



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE MEDICINA
BIOMEDICINA

**ANÁLISIS INTEGRATIVO DE VÍAS ASOCIADAS A APOPTOSIS EN CÁNCER DE
MAMA: UN ENFOQUE DE BIOLOGÍA DE SISTEMAS**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

GARCÍA CAMPOS MIGUEL ANGEL

TUTOR PRINCIPAL DE TESIS: DR. ENRIQUE HERNÁNDEZ LEMUS

INSTITUTO NACIONAL DE MEDICINA GENÓMICA

COMITÉ TUTOR: DR. ALFREDO HIDALGO MIRANDA

INSTITUTO NACIONAL DE MEDICINA GENÓMICA

DRA. MARIANA BENITEZ KEINRAD

INSTITUTO DE ECOLOGÍA, UNAM

MÉXICO, CD. MX., NOVIEMBRE, 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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
POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE MEDICINA
BIOMEDICINA

**ANÁLISIS INTEGRATIVO DE VÍAS ASOCIADAS A APOPTOSIS EN CÁNCER DE
MAMA: UN ENFOQUE DE BIOLOGÍA DE SISTEMAS**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

GARCÍA CAMPOS MIGUEL ANGEL

TUTOR PRINCIPAL DE TESIS: DR. ENRIQUE HERNÁNDEZ LEMUS

INSTITUTO NACIONAL DE MEDICINA GENÓMICA

COMITÉ TUTOR: DR. ALFREDO HIDALGO MIRANDA

INSTITUTO NACIONAL DE MEDICINA GENÓMICA

DRA. MARIANA BENITEZ KEINRAD

INSTITUTO DE ECOLOGÍA, UNAM

MÉXICO, CD. MX., NOVIEMBRE, 2016

Lic. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
Presente

Me permito informar a usted que el Subcomité de Biología Experimental y Biomedicina del Posgrado en Ciencias Biológicas, en su sesión ordinaria del día 19 de septiembre de 2016, aprobó el jurado para la presentación del examen para obtener el grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** del alumno **GARCÍA CAMPOS MIGUEL ANGEL** con número de cuenta **408030452**, con la tesis titulada "**ANÁLISIS INTEGRATIVO DE VÍAS ASOCIADAS A APOPTOSIS EN CÁNCER DE MAMA: UN ENFOQUE DE BIOLOGÍA DE SISTEMAS**", realizada bajo la dirección del **DR. ENRIQUE HERNÁNDEZ LEMUS**:

Presidente: DR. PEDRO EDUARDO MIRAMONTES VIDAL
Vocal: DRA. MARÍA DEL CARMEN MEJÍA VÁZQUEZ
Secretario: DR. ALFREDO HIDALGO MIRANDA
Suplente: DRA. ALMA DELIA CAMPOS PARRA
Suplente: DRA. MARIANA BENITEZ KEINRAD

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, Cd. Mx., a 10 de noviembre de 2016

M. del Coro Arizendi

DRA. MARÍA DEL CORO ARIZMENDI ARRIAGA
COORDINADORA DEL PROGRAMA



AGRADECIMIENTOS

Agradezco al Posgrado en Ciencias Biológicas, UNAM por proveerme de una plataforma sólida para realizar mi investigación con investigadores capaces, productivos y con autoridad en su área a nivel nacional e internacional.

Agradezco el apoyo recibido por parte del Consejo Nacional de Ciencia y Tecnología por medio de la beca número 385288, durante toda la duración de mis estudios. Agradezco el apoyo otorgado por la Coordinación de Estudios de Posgrado de la Universidad Nacional Autónoma de México por medio del Programa de Apoyo a los Estudios de Posgrado para realizar una estancia de investigación en el extranjero.

Agradezco a mi tutor el Dr. Enrique Hernández Lemus y a mi comité tutor el Dr. Alfredo Hidalgo Miranda y la Dra. Mariana Benítez Keinrad por brindarme su experiencia y guía académica, durante mis estudios de maestría y la elaboración del proyecto de investigación.

AGRADECIMIENTOS A TÍTULO PERSONAL

A Rocío Campos, aquella que me dio la vida. Te agradezco tu esfuerzo incansable por darme lo mejor y hacer siempre lo que crees que es bueno para mí. Sé que no ha sido nada fácil y por eso te quiero agradecer con mi trabajo y mis logros, porque también son tuyos. Te quiero mucho.

A Rosa Amelia, la que dio vida a quien me dio la mía. Te agradezco ser una luz en mi camino, los momentos que he vivido contigo y las experiencias que me has dado son tesoros muy valiosos para mí. Espero que mi trabajo te haga orgullosa. Besitos.

A Gerardo Real, el único que me ha soportado tal y como soy. Te agradezco todo el apoyo que me has brindado estos años. Das sentido a mi vida, y me pones los pies en la tierra cada vez que vuelo sin control, sin ti seguramente esta tesis no hubiera sido escrita. Pase lo que pase, siempre te querré.

Al Dr. Enrique Hernández, un guía inspirador. Le agradezco todo su apoyo y guía desde el primer día que puse pie en su laboratorio. Sin lugar a dudas estos dos años con usted me han hecho madurar enormemente como investigador y como persona, y me marcaran de por vida, lo cual agradezco. Gracias por ser un gran mentor y modelo a seguir, estoy seguro que seguiremos colaborando, un abrazo.

Al Dr. Jesús Espinal, amigo y colega. Te agradezco tu apoyo y compañerismo, las correcciones hechas a mis proyectos y las risas que me diste en estos dos años. Espero que sean muchas más, un abrazo.

A los Drs. Alfredo Hidalgo, Mariana Benítez, Pedro Miramontes, Ma. del Carmen Mejía y Alma Parra. Gracias por su tiempo y esfuerzo para corregirme, escucharme y darme sabios consejos. Sus aportaciones no sólo enriquecieron mi proyecto sino también a mí. Gracias.

A todos mis compañeros y amigos: Daniel, Sergio, Hugo, Guillermo, Diana, Tadeo, Rodrigo, Raúl, Joshua, Ollin, Aldo, Ana, Karol, Mireya, Kahorik, Luis, Roselyn, Estefanía, Erandi, Fernanda, Garbiñe, Violeta, Adriana, Miguel, Mónica, Lenin y los que me falten, les agradezco por todo lo que me han compartido: platicas, discusiones, comentarios, comidas, paseos, compañía, apoyo, amistad, alegría y más. Mi trabajo y vida no hubieran sido lo mismo sin todos ustedes. Espero que sigamos compartiendo alegrías. Gracias a todos.

A mis mejores amigos de la vida: Leopoldo, Rubén, Tania, Jorge, Jesús, Ricardo, Eder. Gracias por su amistad, sin ustedes el viaje que ha sido mi vida sería absurdo, y yo no sería la mitad de lo afortunado y feliz sin haberlos conocido.

ÍNDICE

Lista de figuras, tablas y anexos	i
Resumen	1
Abstract	2
Introducción	3
Cáncer	3
Cáncer de mama	3
Apoptosis y cáncer	4
Autofagia y apoptosis	5
Antecedentes	6
Análisis transcriptómicos del cáncer	6
Biología de sistemas	6
Objetivos	9
Metodología	10
Descripción general de la metodología	10
Obtención y curado de los datos experimentales	11
Preprocesamiento de datos de mRNA	12
Información funcional de procesos biológicos	12
Análisis de desregulación y agrupamiento jerárquico	13
Inferencia de RRG	13
Análisis de RRG	14
Análisis de expresión diferencial	14
Visualización de redes	15
Subtipificación PAM50 y comparación de subgrupos	15
Validación de resultados de subtipificación	15
Resultados	17
Información funcional de procesos biológicos	17
Análisis de desregulación y agrupamiento	17
Inferencia y análisis de RRG	24
Densidad	25
Genes altamente influyentes	28
Análisis de expresión diferencial	29
Visualización de RRG	29
Subtipificación PAM50 y enriquecimiento de subtipos intrínsecos	32
Validación de resultados de enriquecimientos de subgrupos	34
Discusión	36
Análisis de desregulación y subgrupos tumorales	36
Densidad de redes, procesos biológicos y subtipos	39
Genes altamente influyentes de apoptosis y autofagia en cáncer de mama	41
Visualización de redes por subgrupo	42
Subtipificación PAM50 y comparación de subgrupos	45
Validación de resultados de enriquecimiento	45
Conclusiones	47
Literatura citada	49
Anexos	56

LISTA DE FIGURAS Y TABLAS

Figura 1. Esquema de la metodología	11
Figura 2. Distribuciones de resultados análisis de desregulación con Pathifier	17
Figura 3. <i>Heatmap</i> del agrupamiento jerárquico Sanos-Enfermos y Apoptosis-Autofagia	18
Figura 4. <i>Heatmap</i> del agrupamiento jerárquico con etiquetas para los 5 subgrupos de muestras	19
Figura 5. <i>Heatmap</i> del PDS promedio por subgrupo	20
Figura 6. Diagrama de cajas y bigotes con los valores de desregulación en PDS promedio para cada subgrupo	21
Figura 7. Los 10 procesos con mayor PDS promedio	22
Figura 8. Los 10 procesos con menor PDS promedio	22
Figura 9. Los 10 procesos con menor desviación estándar	23
Figura 10. Los 10 procesos con mayor desviación estándar	24
Figura 11. Diagramas de cajas y bigotes de valores de p para cada red. Se muestran la distribución de los valores de p de los enlaces en las redes inferidas para cada subgrupo	25
Figura 12. Diagramas de barras de densidad entre los genes de 3 categorías: Apoptosis, Autofagia y genes que participan en ambos procesos	27
Figura 13. Diagrama de Venn de genes influyentes que comparten los subgrupos tumorales	28
Figura 14. Visualización de la red del Subgrupo de muestras de tejido sano	29
Figuras 15A y 15B. Visualizaciones de las redes de los subgrupos 1 y 2	31
Figuras 15C y 15D. Visualizaciones de las redes de los subgrupos 3 y 4	32
Figura 16. Resumen del contraste en los niveles de desregulación de procesos de apoptosis y autofagia en los subgrupos SG1-Basal, SG3-LuminalA y SG4-LuminalB	39
Tabla 1. Muestras empleadas en el estudio	12
Tabla 2. Valores de p para los enlaces de las redes inferidas	25
Tabla 3. Resultados de la clasificación PAM50	33
Tabla 4. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos de desregulación	34
Tabla 5. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos del conjunto de datos de TCGA	35
Tabla 6. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos del conjunto de datos de METABRIC	35
Tabla 7. Genes candidatos a participar en ambos procesos	47
Tabla 8. Genes más influyentes con mayor log Fold-Change por subtipo tumoral	48
Anexo 1. Código SQL para consulta genérica de términos GeneOntology en GOOSE	56
Anexo 2. <i>Heatmap</i> mostrando los resultados del agrupamiento jerárquico a un punto de corte de 6 subgrupos	56
Anexo 3. Gráficos lineales en el que se representa la cobertura por los nodos más influyentes	57
Anexo 4. Tablas que contienen los genes más influyentes para cada red	58
Anexo 5. Visualizaciones de las RRG en mayor tamaño	62
Anexo 6. Agrupamiento jerárquico de genes de apoptosis y autofagia	70

RESUMEN

El siguiente trabajo de tesis presenta el análisis computacional de datos transcriptómicos de un conjunto de datos de cáncer de mama. En búsqueda de relaciones funcionales entre los procesos de apoptosis y autofagia, que sean relevantes para los distintos fenotipos tumorales. El cáncer de mama es una de las neoplasias más comunes del mundo y la relativa facilidad con la que se pueden obtener biopsias de tejido tumoral, permite un análisis detallado de estos procesos en un contexto con múltiples fenotipos, similar al que se obtendría de una población general. Diversos estudios han remarcado las diferencias entre distintos subtipos de tumores, y nuevas herramientas de análisis de datos transcriptómicos permiten predecir el estado de los procesos biológicos y las relaciones entre los genes. Sin embargo, no se han realizado análisis computacionales dirigidos a procesos biológicos específicos, mucho menos a una intersección de los mismos. La presente metodología apela a encontrar diferencias entre los tumores sólo a la luz de los dos procesos: apoptosis y autofagia; además de predecir funciones y patrones no antes identificados para estos procesos y sus genes, o cuya relevancia no ha sido identificada su relevancia en cáncer de mama. Los resultados presentados establecen una relación importante entre los procesos de autofagia y apoptosis, particularmente en los procesos y genes relevantes a la degradación de la mitocondria, y propone genes no identificados en ambos procesos como candidatos a experimentación subsecuente para validar su rol ambivalente. Finalmente establecen una correlación directa entre los subtipos moleculares intrínsecos, ampliamente usados en la investigación clínica, y la desregulación coordinada de los procesos de apoptosis y autofagia.

ABSTRACT

The following thesis work presents the computational analysis of transcriptomic data from breast cancer samples, in the search of functional relationships between the apoptosis and autophagy processes that are relevant to the different tumoral phenotypes. Breast cancer being one of the most common cancers enables a detailed analysis of these processes, nevertheless not all tumors are the same. Different studies have remarked the differences between tumoral subtypes, and new data analysis tools are enabling to predict the integrated state of biological processes and the relationships between their genes. However not many global computational analyses have been directed to specific biological processes, much less to the intersection between processes. The following methodology appeals to find differences between the tumors only at the light of these two processes: apoptosis and autophagy; besides predicting the functions and patterns not yet identified for these processes and their genes, or that have not been identified as important in breast cancer. Our results establish an important relationship between the processes of autophagy and apoptosis, particularly in the processes and genes relevant to mitochondria degradation, and propose genes not yet identified as part of both processes as experimental candidates for further experimentation and validation of their ambivalent role. Finally establishing a direct correlation between molecular intrinsic subtypes, widely used on clinic investigation, and the coordinated deregulation of the apoptotic and autophagic processes.

INTRODUCCIÓN

Cáncer

El cáncer es una enfermedad que ha adquirido gran relevancia a nivel mundial, debido al impacto que ha generado en las economías y sociedades humanas, siendo una de las enfermedades con mayores tasas de morbilidad y mortalidad. Sólo en el año 2012 se registraron alrededor de 14 millones de nuevos casos y 8.2 millones de muertes relacionadas al cáncer (World Health Organization, 2012). Fundamentalmente el cáncer se caracteriza por una proliferación exacerbada de células anormales que invaden otros tejidos y órganos, generando fallas en los sistemas donde se originan y en los que invaden, siendo esta la principal causa de muerte. Sin embargo, el cáncer es un término genérico que se atribuye a un gran grupo de enfermedades denominadas por el tipo celular del cual se origina esta proliferación celular anormal.

Cáncer de mama

El cáncer de mama es el cáncer originado en tejido mamario, más comúnmente se desarrolla en las células que bordean los ductos lácteos y los lóbulos que les proveen la leche a los ductos. El cáncer de mama es uno de los más estudiados debido a su alta incidencia en la población femenina a nivel mundial. El cáncer de mama es el cáncer más frecuente en la población femenina, en el año 2012, se registraron más de un millón y medio de casos (25% de todos los nuevos casos de cáncer) y 522 mil muertes, mientras que más de seis millones de mujeres sobrevivieron el año con la enfermedad (a 5 años de diagnóstico previo) (Ferlay, 2013).

El cáncer de mama ha sido ampliamente caracterizado a nivel molecular y se han identificado diversas fuentes de variación biológica, heredadas y adquiridas, que contribuyen a la iniciación y progresión del cáncer de mamá. Por ejemplo, mutaciones heredadas en los genes BRCA1 y BRCA2 confieren un riesgo de por vida de generar cáncer de mama de 50 a 85 por ciento (Miki et al., 1994; Wooster et al., 1995; Ford et al., 1994; Easton et al., 1995; Ford et al., 1998; Thorlacius et al., 1998). También se ha observado que perfiles de expresión génica pueden predecir el resultado clínico de pacientes con cáncer de mama (Van't Veer et al., 2002; Parker et al., 2009), así como la probabilidad de la existencia de metástasis distales (Wang et al., 2005), proveyendo así de pronósticos generados a partir de información molecular. Más recientemente, uno de los estudios más amplios y ambiciosos en torno al cáncer de mama, analizó datos genómicos y transcriptómicos de alrededor de 2000 tumores, generando una clasificación de subgrupos tumorales de acuerdo a diversos rasgos genéticos heredados y adquiridos, tales como variantes en el número de copias y polimorfismos de un solo nucleótido, así como aberraciones en número de copia (Curtis et al., 2012). Por lo que el entendimiento del cáncer de mama en las últimas dos décadas se ha centrado en entender las diferencias características entre los fenotipos tumorales y sanos a nivel molecular.

Apoptosis y cáncer

Ya desde los años setentas Kerr y colaboradores proponen a la apoptosis como un proceso biológico involucrado en el remplazo de células en tejidos adultos sanos y en la eliminación focal de células durante el desarrollo embrionario, entre otras funciones (Kerr et al., 1972). Además de ser partícipe en procesos tumorales y en la regresión tumoral en respuesta a terapia (Kerr et al., 1972).

La apoptosis es un proceso de muerte celular programada que comienza cuando una célula recibe una señal interna (e.g. daño a DNA) o externa (e.g. un ligando de muerte extracelular), y procede por una serie de eventos bioquímicos (fase de vía de señalización) lo que desencadena una fase de ejecución. La fase de ejecución es el último paso de un proceso apoptótico, y típicamente está caracterizado por un redondeo de la célula, retracción de pseudópodos, reducción del volumen celular (picnosis), condensación de la cromatina, fragmentación nuclear (cariorrhexis), vesicularización de la membrana y fragmentación de la célula en cuerpos apoptóticos, siempre rodeados de membrana plasmática. Cuando la fase de ejecución ha sido completada, la célula ha muerto (Kroemer et al., 2009).

Se ha propuesto que en estado fisiológico la apoptosis se encuentra en equilibrio para mantener la homeostasis del organismo y ser la contraparte de la mitosis que ocurre de manera incesante en el organismo (Elmore, 2007). La maquinaria molecular apoptótica se puede dividir conceptualmente en 2 conjuntos de biomoléculas: sensoras y efectoras (Hanahan & Weinberg, 2000; 2011). Las moléculas sensoras se encargan de integrar distintas señales al interior y exterior de la célula estableciendo un estado de iniciación o inhibición de la ejecución de la apoptosis. Por otro lado, las moléculas efectoras serían las encargadas de ejecutar la muerte por apoptosis. En estado fisiológico existe un equilibrio entre los sensores pro-apoptóticos y anti-apoptóticos de tal manera que las células en los tejidos no inician la apoptosis ante cualquier estímulo, pero lo hacen si la integración de señales indica que es necesario para salvaguardar la integridad del organismo. Los mecanismos de evasión de apoptosis tumoral rompen el equilibrio de las señales promoviendo un estado antiapoptótico.

Estos mecanismos de evasión de la apoptosis conceptualmente pueden dividirse en dos estrategias: un aumento en las señales anti-apoptóticas y en una disminución o falla de las señales pro-apoptóticas; ambas estrategias son encontradas frecuentemente en los tumores. Por ejemplo, mutaciones en el gen p53 evitan la apoptosis debido a que es un gen que codifica para la proteína del mismo nombre cuya función es censar el daño en el DNA y promover la expresión de efectores de la apoptosis tales como BAX (Hanahan & Weinberg, 2000; 2011).

Muchas de las señales que provocan la apoptosis convergen en la mitocondria. Por ejemplo, la vía intrínseca necesita de la disrupción de la membrana mitocondrial y la liberación de proteínas mitocondriales incluyendo a Smac/DIABLO, HtrA2, y citocromo c (Johnstone et al., 2002). Miembros de la familia Bcl-2 controlan la señalización de muerte a través de la liberación del citocromo c, y p53 aumenta la expresión de Bax que a su vez estimula a la mitocondria para liberar citocromo c (Hanahan y Weinber, 2000).

Autofagia y apoptosis

La autofagia es un proceso por el cual las células digieren partes de su propio citoplasma, permitiendo tanto el reciclaje de componentes macromoleculares bajo condiciones de estrés, además de otras condiciones como el remodelado de las estructuras intracelulares en la diferenciación celular (Klionsky & Emr, 2000). Aparte del reciclado de componentes celulares, una función primordial de la autofagia es establecer una respuesta eficaz ante distintos estresores como lo son la falta de energía y nutrientes, sobrecarga del retículo endoplásmico, patógenos, hipoxia, estrés oxidativo y daño mitocondrial (Kroemer et al., 2010). Así mismo, la autofagia mantiene una estrecha relación con el proceso de apoptosis.

Esta estrecha relación entre la autofagia y la apoptosis existe, ya que comparten distintos tipos de entrecruzamiento: tanto en roles funcionales, como la muerte celular programada (Kroemer et al., 2009); como en la maquinaria molecular que emplean para su regulación (Thorburn, 2008). De forma importante una conexión macromolecular para estos dos procesos es la mitocondria, ya que una de las vías principales de iniciación de la apoptosis es a través de modificaciones en este organelo y liberación de citocromo c. Mientras que los procesos autofágicos son capaces de degradar las mitocondrias de forma específica antagonizando la señalización apoptótica (Eisenberg-Lerner et al., 2009). Estos entrecruzamientos pueden llegar a ser más complicados, como se ha observado ante estrés de retículo endoplásmico, la funcionalidad de ambos procesos es necesaria para la sobrevivencia celular. Si la apoptosis no es funcional, la autofagia puede promover necrosis en lugar de mediar el estrés por retículo endoplásmico (Ullman, et al. 2008).

