pharmaceutical design* **20**(15):2648–2666 (2014)
- ANDERSEN, I., KOLODZIEJCZYK, C., THIELEN, K., HEINESEN, E., Y DIDERICHSEN, F. The effect of breast cancer on personal income three years after diagnosis by cancer stage and education: a register-based cohort study among Danish females. *BMC public health* **15**(1):50 (2015)
- AZAD, A., LAWEN, A., Y KEITH, J. Prediction of signaling cross-talks contributing to acquired drug resistance in breast cancer cells by Bayesian statistical modeling. *BMC systems biology* **9**(1):2 (2015)
- AZIMI, I., ROBERTS-THOMSON, S., Y MONTEITH, G. Calcium influx pathways in breast cancer: opportunities for pharmacological intervention. *British journal of pharmacology* **171**(4):945–960 (2014)
- AZMI, A.S. Y MOHAMMAD, R.M. Rectifying cancer drug discovery through network pharmacology. *Future medicinal chemistry* **6**(5):529–539 (2014)
- BACA-LÓPEZ, K., HERNÁNDEZ-LEMUS, E., Y MAYORGA, M. Information-theoretical analysis of gene expression data to infer transcriptional interactions. *Revista mexicana de física* **55**(6):456–466 (2009)

- BACA-LÓPEZ, K., MAYORGA, M., HIDALGO-MIRANDA, A., GUTIÉRREZ-NÁJERA, N., Y HERNÁNDEZ-LEMUS, E. The role of master regulators in the metabolic/transcriptional coupling in breast carcinomas. *PloS one* **7**(8):e42678 (2012)
- BADVE, S., DABBS, D.J., SCHNITT, S.J., BAEHNER, F.L., DECKER, T., EUSEBI, V., FOX, S.B., ICHIHARA, S., JACQUEMIER, J., LAKHANI, S.R. *et al.* Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology* **24**(2):157–167 (2011)
- BAKER, A.T., ZLOBIN, A., Y OSIPO, C. Notch-EGFR/HER2 bidirectional crosstalk in breast cancer. *Frontiers in oncology* **4** (2014a)
- BAKER, A.T., ZLOBIN, A., OSIPO, C. *et al.* Notch-EGFR/HER2 bidirectional crosstalk in breast cancer. *Front in oncol* **4**:360 (2014b)
- BARABASI, A.L. Y OLTVAI, Z.N. Network biology: understanding the cell's functional organization. *Nature reviews genetics* **5**(2):101–113 (2004)
- BASELGA, J. Targeting the phosphoinositide-3 (PI3) kinase pathway in breast cancer. *The oncologist* **16**(Supplement 1):12–19 (2011)
- BAYRAKTAR, S. Y GLÜCK, S. Molecularly targeted therapies for metastatic triple-negative breast cancer. *Breast cancer research and treatment* **138**(1):21–35 (2013)
- BEDARD, P.L., HANSEN, A.R., RATAIN, M.J., Y SIU, L.L. Tumour heterogeneity in the clinic. *Nature* **501**(7467):355–364 (2013)
- BEHJATI, S. Y FRANK, M. The effects of tamoxifen on immunity. *Current medicinal chemistry* **16**(24):3076 (2009)
- BENSON, N. Y VAN DER GRAAF, P.H. The rise of systems pharmacology in drug discovery and development. *Future medicinal chemistry* **6**(16):1731–1734 (2014)

- BONAPACE, L., COISSIEUX, M.M., WYCKOFF, J., MERTZ, K.D., VARGA, Z., JUNT, T., Y BENTIRES-ALJ, M. Cessation of CCL2 inhibition accelerates breast cancer metastasis by promoting angiogenesis. *Nature* **515**(7525):130–133 (2014)
- BRAESS, D. Uber ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12: 258–268, 1968. *English translation in Braess* (2005)
- BRAZHNIK, P., DE LA FUENTE, A., Y MENDES, P. Gene networks: how to put the function in genomics. *TRENDS in Biotechnology* **20**(11):467–472 (2002)
- BURSTEIN, H.J. The distinctive nature of HER2-positive breast cancers. *New England Journal of Medicine* **353**(16):1652–1654 (2005)
- CAMPBELL, J.D. Y RAMSEY, S.D. The costs of treating breast cancer in the US. *Pharmacoeconomics* **27**(3):199–209 (2009)
- CARASCOSSA, S., DUDEK, P., CENNI, B., BRIAND, P.A., Y PICARD, D. CARM1 mediates the ligand-independent and tamoxifen-resistant activation of the estrogen receptor α by cAMP. *Genes & development* **24**(7):708–719 (2010)
- CARTER, H., HOFREE, M., Y IDEKER, T. Genotype to phenotype via network analysis. *Current opinion in genetics & development* **23**(6):611–621 (2013)
- CHAKRAVARTI, B., RAVI, J., Y GANJU, R.K. Cannabinoids as therapeutic agents in cancer: current status and future implications. *Oncotarget* **5**(15):5852 (2014)
- CHAN, S.S.K. Y KYBA, M. What is a Master Regulator? *Journal of stem cell research & therapy* **3** (2013)
- CHÁVARRI-GUERRA, Y., VILLARREAL-GARZA, C., LIEDKE, P.E., KNAUL, F., MOHAR, A., FINKELSTEIN, D.M., Y GOSS, P.E. Breast cancer in Mexico: a growing challenge to health and the health system. *The Lancet Oncology* **13**(8):e335–e343 (2012)