El rol de la autofagia como un mecanismo de reciclado y destrucción específica de organelos ha sido extensamente descrito como un arma de doble filo para la célula, ya que puede ser tanto un mecanismo antitumoral como protumoral. Por ejemplo, en cáncer de mama se ha demostrado que la autofagia puede retrasar la apoptosis mediada por daño a DNA provocada por administración de camptotecina (Abedin et al., 2007). Por otro lado, en ratones la disrupción del gen de autofagia *BECN1*, vía delección monoalélica, ha demostrado tener efectos tumorigénicos, colocando a la autofagia como un proceso antitumoral. Sin embargo, la autofagia también puede promover la sobrevivencia de los tumores. Tal y como se ha observado en cultivos celulares de colon, recto y mama, que ante el tratamiento radioactivo, muestran una alza significativa en su tasa de muerte celular cuando se aplica simultáneamente 3-metiladenina como inhibidor de la autofagia (Paglin et al., 2001). De esta forma, el entrecruzamiento entre los procesos de apoptosis y autofagia es un fenómeno complejo, con muchas lagunas en su conocimiento y con oportunidad para hallazgos y una mejor comprensión de su relevancia en el establecimiento de los fenotipos tumorales.

ANTECEDENTES

El estudio de los procesos de apoptosis y autofagia ha pasado de la observación en el microscopio y la definición de sus elementos descriptivos a la cuantificación y observación de sus moléculas más fundamentales en poco menos de medio de siglo. Entre las tecnologías que han hecho posible estas determinaciones se encuentran los microarreglos de expresión, que han hecho posible la medición cuantitativa de miles de moléculas de RNA mensajero presentes en un tejido de forma simultánea. Con estos datos de expresión en continuo crecimiento ha sido posible realizar comparaciones de perfiles de expresión de RNA en cáncer.

Análisis transcriptómicos del cáncer

Los primeros análisis a gran escala de perfiles de expresión de RNA en cáncer, comenzaron a publicarse a finales de los años noventa (Zhang et al., 1997, Golub et al., 1999). Hallazgos como la identificación de genes con grandes diferencias en sus niveles de expresión entre tejidos tumorales y sanos (Zhang et al., 1997), clasificaciones tumorales basadas únicamente en la expresión de genes en cáncer como leucemia (Golub et al., 1999) y cáncer de mama (Van't Veer et al., 2002). Sin embargo, las conclusiones de estos estudios no consideran las propiedades emergentes que sólo ocurren en el sistema en conjunto, pero no en sus partes individuales (Rhodes & Chinnaiyan, 2005). Esta omisión no fue obstinada, sino que enfoques integrativos que buscaran este tipo de propiedades aún se encontraban en un estado incipiente en la biología molecular y genómica.

Para lograr una integración entre los niveles de expresión génica y las funciones biológicas, fue necesaria la construcción de estos mismos paradigmas. Gene Ontology es un proyecto que busca generar un paradigma unificador para los productos génicos, construyendo un vocabulario estructurado, definido precisamente, común y controlado para describir los roles de los genes y sus productos en todos los organismos (Ashburner et al., 2000). Para lograr este objetivo Gene Ontology ha generado tres ontologías: proceso biológico, función molecular y compartimento celular, en las que se puede argumentar, se ha logrado una unificación del conocimiento biológico común hasta nuestros días. Sin embargo, justamente esta anotación muestra que nuestro conocimiento de los roles de los genes y sus productos se encuentra lejano de ser completo tanto en humano como en otras especies. Además de que la inclusión de nuevas moléculas que regulan el funcionamiento de los procesos biológicos como son los RNAs no codificantes en todas sus diversas formas, las modificaciones epigenómicas y post-transcripcionales, añaden cada vez más capas de complejidad a nuestro conocimiento biológico, que con cada pregunta resuelta, formula cien más. A pesar de que los proyectos de anotación como Gene Ontology no son perfectos, representan la mejor guía para realizar análisis funcionales, como el presente en este trabajo, en los que se establecen relaciones a priori con los conocimientos obtenidos de forma experimental para inferir potenciales nuevos hallazgos biológicos.

Biología de sistemas. Enfoque y herramientas

Encontrando las respuestas más evidentes a lo que hemos observado y la formulación de preguntas cada vez más complejas y difíciles de responder, surge la necesidad de integrar toda la información disponible para reducir el universo posible de búsqueda y por ende experimentación. Puesto que

realizar todos los experimentos posibles a todas las biomoléculas en todas las condiciones es una tarea imposible incluso para toda la humanidad, es necesario generar estrategias que guíen el diseño experimental y nuestra exploración de la forma más informada posible. Esta es una de las razones fundamentales de la emergencia y necesidad de la Biología de Sistemas como una ciencia predictiva, basada no sólo en el conocimiento biológico sino integrando distintas áreas científicas como la física, las matemáticas y los métodos computacionales, en una ciencia interdisciplinaria que busca guiar nuevos hallazgos en la biología, reducir el espacio de posibles respuestas y acelerar la generación de nuevo conocimiento.

Investigaciones desarrolladas bajo el paradigma de la biología de sistemas buscan predecir el comportamiento de los sistemas biológicos o de sus subsistemas. Ejemplos de éxito de estas investigaciones, van desde la modelación del asma crónico y los posibles efectos de fármacos administrados (Lewis et al., 2001), hasta la predicción y comprobación de la emergencia de patrones morfogénicos, como las rayas de las cebras, manchas de los leopardos o estrías en el paladar de los mamíferos; por medio de una ecuación de reacción-difusión impuesta en un sistema homogéneo (Turing, 1952; Economou et al., 2012; Hillmer, 2015). De esta forma, la capacidad de explicar distintos fenómenos sistémicos y ampliar nuestro conocimiento sobre la biología y nosotros mismos, tiene un gran potencial en la biología de sistemas y las herramientas que desarrolla.

Dentro de las herramientas que se han desarrollado en la biología de sistemas se encuentra el análisis de vías. El análisis de vías es un conjunto de herramientas usadas para dar sentido a los datos biológicos de alto rendimiento, mediciones de miles de componentes biológicos, realizadas de forma simultánea en un mismo ensayo; principalmente cuando se hacen comparaciones entre dos estados: casos y controles (García-Campos et al., 2015). Las metodologías del análisis de vías se centran en el uso del conocimiento biológico ya publicado y estructurado en bases de datos, como Gene Ontology, buscando integrarlas con los datos de alto rendimiento con métodos estadísticos y matemáticos. De esta forma, es posible identificar los procesos más desregulados, como es el caso de los métodos de sobre-representación; o en otros casos el nivel de desregulación de una vía en cada una de las muestras, como busca realizar el algoritmo Pathifier (Drier et al., 2013).

Los resultados que se obtienen del análisis de vías permiten identificar patrones de expresión que de forma manual no podrían ser apreciados o sería muy tardado calcularlos, y enfocar los esfuerzos de interpretación y análisis humano a las vías o procesos biológicos con mayor potencial para ser relevantes en el experimento. En el caso de un análisis de vías en el que se pueden obtener los niveles de desregulación de las muestras, es posible realizar subdivisiones en la población muestra con un sentido biológico basado en los niveles de expresión, que pueda ayudar a hacer más homogéneos los conjuntos de muestras y facilitar análisis subsecuentes en datos con alta heterogeneidad, como los obtenidos a partir de biopsias de distintos tipos de cáncer.

Otra herramienta conceptual y práctica utilizada en la biología de sistemas son las redes de regulación genética (RRG) y su inferencia. Partiendo de que una red está definida por un conjunto de artículos, llamados nodos, con conexiones entre ellos llamadas aristas (Newman, 2003). Las RRG son aquellas redes en las que los nodos representan biomoléculas y las aristas las relaciones que tienen entre ellas. La inferencia de las RRG consiste en representar el conjunto de dependencias estadísticas entre las

variables de un conjunto (el cual puede ser todo el conjunto de datos o un subconjunto seleccionado) por medio de una red. Cuando es aplicada a datos de expresión génica es posible realizar ingeniería reversa en las RRG de la célula relacionada (Hernández-Lemus & Rangel-Escareño, 2011). A partir de la inferencia y construcción de las RRG es posible estudiar las relaciones que se dan a nivel gen-gen en condiciones específicas, y cómo pueden cambiar entre fenotipos.

Estas herramientas han sido aplicadas para generar conocimiento básico de la fisiología molecular y patofisiología del cáncer. Así mismo, los resultados del enfoque de sistemas en la genómica básica han permitido el desarrollo de metodologías aplicadas en la clínica, que ayudan a predecir el pronóstico de los pacientes y sugerir el mejor tratamiento terapéutico. Un ejemplo de estas aplicaciones es el algoritmo de clasificación de subtipos moleculares en cáncer de mama PAM50 (Parker et al., 2009), el cual clasifica a los tumores en los 5 subtipos intrínsecos Luminal A, Luminal B, Her2, Basal y normal, con resultados que pueden asistir en la determinación del pronóstico del paciente y guiar su terapia. Otro algoritmo El algoritmo de inferencia de redes ARACNE (Margolin et al., 2006) con el cual es posible realizar ingeniería reversa de las RRG, tomando en cuenta todas las dependencias estadísticas lineales y no-lineales, es de especial utilidad en el análisis de redes de organismos eucariotas. Y el algoritmo Pathifier (Drier et al., 2013) con el cual ha sido posible identificar una mayor estratificación de los subtipos tumorales de cáncer de mama ya reportados, y encontrar altas correlaciones de esta estratificación adicional con la clínica relevante para los pacientes (Livshits et al., 2015).

Este cuerpo de antecedentes ha promovido el desarrollo del presente estudio. Aquí se plantea analizar a los procesos de apoptosis, autofagia y la interacción que tienen estos dos procesos, definidos como independientes pero con amplias coincidencias, y con una maquinaria molecular en la que muchas veces es difícil de definir un inicio, un final y los componentes que comparten.

OBJETIVOS

Objetivo general:

Identificar patrones de expresión en los genes asociados con los procesos de apoptosis y autofagia que puedan ser relevantes para el establecimiento de distintos fenotipos tumorales en cáncer de mama.

Objetivos particulares:

1. Identificación de subgrupos tumorales con características homogéneas a nivel genético para los procesos biológicos de apoptosis y autofagia en cáncer de mama.
2. Evaluar el potencial clínico de los diferentes subgrupos establecidos a partir de los patrones de expresión de genes en los procesos de apoptosis y autofagia.
3. Inferir las correlaciones más plausibles entre los genes de apoptosis y autofagia en los subgrupos ya identificados.
4. Sugerir genes y procesos biológicos con roles prioritarios e influyente en el establecimiento de los distintos fenotipos tumorales en los procesos de apoptosis y autofagia.
5. Sugerir potenciales roles funcionales de los genes de apoptosis y autofagia que no han sido confirmados de forma experimental, al no estar anotados de esta forma en las bases de datos de Gene Ontology.

METODOLOGÍA

Descripción general de la metodología

La metodología general para realizar el análisis integral de los procesos biológicos de apoptosis y autofagia en cáncer de mama es la siguiente:

- 1) **Obtención y curado de los datos experimentales de microarreglos.** Los datos experimentales empleados para el estudio se obtuvieron del repositorio Gene Expression Omnibus (GEO). GEO es un repositorio público internacional, en el que se recaba la información experimental de estudios genómicos publicados en diversas revistas científicas, esto bajo protocolos que hacen comparables los datos en términos de procesamiento experimental.
- 2) **Recabado de información funcional.** Esta información se obtuvo de bases de datos curadas por expertos, en las que se realizan anotaciones exhaustivas asignando pertenencias o funciones a genes con evidencias experimentales e inferidas computacionalmente.
- 3) **Análisis de desregulación y agrupamiento jerárquico.** Este tipo de análisis plantea el estudio de la información genómica por conjuntos de genes con una relación funcional, en contraposición al estudio de genes de forma individual. A través de estos métodos es posible realizar agrupamientos de las muestras con un sentido biológico.
- 4) **Inferencia y análisis de redes de regulación genética (RRG).** Se usaron métodos basados en teoría de la información y estadística para inferir las relaciones funcionales entre los genes. Se analizó las relaciones entre los genes de apoptosis y autofagia, así como su influencia en la red.
- 5) **Análisis complementarios.** Se incorporaron los resultados de los análisis de desregulación y de redes, con análisis complementarios de subtipificación tumoral molecular PAM50 y de expresión diferencial, buscando una síntesis coherente.
- 6) **Integración de resultados.** Se cotejaron los resultados buscando coherencia entre los análisis realizados. Los subgrupos obtenidos a partir del análisis de desregulación se compararon con la clasificación de subtipos tumorales intrínsecos de PAM50, mientras que el análisis de expresión diferencial sirvió para encontrar patrones de expresión en las RRG.

Un esquema de la metodología se encuentra en la figura 1. La metodología antes descrita es detallada a continuación.

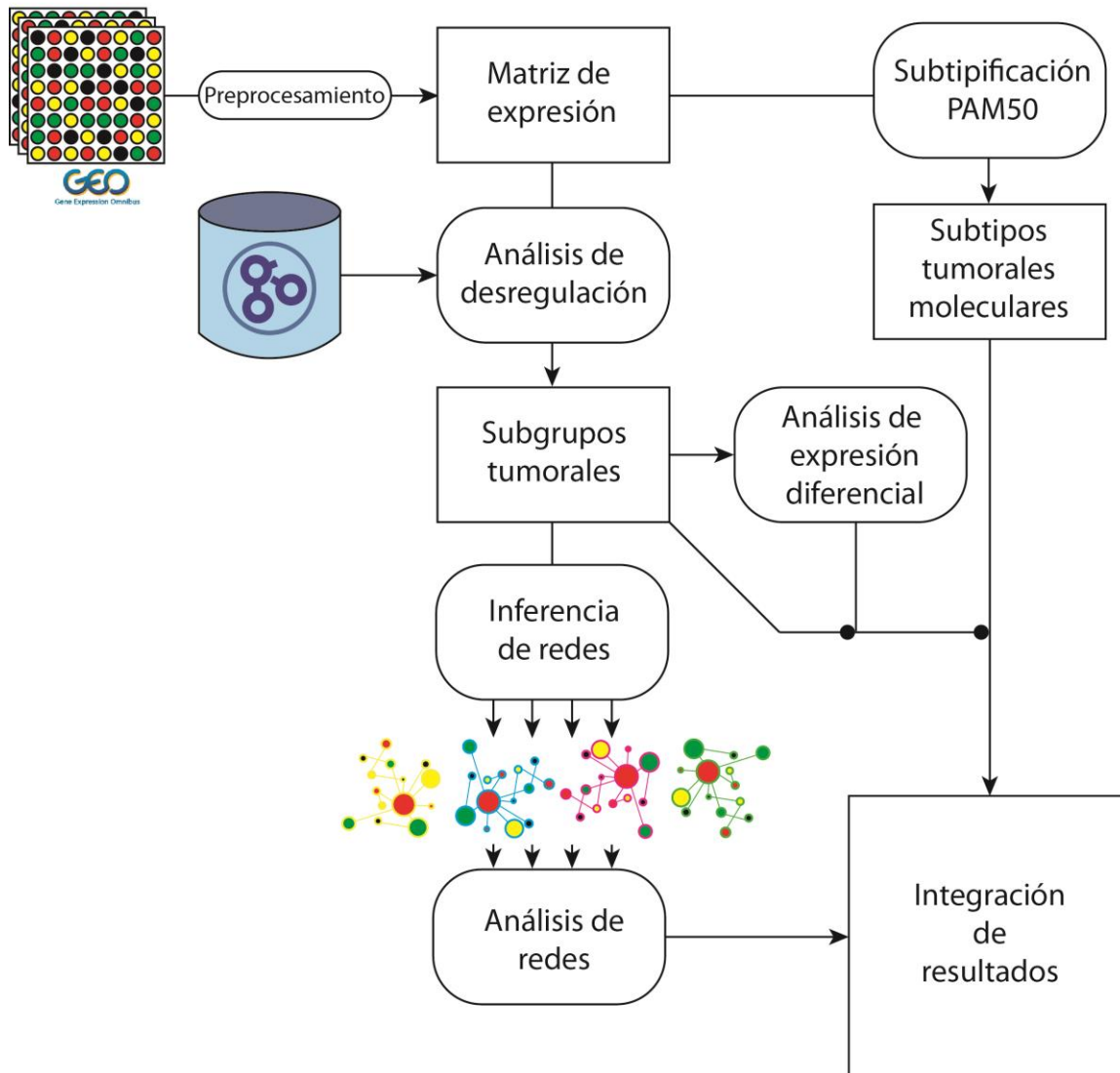


Figura 1. Esquema de la metodología. El análisis inicia con el preprocesamiento de los datos de microarreglos de las muestras seleccionadas del repositorio GEO, obteniendo una matriz de expresión conteniendo los datos de expresión de cada gen para todas las muestras. El cilindro con el logotipo de Gene Ontology representa la información de procesos biológicos seleccionados para el análisis de desregulación de procesos de apoptosis y autofagia. En el análisis de desregulación se obtienen grupos de muestras con patrones de desregulación similares entre sí, utilizando la información de los procesos biológicos de GO. Usando esta agrupación para cada subgrupo se infirió una RRG y se realizó análisis de expresión diferencial. Por otro lado, a partir de la matriz de expresión inicial se realizó subtipificación de las muestras por PAM50 (Parker et al., 2009). Finalmente se integraron los resultados del análisis de desregulación, análisis de redes, subtipificación por PAM50 y expresión diferencial.

Obtención y curado de datos experimentales

La obtención de datos genómicos es difícil debido a su alto costo y a otros factores sociales y epidemiológicos, por lo que el uso de información recabada en bases de datos públicas es óptimo para un análisis extensivo de la genómica del cáncer de mama. La información experimental se obtuvo del repositorio público de datos de expresión genómica *Gene Expression Omnibus* (GEO). La plataforma de

medición seleccionada fue el microarreglo HGU-133A V.1 de Affymetrix, abarcando los datos de 10 experimentos independientes. El tamaño total de muestra fue de 880 individuos (Baca-López et al., 2012). El total de las muestras tumorales es de 819 y las muestras de tejido sano 61, sus respectivos números de identificador GEO y referencias a su publicación original se presentan en la Tabla 1. Todos los individuos muestreados son de sexo femenino. Las biopsias tumorales son de cáncer de mama primario sin exposición a tratamiento farmacológico previo. Los datos control se obtuvieron de biopsias de tejido histológicamente sano adyacente al tumoral (GSE15852), operaciones por mamoplastía sin evidencias aparentes de cáncer de mama (GSE9574) y de biopsias de tejido epitelial de individuos sanos (GSE6883).

Identificador GEO	Tumores	Controles	Referencias
GSE1456	159	-	Pawitan et al., 2005
GSE4922	249	-	Ivshina et al., 2006
GSE7390	198	-	Desmedt et al., 2007
GSE2603	99	-	Minn et al., 2005
GSE2990	61	-	Sotiriou et al., 2006
GSE3494	4	-	Miller et al., 2005
GSE1561	49	-	Farmer et al., 2005
GSE15852	-	43	Farmer et al., 2005
GSE9574	-	15	Tripathi et al., 2007
GSE6883	-	3	Liu et al., 2007
TOTAL	819	61	

Tabla 1. Muestras empleadas en el estudio. Desglose del origen de las muestras con respecto a su identificador GEO y al tipo de muestra.

Preprocesamiento de datos de mRNA

Para realizar el análisis de los datos genómicos recabados, es necesario su preprocesamiento haciendo comparables los datos entre distintos experimentos. Los datos crudos se obtuvieron en formato CEL del repositorio de GEO. Se implementó el algoritmo RMA (Robust-Multi-array Average) con el paquete *Affy* (Gautier, 2004) alojado en el proyecto Bioconductor (Gentleman, 2004). El preprocesamiento se realizó en una unidad de procesamiento múltiple simétrico dual-core 8-Power5+ de IBM, con 128 GB disponibles en RAM. Los parámetros usados fueron: sin corrección de fondo, normalización por cuantiles y ajuste por *Median Polish*, finalmente resumiendo la información de distintos juegos de sondas a un valor para cada gen, en formato Gene Symbol para cada microarreglo.

Información funcional de procesos biológicos

Para realizar los análisis funcionales de los datos genómicos se utilizaron los términos de Gene Ontology pertenecientes a la ontología Proceso Biológico. Específicamente, se seleccionaron los términos de GO descendientes de los procesos biológicos de apoptosis, *Apoptotic process* (GO:0006915) y autofagia, *Autophagy* (GO:0006914), removiendo los procesos biológicos pertenecientes a apoptosis específica de tipo celular, *Cell-type specific apoptotic process*

(GO:0097285). Esto se realizó con el fin de reducir el número de procesos biológicos analizados y dar prioridad a los procesos generales en lugar de a los procesos específicos a tipos celulares, los cuales en su mayoría no serán relevantes para la interpretación de resultados en este estudio. Los términos de GeneOntology se descargaron accediendo al servicio de consulta en línea GOOSE (Gene Ontology Consortium, 2008), realizándose las consultas en lenguaje SQL para obtener los términos descendientes de los términos “apoptotic process”, “autophagy” y “cell-type specific apoptotic process”. El código SQL para realizar una búsqueda genérica de los términos descendientes de un término GO se encuentra en Anexo 1.

Para obtener los genes que conforman cada término GO seleccionado, se consultó la base de datos del laboratorio Bader de la Universidad de Toronto (Merico et al., 2010), el cual realiza minería de datos mensualmente en diversas fuentes electrónicas de información biológica como Gene Ontology, y las distribuye en formatos de texto plano. La versión de base de datos empleada para los análisis subsecuentes con términos GO es la de Marzo 25 del 2015, sin tomar en cuenta las anotaciones inferidas electrónicamente (IEA). Adicionalmente no se tomaron en cuenta términos con menos de tres genes anotados, ya que la metodología de Pathifier requiere mínimamente tres genes para realizar su análisis. La base de datos se encuentra disponible en: http://download.baderlab.org/EM_Genesets/March_24_2015/Human/symbol/Human_GOBP_AllPathways_no_GO_iea_March_24_2015_symbol.gmt.

Análisis de desregulación y agrupamiento jerárquico

Para calcular los niveles de desregulación de todos los procesos biológicos pertenecientes a apoptosis y autofagia en cada muestra se empleó el algoritmo Pathifier (Drier et al., 2013). Pathifier es un algoritmo semi-supervisado que integra los datos de expresión de todos los genes que conforman a un proceso biológico en un solo valor de desregulación para cada muestra. Para cada muestra Pathifier asigna una calificación de desregulación *Pathway Deregluation Score* (PDS) de 0 a 1. Valores cercanos a 0 son calculados para muestras con niveles de expresión similares a los del promedio de las muestras de tejido sano, mientras que las muestras con valores de PDS más altos poseen diferencias mayores en sus niveles de expresión en comparación con el grupo control. Para realizar este análisis se empleó la versión *stand-alone* disponible para descarga en la página del autor (véase referencia a Drier et al., 2013). El equipo empleado para este análisis fue un AMD Phenom X2 550 con 4 GB disponibles en RAM. Posterior al cálculo del PDS se emplearon scripts personalizados (García-Campos, 2016) para el agrupamiento jerárquico de las muestras en subgrupos, empleando el paquete base de R *stats* (R core team, 2013), y utilizando, para todos los agrupamientos, la distancia euclidiana de los PDSs y el método de agrupamiento aglomerativo jerárquico de Ward (Murtagh & Legendre, 2014).

Inferencia de RRG

Posterior a la agrupación por características biológicas por medio del algoritmo Pathifier, se procedió a inferir las redes de regulación genética para cada subgrupo establecido. Esto nos permitirá inferir a partir de los datos de expresión relaciones no anotadas en la literatura y bases de datos. Para realizar la inferencia de las RRG se empleó el algoritmo ARACNE (Basso et al., 2005). ARACNE es un algoritmo basado en teoría de la información el cual calcula la Información Mutua (IM), una medida de

dependencia estadística entre dos variables, en este caso genes. La matriz resultante del cálculo de IM para cada par de genes es una matriz cuadrada de tamaño $n \times n$ (n = número de genes analizados).

Para la inferencia de las redes se usaron sólo los valores de expresión de los genes de los procesos de apoptosis y autofagia anotados en el microarreglo. El algoritmo se ejecutó en una unidad de súper-computo con un CPU Intel Xeon E5-2650 v2 con 32 núcleos a 2.60GHz con 384 GB disponibles en RAM. Para reducir el tamaño de la red y reducir significativamente el número de relaciones falsas positivas, posterior al cálculo de IM se realizó un corte al percentil 1% de IM más alto, extrayendo las relaciones más confiables. Aunado al corte por percentil, la metodología descrita por Basso y colaboradores permite diezmar la red usando el concepto de teoría de la información Data-Processing-Inequality (DPI), con el cual se disminuyen las relaciones indirectas, esto se realizó implementando el paquete de R *minet* (Meyer et al., 2008) con scripts personalizados (García-Campos, 2016). El equipo empleado para este análisis fue un procesador quad-core APU A10-4655M de AMD con 12 GB disponibles en RAM.

Análisis de RRG.

De las redes resultantes del proceso de inferencia antes descrito, se procedió a extraer el componente gigante, la red de mayor tamaño, para los casos en los que se generaban subredes. Esto se realizó de forma manual con el software *Cytoscape* (Shannon et al., 2003). Con el mismo software, se realizaron análisis topológicos estándar usando la herramienta *NetworkAnalyzer*, para obtener el grado de conexión de cada gen y otros parámetros de la topología de las redes.

Para identificar los genes con mayor influencia en las RRG se usaron tres medidas de centralidad topológica distintas. El grado, el cual se obtuvo a partir del análisis topológico estándar de *Cytoscape*. La unidad híbrida de centralidad Degree-Betweenness Centrality (DBC) (Abasi & Hossain, 2006), implementada en R. Y la tercera el Collective Influence (CI) (Morone & Makse, 2015), implementado de forma interna por mi grupo de investigación (Langle, 2016). Se seleccionaron el 5% de los genes más influyentes de acuerdo a las tres medidas, y se procedió a hacer pruebas manuales en *Cytoscape*, determinando el número de vecinos a primer y segundo grado. Esto con la finalidad de elegir la medida de centralidad que es capaz de identificar los nodos más influyentes en la red. El equipo empleado para este análisis fue un procesador *quad-core* APU A10-4655M de AMD con 12 GB disponibles en RAM.

Análisis de Expresión diferencial

Buscando complementar los resultados del análisis de redes con la información de expresión para cada gen seleccionado, se procedió a hacer un análisis de expresión diferencial. Un análisis de expresión diferencial busca establecer si existe una diferencia significativa entre la expresión de los genes en un estado en comparación de otro; considerando además la estabilidad de esta diferencia entre las muestras de ambos grupos y proporcionando una magnitud y confianza estadística de esta diferencia. El análisis de expresión diferencial se realizó comparando la expresión de mRNA de cada subgrupo tumoral obtenido en el análisis de desregulación, en comparación con el grupo control. Para esto se implementó el paquete de R *limma* (Smyth, 2005), para realizar el análisis y los paquetes *gplots* (Warnes et al., 2009) y *dplyr* (Wickham & Francois, 2014) para visualización de resultados y manejo de datos respectivamente, en scripts personalizados (García-Campos, 2016). Del análisis resultante se

extrajo la información de *log-fold change*, expresión promedio, valor de p , valor de p ajustado, y estadístico B para los genes seleccionados por CI, la definición de estas métricas se encuentra en el glosario. El equipo empleado para este análisis fue un procesador Quad-core APU A10-4655M de AMD con 12 GB disponibles en RAM.

Visualización de redes

Para la visualización de redes se empleó el software *Cytoscape*. Cada red fue visualizada de forma independiente y los ajustes se realizaron de forma manual. Se usó el algoritmo de disposición *Weak Clustering* del plug-in de *Cytoscape Allegro Layout 2.2*, para realizar el acomodo de los nodos en las redes, usando los valores de IM entre los nodos para dar peso a las interacciones. Los nodos más influyentes fueron resaltados en las redes de la siguiente forma: el tamaño de los nodos seleccionados es relativo a su nivel de influencia basado en el score de CI, el color de los nodos es relativo a su valor de *log-fold change*, verde para valores negativos y rojo para valores positivos. Sólo se usaron los valores de *log-fold change* de los genes con un estadístico B superior a 5. La transparencia de las aristas es relativa al nivel de IM que tiene cada interacción, aristas más oscuras representan valores de IM más altos.

Subtipificación PAM50 y comparación de subgrupos

Para comparar la relevancia clínica del agrupamiento por desregulación de los procesos de apoptosis y autofagia se implementó el algoritmo PAM50 (Parker et al., 2009). PAM50 analiza los niveles de expresión de 50 genes y su resultado final es la clasificación de cada muestra en uno de 5 subtipos con relevancia clínica, denominados luminal A, luminal B, Her2, basal y normal (o normal-like si una muestra tumoral es clasificada como normal). El algoritmo PAM50 fue implementado en lenguaje R, disponible en la página de los autores (<https://genome.unc.edu/pubsup/breastGEO/PAM50.zip>). El equipo empleado para este análisis fue un procesador quad-core APU A10-4655M de AMD con 12 GB disponibles en RAM.

Para verificar el enriquecimiento de subtipos moleculares determinados con PAM50, en los subgrupos resultantes del análisis de desregulación con Pathifier, se realizaron pruebas hipergeométricas. La prueba hipergeométrica consiste en determinar la probabilidad de obtener un número de éxitos específico en un tamaño de muestra determinado, muestreando al azar y sin recambio, cuando se conoce el número de éxitos en la población. De esta forma una baja probabilidad obtenida en la prueba, respalda una relación no azarosa entre un subgrupo y un subtipo de PAM50. Las pruebas hipergeométricas se realizaron en R con scripts personalizados (García-Campos, 2016) usando el paquete stats (R core team, 2013), el equipo empleado para este análisis fue un procesador quad-core APU A10-4655M de AMD con 12 GB disponibles en RAM.

Validación de resultados de subtipificación

Para dar mayor sustento a los resultados del enriquecimiento de los subgrupos obtenidos, se realizó un análisis de desregulación idéntico con otros dos conjuntos de datos de cáncer de mama. Los dos conjuntos de datos nuevos provienen de la tecnología de secuenciación de RNA. El primero de ellos es el perteneciente al proyecto TCGA (Cancer Genome Atlas Network, 2012), el cual empleó la plataforma

de secuenciación de mRNA Illumina HiSeq 2000 RNA Sequencing Version 2 y el segundo pertenece al proyecto METABRIC (Curtis et al., 2012), el cual empleó la plataforma de microarreglo de expresión *HumanHT-12 v3.0 Gene Expression BeadChip*. Ambos conjuntos de datos fueron obtenidos a partir de biopsias de cáncer primario sin tratamiento.

Los dos conjuntos de datos fueron analizados usando el mayor número posible de genes para cada proceso, debido a que cada plataforma analiza distintos genes, el número de genes analizados fueron ligeramente diferentes en cada uno de ellos. Posteriormente se realizó el agrupamiento jerárquico de la misma forma que con los datos originales para llegar a una subdivisión de todas las muestras.

De forma paralela se realizó el análisis de subtipos tumorales intrínsecos de PAM50 con el mismo script proporcionado por Parker y colaboradores (Parker et al., 2009). Y se realizó el enriquecimiento de la misma forma que con los datos originales. Obteniendo un valor de p corregido para cada grupo y subtipos tumoral intrínseco.

RESULTADOS

Información funcional de procesos biológicos

Se usó la base de datos Gene Ontology (GO) para obtener la información funcional de los genes cuantificados en el microarreglo. Se usaron todos los términos GO que fueran descendientes de los términos “*apoptotic process*” y “*autophagy*”. El número de términos GO resultantes del curado de la base de datos de Gene Ontology fue de 88 para apoptosis y 16 para autofagia, conteniendo en total 1105 y 278 genes respectivamente, compartiendo 51 genes entre ambos procesos.

Análisis de desregulación y agrupamiento

El análisis de desregulación con Pathifier se logró realizar en 96 procesos biológicos, 15 de autofagia y 81 de apoptosis. Debido a que el número mínimo de genes que debe tener un procesos para ser analizado por el algoritmo es de tres, ocho procesos con menor número de genes en el microarreglo fueron omitidos del análisis. En total se analizaron 1144 genes pertenecientes a los dos procesos.

Las calificaciones de desregulación de vía (PDS) para las muestras de tejido sano se mantuvieron en niveles bajos, en promedio 0.1768; mientras que las de las muestras tumorales calificaron con niveles altos, en promedio 0.6408. Las distribuciones de los PDS para los dos grupos se muestran en la figura 2.

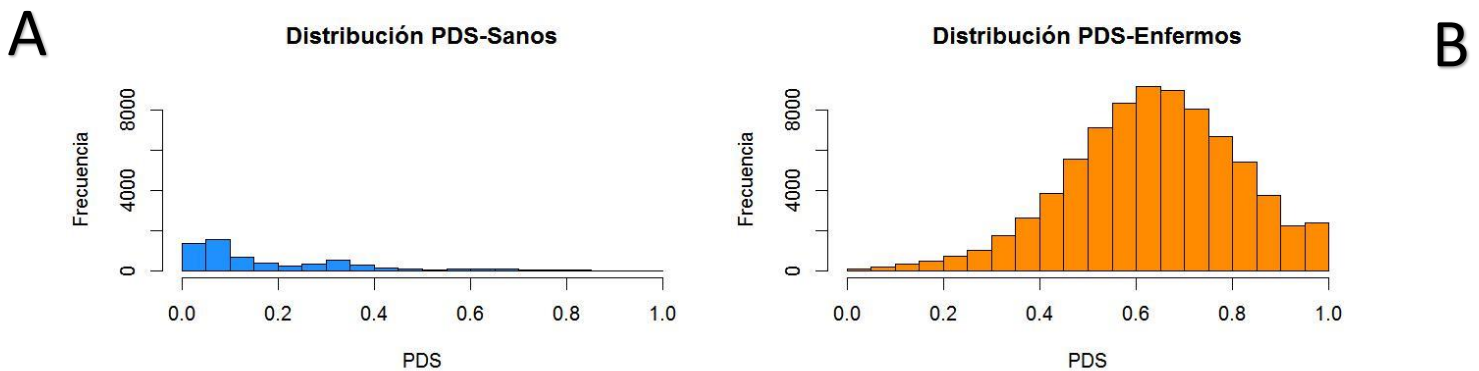


Figura 2. Distribuciones de resultados análisis de desregulación con Pathifier. A) Distribución de frecuencia de valores de desregulación para las muestras de tejido sano. B) Distribución de frecuencia de valores de desregulación para las muestras de tejido tumoral.

El análisis de desregulación resultó en una matriz que representa el nivel de desregulación para cada una de las 880 muestras (columnas) en cada uno de los 96 procesos biológicos analizados (filas) con una calificación de 0 a 1. Con esta matriz fue posible realizar un agrupamiento jerárquico usando la información integrada de los procesos biológicos que contienen los PDS.

La agrupación jerárquica por medio del algoritmo de Ward, dio los siguientes resultados. Cuando se toma la división de las muestras para 2 grupos, el algoritmo separó a todas las muestras de tejido sano en el mismo subgrupo, mientras que sólo una muestra tumoral se agrupó con las muestras de tejido

sano (esta muestra fue descartada para análisis posteriores). Una visualización de estos resultados representados por un *heatmap* se aprecia en la Figura 3.

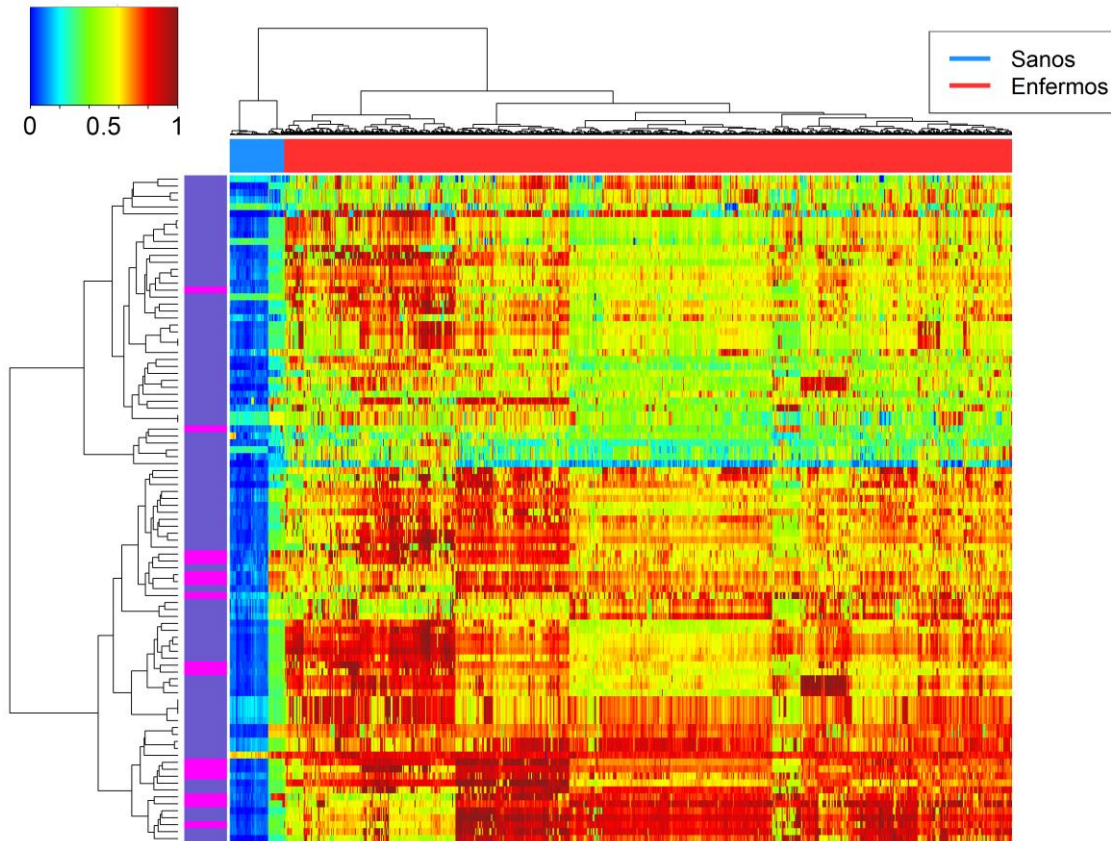


Figura 3. *Heatmap* del agrupamiento jerárquico con etiquetas para Sanos-Enfermos y Apoptosis-Autofagia. En el *heatmap* se representan los resultados del análisis de desregulación por medio del algoritmo Pathifier, los resultados se representan ordenados por columnas para cada muestra y en filas para cada proceso biológico. Valores bajos de desregulación se representan en color azul, mientras que valores altos en color rojo oscuro. En la parte superior del *heatmap* se encuentran etiquetadas, a modo de barra, las muestras por color: rojo para muestras tumorales y azul para muestras de tejido sano (la única muestra tumoral que se agrupa con el subgrupo de muestras de tejido sano se encuentra adyacente al grupo de muestras tumorales, por lo que no es distinguible a simple vista). La barra en la parte izquierda del *heatmap* etiqueta los procesos con base a su pertenencia, en morado para apoptosis y rosa para autofagia. Todos los agrupamientos jerárquicos se realizaron con el algoritmo aglomerativo de Ward y la distancia euclidiana de los valores de PDS.

El agrupamiento jerárquico a un punto de corte de 6 grupos dividió a las muestras tumorales en 4 subgrupos, ya que las muestras control se dividieron en 2 grupos. Sin embargo, se consideraron como un solo grupo debido a su efecto nulo en los resultados de los grupos tumorales. De forma eficiente, los subgrupos de muestras tumorales generados a partir de los PDS están conformados en su totalidad por muestras tumorales y presentan patrones de desregulación que los distinguen entre sí. Debido a que el estudio se centra en el estudio de las muestras tumorales y que la metodología de inferencia de redes necesita de un número de muestras amplio, el grupo control comprendió a todas las muestras de tejido sano en un solo subgrupo (una visualización de los resultados intermedios se encuentra en el

Anexo 2). La subdivisión final de las muestras se delimitó a 5 subgrupos, etiquetados subsecuentemente como subgrupos del 0 al 4 (Subgrupo 0 = SG0, Subgrupo 1 = SG1, Subgrupo 2 = SG2, Subgrupo 3 = SG3, Subgrupo 4 = SG4), siendo el SG0 el de muestras de tejido sano. Los resultados finales del agrupamiento se pueden observar en el *heatmap* de la figura 4.