- CHEN, Y.F., CHEN, Y.T., CHIU, W.T., Y SHEN, M.R. Remodeling of calcium signaling in tumor progression. *J Biomed Sci* **20**:23 (2013)
- CHIA, K., OBRIEN, M., BROWN, M., Y LIM, E. Targeting the Androgen Receptor in Breast Cancer. *Current oncology reports* **17**(2):1–6 (2015)
- CLARKE, R., COOK, K.L., HU, R., FACEY, C.O., TAVASSOLY, I., SCHWARTZ, J.L., BAUMANN, W.T., TYSON, J.J., XUAN, J., WANG, Y. *et al.* Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. *Cancer research* **72**(6):1321–1331 (2012)
- COCHRANE, D.R., BERNALES, S., JACOBSEN, B.M., CITTELLY, D.M., HOWE, E.N., D AMATO, N.C., SPOELSTRA, N.S., EDGERTON, S.M., JEAN, A., GUERRERO, J. *et al.* Role of the androgen receptor in breast cancer and preclinical analysis of enzalutamide. *Breast Cancer Res* **16**(1):R7 (2014)
- COOK, K.L., WÄRRI, A., SOTO-PANTOJA, D.R., CLARKE, P.A., CRUZ, M.I., ZWART, A., Y CLARKE, R. Hydroxychloroquine inhibits autophagy to potentiate anti-estrogen responsiveness in ER+ breast cancer. *Clinical Cancer Research* **20**(12):3222–3232 (2014)
- CUI, J., GERMER, K., WU, T., WANG, J., LUO, J., WANG, S.C., WANG, Q., Y ZHANG, X. Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer research* **72**(21):5625–5634 (2012)
- DE ANDA-JÁUREGUI, G., MEJÍA-PEDROZA, R.A., ESPINAL-ENRÍQUEZ, J., Y HERNÁNDEZ-LEMUS, E. Crosstalk events in the estrogen signaling pathway may affect tamoxifen efficacy in breast cancer molecular subtypes. *Computational biology and chemistry* (2015)
- DESMEDT, C., PIETTE, F., LOI, S., WANG, Y., LALLEMAND, F., HAIBE-KAINS, B., VIALE, G., DELORENZI, M., ZHANG, Y., D’ASSIGNIES, M.S. *et al.* Strong

- time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clinical cancer research* **13**(11):3207–3214 (2007)
- DONATO, M., XU, Z., TOMOIAGA, A., GRANNEMAN, J.G., MACKENZIE, R.G., BAO, R., THAN, N.G., WESTFALL, P.H., ROMERO, R., Y DRAGHICI, S. Analysis and correction of crosstalk effects in pathway analysis. *Genome research* **23**(11):1885–1893 (2013)
- DRIER, Y., SHEFFER, M., Y DOMANY, E. Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences* **110**(16):6388–6393 (2013)
- DUNNWALD, L.K., ROSSING, M.A., LI, C.I. *et al.* Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* **9**(1):R6 (2007)
- EDGAR, R., DOMRACHEV, M., Y LASH, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* **30**(1):207–210 (2002)
- EGNER, J.R. AJCC cancer staging manual. *JAMA* **304**(15):1726–1727 (2010)
- ELSTON, C.W., ELLIS, I.O. *et al.* Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**(5):403–410 (1991)
- ESPINAL-ENRÍQUEZ, J., MUÑOZ-MONTERO, S., IMAZ-ROSSHANDLER, I., HUERTA-VERDE, A., MEJÍA, C., Y HERNÁNDEZ-LEMUS, E. Genome-wide expression analysis suggests a crucial role of dysregulation of matrix metalloproteinases pathway in undifferentiated thyroid carcinoma. *BMC genomics* **16**(1):207 (2015)
- FANG, W.B., YAO, M., JOKAR, I., ALHAKAMY, N., BERKLAND, C., CHEN, J., BRANTLEY-SIEDERS, D., Y CHENG, N. The CCL2 chemokine is a negative regulator

- of autophagy and necrosis in luminal B breast cancer cells. *Breast cancer research and treatment* **150**(2):309–320 (2015)
- FARMER, P., BONNEFOI, H., BECETTE, V., TUBIANA-HULIN, M., FUMOLEAU, P., LARSIMONT, D., MACGROGAN, G., BERGH, J., CAMERON, D., GOLDSTEIN, D. *et al.* Identification of molecular apocrine breast tumours by microarray analysis. *Breast Cancer Research* **7**(Suppl 2):P2–11 (2005)
- FERLAY, J., SOERJOMATARAM, I., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D.M., FORMAN, D., Y BRAY, F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**(5):E359–E386 (2015)
- FIORIO, E., MERCANTI, A., TERRASI, M., MICCIOLO, R., REMO, A., AURIEMMA, A., MOLINO, A., PAROLIN, V., STEFANO, B., BONETTI, F. *et al.* Leptin/HER2 crosstalk in breast cancer: in vitro study and preliminary in vivo analysis. *BMC cancer* **8**(1):1 (2008)
- GARCÍA-CAMPOS, M.A., ESPINAL-ENRÍQUEZ, J., Y HERNÁNDEZ-LEMUS, E. Pathway analysis: State of the art. *Frontiers in physiology* **6** (2015)
- GASPARINI, P., FASSAN, M., CASCIONE, L., GULER, G., BALCI, S., IRKKAN, C., PAISIE, C., LOVAT, F., MORRISON, C., ZHANG, J. *et al.* Androgen receptor status is a prognostic marker in non-basal triple negative breast cancers and determines novel therapeutic options. *PLoS One* **9**(2):e88525 (2014)
- GIANCOTTI, F.G. Deregulation of cell signaling in cancer. *FEBS letters* **588**(16):2558–2570 (2014)
- GRIFFITH, M., GRIFFITH, O.L., COFFMAN, A.C., WEIBLE, J.V., MCMICHAEL, J.F., SPIES, N.C., KOVAL, J., DAS, I., CALLAWAY, M.B., ELDRED, J.M. *et al.* DGIdb: mining the druggable genome. *Nature methods* **10**(12):1209–1210 (2013)

- GUEDJ, M., MARISA, L., DE REYNIES, A., ORSETTI, B., SCHIAPPA, R., BIBEAU, F., MACGROGAN, G., LEREBOURS, F., FINETTI, P., LONGY, M. *et al.* A refined molecular taxonomy of breast cancer. *Oncogene* **31**(9):1196–1206 (2012)
- GUIDA, M., LIGRESTI, A., DE FILIPPIS, D., D’AMICO, A., PETROSINO, S., CIPRIANO, M., BIFULCO, G., SIMONETTI, S., ORLANDO, P., INSABATO, L. *et al.* The levels of the endocannabinoid receptor CB2 and its ligand 2-arachidonoylglycerol are elevated in endometrial carcinoma. *Endocrinology* **151**(3):921–928 (2010)
- GUO, S., LIU, M., Y GONZALEZ-PEREZ, R.R. Role of Notch and its oncogenic signaling crosstalk in breast cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **1815**(2):197–213 (2011)
- HANAHAN, D. Y WEINBERG, R.A. The hallmarks of cancer. *cell* **100**(1):57–70 (2000)
- HANAHAN, D. Y WEINBERG, R.A. Hallmarks of cancer: the next generation. *cell* **144**(5):646–674 (2011)
- HARR, M., CAIMI, P., MCCOLL, K., ZHONG, F., PATEL, S., BARR, P., Y DISTELHORST, C. Inhibition of Lck enhances glucocorticoid sensitivity and apoptosis in lymphoid cell lines and in chronic lymphocytic leukemia. *Cell Death & Differentiation* **17**(9):1381–1391 (2010)
- HASTINGS, J., DE MATOS, P., DEKKER, A., ENNIS, M., HARSHA, B., KALE, N., MUTHUKRISHNAN, V., OWEN, G., TURNER, S., WILLIAMS, M. *et al.* The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic acids research* **41**(D1):D456–D463 (2013)
- HEINLEIN, C.A. Y CHANG, C. The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions. *Molecular Endocrinology* **16**(10):2181–2187 (2002)

- HELLEMAN, J., JANSEN, M.P., RUIGROK-RITSTIER, K., VAN STAVEREN, I.L., LOOK, M.P., MEIJER-VAN GELDER, M.E., SIEUWERTS, A.M., KLIJN, J.G., SLEIJFER, S., FOEKENS, J.A. *et al.* Association of an extracellular matrix gene cluster with breast cancer prognosis and endocrine therapy response. *Clinical Cancer Research* **14**(17):5555–5564 (2008)
- HERNÁNDEZ-LEMUS, E. Systems Biology and Integrative Omics in Breast Cancer. En *Omics Approaches in Breast Cancer*, págs. 333–352. Springer (2014)
- HERNÁNDEZ-LEMUS, E., LI, W., Y MEYER, P. Advances in systems biology-New trends and perspectives. *Computational Biology and Chemistry* **59**(PB):1–2 (2015)
- HOLLMÉN, M., ROUDNICKY, F., KARAMAN, S., Y DETMAR, M. Characterization of macrophage-cancer cell crosstalk in estrogen receptor positive and triple-negative breast cancer. *Scientific reports* **5** (2015)
- HOLME, P., HUSS, M., Y JEONG, H. Subnetwork hierarchies of biochemical pathways. *Bioinformatics* **19**(4):532–538 (2003)
- HU, Z., FAN, C., OH, D.S., MARRON, J., HE, X., QAQISH, B.F., LIVASY, C., CAREY, L.A., REYNOLDS, E., DRESSLER, L. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC genomics* **7**(1):96 (2006)
- IARC. Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed. *International Agency for Research on Cancer, World Health Organization* **12** (2013)
- IRIZARRY, R.A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y.D., ANTONELLIS, K.J., SCHERF, U., Y SPEED, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**(2):249–264 (2003)
- IVSHINA, A.V., GEORGE, J., SENKO, O., MOW, B., PUTTI, T.C., SMEDS, J., LINDAHL, T., PAWITAN, Y., HALL, P., NORDGREN, H. *et al.* Genetic reclassification

- of histologic grade delineates new clinical subtypes of breast cancer. *Cancer research* **66**(21):10292–10301 (2006)
- JHA, S., JONES, V., BURRIDGE, K., Y MUKHOPADHYAY, S. CB2 receptor-mediated Regulation of Prostate Cancer Cell Migration: Involvement of RhoA and Stress fiber formation. *The FASEB Journal* **26**(1.MeetingAbstracts):782–11 (2012)
- JIA, J., ZHU, F., MA, X., CAO, Z.W., LI, Y.X., Y CHEN, Y.Z. Mechanisms of drug combinations: interaction and network perspectives. *Nature reviews Drug discovery* **8**(2):111–128 (2009)
- JOSHI-TOPE, G., GILLESPIE, M., VASTRIK, I., D'EUSTACHIO, P., SCHMIDT, E., DE BONO, B., JASSAL, B., GOPINATH, G., WU, G., MATTHEWS, L. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic acids research* **33**(suppl 1):D428–D432 (2005)
- KANEHISA, M., GOTO, S., SATO, Y., KAWASHIMA, M., FURUMICHI, M., Y TANABE, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research* **42**(D1):D199–D205 (2014)
- KARAMOUZIS, M.V., PAPAVALASSIOU, K.A., ADAMOPOULOS, C., Y PAPAVALASSIOU, A.G. Targeting Androgen/Estrogen Receptors Crosstalk in Cancer. *Trends in Cancer* (2015)
- KHATRI, P., SIROTA, M., Y BUTTE, A.J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* **8**(2):e1002375 (2012)
- KITTANEH, M., MONTERO, A.J., Y GLÜCK, S. Molecular profiling for breast cancer: a comprehensive review. *Biomarkers in cancer* **5**:61 (2013)
- KNAUL, F.M., ARREOLA-ORNELAS, H., VELÁZQUEZ, E., DORANTES, J., MÉNDEZ, Ó., Y ÁVILA-BURGOS, L. El costo de la atención médica del cáncer mamario: el caso del Instituto Mexicano del Seguro Social. *Salud pública de México* **51**:s286–s295 (2009)

- KNOX, S.S. From omics to complex disease: a systems biology approach to gene-environment interactions in cancer. *Cancer Cell Int* **10**(11) (2010)
- KREEGER, P.K. Y LAUFFENBURGER, D.A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**(1):2–8 (2010)
- LEE, K., OUWEHAND, I., GIANNINI, A., THOMAS, N., DIBB, N., Y BIJLMAKERS, M. Lck is a key target of imatinib and dasatinib in T-cell activation. *Leukemia* **24**(4):896–900 (2010)
- LEVIN, E.R. Integration of the extranuclear and nuclear actions of estrogen. *Mol Endocrinol* **19**(8):1951–1959 (2005)
- LIANG, J. Y SHANG, Y. Estrogen and cancer. *Annual review of physiology* **75**:225–240 (2013)
- LIU, C.Y., HUNG, M.H., WANG, D.S., CHU, P.Y., SU, J.C., TENG, T.H., HUANG, C.T., CHAO, T.T., WANG, C.Y., SHIAU, C.W. *et al.* Tamoxifen induces apoptosis through cancerous inhibitor of protein phosphatase 2A-dependent phospho-Akt inactivation in estrogen receptor-negative human breast cancer cells. *Breast Cancer Research* **16**(5):431 (2014)
- LIU, R., WANG, X., CHEN, G.Y., DALERBA, P., GURNEY, A., HOEY, T., SHERLOCK, G., LEWICKI, J., SHEDDEN, K., Y CLARKE, M.F. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *New England Journal of Medicine* **356**(3):217–226 (2007)
- LIVSHITS, A., GIT, A., FUKS, G., CALDAS, C., Y DOMANY, E. Pathway-based personalized analysis of breast cancer expression data. *Molecular oncology* **9**(7):1471–1483 (2015)
- LOUNEVA, N., SAITTA, B., HERRICK, D.J., Y JIMENEZ, S.A. Transcriptional inhibi-