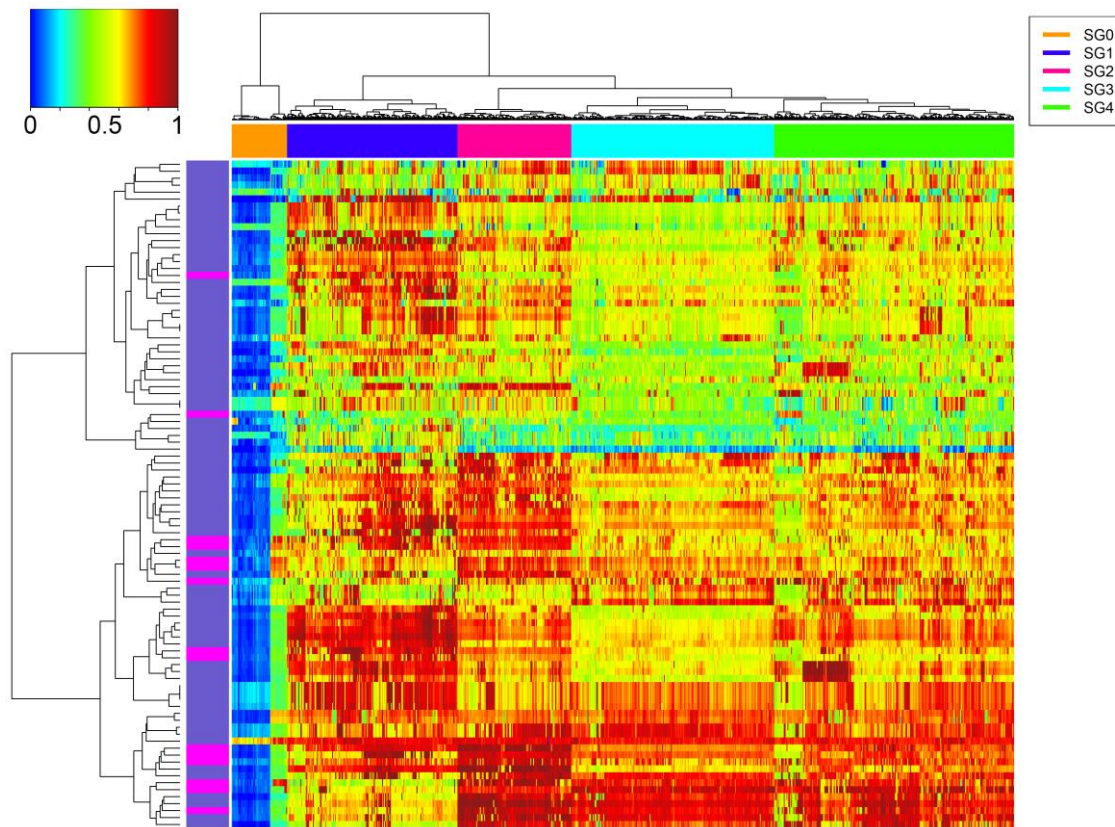


Figura 4. *Heatmap* del agrupamiento jerárquico con etiquetas para los 5 subgrupos de muestras. En el *heatmap* se representan los resultados del análisis de desregulación, cada columna representa los datos de una muestra, mientras que las filas representan a los procesos biológicos analizados. Valores bajos de desregulación se representan en color azul, mientras que valores altos se representan en color rojo oscuro. En la barra superior del *heatmap* se encuentran etiquetadas las muestras de acuerdo a su pertenencia a 5 subgrupos, y en la barra izquierda se etiquetan los procesos biológicos de acuerdo a su pertenencia a apoptosis (morado) y autofagia (rosa) de forma análoga a la figura 3.

Para facilitar la interpretación de los resultados se realizó un promedio por proceso biológico de los PDS de las muestras pertenecientes a cada subgrupo. Los PDS promedio para cada subgrupo se representan en un *heatmap* en la Figura 5. Considerando estos últimos resultados, el agrupamiento jerárquico de los procesos biológicos realiza una primera división en dos grupos, el bloque superior presenta valores predominantemente bajos de desregulación en todos los subgrupos, exceptuando bloques de procesos en los SG1 y SG2. Por otro lado, el grupo inferior de procesos presentan valores predominantemente altos de desregulación.

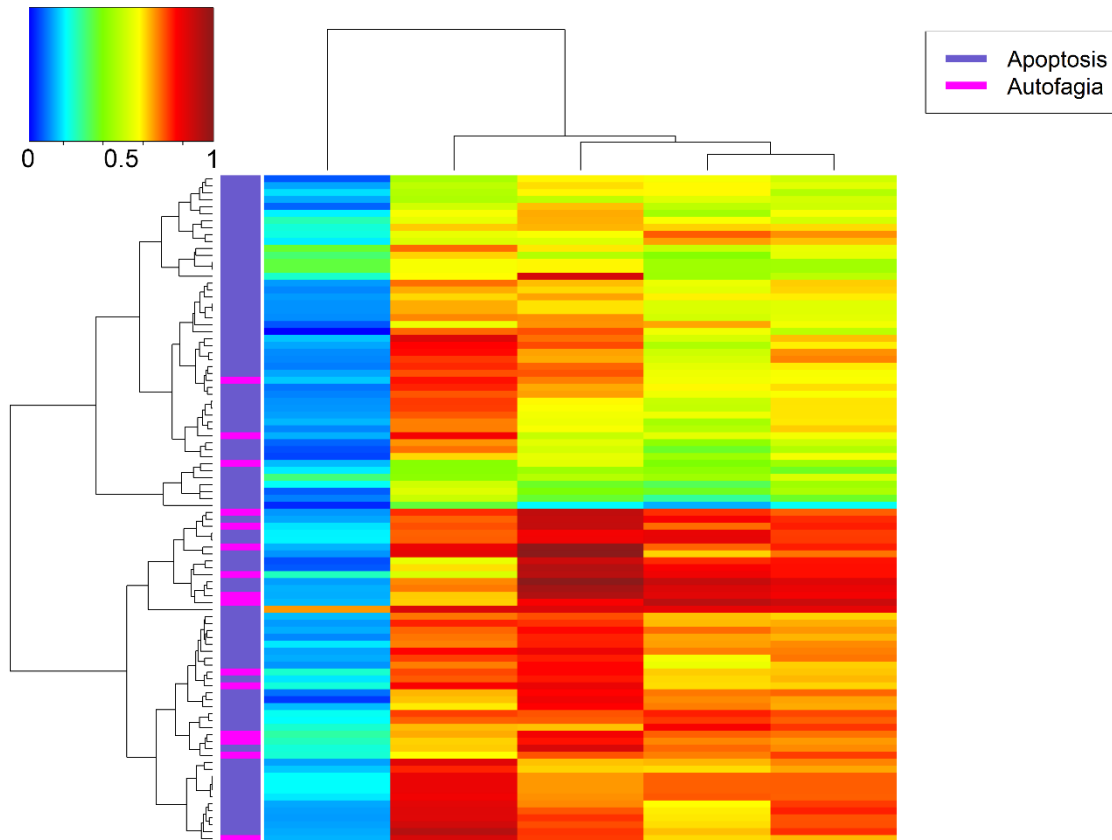


Figura 5. *Heatmap* del PDS promedio por subgrupo. En este *heatmap* se encuentra representados los valores de desregulación para cada proceso biológico analizado. Los colores de la barra izquierda denotan el tipo de proceso biológico, rosa para autofagia y morado para apoptosis. Los colores verdes del grupo superior de procesos denotan un menor nivel de desregulación promedio en comparación con los procesos de la parte inferior, en los cuales predominan valores más altos de PDS.

La distribución de los valores promedio de desregulación de PDS para cada subgrupo refleja los niveles de desregulación que posee, las distribuciones de PDS para los 5 subgrupos se pueden apreciar en la Figura 6 como diagramas de cajas y bigotes. En esta imagen resulta interesante observar que aun separando a las muestras en grupos con distintos patrones de desregulación, los subgrupos mantienen niveles de desregulación similares entre sí. Sin embargo, se pueden apreciar diferencias como el hecho de que los subgrupos 3 y 4 presenten valores inferiores a los de los subgrupos 1 y 2. De las distribuciones de PDS de muestras control y tumorales por separado, es relevante la consistencia de valores bajos de PDS para el grupo de muestras de tejido sano y valores altos para los subgrupos tumorales (Figura 6). Cabe resaltar que el *outlier* con valores bajos de PDS, el menos desregulado, para todos los grupos tumorales resultó siempre ser el mismo proceso, GO:0008626: “Vía de señalización de apoptosis mediada por granzimas”.

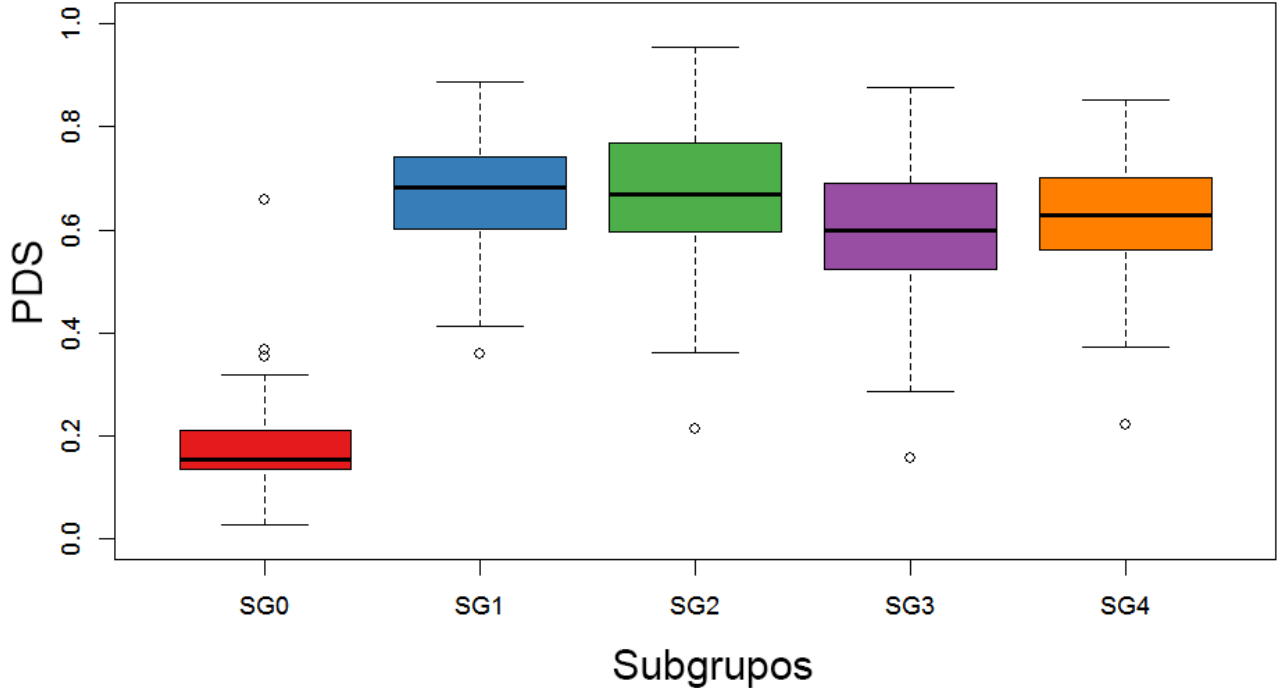


Figura 6. Diagrama de cajas y bigotes con los valores de desregulación en PDS promedio para cada subgrupo. En este diagrama se muestran las distribuciones de las calificaciones PDS para cada subgrupo, el subgrupo 0 (SG0) conformado por muestras de tejido sano presenta valores relativamente bajos en comparación con los de los subgrupos restantes, conformados en su totalidad por muestras tumorales.

Cada subgrupo presenta distintos patrones de desregulación, siendo de importancia evidente las diferencias y similitudes en los niveles de desregulación entre los subgrupos. En la Figura 7 se muestran los 10 procesos con mayor desregulación en promedio de todos los subgrupos. Pese a su bajo número con respecto a los procesos de apoptosis, los procesos de autofagia figuran entre los que tienen valores altos de desregulación, en este caso con 4 procesos, todos de regulación. Por otro lado, en la Figura 8 se muestran los 10 procesos con menor desregulación promedio de todos los subgrupos. Estos últimos procesos son predominantemente de apoptosis, cabe mencionar que varios de estos pertenecen a la regulación de actividad de endopeptidasas tipo cisteína involucradas en señalización de apoptosis. El valor más bajo de desregulación para todos los subgrupos fue el proceso de señalización de apoptosis mediada por granzimas.

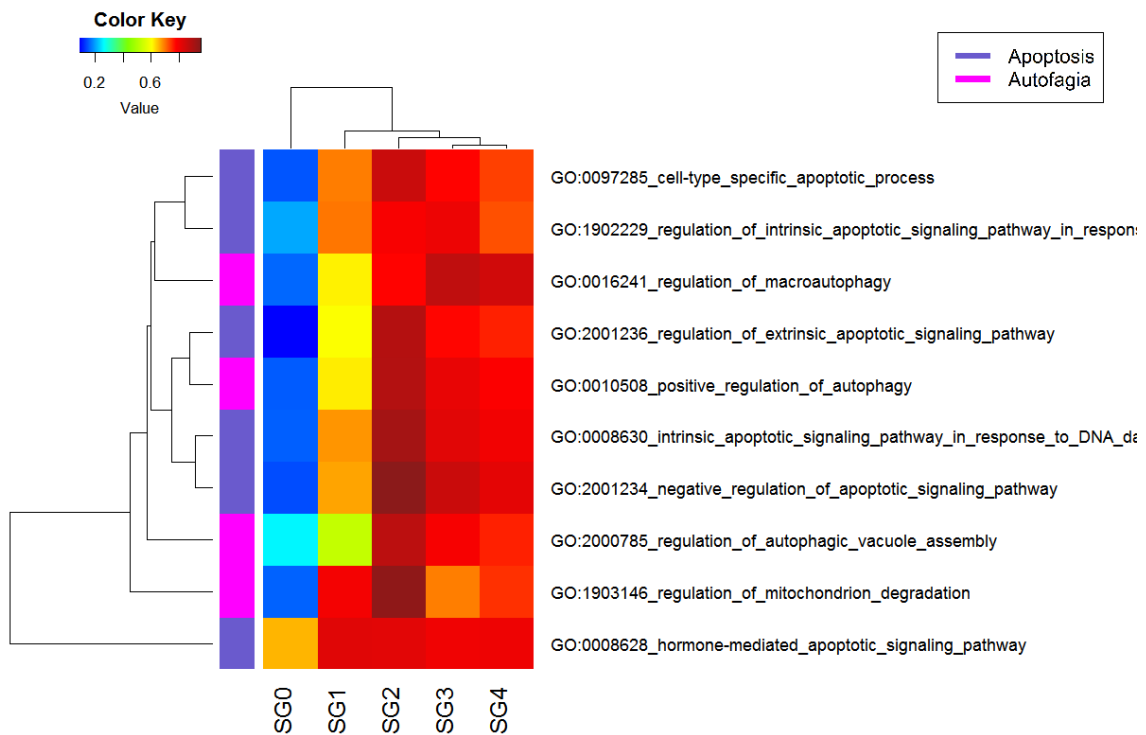


Figura 7. Los 10 procesos con mayor PDS promedio. Se muestran los 10 procesos con valores mayores de desregulación promedio para todos los subgrupos.

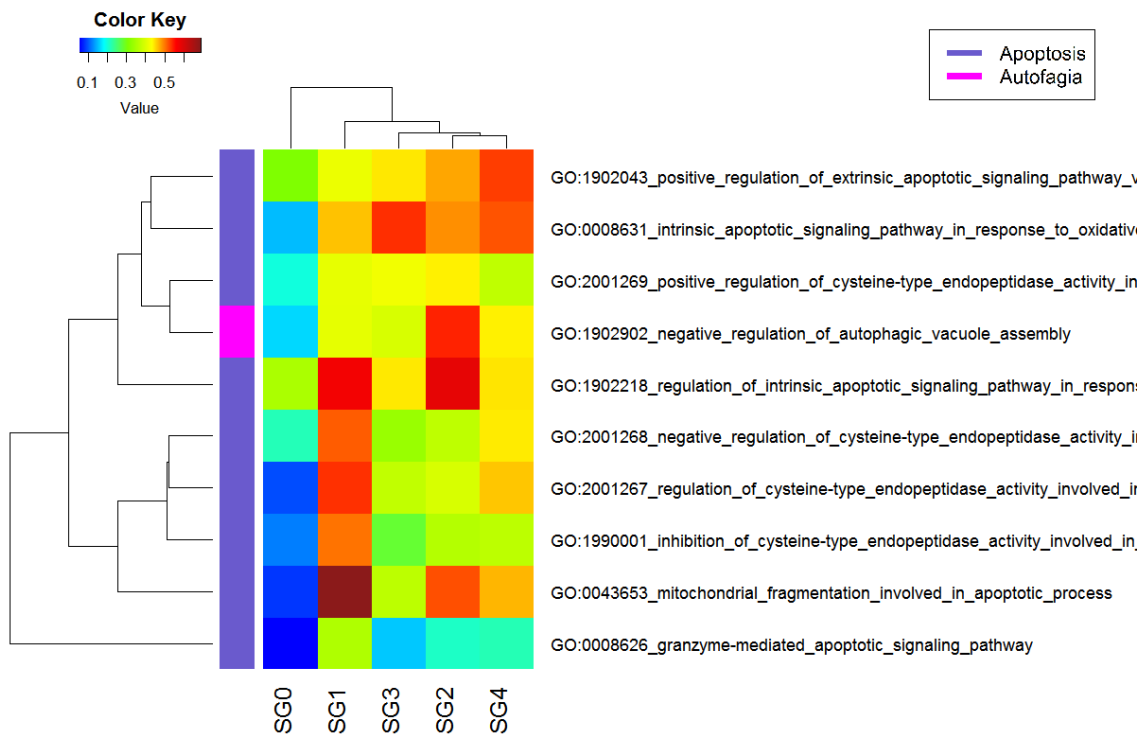


Figura 8. Los 10 procesos con menor PDS promedio. Se muestran los 10 procesos con valores menores de desregulación promedio para todos los subgrupos.

Para observar los procesos biológicos con menor y mayor variación entre subgrupos se calculó la desviación estándar de los PDS, de esta forma se identificaron los 10 procesos con mayor y menor variación en sus valores de PDS. Estos resultados se aprecian en los *heatmaps* de los procesos seleccionados en las Figuras 9 y 10.

En la Figura 9 se puede observar que, tomando en cuenta los PDS de los 10 procesos menos variables, la separación entre los subgrupos se reduce y prácticamente los PDS promedio son idénticos entre subgrupos. De los procesos con menor variación es de destacarse que todos pertenecen a los procesos de apoptosis; y entre los de alta desregulación uniforme se encuentran los procesos de apoptosis canónicos: “vía de señalización de apoptosis extrínseca vía dominios de muerte” y la “vía de señalización de apoptosis intrínseca por mediador clase p53”. En contraparte la “vía de señalización de apoptosis en respuesta al estrés oxidativo” está poco desregulada en todos los subgrupos.

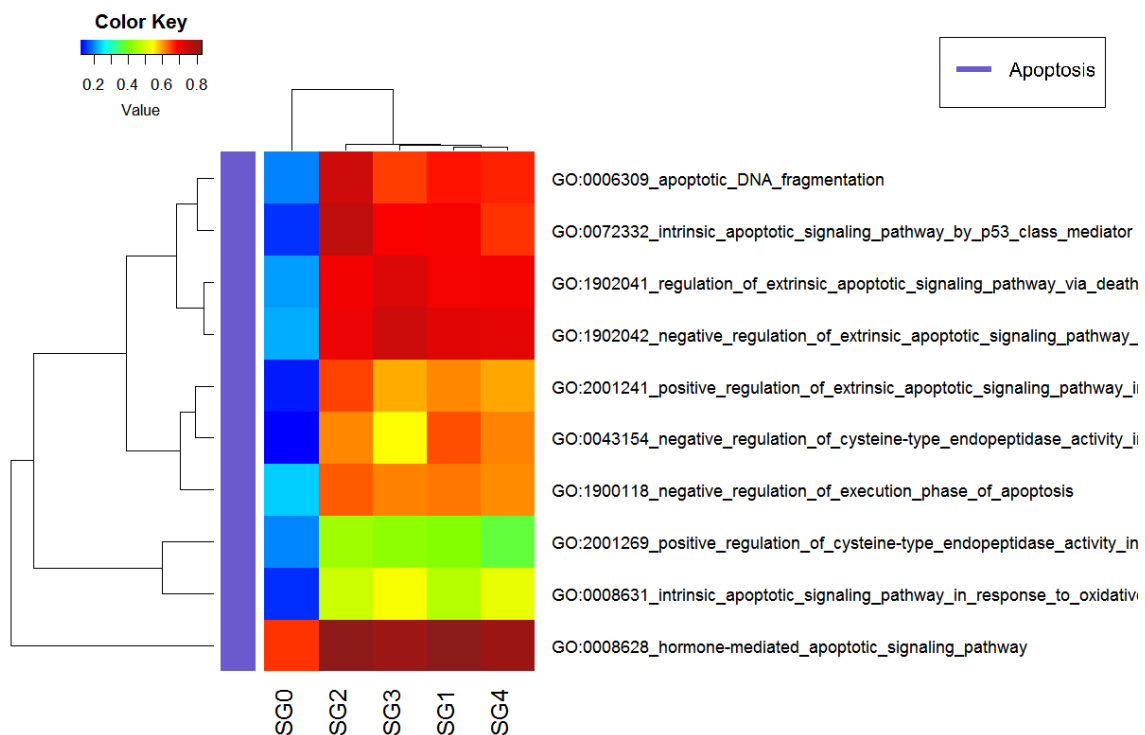


Figura 9. Los 10 procesos con menor desviación estándar. Se muestran los 10 procesos menos variantes en sus PDS por subgrupos.