- tion of type I collagen gene expression in scleroderma fibroblasts by the antineoplastic drug ecteinascidin 743. *Journal of Biological Chemistry* **278**(41):40400–40407 (2003)
- LUO, W., FRIEDMAN, M.S., SHEDDEN, K., HANKENSON, K.D., Y WOOLF, P.J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics* **10**(1):161 (2009)
- MACNEIL, L.T. Y WALHOUT, A.J. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research* **21**(5):645–657 (2011)
- MALHOTRA, G.K., ZHAO, X., BAND, H., Y BAND, V. Histological, molecular and functional subtypes of breast cancers. *Cancer biology & therapy* **10**(10):955–960 (2010)
- MANDLEKAR, S. Y KONG, A.N. Mechanisms of tamoxifen-induced apoptosis. *Apoptosis* **6**(6):469–477 (2001)
- MARGOLIN, A.A., NEMENMAN, I., BASSO, K., WIGGINS, C., STOLOVITZKY, G., FAVERA, R.D., Y CALIFANO, A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* **7**(Suppl 1):S7 (2006)
- MATSEN, C.B. Y NEUMAYER, L.A. Breast cancer: a review for the general surgeon. *JAMA surgery* **148**(10):971–980 (2013)
- MILLER, L.D., SMEDS, J., GEORGE, J., VEGA, V.B., VERGARA, L., PLONER, A., PAWITAN, Y., HALL, P., KLAAR, S., LIU, E.T. *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proceedings of the National Academy of Sciences of the United States of America* **102**(38):13550–13555 (2005)
- MINN, A.J., GUPTA, G.P., SIEGEL, P.M., BOS, P.D., SHU, W., GIRI, D.D., VIALE,

- A., OLSHEN, A.B., GERALD, W.L., Y MASSAGUÉ, J. Genes that mediate breast cancer metastasis to lung. *Nature* **436**(7050):518–524 (2005)
- MISHRA, A.K., AGRAWAL, U., NEGI, S., BANSAL, A., MOHIL, R., CHINTAMANI, C., BHATNAGAR, A., BHATNAGAR, D., SAXENA, S. *et al.* Expression of androgen receptor in breast cancer & its correlation with other steroid receptors & growth factors. *The Indian journal of medical research* **135**(6):843 (2012)
- MORALES, P., BLASCO-BENITO, S., ANDRADAS, C., GÓMEZ-CAÑAS, M., FLORES, J.M., GOYA, P., FERNANDEZ-RUIZ, J., SÁNCHEZ, C., Y JAGEROVIC, N. Selective, Nontoxic CB2 Cannabinoid o-Quinone with in Vivo Activity against Triple-Negative Breast Cancer. *Journal of medicinal chemistry* **58**(5):2256–2264 (2015)
- MOREIRA-FILHO, C., BANDO, S., BERTONHA, F., IAMASHITA, P., SILVA, F., COSTA, L.F., SILVA, A., CASTRO, L., Y WEN, H. Community structure analysis of transcriptional networks reveals distinct molecular pathways for early-and late-onset temporal lobe epilepsy with childhood febrile seizures. *PloS one* **10**(5):e0128174–e0128174 (2014)
- MUNRO, S., THOMAS, K.L., Y ABU-SHAAR, M. Molecular characterization of a peripheral receptor for cannabinoids (1993)
- NADERI, A. Y HUGHES-DAVIES, L. A functionally significant cross-talk between androgen receptor and ErbB2 pathways in estrogen receptor negative breast cancer. *Neoplasia* **10**(6):542–548 (2008)
- NAHTA, R. Pharmacological strategies to overcome HER2 cross-talk and Trastuzumab resistance. *Current medicinal chemistry* **19**(7):1065 (2012)
- NASSER, M.W., QAMRI, Z., DEOL, Y.S., SMITH, D., SHILO, K., ZOU, X., Y GANJU, R.K. Crosstalk between chemokine receptor CXCR4 and cannabinoid receptor CB 2 in modulating breast cancer growth and invasion. *PLoS One* **6**(9):e23901 (2011)

- NI, I.B.P., ZAKARIA, Z., MUHAMMAD, R., ABDULLAH, N., IBRAHIM, N., EMRAN, N.A., ABDULLAH, N.H., Y HUSSAIN, S.N.A.S. Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context. *Pathology-Research and Practice* **206**(4):223–228 (2010)
- NIKOLSKY, Y., NIKOLSKAYA, T., Y BUGRIM, A. Biological networks and analysis of experimental data in drug discovery. *Drug discovery today* **10**(9):653–662 (2005)
- NISHIMURA, D. BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient* **2**(3):117–120 (2001)
- NOVAC, N. Challenges and opportunities of drug repositioning. *Trends in pharmacological sciences* **34**(5):267–272 (2013)
- PARISE, C.A. Y CAGGIANO, V. Breast cancer survival defined by the ER/PR/HER2 subtypes and a surrogate classification according to tumor grade and immunohistochemical biomarkers. *Journal of cancer epidemiology* **2014** (2014)
- PARKER, J.S., MULLINS, M., CHEANG, M.C., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X., HU, Z. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology* **27**(8):1160–1167 (2009)
- PAVLOPOULOS, G.A., SECRIER, M., MOSCHOPOULOS, C.N., SOLDATOS, T.G., KOSIDA, S., AERTS, J., SCHNEIDER, R., BAGOS, P.G. *et al.* Using graph theory to analyze biological networks. *BioData mining* **4**(10):1–27 (2011)
- PAWITAN, Y., BJÖHLE, J., AMLER, L., BORG, A.L., EGYHAZI, S., HALL, P., HAN, X., HOLMBERG, L., HUANG, F., KLAAR, S. *et al.* Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Research* **7**(6):R953 (2005)
- PÉREZ-GÓMEZ, E., ANDRADAS, C., BLASCO-BENITO, S., CAFFAREL, M.M., GARCÍA-TABOADA, E., VILLA-MORALES, M., MORENO, E., HAMANN, S.,

- MARTÍN-VILLAR, E., FLORES, J.M. *et al.* Role of cannabinoid receptor CB2 in HER2 Pro-oncogenic signaling in breast cancer. *Journal of the National Cancer Institute* **107**(6):dju077 (2015)
- PEROU, C.M., SØRLIE, T., EISEN, M.B., VAN DE RIJN, M., JEFFREY, S.S., REES, C.A., POLLACK, J.R., ROSS, D.T., JOHNSEN, H., AKSLEN, L.A. *et al.* Molecular portraits of human breast tumours. *Nature* **406**(6797):747–752 (2000)
- PISANTI, S., PICARDI, P., D ALESSANDRO, A., LAEZZA, C., Y BIFULCO, M. The endocannabinoid signaling system in cancer. *Trends in pharmacological sciences* **34**(5):273–282 (2013)
- POLYAK, K. Heterogeneity in breast cancer. *The Journal of clinical investigation* **121**(10):3786 (2011)
- PRAT, A. Y PEROU, C.M. Deconstructing the molecular portraits of breast cancer. *Molecular oncology* **5**(1):5–23 (2011)
- QAMRI, Z., PREET, A., NASSER, M.W., BASS, C.E., LEONE, G., BARSKY, S.H., Y GANJU, R.K. Synthetic cannabinoid receptor agonists inhibit tumor growth and metastasis of breast cancer. *Molecular cancer therapeutics* **8**(11):3117–3129 (2009)
- RAVASZ, E., SOMERA, A.L., MONGRU, D.A., OLTVAI, Z.N., Y BARABÁSI, A.L. Hierarchical organization of modularity in metabolic networks. *science* **297**(5586):1551–1555 (2002)
- RIVENBARK, A.G., O CONNOR, S.M., Y COLEMAN, W.B. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *The American journal of pathology* **183**(4):1113–1124 (2013)
- RODY, A., HOLTRICH, U., PUSZTAI, L., LIEDTKE, C., GAETJE, R., RUCKHAEBERLE, E., SOLBACH, C., HANKER, L., AHR, A., METZLER, D. *et al.* T-cell metagene

- predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res* **11**(2):R15 (2009)
- SAEEDI-SAEEDI, H., SHAHIDSALES, S., KOOCHAK-POUR, M., SABAHI, E., Y MORIDI, I. Evaluation of Emotional Distress in Breast Cancer Patients. *Iranian journal of cancer prevention* **8**(1):36 (2015)
- SAFARPOUR, D. Y TAVASSOLI, F.A. A targetable androgen receptor-positive breast cancer subtype hidden among the triple-negative cancers. *Archives of Pathology and Laboratory Medicine* **139**(5):612–617 (2014)
- SCHAEFER, C.F., ANTHONY, K., KRUPA, S., BUCHOFF, J., DAY, M., HANNAY, T., Y BUETOW, K.H. PID: the pathway interaction database. *Nucleic acids research* **37**(suppl 1):D674–D679 (2009)
- SHAH, P.D., GUCALP, A., Y TRAINA, T.A. The role of the androgen receptor in triple-negative breast cancer. *Women's health* **9**(4):351–360 (2013)
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N.S., WANG, J.T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B., Y IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**(11):2498–2504 (2003)
- SMYTH, G.K. Limma: linear models for microarray data. En *Bioinformatics and computational biology solutions using R and Bioconductor*, págs. 397–420. Springer (2005)
- SOPHOCLEOUS, A., MARINO, S., LOGAN, J.G., MOLLAT, P., RALSTON, S.H., Y IDRIS, A.I. Bone Cell-autonomous Contribution of Type 2 Cannabinoid Receptor to Breast Cancer-induced Osteolysis. *Journal of Biological Chemistry* **290**(36):22049–22060 (2015)
- SØRLIE, T., PEROU, C.M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M.B., VAN DE RIJN, M., JEFFREY, S.S. *et al.* Gene expression

- patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences* **98**(19):10869–10874 (2001)
- SØRLIE, T., TIBSHIRANI, R., PARKER, J., HASTIE, T., MARRON, J., NOBEL, A., DENG, S., JOHNSEN, H., PESICH, R., GEISLER, S. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences* **100**(14):8418–8423 (2003)
- SOTIRIOU, C., WIRAPATI, P., LOI, S., HARRIS, A., FOX, S., SMEDS, J., NORDGREN, H., FARMER, P., PRAZ, V., HAIBE-KAINS, B. *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *Journal of the National Cancer Institute* **98**(4):262–272 (2006)
- STRITTMATTER, S.M. Overcoming drug development bottlenecks with repurposing: old drugs learn new tricks. *Nature medicine* **20**(6):590–591 (2014)
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V.K., MUKHERJEE, S., EBERT, B.L., GILLETTE, M.A., PAULOVICH, A., POMEROY, S.L., GOLUB, T.R., LANDER, E.S. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**(43):15545–15550 (2005)
- SVENSSON, S., ABRAHAMSSON, A., RODRIGUEZ, G.V., OLSSON, A.K., JENSEN, L., CAO, Y., Y DABROSIN, C. CCL2 and CCL5 are novel therapeutic targets for estrogen-dependent breast cancer. *Clinical Cancer Research* **21**(16):3794–3805 (2015)
- SWAMI, M. Signalling: the calcium connection. *Nature Reviews Cancer* **10**(11):738–739 (2010)
- TAVASSOLI, F.A. Y DEVILEE, P. *Pathology and genetics of tumours of the breast and female genital organs*, tomo 4. Iarc (2003)