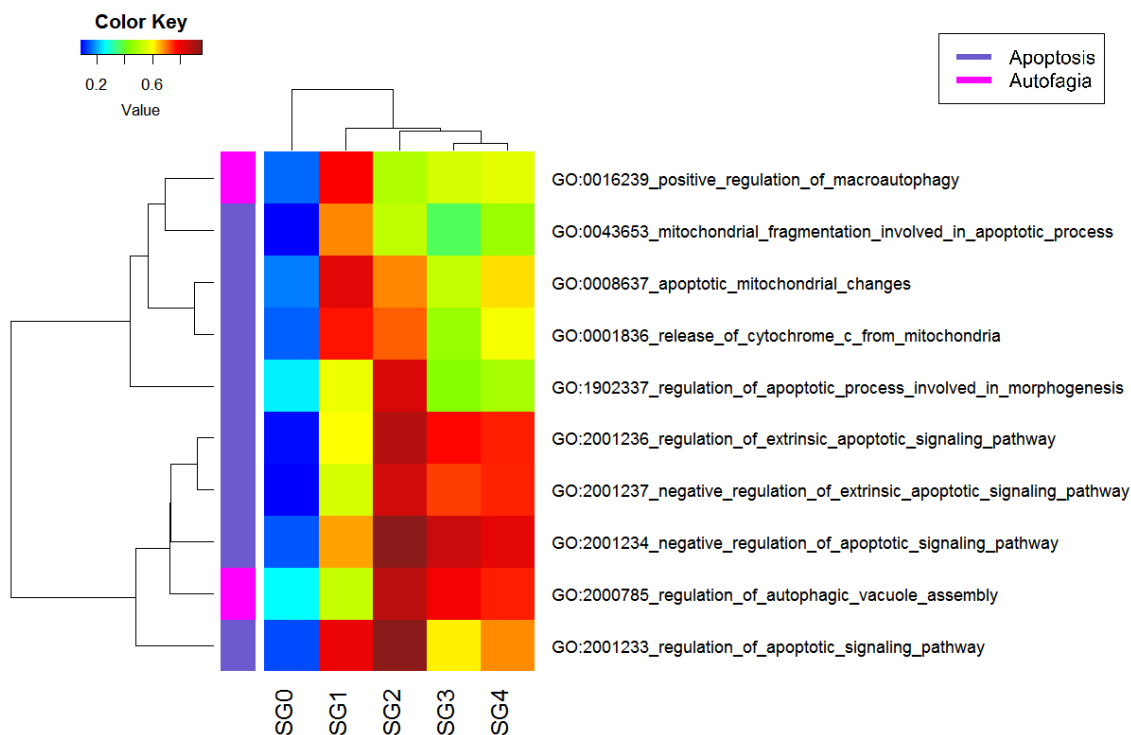


Figura 10. Los 10 procesos con mayor desviación estándar. Se muestran los 10 procesos más variantes en sus PDS por subgrupos.

En la Figura 10 se observa que el SG1 es el que más difiere de los demás subgrupos tumorales, mientras que el SG3 y SG4 son los más similares entre sí, tal y como sucede en el agrupamiento jerárquico usando todas las muestras y todos los procesos, además de que se presentan de igual forma 2 bloques, el bloque superior presenta valores altos de desregulación tanto en el SG1 y en menor proporción en el SG2; mientras que el bloque inferior presenta valores bajos sólo en el SG1.

Inferencia y análisis de RRG

La inferencia de RRG con la información de expresión de mRNA usando el algoritmo ARACNE y la metodología descrita anteriormente, resultó en 5 RRG, una por subgrupo. Los valores de p calculados para todas las interacciones inferidas son menores a 1×10^{-6} . La conformación de las redes varió para cada subgrupo, pero manteniendo relaciones que las hacen comparables entre sí. La red con valor promedio de mayor significancia estadística es la del subgrupo de muestras de tejido sano (SG0), al mismo tiempo esta red resultó ser la de menor tamaño, en relación al número de nodos. Las redes de subgrupos tumorales se caracterizaron por tener colas largas hacia los valores de p más bajos, llegando a valores como 6.9×10^{-62} para la red del subgrupo 1. Un resumen de los resultados de los valores de p mínimos, promedio y máximos para cada red se describe en la Figura 11 en la que se muestran su distribución como diagramas de cajas y bigotes. La composición de las redes, por tipo de proceso biológico al que pertenecen los genes, no varió en cada red. En promedio cada red contiene 80% de genes anotados a apoptosis, 16 % genes anotados a autofagia y 4 % de genes anotados en ambos procesos. El resumen de la pertenencia de los genes de cada red por proceso biológico se encuentra en la Tabla 2.

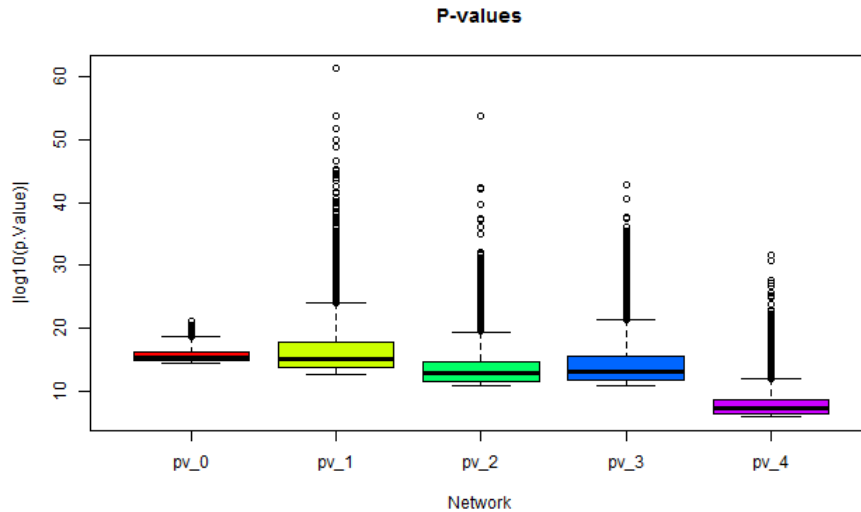


Figura 11. Diagramas de cajas y bigotes de valores de p para cada red. Se muestran la distribución de los valores de p de los enlaces en las redes inferidas para cada subgrupo.

	Apoptosis	Autofagia	Ambivalentes	Nodos totales	Aristas totales
Red 0	349	64	11	424	6543
Red 1	660	131	29	820	6162
Red 2	662	139	33	834	6418
Red 3	574	104	23	714	5901
Red 4	701	144	36	881	6201

Tabla 2. Número y pertenencia de nodos para cada red inferida. Se presentan el número de genes totales en cada red inferida y su pertenencia a los procesos biológicos de Apoptosis, Autofagia o ambos.

Densidad

De las RRG uno de los indicadores más generales que se puede medir es la proporción de conexiones potenciales en una red, que son conexiones existentes, a esta proporción se le conoce como densidad. Se ha propuesto que la densidad es una característica que permite a las redes complejas ser resilientes, adaptándose a perturbaciones y manteniendo su funcionalidad básica (Gao et al., 2016). Para determinar la densidad entre los procesos (interprocesos) y al interior de los procesos (intraproceso), se realizó el cálculo de la densidad de conexiones de los genes de cada grupo consigo mismo y con los de los otros grupos. Para esto se dividieron los genes de cada red en los tres grupos antes mencionados: Apoptosis, Autofagia y genes compartidos por ambos procesos.

El cálculo de la densidad de una red se obtiene al dividir el número de vínculos existentes entre el número de vínculos posibles (Fórmula 1). Para el cálculo de la densidad entre los genes de un mismo proceso se usó la fórmula clásica para obtener el número de vínculos posibles (Fórmula 2), donde n es igual al número de nodos de la red. Mientras que para calcular el número de vínculos posibles entre dos grupos de nodos diferentes, se usó una fórmula modificada (Fórmula 3), donde n_a es el número de nodos del grupo A y n_b el número de nodos del grupo B. Este resultado nos permitirá vislumbrar el

nivel de conexión que tienen los procesos entre sí, haciendo posible comparar la densidad o intensidad de conexión entre distintas condiciones como los subgrupos antes inferidos.

$$Densidad = \frac{vínculos\ existentes}{vínculos\ posibles} \quad (\text{Formula 1})$$

$$VP = \frac{n \cdot (n-1)}{2} \quad (\text{Formula 2})$$

$$VP = n_A \cdot n_B \quad (\text{Fórmula 3})$$

Los resultados obtenidos del cálculo de densidad para cada red se muestran en la Figura 12 con diagramas de barras representando la densidad de conexión entre los genes de los tres grupos, en los 6 posibles tipos de conexiones: intraproceso de apoptosis, autofagia, ambivalentes; y entre los procesos de apoptosis-autofagia, apoptosis-ambivalentes y autofagia-ambivalentes. Los resultados muestran que los valores de densidad en la red del grupo control son en general mucho mayores que en los subgrupos tumorales. Es remarcable que en la red del grupo control (Figura 12B) la densidad de conexión entre los genes ambivalentes es la mayor, mientras que la conexión entre los genes de autofagia es la menor. La densidad de conexión entre los genes de autofagia y apoptosis es mayor que la que poseen entre sí los genes de autofagia. A su vez los genes ambivalentes tienen mayor densidad de conexión con cada proceso que la densidad intra-proceso de apoptosis y autofagia.

Para la red del SG1 (Figura 12C) los valores de densidad entre los procesos es muy similar entre sí, ahora con la densidad de conexión entre los genes de apoptosis y autofagia como la menor. De forma similar los valores de densidad para la red del SG2 (Figura 12D) son similares entre sí rondando en la densidad 0.02, lo que significa que sólo ocurren alrededor del 2% de las posibles conexiones en toda la red sin sesgo evidente hacia ningún tipo de gen. Por otro lado, los valores de densidad del SG3 (Figura 12E) desfavorecen a la conexión de los genes ambivalentes, siendo la única red en el que la conexión entre los genes de autofagia es la mayor, seguida de la conexión entre genes de apoptosis y autofagia y las conexiones intra-apoptosis; mientras que la conexión entre los genes ambivalentes casi desaparece. En la red del SG4 (Figura 12F), nuevamente los genes del grupo ambivalente son los que poseen mayor densidad de conexión entre, guardando proporciones más similares a las de la red del grupo control pero en niveles de densidad más bajos, por un orden de magnitud aproximadamente.

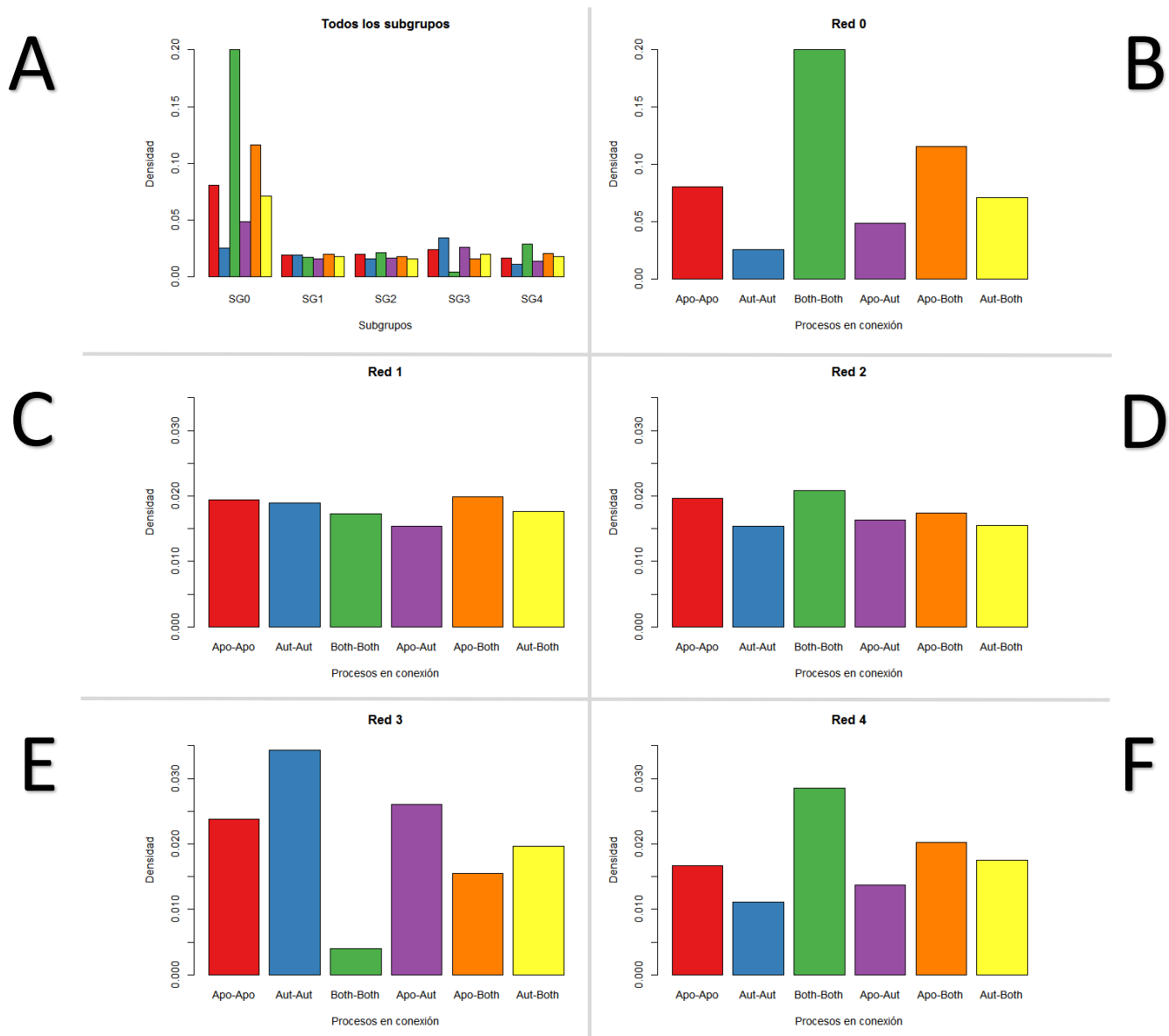


Figura 12. Diagramas de barras de densidad entre los genes de 3 categorías: Apoptosis, Autofagia y genes que participan en ambos procesos. A) En este diagrama se muestra la diferencia entre los valores de densidad de los grupos control (SG0) y tumorales (SG1, SG2, SG3, SG4), usando el mismo código de colores para las densidades entre los genes. B) Valores de densidad para la red del grupo control sano. C-F) Valores de densidad para las redes de los subgrupos tumorales SG1, SG2, SG3, SG4 respectivamente.

Así como por medio de la densidad es posible apreciar patrones generales de conexión entre los genes de los procesos biológicos, cabe mencionar que no todos los genes tienen el mismo nivel de conexión entre sí. La mayoría de las conexiones en las RRG son capitalizadas por un conjunto pequeño de genes, por lo que se buscó identificar estos genes a través de medidas de centralidad.

Genes altamente influyentes

De las 5 RRG inferidas se identificaron y compararon el 5% de nodos más influyentes, de acuerdo a las 3 medidas de centralidad mencionadas en la metodología de análisis de RRG: grado, Degree-Betweenes Centrality y Collective Influence. Posteriormente, se calculó el número de vecinos a primer y segundo grado de estos genes forma manual en *Cytoscape*, logrando apreciar el potencial de los nodos seleccionados para transmitir información a un mayor número de vecinos en el menor número de “pasos” posible. Siendo identificados los nodos seleccionados por *Collective Influence* como los que mejor comunicarían una señal, por lo que fue elegida esta centralidad para determinar los nodos más influyentes para todas las RRG. Resultados detallados de la comparación de las medidas de centralidad se pueden encontrar en el Anexo 3. Cabe resaltar que los nodos seleccionados por CI no son siempre los de mayor grado, pero en conjunto sobrepasan la influencia que tienen los nodos seleccionados por otra centralidad, desde el primer nivel de vecinos. Los nodos seleccionados para cada red e información complementaria de los genes se muestran en el Anexo 4.

Se observó que los genes influyentes en los subgrupos tumorales tienen un empalme considerable entre sí, mientras que sólo tres genes influyentes del grupo control fueron influyentes en las redes tumorales: TARDBP en la red del Subgrupo 1, RBM5 en la red del subgrupo 2 y PSMA3 en la red del subgrupo 4. En la Figura 13 se muestra el empalme de los genes influyentes en las 4 redes de los subgrupos tumorales.

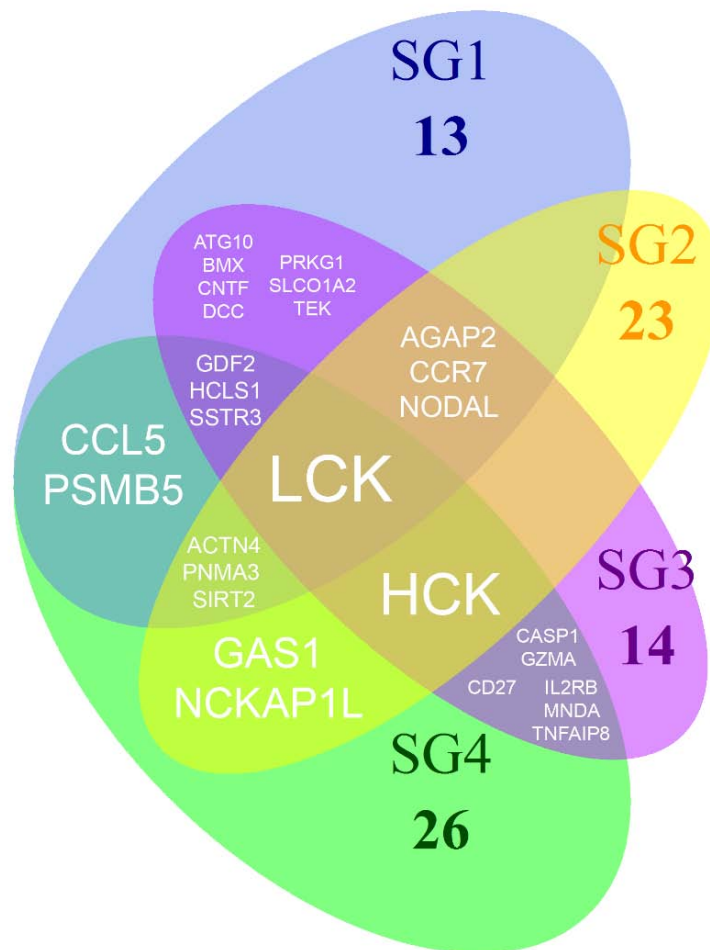


Figura 13. Diagrama de Venn de genes influyentes que comparten los subgrupos tumorales.

Análisis de expresión diferencial

Para este análisis se usó la expresión de todos los genes, pero sólo se extrajo la información de los nodos seleccionados por CI. La mayor parte de los nodos seleccionados de esta forma presentan cambios consistentes en sus niveles de expresión, teniendo valores de estadístico B mayores a 5. Los genes seleccionados por CI y sus niveles de expresión diferencial en \log *Fold-Change* se muestra en el Anexo 4.

Visualización de RRG

Las redes fueron visualizadas con parámetros que resaltan los genes con mayor influencia sobre la red de acuerdo a su CI, así como sus niveles de expresión diferencial en \log *Fold-Change*. De forma espacial los nodos están localizados de acuerdo a un modelo en el que los nodos se repelen entre sí pero las aristas ejercen una fuerza de atracción entre ellos, por lo que genes cercanos en la visualización no sólo están conectado entre sí, sino que también comparten vecinos formando grupos. En las RRG de los subgrupos tumorales es posible observar no sólo que los genes altamente influyentes se encuentran agrupados en zonas altamente conectadas, sino que también presentan expresión diferencial de forma coordinada, presentando principalmente sub-expresión o sobre expresión.

A continuación se presentan las visualizaciones obtenidas para cada red empezando por la obtenida para las muestras de tejido sano (Figura 14). En la red de muestras de individuos sanos se observa una agrupación muy estrecha de todos los genes en un solo grupo principal, del cual los genes más influyentes son principalmente del proceso de apoptosis, con el factor de transcripción SOX2 como el gen más influyente de la red, mientras que sólo un gen de autofagia se hace visible en este grupo MRPS10.

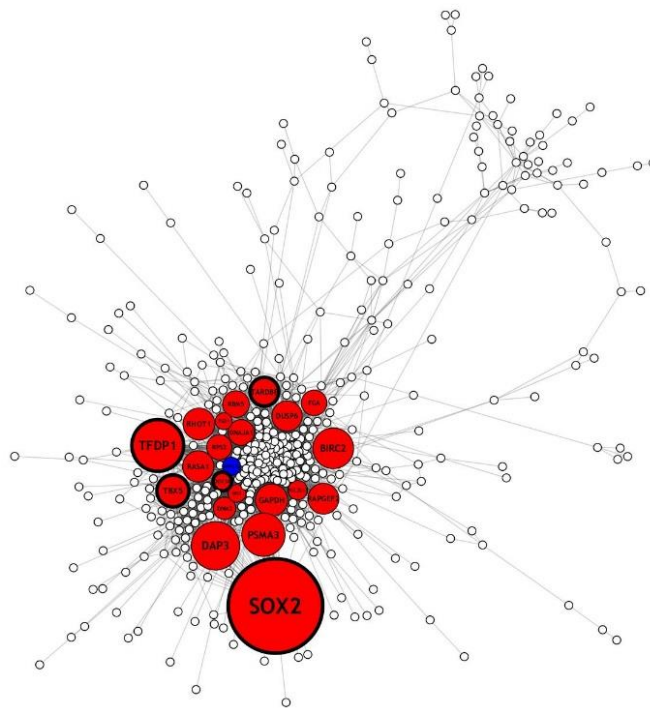


Figura 14. Visualización de la red del Subgrupo de muestras de tejido sano. El tamaño de los nodos refleja su influencia de acuerdo al CI. Los genes de apoptosis se marcan en rojo, mientras que los de autofagia en azul.