- TAYLOR, J.E., HATCHER, P.E., Y PAUL, N.D. Crosstalk between plant responses to pathogens and herbivores: a view from the outside in. *Journal of experimental botany* **55**(395):159–168 (2004)
- THIKE, A.A., CHONG, L.Y.Z., CHEOK, P.Y., LI, H.H., YIP, G.W.C., BAY, B.H., TSE, G.M.K., IQBAL, J., Y TAN, P.H. Loss of androgen receptor expression predicts early recurrence in triple-negative and basal-like breast cancer. *Modern Pathology* **27**(3):352–360 (2014)
- TOVAR, H., GARCÍA-HERRERA, R., ESPINAL-ENRÍQUEZ, J., Y HERNÁNDEZ-LEMUS, E. Transcriptional master regulator analysis in breast cancer genetic networks. *Computational biology and chemistry* **59**:67–77 (2015)
- TRIPATHI, A., KING, C., DE LA MORENAS, A., PERRY, V.K., BURKE, B., ANTOINE, G.A., HIRSCH, E.F., KAVANAH, M., MENDEZ, J., STONE, M., GERRY, N.P., LENBURG, M.E., Y ROSENBERG, C.L. Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *Int J Cancer* **122**(7):1557–1566 (2008)
- VANHOOK, A.M. Ligand-Independent ER Activation. *Science Signaling* **3**(117):ec112–ec112 (2010)
- VAQUERIZAS, J.M., KUMMERFELD, S.K., TEICHMANN, S.A., Y LUSCOMBE, N.M. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* **10**(4):252–263 (2009)
- VELASCO, G., HERNANDEZ-TIEDRA, S., DAVILA, D., Y LORENTE, M. The use of cannabinoids as anticancer agents. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **64**:259–266 (2016)
- VIALE, G. The current state of breast cancer classification. *Annals of Oncology* **23**(suppl 10):x207–x210 (2012)

- VIEDMA-RODRÍGUEZ, R., BAIZA-GUTMAN, L., SALAMANCA-GÓMEZ, F., DIAZ-ZARAGOZA, M., MARTÍNEZ-HERNÁNDEZ, G., RUIZ ESPARZA-GARRIDO, R., VELÁZQUEZ-FLORES, M.A., Y ARENAS-ARANDA, D. Mechanisms associated with resistance to tamoxifen in estrogen receptor-positive breast cancer (Review). *Oncology reports* **32**(1):3–15 (2014)
- VIPHAKONE, N., CUMBERBATCH, M.G., LIVINGSTONE, M.J., HEATH, P.R., DICKMAN, M.J., CATTO, J.W., Y WILSON, S.A. Luszp4 defines a new mRNA export pathway in cancer cells. *Nucleic acids research* **43**(4):2353–2366 (2015)
- WAGNER, A.H., COFFMAN, A.C., AINSCOUGH, B.J., SPIES, N.C., SKIDMORE, Z.L., CAMPBELL, K.M., KRYSIAK, K., PAN, D., MCMICHAEL, J.F., ELDRED, J.M. *et al.* DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic acids research* pág. gkv1165 (2015)
- WANG, J., LIU, Q., Y SHYR, Y. Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC genomics* **16**(Suppl 7):S5 (2015)
- WANG, L., LI, J., ZHAO, H., HU, J., PING, Y., LI, F., LAN, Y., XU, C., XIAO, Y., Y LI, X. Identifying the crosstalk of dysfunctional pathways mediated by lncRNAs in breast cancer subtypes. *Molecular BioSystems* (2016)
- YU, P.N., YAN, M.D., LAI, H.C., HUANG, R.L., CHOU, Y.C., LIN, W.C., YEH, L.T., Y LIN, Y.W. Downregulation of miR-29 contributes to cisplatin resistance of ovarian cancer cells. *International Journal of Cancer* **134**(3):542–551 (2014)
- ZAHA, D.C. Significance of immunohistochemistry in breast cancer. *World journal of clinical oncology* **5**(3):382 (2014)
- ZAKARIA, F., EL-MASHAD, N., Y MOHAMED, D. Androgen receptor expression as a prognostic and predictive marker in triple-negative breast cancer patients. *Alexandria Journal of Medicine* (2015)

- ZARDAVAS, D., IRRTHUM, A., SWANTON, C., Y PICCART, M. Clinical management of breast cancer heterogeneity. *Nature Reviews Clinical Oncology* (2015)
- ZHANG, L., ZHOU, W., VELCULESCU, V.E., KERN, S.E., HRUBAN, R.H., HAMILTON, S.R., VOGELSTEIN, B., Y KINZLER, K.W. Gene expression profiles in normal and cancer cells. *Science* **276**(5316):1268–1272 (1997)
- ZWART, W., THEODOROU, V., Y CARROLL, J.S. Estrogen receptor-positive breast cancer: a multidisciplinary challenge. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **3**(2):216–230 (2011)

Apéndice A

Análisis de datos de microarreglos

```
1 #codigo ejecutable en R
2 # Procesamiento de datos de microarreglos
3 # Bibliotecas utilizadas
4 library(affy)
5 library(frma)
6 #posado en un directorio con archivos CEL
7 Data = ReadAffy()
8 # Robust Multi-Array summarization/normalization
9 #Esta funcion corrige fondo, y normaliza cada signal por
   cuantiles
10 frmaData <- frma(Data, summarize="robust_weighted_average") #
   tardado
11 edata<-exprs(frmaData) # MATRIZ DE EXPRESION
12           # cada columna es un microarreglo
13           # cada renglon un transcrito
14           # y cada celda una intensidad de signal
           normalizada
15
```

```

16 # Correccion de batch effect
17 # Bibliotecas utilizadas
18 library(sva)
19 N=length(Data@phenoData@data$sample)
20 GSMnames <- colnames(edata)
21 GSMnames <- data.frame(lapply(GSMnames, function(v) {
22   if (is.character(v)) return(toupper(v))
23   else return(v)
24 })))
25 GSMnames <- as.character(GSMnames)
26 colnames(edata) <- GSMnames
27 batch <- c((rep(0,N))) # lotes
28 # Enfermos
29 GSE1456<-read.table("/lists_of_names/all_GSE1456.txt",
   colClasses = "character") # This .txt files contain the list
   of GSM ID for any GSE
30 GSE1561<-read.table("/lists_of_names/all_GSE1561.txt",
   colClasses = "character") # This .txt files contain the list
   of GSM ID for any GSE
31 GSE2603<-read.table("/lists_of_names/all_GSE2603.txt",
   colClasses = "character") # This .txt files contain the list
   of GSM ID for any GSE
32 GSE2990<-read.table("/lists_of_names/all_GSE2990.txt",
   colClasses = "character") # This .txt files contain the list
   of GSM ID for any GSE
33 GSE3494<-read.table("/lists_of_names/all_GSE3494.txt",
   colClasses = "character") # This .txt files contain the list
   of GSM ID for any GSE

```



```

34 GSE4922<-read.table("/lists_of_names/all_GSE4922.txt",
    colClasses = "character") # This .txt files contain the list
    of GSM ID for any GSE
35 GSE7390<-read.table("/lists_of_names/all_GSE7390.txt",
    colClasses = "character") # This .txt files contain the list
    of GSM ID for any GSE
36 GSE1456 <- GSE1456 [[1]]
37 GSE1561 <- GSE1561 [[1]]
38 GSE2603 <- GSE2603 [[1]]
39 GSE2990 <- GSE2990 [[1]]
40 GSE3494 <- GSE3494 [[1]]
41 GSE4922 <- GSE4922 [[1]]
42 GSE7390 <- GSE7390 [[1]]
43 n1456 <- which(GSMnames %in% GSE1456)
44 n1561 <- which(GSMnames %in% GSE1561)
45 n2603 <- which(GSMnames %in% GSE2603)
46 n2990 <- which(GSMnames %in% GSE2990)
47 n3494 <- which(GSMnames %in% GSE3494)
48 n4922 <- which(GSMnames %in% GSE4922)
49 n7390 <- which(GSMnames %in% GSE7390)
50 batch[n1456] = 1
51 batch[n1561] = 2
52 batch[n2603] = 3
53 batch[n2990] = 4
54 batch[n3494] = 5
55 batch[n4922] = 6
56 batch[n7390] = 7
57 case <- c(n1456, n1561, n2603, n2990, n3494, n4922, n7390)

```

```

58 # Sanos
59 GSE15852 <-read.table("/lists_of_names/all_GSE15852.txt",
      colClasses = "character") # This .txt files contain the list
      of GSM ID for any GSE
60 GSE6883 <-read.table("/lists_of_names/all_GSE6883.txt",
      colClasses = "character") # This .txt files contain the list
      of GSM ID for any GSE
61 GSE9574 <-read.table("/lists_of_names/all_GSE9574.txt",
      colClasses = "character") # This .txt files contain the list
      of GSM ID for any GSE
62 GSE15852 <- GSE15852 [[1]]
63 GSE6883 <- GSE6883 [[1]]
64 GSE9574 <- GSE9574 [[1]]
65 n15852 <- which(GSMnames %a n % GSE15852)
66 n6883 <- which(GSMnames %a n % GSE6883)
67 n9574 <- which(GSMnames %a n % GSE9574)
68 batch[n15852] = 8
69 batch[n6883] = 9
70 batch[n9574] = 10
71 control <- c(n15852, n6883, n9574)
72 # Separar las matriz de expresion en sanos y enfermos
73 caseExp <- edata[,-control]
74 healtExp <- edata[,-case]
75 #
76 caseBatch <- c(batch[-control])
77 healtBatch <- c(batch[-case])
78 # ComBat para la correccion de efecto de lote
79 case_combat = ComBat(dat=caseExp, batch=caseBatch, mod=NULL)

```

```

80 healt_combat = ComBat(dat=healtExp, batch=healtBatch, mod=NULL)
81 # "combat_2ways" es la matriz corregida
82 combat_2ways <- matrix(rep(0,19609040), ncol=880)
83 colnames(combat_2ways) <- colnames(edata)
84 rownames(combat_2ways) <- rownames(edata)
85 x<- combat_2ways
86 write.csv(x, file = "EXPMATRIX_AFFYID.txt", sep ="\t")
87 #anotar sondas con gene symbol
88 library("annotate")
89 library("hgu133a.db")
90 x<-as.data.frame(x)
91 genesymbols<-data.frame()
92 genesymbols<-(getSYMBOL(as.character(rownames(x)), "hgu133a.db")
    )
93 sym<-as.data.frame(genesymbols)
94 rownames(x)<-rownames(sym)
95 x$GS<-genesymbols
96 x$GS<-ifelse(is.na(x$GS), as.vector(rownames(x)), as.vector(x$GS
    ))
97 x.med <- aggregate(. ~ GS, data = x, median) #Agregar valores de
    Genesymbol por mediana#
98 write.csv(x, file = "EXPMATRIX.txt", sep ="\t")
99 #output EXPMATRIX con un valor por GeneSymbol

```

Apéndice B

Subtipificacion molecular

PAM50

```
1 ###
2 # clasificacion con PAM50. Adaptada del material en https://
   genome.unc.edu/pubsup/breastGEO/PAM50.zip
3 # para ser ejecutado como un Rscript
4 ###
5 library(ctc)
6 library(heatmap.plus)
7
8 paramDir<- "/home/biclassifier_R" # localizacion de las
   funciones distribuidas por
9                                     #Parker en https://genome.unc
                                       .edu/pubsup/breastGEO/
                                       PAM50.zip
10 inputDir<- "/home" # localizacion de la matriz de expresion
11
```

```

12 inputFile<- "EXPMATRIX.txt" # matriz de expresion separada por
    tabuladores
13 short<-"KDB" # nombre para los archivos de salida
14
15 calibrationParameters<- NA      # ajustar los centroides de
    acuerdo al conjunto de entrenamiento
16
17 hasClinical<-FALSE      # No se cuenta con informacion de tamano
    de tumor
18 collapseMethod<-"median" # la matriz obtenida ya esta colapsada
19
20 #####
21 # run the assignment algorithm
22 #####
23
24 source(paste(paramDir," subtypePrediction_functions.R",sep="/"))
25 source(paste(paramDir," subtypePrediction_distributed.R",sep="/")
    )
26
27 ##Output: Tabla con la clasificacion de cada muestra en un
    subtipo molecular
28 ##      y una visualizacion en PCA de las asignaciones de las
    muestras

```

CITBCMST

```

1 #Asignacion de clasificaciones de citbcmst
2 #codigo ejecutable en R
3 library(citbcmst)

```

```

4 expmat_affy <- read.table(file = "EXPMATRIX_AFFYID.txt") #
   matriz de expresion
5 anotacion <- data.frame(id=rownames(exp.norm.bertheau07),
   stringsAsFactors=FALSE,
6
   row.names=rownames(exp.norm.
   bertheau07) ) #anotacion
   proporcionada por Marisa
   et al
7
8 asignacion_citbcmst <- cit.assignBcmst( data = expmat_affy ,
9
   data.annot = anotacion ,
10
   data.colId="id" ,
11
   data.colMap="id" ,
12
   citbcmst.annot=NULL,
13
   dist.method="dlda" ,
14
   plot=TRUE)
15 #regresa un data frame con los datos de asignacion para cada
   muestra