Las redes subsecuentes se presentan con 2 visualizaciones, usando la misma disposición de los nodos. En una visualización se representa la pertenencia de los genes más influyentes a los procesos de apoptosis (en rojo), autofagia (en azul) o ambos procesos (en magenta), de acuerdo a sus anotaciones en Gene Ontology; y otra en la que se aprecia su expresión diferencial con respecto al grupo control: verde para sub-expresión y rojo para sobre expresión. Para facilitar la visualización de las redes en el Anexo 5 se presentan las redes a hoja completa. A continuación se describen las características más relevantes para cada una de las redes y la integración de sus datos con las visualizaciones.

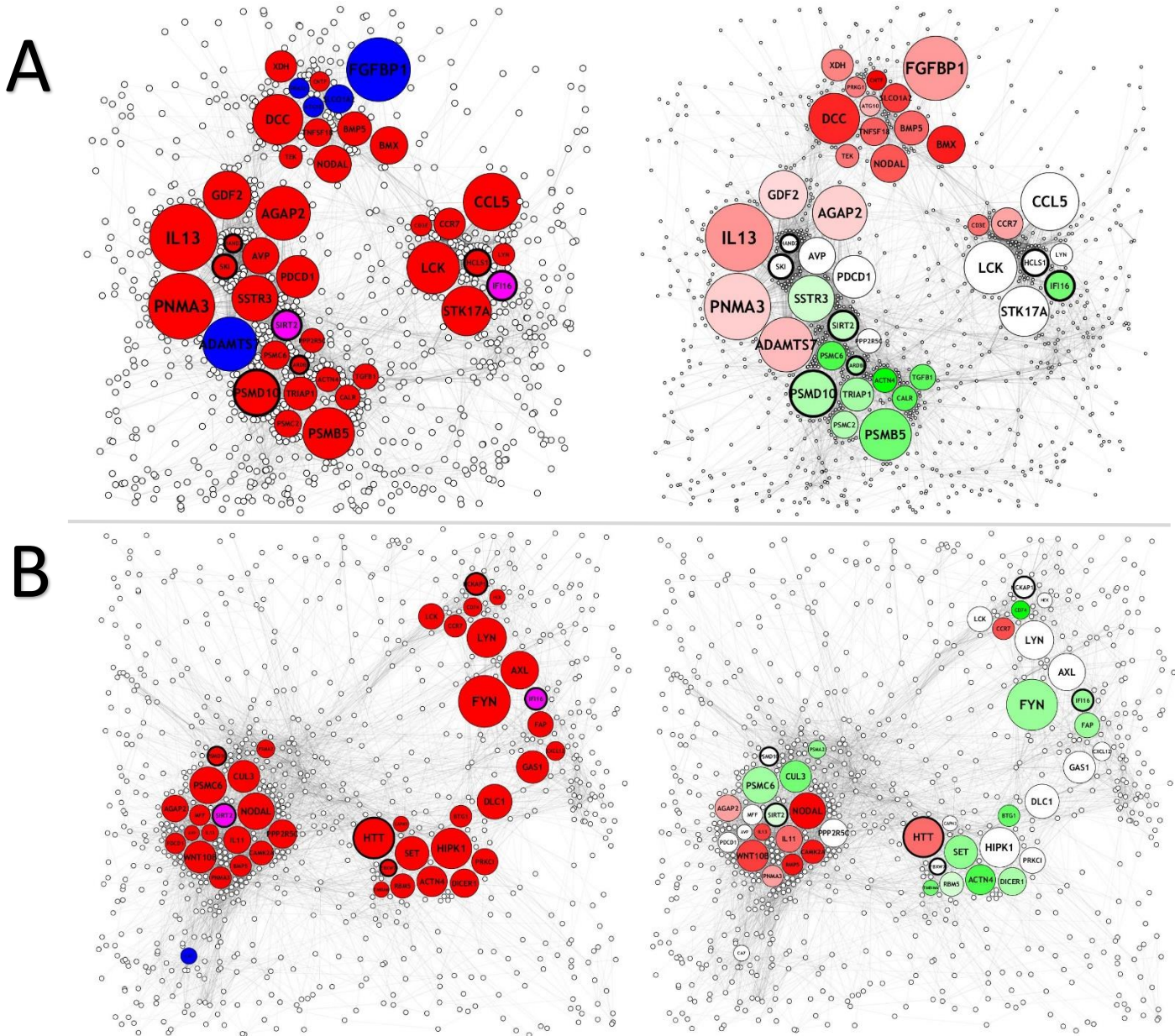
En la red del subgrupo 1 (Figura 15A), se pueden apreciar la división de los genes influyentes en 3 grupos. En el grupo de la parte superior los genes de autofagia y apoptosis se encuentran a una proporción 2:1, y de forma interesante todos los genes se encuentran sobre-expresados en este subgrupo de tumores. Por otro lado, el grupo de genes del lado izquierdo presenta un comportamiento distinto, en el que los genes influyentes quedan divididos en genes sub-expresados y sobre expresados con un gen que participa en ambos procesos SIRT2. El último grupo de genes a la derecha, presenta pocos genes diferencialmente expresados, pero con un comportamiento similar al antes descrito, genes sobre expresados y sub expresados en el mismo grupo con la presencia del factor de transcripción IFI16 que pertenece tanto a apoptosis como a autofagia.

En la red del subgrupo 2 (Figura 15B), se presenta de forma similar una separación entre genes influyentes, pero no es tan clara como en el caso del subgrupo 1. En esta visualización se aprecia un grupo compacto, altamente conectado entre sí, y otro grupo más disperso, con menor conexión, mientras que un sólo gen altamente influyente de autofagia está presente, pero aislado de los demás. Nuevamente, los factores de transcripción SIRT2 e IFI16 se presentan como influyentes en la red, cada uno en un grupo de genes distinto, pero ambos con expresión a la baja respecto a los controles. El grupo que se encuentra a la izquierda, con los genes NODAL, WNT108 y PSMC6 como más influyentes, presenta un comportamiento de expresión diferencial dividido. Por otro lado, el segundo grupo presenta genes mayormente sub-expresados, pero sus interrelaciones no son tan estrechas como las del grupo anterior.

En la red del subgrupo 3 (Figura 15C), los genes influyentes se dividen en 3 grupos, 2 de mayor tamaño. El grupo de la izquierda presenta genes tanto de apoptosis como de autofagia en una proporción 11:4, mientras que los otros dos grupos sólo presentan genes de apoptosis. Del grupo mixto es interesante que se repita la triada de genes de autofagia ATG10, PRKG1 y SLCO1A2, presente en el SG1, y que de forma reiterada todos los genes del grupo presenten sobre expresión con respecto a los controles. Sin embargo, el contexto en el que se encuentra en esta red parece ser muy distinto al de la red del subgrupo 1. Por otro lado, en el grupo de genes de la parte izquierda, el factor de transcripción HCLS1 se presenta como el gen más influyente de un grupo totalmente dominado por genes de apoptosis, con niveles de expresión principalmente a la baja. Cabe resaltar que la red del subgrupo 3 es la única en la que se pierde la fuerte influencia del gen ambivalente SIRT2.

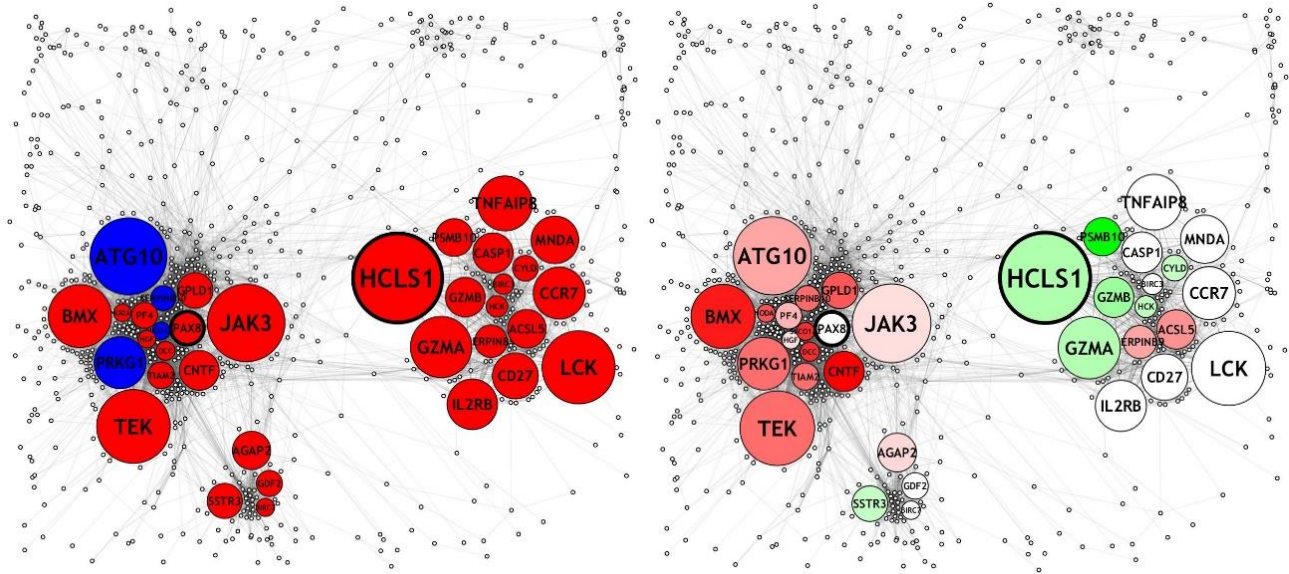
En la red del subgrupo 4 (Figura 15D), los genes influyentes se encuentran divididos en dos grupos: uno compacto de genes principalmente de apoptosis con SLC1A3 como único gen de autofagia; y otro grupo de genes formado por un conjunto compacto en la parte superior que se extiende hacia otro conjunto de menor tamaño en la parte inferior derecha de la visualización. En este caso, a diferencia de

los subgrupos anteriores, los genes influyentes se caracterizan por la ausencia de un grupo de genes con expresión diferencial a la alta, con sólo 4 genes influyentes con sobre expresión dispersos en la red. Por otro lado, la mayoría de los genes del grupo superior se encuentran sub-expresados, con genes tanto de apoptosis como de autofagia y ambivalentes.

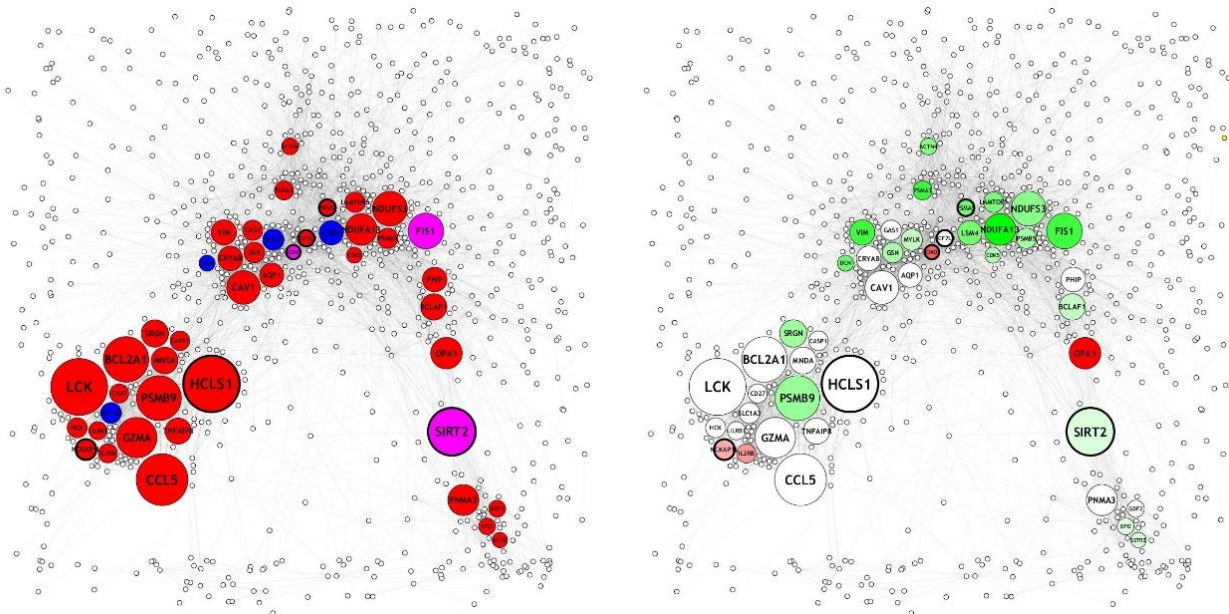


Figuras 15A y 15B. Visualizaciones de las redes de los subgrupos 1 y 2. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. En la derecha el color del nodo representa su nivel de expresión diferencial en comparación con el grupo control. En la izquierda el color del nodo representa el proceso al que pertenece: apoptosis en rojo, autofagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados. A) Red Subgrupo 1, B) Red subgrupo 2.

C



D



Figuras 15C y 15D. Visualizaciones de las redes de los subgrupos 3 y 4. El tamaño de los nodos refleja su importancia de acuerdo al CI. En la derecha el color del nodo representa su nivel de expresión diferencial en comparación con el grupo control. En la izquierda el color del nodo representa el proceso al que pertenece: apoptosis en rojo, autofagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados. C) Red Subgrupo 3, D) Red subgrupo 4

Subtipificación PAM50 y enriquecimiento de subtipos intrínsecos

Para comparar los subgrupos obtenidos a partir del análisis de desregulación con conocimiento previo sobre el cáncer de mama se usó el algoritmo PAM50. El análisis de las muestras, tanto casos como controles, con el algoritmo PAM50 dio como resultado su clasificación en uno de los 5 subtipos tumorales con relevancia clínica descritos por Parker y colaboradores: lumA, lumB, Her2, Basal y

normal-like. De manera poco eficiente, PAM50 clasificó 43 de 61 muestras de tejido sano como basales, mientras que clasificó 82 muestras tumorales como “normales”. Los resultados del análisis de PAM50 se emparejaron con los resultados del agrupamiento por desregulación, de forma que en cada subgrupo se puede desglosar cuantas muestras son clasificadas por PAM50 para cada subtipo tumoral. El resumen de los resultados de clasificación por PAM50 emparejados con el agrupamiento por subgrupos de desregulación se encuentra en la Tabla 3. Es importante recordar, que el gran total de las muestras en el análisis fue de 879, debido a que una muestra tumoral fue eliminada del estudio por ser clasificada como normal tanto por el algoritmo de PAM50, como por el agrupamiento jerárquico de los PDS.

	Basal	lumA	lumB	her2	normal	TOTAL
SG0	42	0	0	0	19	61
SG1	139	0	10	41	2	192
SG2	3	51	35	16	23	128
SG3	1	162	10	9	46	228
SG4	20	65	110	64	11	270
GRAN TOTAL:						879

Tabla 3. Resultados de la clasificación PAM50. Se muestran el número de muestras para cada subgrupo que fueron clasificadas en uno de los 5 subtipos clínicos. lumA = Luminal A, lumB = Luminal B, her2 = Her2-enriched, normal = normal-like.

El siguiente paso para establecer la significancia de esta integración de resultados es calcular la significancia estadística del enriquecimiento de los subtipos tumorales en los subgrupos de desregulación, esto es posible por medio de la prueba hipergeométrica. La prueba hipergeométrica consiste en calcular la significancia estadística de obtener un número k de éxitos cuando se hacen n muestreos en una población de tamaño N conteniendo K éxitos. Lo que se busca calcular con la prueba hipergeométrica es la probabilidad de que de forma aleatoria se haya dado el acomodo de muestras por subtipos intrínsecos en los subgrupos obtenidos por el análisis de desregulación, cuando esta probabilidad es extremadamente pequeña significa que obtener un enriquecimiento de los subgrupos por azar es en extremo poco probable. El enriquecimiento dio como resultado la identificación de los subtipos tumorales con presencia más significativa. Para todos los subgrupos se identificaron subtipos enriquecidos con distintos valores de significancia estadística, todos con valores de $p < 0.05$. Sin embargo, para algunos subgrupos se identificaron más de un subtipo tumoral enriquecido. Tal es el caso del Subgrupo 4 en el que se detectó enriquecimiento para los subtipos Her2 y luminal B; no obstante, la confianza estadística del enriquecimiento de luminal B es mucho mayor que el de Her2. Los resultados del enriquecimiento corregidos por el método de Bonferroni se encuentran en la Tabla 4.

	basal	lumA	lumB	her2	normal
SG0	7.64E-16	1	1	1	2.15E-05
SG1	8.01E-66	1	1	0.008668	1
SG2	1	0.06198	0.016483	1	0.029679
SG3	1	2.33E-48	1	1	7.41E-06
SG4	1	1	7.90E-27	2.22E-06	1

Tabla 4. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos de desregulación. Se presentan los valores de significancia estadística en valores de p con corrección por Bonferroni para las pruebas hipergeométricas realizadas. Valores significativos (valor de $p < 0.05$) se resaltan en negritas.

Con base en estos resultados, en secciones subsecuentes se hace referencia a tres de los subgrupos tumorales de la siguiente forma: SG1-Basal, SG3-LuminalA, SG4-LuminalB.

Validación de resultados de enriquecimiento de subgrupos

Para verificar la reproducibilidad del enriquecimiento obtenido en los subgrupos de desregulación, se realizó un análisis con 2 conjuntos de datos alternos obtenidos de las bases de datos de los proyectos The Cancer Genome Atlas (TCGA) y Molecular Taxonomy of Breast Cancer International Consortium (METABRIC). Los resultados obtenidos con la metodología propuesta muestran una gran consistencia entre los resultados obtenidos con los datos originales y los obtenidos con los conjuntos de datos alternos. Las muestras se agruparon nuevamente en grupos enriquecidos con los subtipos tumorales intrínsecos descritos por Parker.

Para el conjunto de datos de TCGA el enriquecimiento de los subgrupos se dio de forma específica para cada uno, sin tener subgrupos que tengan más de un subtipo intrínseco enriquecido. Mientras que para el conjunto de datos de METABRIC el enriquecimiento fue no único sólo para el subgrupo 4, el cual presentó enriquecimiento para el subtipo basal y Her2. Sin embargo, el enriquecimiento más significativo fue para el subtipo her2 con un valor de p de 4.27×10^{-42} , en comparación del enriquecimiento con valor de p de 0.000848 para el subtipo basal; por lo que el enriquecimiento del subgrupo 4 sería mejor asignado al subtipo Her2. Todos los resultados de los enriquecimientos de subtipo tumoral intrínseco para los conjuntos de datos de TCGA y METABRIC se muestran en las Tablas 5 y 6. Cabe aclarar que la numeración de los subgrupos en ningún conjunto de datos alternos es específica para un subtipo tumoral, los subgrupos fueron numerados sólo de forma arbitraria y ordenados para reflejar la similitud de resultados.

Es importante resaltar que para los dos conjuntos de datos de validación se encontró un grupo de muestras con altos valores de desregulación, denominado Subgrupo 5 (SG5), el cual no enriqueció con ninguno de los subtipos tumorales intrínsecos. Este grupo de muestras difiere en gran medida con las demás muestras, resultando en la generación de un grupo independiente de forma temprana en el cladograma, por lo que se tuvo que modificar la subdivisión inicial de 4 subgrupos a 5 subgrupos tumorales en estos conjuntos de datos para obtener el enriquecimiento de los 4 subtipos intrínsecos de Parker.

	basal	lumA	lumB	her2	normal
SG0	1	1	1	1	< 0.001
SG1	< 0.001	1	1	1	1
SG2	1	< 0.001	1	1	1
SG3	1	1	< 0.001	1	1
SG4	0.372702	1	1	< 0.001	1
SG5	1	1	1	0.78654	1

Tabla 5. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos del conjunto de datos de TCGA. Se presentan los valores de p con corrección por Bonferroni para las pruebas hipergeométricas realizadas en los subgrupos resultantes del análisis de desregulación en el conjunto de datos de TCGA. Valores significativos (valor de $p < 0.05$) se resaltan en negritas.

	basal	lumA	lumB	her2	normal
SG0	1	1	1	1	< 0.001
SG1	< 0.001	1	1	1	1
SG2	1	< 0.001	1	1	1
SG3	1	0.090264	< 0.001	1	1
SG4	< 0.001	1	1	< 0.001	1
SG5	0.78162	1	1	1	0.260568

Tabla 6. Valores de p del enriquecimiento de subtipos tumorales intrínsecos en los subgrupos del conjunto de datos de METABRIC. Se presentan los valores de p con corrección por Bonferroni para las pruebas hipergeométricas realizadas en los subgrupos resultantes del conjunto de datos de METABRIC. Valores significativos (valor de $p < 0.05$) se resaltan en negritas.