```

Apéndice C

Expresión Diferencial

```
1 #Análisis de expresion diferencial
2 #Codigo adaptable para ejecutarse en R
3 library(limma)
4
5 design<-read.table(file = "subtipos_design.txt") #un archivo con
   las muestras y sus subtipos
6 ####De la siguiente forma####
7 # muestra sanos lumA lumB Basal HER2
8 # EJE000 1 0 0 0 0
9 # EJE001 0 1 0 0 0
10 # EJE003 0 0 0 1 0
11 # EJE002 0 0 1 0 0
12 # EJE004 0 0 0 0 1
13
14 matriz_expresion <- read.table(file = "EXPMATRIX_AFFYID.txt") #
   matriz de expresion
15
```

```

16 #hacer una matriz de contrastes que defina las comparaciones
    entre grupos a realizarse
17 cont.matrix<- makeContrasts('lumA - sanos', 'lumB - sanos', '
    Basal - sanos', 'HER2 - sanos', levels = design)
18
19 #ajuste de datos de expresion a un modelo lineal
20 fit <- lmFit(object = matriz_expresion, design = design, method
    = robust)
21 #reorientar el modelo lineal en funcion de las comparaciones y
    calcular estadisticos
22 final_fit<-eBayes(contrasts.fit(fit = fit, contrasts = cont.
    matrix))
23 # Escribir archivos con los valores de comparacion
24 #recorrer los valores de coef para tener todas las comparaciones
    realizadas
25 write.table(file = "contrastos.txt", x = topTable(final_fit,
    coef = 1, adjust = 'fdr', n = length(row.names(
    matriz_expresion))))

```


Apéndice D

Enriquecimiento de vías

```
1 #Enriquecimiento de vias con GAGE
2
3 library("gage")
4 matriz_expresion <- read.table(file = "EXPMATRIX.txt") #matriz
  de expresion
5 subtipos<-read.table(file = "subtipos_factores.txt") #tabla con
  nombres de muestra y subtipo (o sanos)
6 #puede usarse una lista de conjuntos de genes
7 #por ejemplo
8 #en un directorio con unicamente pathways en formato .txt
9 file_list <- list.files()
10 Pathways<-list()
11 for(i in file_list){
12   x<-as.matrix(read.table(i, header=TRUE))
13   assign(paste(colnames(x)), i) ##nombre de columna = nombre
     variable
14   y<-(read.table(i, header=TRUE))
15   Pathways<-c(Pathways, y)
```

```

16  rm(x)
17  rm(i)
18  }
19
20  #grupos
21  sanos = which(colnames(matriz_expresion) %in% subtipos [
      subtipos$subtipos=="SANOS" ,1])
22  lumA = which(colnames(matriz_expresion) %in% subtipos [
      subtipos$subtipos=="lumA" ,1])
23
24  #El enriquecimiento se hace con la funcion GAGE
25  #a nivel de comparaciones de niveles de expresion muestra-a-
      muestra
26  luma_v_sanos <-gage(
27  data = matriz_expresion , #la matriz de expresion
28  gsets = Pathways, #conjuntos de genes a evaluarse
29  ref = sanos ,
30  samp = lumA ,
31  compare = "unpaired" #dado que los grupos no son de igual
      tamano ,
32
      #unpaired realiza comparaciones muestra-a-
      -muestra
33
      #y sumaliza los valores promedio de
      comparaciones
34
      #recomendado para este caso por los
      autores
35  )
36

```

```
37 #y se descartan los pathways no significativos
38 luma_v_sanos_sig<-sigGeneSet(luma_v_sanos)
39 #repetir para otros subtipos
```

Apéndice E

Redes de Entrecruzamiento

```
1 #Funcion jaccard_matrix
2 #Calcula indice de Jaccard entre dos listas de conjuntos (ej, de
   genes)
3 jaccard_matrix<-function(x,y){
4   matriz<-matrix(nrow = length(x), ncol = length(y))
5   colnames(matriz)<-names(y)
6   rownames(matriz)<-names(x)
7   for (i in seq_along(x)){
8     for (j in seq_along(y)){
9       alfa<-x[[i]]
10      beta<-y[[j]]
11      matriz[i,j]<-jaccard(alfa,beta)
12    }
13  }
14  print(matriz)
15  return(matriz)
16 }
17
```

```

18 ##jaccard
19 jaccard<-function(a,b)
20 {
21   x<-intersect(a,b)
22   y<-union(a,b)
23   nx<-length(x)
24   ny<-length(y)
25   J<-as.numeric(nx/ny)
26   print(J)
27 }
28
29 ##
30 #leer archivos de pathways
31 #en un directorio con unicamente pathways en formato .txt
32 file_list<-list.files()
33 Pathways<-list()
34 for(i in file_list){
35   x<-as.matrix(read.table(i, header=TRUE))
36   assign(paste(colnames(x), i) ##nombre de columna = nombre
          variable
37   y<-(read.table(i, header=TRUE))
38   Pathways<-c(Pathways, y)
39   rm(x)
40   rm(i)
41 }
42
43 jaccard_matrix(x = Pathways, y = Pathways)

```

```
44 #regresa una matriz de Jaccard de todos los pathways contra si
    mismos
45 #Equivalente a la matriz de ayacencia de red de entrecruzamiento
    de vias pesada por Jaccard
```

Apéndice F

Minería de DGIdb

```
1 #Interrogar sistemáticamente la DGIdb
2
3 #funcion que interroga la DGIdb desde R
4 library(jsonlite)
5
6 gene_drug_interactions<-function(gen){
7   interrogador<-paste0("http://dgidb.genome.wustl.edu/api/v1/
8     interactions.json?genes=", gen)
9   query_dgidb<-fromJSON(interrogador) #USE jsonlite version
10  if(length(query_dgidb$matchedTerms)==0){
11    return(NULL)
12  }
13  name_drugs<-character()
14  drug_interaction<-character()
15  for(i in seq_along(query_dgidb$matchedTerms$interactions[[1]]
16    $drugName)){
17    name_drugs<-c(name_drugs,
18      query_dgidb$matchedTerms$interactions[[1]]$drugName[i])
19  }
```

```

16     drug_interaction<-c(drug_interaction ,
17         query_dgidb$matchedTerms$interactions [[1]]
18         $interactionType[i])
19     }
20     names(drug_interaction)<-name_drugs
21     return(drug_interaction)
22 }
23 #Si se tiene un conjunto de genes de interes
24 #Por ejemplo, aquellos genes desregulados en cancer
25 genes # vector de caracteres con los genes
26 interacciones <- apply(X = genes, FUN = gene_drug_interactions)

```