DISCUSIÓN

Análisis de desregulación y subgrupos tumorales

El análisis de desregulación y separación de las muestras tumorales en subgrupos tuvo la función de identificar la heterogeneidad tumoral a nivel transcripcional de los procesos de apoptosis y autofagia. Y al mismo tiempo, permitió apreciar similitudes entre distintos pacientes y las posibles correlaciones entre los procesos. Ya que los niveles de desregulación calculados parecen no ser azarosos y forman patrones como grupos con desregulación alta para grupos de procesos biológicos.

El primer resultado del análisis de desregulación con Pathifier es la matriz de desregulación, ilustrada en el *heatmap* de la Figura 3. Con estos resultados se determinaron 4 subgrupos tumorales, los cuales poseen cualidades distintivas en su desregulación de los procesos de apoptosis y autofagia (Figura 4). Una de las primeras preguntas que se generan a partir de esta división, es si tiene relevancia a la luz de lo ya conocido; por lo que el análisis PAM50 (Parker et al., 2009) fue un procedimiento obligado para contrastar esta división. Notablemente, la división de subgrupos tumorales basada en los PDS de Pathifier mostró un alto enriquecimiento con los subtipos intrínsecos de PAM50, de forma que es posible extrapolar los resultados obtenidos de cada subgrupo a los subtipos clínicos ya estudiados. De los 4 subgrupos tumorales identificados los subgrupos SG1, SG3 y SG4 obtuvieron la mejor correlación estadística, siendo así identificados como los subtipos Basal, Luminal A y Luminal B respectivamente (referidos en secciones subsecuentes como: SG1-Basal, SG3-LuminalA, SG4-LuminalB). Al tener un enriquecimiento del subtipo Her2 tanto en el SG1 y el SG4 con menos confianza estadística que los subtipos anteriores, se ha decidido no considerar este subtipo como relevante para la subdivisión establecida. Además, tanto el SG2 como el SG4 obtuvieron un enriquecimiento para el subtipo Luminal B, debido a que la confianza estadística del enriquecimiento fue mayor para el SG4, se prefirió usar a este para extrapolar los resultados del subtipo Luminal B.

Para resumir el nivel de desregulación de los procesos biológicos a nivel de subtipo se promediaron los valores de PDS para todas las muestras de cada subgrupo por proceso biológico, conformando una matriz que se muestra en forma de *heatmap* la Figura 5. Esta representación de los resultados hace evidente la separación de los procesos biológicos en 2 grupos de tamaño similar, uno en el que predominan altos niveles de desregulación en todos los subgrupos de cáncer de mama (inferior) y otro de baja desregulación (superior), en el que sólo los SG1-Basal y el SG2 presentan valores altos de desregulación en ciertos procesos biológicos. Estos altos valores de desregulación de los SG1-basal y SG2 también se hacen evidente en el gráfico de *boxplot* de la Figura 6. De forma interesante una mayor proporción de procesos de autofagia se encuentran en el grupo de alta desregulación.

Debido al gran número de procesos que conforman tanto la apoptosis y la autofagia de acuerdo a Gene Ontology se decidió observar a detalle sólo aquellos que poseen características sobresalientes sobre los demás. Cabe resaltar que sin realizar el análisis de todos los procesos por igual, aquellos con características distintivas no podrían ser identificados. En primera instancia en la Figura 7 se puede observar un acercamiento a los 10 procesos más desregulados en los subgrupos, de los que se puede apreciar que 4 de ellos pertenecen a autofagia, y más específicamente a sus propios procesos de regulación. Esto argumenta que a la par de la desregulación de apoptosis, la desregulación de la

autofagia tiene una repercusión en el establecimiento del fenotipo tumoral. Más aun, el fenotipo SG1-Basal se diferencia de los demás en presentar niveles bajos de desregulación en estos 10 procesos, a comparación de los demás subgrupos.

Por otro lado, los procesos con menor desregulación en todos los subgrupos pueden dar pauta de cuales componentes de la maquinaria celular se encuentran en un estado similar al fisiológico. En la Figura 8 se visualizan los 10 procesos menos desregulados en los subgrupos de cáncer. El proceso con menor desregulación es la “vía de señalización de apoptosis mediada por granzimas”, siendo el proceso con menor desregulación en todos los subgrupos. Esta vía se encarga de activar la fase de ejecución de la apoptosis tras la secreción de granzimas, un tipo específico de proteasas de serina, por parte de las células *natural-killer* y linfocitos T, que finaliza con la iniciación de la fase de ejecución de la apoptosis. Lo que sugiere que este elemento de la maquinaria de respuesta inmune antitumoral no es responsable de la falla por controlar a la enfermedad. Tratamientos que busquen aprovechar la funcionalidad casi fisiológica de la vía de señalización de apoptosis mediada por granzimas, podrían ser más eficientes en el control del cáncer de mama debido a que es poco desregulada en todos los subtipos tumorales. El Traztuzumab es un medicamento usado en el tratamiento del cáncer de mama, que podría deber su efectividad al actuar en esta vía, ya que se ha reportado una asociación significativa de su aplicación con el incremento de expresión de la Granzima B en linfocitos y una mayor presencia de células NK asociadas al tumor (Arnould et al., 2006).

Prosiguiendo con los procesos poco desregulados, destacan los procesos apoptóticos relacionados con la “actividad de endopeptidasa tipo cisteína involucrada en la vía de señalización de apoptosis”, ya que estos 4 procesos aparecen en este grupo de baja desregulación. Estas endopeptidasas son mejor conocidas como caspasas, y los PDS observados sugieren que por lo menos en 3 de los subgrupos sus niveles de expresión son similares a los fisiológicos. Sin embargo, para el caso del SG1-Basal los procesos de regulación, regulación negativa e inhibición se encuentran alterados en mayor medida, muy probablemente promoviendo la evasión de la apoptosis al deshabilitar estos mecanismos de activación de apoptosis.

Algo que resulta interesante al calcular los PDS es identificar las similitudes y diferencias entre los subgrupos identificados. Para este fin se identificaron los procesos con menor y mayor desviación estándar en sus PDS por subgrupo, ya que es un método ampliamente usado para medir la variación en un conjunto de datos. De esta forma, los procesos con menor desviación estándar poseen valores de desregulación muy similares en todos los subgrupos, mientras que aquellos con mayor desviación estándar poseen mayor variación en sus valores de desregulación entre los subgrupos. Identificar estos procesos puede ayudar a establecer particularidades de cada subgrupo y subtipo intrínseco. En la Figura 9 y en la Figura 10 se muestran los 10 procesos con menor y mayor desviación estándar respectivamente. De forma interesante, aquellos procesos con poca variación entre subgrupos pertenecen sólo a apoptosis, mientras que en los que poseen mayor variación, sí figuran procesos autofágicos, sugiriendo que la desregulación de los procesos de autofagia tiene un carácter más importante para la diferenciación entre subtipos tumorales que algunos procesos apoptóticos. Esta importancia ha sido reportada anteriormente, como lo es la asociación de marcadores de autofagia beclin-1, LC3B, LC3A a los subtipos de cáncer de mama, sin embargo, para ese caso sólo se

identificaron asociaciones para los tumores triples negativos y sólo a genes individuales (Choi et al., 2013).

Por un lado, de los procesos con poca variación, se puede argumentar que aquellos con alta desregulación son ampliamente conocidos por su importante papel como mecanismos antitumorales. Mientras que los procesos con bajos niveles de desregulación implicarían que estos no son tan relevantes para la tumorigénesis, ya que. Sin embargo, estos pueden serlo para generar estrategias antitumorales, tomando como prerrogativa que la baja desregulación puede indicar que es necesario mantener los niveles fisiológicos de estos componentes para la sobrevivencia de las células tumorales suponiendo un talón de Aquiles. Tal es el caso de la “señalización intrínseca de apoptosis en respuesta a estrés oxidativo”, proceso que se encuentra poco desregulado en todos los subgrupos. Sin embargo, esta vía podría ser de utilidad como blanco terapéutico para desencadenar una mayor tasa de apoptosis en el cáncer de mama al ser potencialmente más funcional que otros elementos de la maquinaria apoptótica. Se especula que el uso de diversos fármacos que generan estrés oxidativo en los tumores, tales como NOV-002 y la butionina sulfoximina, (Gorrini et al., 2013), pueden ser un mecanismo efectivo como terapia neo-adyuvante. Sin embargo, el uso de suplementación anti-oxidante en pacientes con cáncer para abrogar los efectos secundarios de la quimioterapia sigue siendo una práctica médica común. En su revisión del estrés oxidativo como terapia anti-cáncer, Gorrini y colaboradores argumentan en contra del uso de suplementación anti-oxidante en pacientes bajo tratamiento de quimioterapia, ya que diversos estudios sugieren que estos proveen protección a los tumores (Gorrini et al., 2013).

En contraparte, los procesos con alta variación en sus valores de PDS son interesantes por mostrar patrones que muestran alta especificidad al subtipo tumoral. De esta forma se hacen evidentes las diferencias entre el SG1-Basal en comparación a los demás subgrupos. Presentando alta desregulación en procesos de autofagia y mitocondriales que tanto el SG3-LumA y el SG4-LumB tienden a la baja: “regulación positiva de la macroautofagia”, “fragmentación mitocondrial involucrada en el proceso apoptótico”, “liberación de citocromo-c de la mitocondria” y “cambios apoptóticos mitocondriales”. Por otro lado los SG3-LumA y SG4-LumB poseen procesos altamente desregulados, que en el SG1-Basal se encuentran poco desregulados, como lo son la “regulación de la apoptosis vía extrínseca” y el “ensamblado de la vacuola autofágica”. Estas observaciones sugieren que la evasión de apoptosis en el cáncer de mama ocurre tras la desregulación de diferentes mecanismos en distintos subtipos tumorales. Por un lado en el subtipo Basal la descoordinación de la mitocondria, como parte esencial de la apoptosis, sería el mecanismo preferido de evasión de apoptosis. Mientras que en los subtipos Luminal A y B la desregulación de las “vías de regulación extrínseca” son el mecanismo por el cual es más probable que evadan la apoptosis. Un resumen de estos resultados se presenta en la Figura 16. El estudio experimental de estos procesos en los distintos subtipos de cáncer podría dar pie a entender la heterogeneidad tumoral en cáncer de mama.

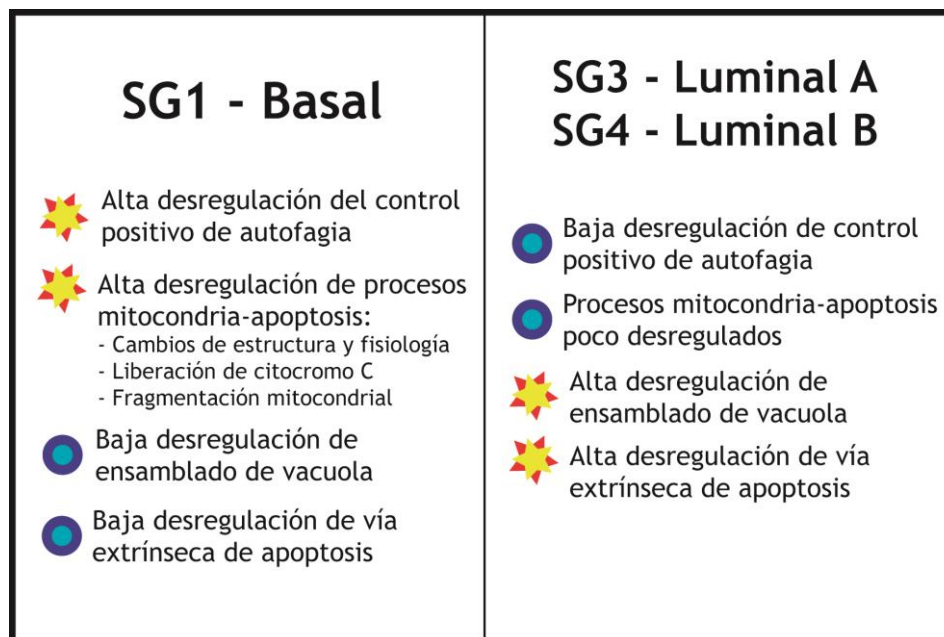


Figura 16. Resumen del contraste en los niveles de desregulación de procesos de apoptosis y autofagia en los subgrupos SG1-Basal, SG3-LuminalA y SG4-LuminalB.

El análisis de desregulación de vías permitió realizar una clasificación de todas las muestras en base a los niveles de expresión de los genes involucrados en los procesos estudiados. De esta forma es posible realizar análisis subsecuentes más detallados y sólo en conjuntos de muestras más homogéneas entre sí. Resulta interesante que esta misma clasificación converja con la clasificación por excelencia del cáncer de mama la PAM50. Estos resultados sugieren que el desarrollo del cáncer tiene predisposición a establecer fenotipos definidos y no sólo a una desregulación azarosa de los mecanismos de evasión de apoptosis; y tal vez en general para todas los eventos patofisiológicos que caracterizan al cáncer.

De manera suplementaria se evaluó la diferencia de usar los valores de PDS en comparación con los niveles de expresión de los genes aislados. Para este fin se usó la misma metodología de agrupamiento y enriquecimiento usando sólo los valores de expresión de genes, sin estar integrados por proceso biológico. Los resultados obtenidos fueron difíciles de interpretar pues los niveles de expresión no difieren de forma evidente en las muestras, además el agrupamiento clasificó 15 muestras de tejido sano como tumorales. Sin embargo, el enriquecimiento de los subgrupos resultantes fue estadísticamente significativo para 2 subgrupos (Luminal A y basal), mientras que se generaron 2 grupos aislados de muestras de tejido sano. Cabe resaltar que la mayor dificultad con este tipo de resultados es el de interpretación, ya que al tener más de 1000 variables la interpretación del resultado es más complicada, que con decenas de procesos biológicos; además de que el análisis estadístico de tantas variables se vuelve problemático por la llamada “maldición de la dimensión”. Por estos motivos se prefirió el uso de los PDS para modelar el estado de las muestras. El *heatmap* resultante de este análisis suplementario se encuentra en el Anexo 6.

Densidad de redes, procesos biológicos y subtipos

Una vez clasificadas las muestras de acuerdo a la desregulación de los proceso de apoptosis y autofagia, es posible realizar análisis dirigidos a conjuntos de muestras más homogéneos. De esta

forma señales como las correlaciones entre los genes deberán ser amplificadas, en comparación a tomar todas las muestras como un solo conjunto.

Las redes obtenidas por la metodología descrita en secciones anteriores para cada subgrupo fueron similares en tamaño y proporción de genes de apoptosis, autofagia y ambivalentes. Además de poseer correlaciones con valores de alta confianza estadística. Esto da certeza de un buen poder de comparación entre las redes y de confianza sobre las correlaciones, dado que el método proporciona resultados análogos.

De forma inocente los resultados esperados de una red de correlación entre los genes que han sido anotados a dos procesos biológicos “independientes” supondría la generación de una red en la que los genes de un proceso dado se encuentran más correlacionados entre sí que con los genes del otro proceso. Sin embargo, la realidad biológica puede ser otra. Los componentes de los procesos biológicos pueden estar más interconectados entre sí de lo que suponemos, y más de lo que conocemos. Es decir, la anotación de los genes a los procesos en los que participa todavía es incompleta.

Las correlaciones entre genes encontradas en cada subgrupo son ampliamente distintas. Y esto es claro desde el análisis de densidad de los procesos biológicos. Uno de los resultados más interesantes de las densidades de las redes se presenta en el SG3-LumA (Figura 12E), en el que los genes que participan en los dos procesos se desconectan dramáticamente entre sí, mientras que las conexiones al interior de cada proceso se ve aumentada y también la densidad de conexión entre los genes de apoptosis y autofagia de forma “directa”. Este resultado es único para el SG3-LumA y supondría una pérdida de la función de genes de intersección que son necesarios para la comunicación entre los dos procesos. Sugiriendo que al dejar de funcionar este vínculo, es necesaria la generación de otros mecanismos que compensen esta desconexión, resultando en una mayor conexión “directa” entre los procesos. Esto supondría una coordinación de los procesos de apoptosis y autofagia a través de genes aún no anotados como participantes de ambos procesos. El análisis a detalle de las interacciones da candidatos potenciales para este papel ambivalente se plantea en párrafos subsecuentes.

Resulta interesante que los niveles de densidad en la red del grupo control sean mucho mayores en comparación con los niveles de densidad de las redes de los grupos tumorales. Este evento puede ocurrir debido a que el número de genes que poseen las redes no es controlado en la metodología, pero sí el número de interacciones por red, por lo que al poseer menos genes el número posible de conexiones se reduce y esto hace que la densidad aumente. Sin embargo, esta diferencia puede no ser meramente metodológica, si no que tenga una explicación a nivel biológico. Gao y colaboradores sugieren que los sistemas resilientes tienen niveles de densidad altos en sus redes (Gao et al., 2016), por lo que los resultados indicarían que las células tumorales son menos resilientes a las perturbaciones en comparación a las células normales. Otra hipótesis que sugieren los resultados es la posibilidad de que relaciones entre los genes de apoptosis y autofagia en condiciones fisiológicas sean más estrechas entre sí, mientras que las relaciones en el estado patológico del cáncer son más dispersas, teniendo menor impacto afectar un proceso en el otro. Sin embargo, las correlaciones entre los genes de los subgrupos tumorales tienden a niveles mayores que los del grupo control, lo cual puede indicar una gran coordinación de los genes y funciones específicas en las condiciones patológicas del cáncer, a pesar de que a nivel de apoptosis y autofagia sean más dispersas.

Genes altamente influyentes de apoptosis y autofagia en cáncer de mama

La búsqueda de genes relevantes en el establecimiento de un fenotipo tumoral conduce a la identificación de genes con un papel relevante en las redes inferidas. Para esto se emplearon medidas de centralidad que calculan la relevancia de cada nodo, considerando las conexiones que poseen con los demás genes. La medida de centralidad seleccionada para lograr este objetivo fue el *Collective Influence*. Ya que esta medida seleccionó a los nodos más influyentes de cada red, los cuales dispersan un mensaje a través de toda la red de la forma más eficiente y rápida, en comparación con nodos seleccionados con otras medidas como el grado. Observando la localización y comportamiento de la dispersión de un mensaje a través de estos nodos, se puede argumentar que el CI es capaz de elegir los genes que tendrán mayor impacto en la red, además de posiblemente encontrar conjuntos de genes con papeles no redundantes en la red y con una distribución más dispersa, en comparación con los genes seleccionados simplemente por su grado, los cuales sólo se agruparon en una sola región de la red. Los genes seleccionados por su grado se agrupan todos en la zona visualmente más densa de la red, mientras que los nodos seleccionados por CI siempre se sitúan en más de una zona con alto grado de conexión. De esta forma las conexiones que comparten los genes influyentes por CI probablemente son menos redundantes y pueden comunicar el mensaje de forma más eficiente que los de mayor grado.

Identificando los genes influyentes para cada subgrupo tumoral y realizando un diagrama de Venn de su participación en los subgrupos, los genes altamente influyentes presentan un empalme participando en más de un subgrupo tumoral o hasta en los 4 como altamente influyentes, lo cual podría suponer similitudes funcionales entre los subtipos. Del empalme de los genes altamente influyentes en los subgrupos tumorales, mostrados en el diagrama de Venn en la Figura 13, resaltan los genes LCK y HCK. De forma interesante estos genes pertenecen a la familia de tirosina quinasas Src, y son identificados como influyentes en casi todos los subgrupos tumorales. LCK es identificado como un gen altamente influyente en todos los subtipos tumorales, y es responsable de la maduración de las células T (Molina et al. 1992), así como un regulador positivo de la apoptosis a través de la activación de las caspasas 3 y 9, esenciales para la iniciación de la apoptosis (Samraj et al., 2006). En valores de expresión LCK no tiene cambios significativos respecto al grupo control en ningún subgrupo tumoral. Es posible que los niveles fisiológicos sean suficientes para ejercer su influencia sobre otros genes y que su influencia dependa de su contexto, o que su mecanismo de patogenicidad sea independiente a su nivel de expresión e.g. mutación. De forma similar HCK es identificado como un gen influyente en los SG2, SG3-LumA y SG4-LumB. Se conoce que HCK participa en la regulación de la forma de las células (Howlett et al., 1999) así como es un regulador negativo de la apoptosis (Paliwal et al., 2007); asociado a carcinoma epinocelular y cervico adenoescamoso (Pletscher-Frankild, et al., 2015). Los valores de expresión de HCK sólo son significativamente más bajos en el SG3 ($\log_{FC} = -0.5765$), lo cual significaría alza en la tasa de apoptosis. Los roles específicos de LCK y HCK en cáncer de mama todavía no son claros, pero de acuerdo al análisis estos genes deben poseer roles altamente influyentes en la generación del fenotipo tumoral de los subtipos Lumina A, Luminal B. Una consulta de la base de datos de cBioPortal (Gao et al., 2013, Cerami et al., 2013) indica que de estos dos genes solo LCK presenta aberraciones en el número de copia, con amplificaciones en el 0.64% de los tumores del estudio de METABRIC, mientras que no se reportan mutaciones significativas en ninguno de los dos genes.

Visualización de redes por subgrupo.

Las visualizaciones mostradas en las Figuras 9 y 10 A-D nos muestran en primer lugar que las relaciones entre los genes de apoptosis y autofagia a nivel transcripcional cambian dramáticamente del estado fisiológico al estado tumoral, ya que las topologías y genes relevantes son completamente diferentes entre el estado fisiológico y los tumorales. Y no solo eso, sino que los genes que juegan un papel fundamental en el cambio de topología pueden tener cambios significativos en sus niveles de expresión. Cabe resaltar, que genes correlacionados, cercanos en posición en las visualizaciones, tienden a tener expresión diferencial coordinada, lo cual no es imperante para establecer una correlación entre ellos. Los genes que más influyen en las redes de distintos subgrupos tumorales son compartidos pero a la vez cada subgrupo posee un conjunto de genes altamente influyentes independiente. Por lo que es posible que compartan mecanismos tumorales subyacentes, aunque cada subtipo tumoral posea mecanismos tumorigénicos específicos, que los diferencie del resto.

En la RRG inferida del SGO-Control (Figura 14), casi la totalidad de los genes identificados como influyentes son de apoptosis y todos se encuentran aglomerados en la red formando un solo grupo, lo cual indicaría que estos genes comparten muchas relaciones entre sí formando un grupo compacto, lo que empata con los altos niveles de densidad observados anteriormente para el grupo control en comparación con los de los subgrupos tumorales. El gen identificado como el más influyente en esta red es SOX2, un factor transcripcional involucrado principalmente con el desarrollo embrionario y la determinación de destino celular (Rebhan et al., 1997), su participación en apoptosis sólo ha sido anotada en el proceso de regulación de caspasas gracias a un estudio en el que se evidencia su baja expresión en cáncer gástrico (Otsubo et al., 2008). Consultando la base de datos de cBioPortal (Gao et al., 2013, Cerami et al., 2013) se ha reportado que el gen SOX2 se encuentra amplificado en el 3.8% de los tumores del estudio de TCGA (Ciriello et al., 2015). De esta forma, estudios subsecuentes que identifiquen los mecanismos de funcionamiento de SOX2 en la apoptosis en estado fisiológico suponen una oportunidad de investigación. Por otro lado, el único gen influyente anotado en autofagia es MRPS10, un gen que codifica para una proteína ribosomal de mamífero y participa en el proceso de mitofagia en respuesta a la despolarización de la mitocondria (Orvedahl et al., 2011). Este gen puede ser un gen clave entre la comunicación de los procesos de apoptosis y autofagia en estado fisiológico o participar en ambos procesos y aún no ser anotado con evidencias experimentales.

En el caso del SG1-Basal la visualización de su RRG (Figura 15A) nos muestra que tanto los genes de apoptosis y autofagia juegan un papel importante en el estado tumoral de estos procesos. De estos genes ATG10 sólo se encuentra anotado a "Autofagia en respuesta a sobrecarga de retículo endoplásmico". Sin embargo, resulta muy interesante que los otros 4 genes más influyentes de autofagia (FGFBP1, SLCO1A2, ADAMTS7 y PRKG1) presentan un patrón muy específico: todos están agrupados de forma cercana en la visualización, lo que indica que comparten interacciones, todos están sobre-expresados, y además sólo están anotados para los mismos 2 términos GO: "Regulación de la autofagia" y "Regulación de la degradación mitocondrial". Es posible que la sobre-expresión de estos genes indiquen una alta degradación de mitocondrias. Lo cual también está en línea con la alta desregulación únicamente del SG1-Basal en los procesos relacionados a mitocondria, como se observa en el *heatmap* de la Figura 10. Estas evidencias sugieren que estos genes pueden ser responsables de

un entrecruzamiento de regulación de la apoptosis a través de la degradación de las mitocondrias vía autofagia, como ha sido propuesto anteriormente (Kim et al., 2007, Dieter & Gustaffson, 2012), y que este mecanismo se encuentra sobre-regulado en el SG1-Basal, promoviendo la evasión de la apoptosis por vía de la mitofagia. Cabe mencionar que de este conjunto de genes FGFBP1 posee asociación importante con otros tipos de carcinomas en la literatura: en células escamosas, de colon y con el mismo de mama (Harris et al., 2000), además de ser propuesto como un gen marcador importante para el pronóstico de cáncer de mama sin aludir a su función biológica en el trabajo de Martin y colaboradores (Martin et al., 2008). Adicionalmente, consultando la base de datos de Gene Ontology a través del servicio amiGO (Carbon et al., 2009), se encontró que ADAMTS7 ha sido anotado en el proceso de “mitofagia en respuesta a la despolarización de membrana” en las especies *R. Norvegicus*, *B. Taurus* y *S. scrofa*, con evidencia inferida sólo de forma electrónica (transferencia de anotación por ortología); mientras que el gen FGFBP1, tiene esta misma anotación pero sólo para la especie *R. norvegicus*. Estas evidencias apuntan a que el estudio experimental estos genes podría validar su funcionamiento en la promoción de mitofagia y evasión de la apoptosis en humano y en el subtipo basal de cáncer de mama.

En la red del SG3-LumA (Figura 15C) los genes de autofagia identificados con alta influencia son muy similares a los de la red del SG1-Basal, ya que se encuentran en un solo grupo de genes que están completamente sobre-expresados. De los 4 genes identificados sólo SERPINB10 no se encuentra como un gen influyente en el SG1-Basal. PRKG1, SLCO1A2 y SERPINB10 están anotados simultáneamente para el proceso de “regulación de la degradación de la mitocondria”. Su papel como reguladores todavía no ha sido definido como positivo ni negativo; sin embargo, con la información experimental de sobre-expresión de estos genes, y el papel de la mitofagia como un protector contra la apoptosis al degradar mitocondrias de forma selectiva (Dieter & Gustaffson, 2012), es probable que su naturaleza sea de reguladores positivos de la mitofagia, y al estar sobre expresados inhiban la apoptosis a través de promover la degradación de mitocondrias despolarizadas. Consultando la base de datos de Gene Ontology a través del servicio amiGO (Carbon et al., 2009), se encontró de forma similar que con los genes del SG1-basal, PRKG1, SLCO1A2 y SERPINB10 han sido anotados en el proceso de “mitofagia en respuesta a la despolarización de membrana” en las especies *B. taurus*, *C. lupus familiaris* y *S. scrofa*, con evidencias inferidas sólo de forma electrónica (transferencia de anotación por ortología), incrementando la posibilidad de que estos genes en humano tengan esta misma función, por lo que su validación experimental es especialmente prometedora.

Una característica relevante de la naturaleza de los genes influyentes de la red del SG3-LumA, es la completa ausencia de genes anotados en ambos procesos. Este resultado coincide con la baja densidad de conexión calculada entre genes ambivalentes para la red del SG3-LumA (Figura 12E). SIRT2 es un gen ambivalente identificado como altamente influyente en todos los grupos, exceptuando al SG3-LumA. Es posible que debido a la falta de participación de este único gen las relaciones en estos procesos cambien dramáticamente en el subtipo Luminal A. SIRT2 es un gen que participa como un regulador positivo de la apoptosis y un regulador negativo de la autofagia. Por lo que sus niveles de sub-expresión en los demás subgrupos es compatible con la hipótesis de que su papel sea de supresor de tumores. Evidencia experimental que soporta esta hipótesis y su sub-expresión en tumores se ha obtenido en gliomas humanos, sin embargo, su efecto funcional es atribuido a la regulación de la red

de microtúbulos (Hiratsuka et al., 2003) y no a su participación directa en la apoptosis y autofagia. La validación de SIRT2 como un supresor de tumores supone un experimento interesante, el cual podría realizarse en conjunción con la hipótesis de que esta función antitumoral está ligada a su participación en ambos procesos.

Un resultado común a las primeras tres redes de los subgrupos tumorales es el gen NODAL, como un gen con alta influencia y sobre-expresión, ya que se encuentra anotado en *Gene Ontology* como un “regulador positivo de la apoptosis”. Esto indicaría que el proceso de apoptosis debería incrementar su tasa. Sin embargo, el papel de NODAL en la apoptosis es quizá poco comprendido, ya que fue identificado como promotor de la apoptosis sólo en células del trofoblasto (Munir et al., 2004). Además, evidencia experimental confirma que el papel de NODAL en el cáncer de mama es diferente ya que experimentos de *knockdown* en líneas celulares agresivas de cáncer de mama incrementaron la tasa de apoptosis y detuvieron su ciclo celular (Kirsammer et al., 2014). Este hallazgo está en línea con los resultados obtenidos en las RRG y la hipótesis de que se trate de un proto-oncogen. Subsecuentes estudios experimentales del gen NODAL y su proteína podrían validar su papel como proto-oncogen influyendo en los procesos tanto de apoptosis como autofagia.

La red del SG4-LumB posee 11 genes influyentes compartidos con el SG3-LumA, sin embargo, poseen topologías muy distintas por lo que genes influyentes en un subtipo, no comparten el mismo contexto en el otro. Esto es evidente al identificar 3 genes como altamente influyentes anotados a ambos procesos en este subtipo: SIRT2, FIS1 y FOXO1. De estos genes FIS1 y FOXO1 sólo son identificados como influyentes en el SG4-LumB. FIS1 es un gen que codifica para una proteína que media la fisión de las mitocondrias y también puede promover la liberación de citocromo-C que finalmente lleva a la apoptosis (James et al., 2003; UniProt Consortium, 2014). Además, se encuentra anotado como “regulador positivo” de diversos procesos apoptóticos y como participante del proceso de mitofagia. El análisis de expresión diferencial muestra que la expresión de FIS1 se encuentra a la baja ($\log\text{-FoldChange} = -2.7054$), por lo que su poca expresión en consecuencia disminuiría la tasa de apoptosis. Por otro lado, FOXO1 ha sido validado experimentalmente como un supresor de tumores induciendo autofagia que lleva a la muerte celular en tumores de colon humano y en xenotransplante en ratón (Zhao et al., 2010). Adicionalmente, OPA1 es el gen con mayor sobre-expresión en los genes más influyentes en la RRG del SG4-LumB ($\log\text{-FoldChange} = 1.559$), y es un regulador negativo de la apoptosis así como regulador de la fusión mitocondrial, por lo que su presencia debe ser un gran impedimento para lograr la apoptosis en las células tumorales, dada su potencial gran influencia. OPA1 es un gen fuertemente asociado a atrofia óptica dominante (Delettre et al., 2000), y su participación en cáncer apenas está siendo vislumbrada en cáncer de pulmón (Chen et al., 2011). Lo cual lo coloca como un gen candidato a validación subsecuente, por ejemplo promoviendo su degradación con miRNAs dirigidos en líneas celulares similares al subtipo Luminal B y medir si existe un incremento significativo en la tasa de apoptosis *per se* y/o si esto ocurre en sinergia con fármacos antitumorales existentes.

Los genes influyentes de cada subgrupo también pueden ser tratados como potenciales blancos farmacológicos, de los cuales se esperaría el mayor efecto para los procesos de apoptosis y autofagia. Es posible argumentar que restableciendo la expresión fisiológica de estos genes la expresión de los demás genes en la red debería revertir al estado fisiológico, por lo que el estado tumoral se revertiría.

Subsecuentes estudios experimentales podrían seguir esta hipótesis para validarla. Los genes más influyentes de las RRG e información suplementaria se encuentran en el Anexo 4. En conclusión las visualizaciones de cada una de las redes identifican genes que pueden ser de gran relevancia en el diseño experimental de metodologías que confirmen sus papeles en los procesos de apoptosis y autofagia en cáncer de mama. Especial atención se debe dar a los genes que poseen tanto un nivel alto de influencia en la red como de expresión diferencial, pues estos sugieren tanto un potencial funcional como de maniobrabilidad experimental.

Subtipificación PAM50 y comparación de subgrupos

Uno de los resultados más interesantes del estudio es el enriquecimiento de los subtipos tumorales intrínsecos (LumA, LumB y Basal) en los subgrupos de desregulación de apoptosis y autofagia. Este resultado implica que todas las observaciones y resultados anteriormente mencionados, de los subgrupos con enriquecimiento (niveles de desregulación, sus RRG y los genes que potencialmente afectan más al establecimiento del fenotipo tumoral) pueden ser extrapolados a otros tumores identificados por subtipo intrínseco y que los resultados tengan relevancia clínica. Ya que se ha demostrado que los subtipos intrínsecos tumorales tienen relevancia para designar el pronóstico del pacientes y predecir la respuesta terapéutica en cáncer de mama primario (Senkus et al., 2013). Además, es importante recalcar que de los 50 genes que analiza PAM50 sólo 8 son analizados con la metodología aquí presentada, ya que entran en la categoría de procesos de apoptosis o autofagia. Por un lado los genes BAG1, BIRC5, EGFR, MELK, MYC y SFRP1 pertenecen a los procesos de apoptosis, mientras que BCL2 y MAPT pertenecen tanto a los procesos de autofagia como a los de apoptosis.

En el conjunto de datos principales del estudio sólo se lograron identificar 3 subgrupos de desregulación con alta confianza estadística y sin empalmes entre grupos. Razón por la cual se dio prioridad a los SG1-Basal, SG3-LumA y SG4-LumB para la interpretación y discusión de sus resultados a detalle. Una de las explicaciones más plausibles de este evento es la construcción del conjunto de datos a partir de distintos experimentos y el hecho de que estos datos no fueron construidos con el propósito de un análisis global del cáncer de mama. Pese a que todas las muestras fueron analizadas con los mismos protocolos y tecnologías, diferencias mínimas en la ejecución pueden ser la causa de no poder identificar con alta confianza a un conjunto de muestras enriquecidas con el subtipo Her2. Sin embargo, los resultados obtenidos son valiosos, ya que aun siendo datos obtenidos a partir de experimentos independientes se logran encontrar patrones coherentes.

Validación de resultados de enriquecimiento

La búsqueda de reproducibilidad de los resultados con la misma metodología con conjuntos de datos independientes fue positiva, ya que se obtuvieron resultados similares y adicionales. En los conjuntos de datos de TCGA y METABRIC se identificaron con una alta significancia estadística enriquecimientos individuales para los cuatro subgrupos de desregulación sin tener empalmes de enriquecimiento entre los subgrupos. Estos resultados demuestran que la metodología fue capaz de extraer significado biológico de una muestra representativa de la población de forma reproducible.

Adicionalmente se identificó en ambos conjuntos de datos un subgrupo de muestras que difieren ampliamente de los 4 subgrupos identificados como subtipos intrínsecos. Este subgrupo de muestras

puede suponer un hallazgo importante al tratarse de un subtipo difícil de clasificar. Ya que se puede tratar de un grupo con características distintivas e independientes de los subtipos intrínsecos establecidos actualmente.

El encontrar grupos de muestras que no son clasificados consistentemente con los subtipos tumorales descritos anteriormente tiene precedentes. Por ejemplo, el estudio de Curtis y colaboradores en el proyecto METABRIC propone una clasificación dividida en 10 subgrupos (Curtis et al., 2012). Por otro lado, Milioli y colaboradores desarrollaron una metodología en la que refinan la clasificación de PAM50 subdividiendo a las muestras en 6 grupos, logrando mayor consistencia entre distintos clasificadores (Milioli et al., 2016). La clasificación refinada de Milioli incluye los 4 grupos tradicionales de Parker más un grupo denominado *normal-like* y otro denominado *inconsistente*, debido a que usando diversos clasificadores un grupo de muestras resulta inconsistente en su clasificación, además de poseer niveles de expresión distintos en comparación a los otros sub grupos (Milioli et al., 2016). Esto argumenta a la heterogeneidad tumoral aun no caracterizada adecuadamente en el cáncer de mama.

Finalmente, la obtención de resultados más significativos y reproducibles en los conjuntos de datos alternos puede ser debida a un mayor control de las condiciones experimentales, ya que todos los experimentos y el diseño experimental fueron realizados considerando que los datos finales serían empleados para realizar un análisis global del cáncer de mama. Además, el uso de tecnologías y plataformas más modernas como lo es la secuenciación de RNA en el caso de los datos de TCGA puede proveer resultados más reproducibles y con menor ruido técnico.

CONCLUSIONES

A partir del cúmulo de resultados presentados y su discusión se concluye:

- **Existe una fuerte correlación entre la desregulación de los procesos de apoptosis y autofagia con los subtipos tumorales moleculares de cáncer de mama.** El estudio realizado logró identificar, a partir del análisis de desregulación de los procesos de apoptosis y autofagia, grupos tumorales correspondientes con los subtipos moleculares ampliamente aceptados. Al ser posible esta identificación se establece un vínculo directo entre el nivel de desregulación de estos procesos y los distintos fenotipos tumorales en cáncer de mama.
- **Lo autofagia es un proceso que presenta una alta desregulación en cáncer de mama, casi tan alta como los procesos de apoptosis y su desregulación es subtipo específica, probablemente siendo un determinante del fenotipo tumoral.** Los resultados sugieren que el proceso de autofagia funge un papel relevante en el establecimiento del fenotipo tumoral, ya que los niveles de desregulación de sus procesos difieren entre subtipos tumorales. Esto puede conllevar a que la generación de distintos subtipos dependa de la desregulación de procesos específicos. Como lo es la desregulación del control positivo de la macroautofagia en el caso del subtipo Basal, o el ensamblado de la vacuola en los subtipos luminales. Por lo que su importancia en la tumorigénesis puede ser menospreciada actualmente.
- **Los resultados sugieren que los genes que participan en ambos procesos son genes clave en la comunicación de la apoptosis y la autofagia.** Además se plantea que la intersección entre estos dos procesos debe ser mucho mayor. Por lo que se propone a un grupo de genes como los mejores candidatos para este papel de intersección y comunicación entre los dos procesos en el estado tumoral. Los genes propuestos se muestran en la Tabla 7.

Gen	Proceso biológico	Resultados
SOX2	Apoptosis	Es el gen más influyente en la RRG inferida a partir de los datos del grupo control y posee el número más alto de conexiones con genes de autofagia. Se encuentra amplificado en el 3.8% de los tumores del proyecto TCGA.
MRPS10	Autofagia	Es el único gen autofágico altamente influyente en la RRG del grupo control y posee el número más alto de conexiones con genes de apoptosis. Probablemente tiene un papel relevante en apoptosis.
FGFBP1		Sobre-expresión coordinada en el subtipo basal, asociada a desregulación de procesos mitocondriales apoptóticos. Anotación vía inferencia electrónica en “mitofagia en respuesta a despolarización de mitocondria”. Reducir los niveles de expresión de estos genes podría restablecer la apoptosis reduciendo la mitofagia.
ADAMTS7		
SLCO1A2		
PRKG1		
SERPINB10		Sobre-expresión coordinada en el subtipo luminal A asociada a degradación autofágica de mitocondria. Anotados vía inferencia electrónica en “mitofagia en respuesta a despolarización de mitocondria”. Reducir los niveles de expresión de estos genes podría restablecer la apoptosis.
PRKG1		
SLCO1A2		

Tabla 7. Genes candidatos a participar en ambos procesos.

Por otro lado, al nivel de genes individuales, la inferencia de RRG integrada con los niveles de expresión diferencial permitió identificar patrones de genes con mayor potencial en el establecimiento de la patofisiología tumoral en los distintos subtipos. Ya que muchos de los genes influyentes identificados en “islas” de las RRG presentan patrones de expresión diferencial coordinada. Sugiriendo que estos genes no solo funcionan en concierto sino que cambios en sus niveles de expresión son relevantes en la tumorigénesis y podrían formar parte de los mismos procesos tanto en el fenotipo tumoral como en el estado fisiológico. Se propone que el restablecimiento de los niveles de expresión de los genes más influyentes de cada subtipo podría restablecer a estado fisiológico los procesos biológicos de apoptosis y autofagia afectados en cada subtipo tumoral. En la Tabla 8 se presentan los genes influyentes en los subgrupos SG1-Basal, SG3-LuminalA y SG4-LuminalB con mayor cambio en su tasa de expresión como genes candidatos para estudios experimentales dirigidos a validar esta hipótesis.

SG1-Basal		SG3-LumA		SG4-LumB	
ID	LogFCh.	ID	LogFCh.	ID	LogFCh.
ACTN4	-2.15848	PSMB10	-2.36521	NDUFA13	-3.61886
PSMC6	-1.50147	GZMB	-0.91163	PSMA3	-2.80454
CALR	-1.45938	HCLS1	-0.74355	VIM	-2.72054
TGFB1	-1.382	GZMA	-0.68646	FIS1	-2.7054
PSMB5	-1.21203	HCK	-0.57653	DCN	-2.07561
NODAL	1.219388	NODAL	1.381282	NCKAP1L	0.495957
SLCO1A2	1.42891	DCC	1.687117	IL2RB	0.630864
DCC	1.575055	SLCO1A2	1.695183	FOXO1	0.996433
BMX	1.640537	BMX	1.841147	OPA1	1.559525
CNTF	1.843108	CNTF	2.054101		

Tabla 8. Genes más influyentes con mayor *log Fold-Change* por subtipo tumoral. Se propone que restablecer los niveles de expresión de estos genes podría restablecer el funcionamiento fisiológico de los procesos de apoptosis y autofagia en los subtipos tumorales identificados.

Además de identificar genes coordinados y potencialmente de intersección entre los procesos de apoptosis y autofagia, al tener anotaciones parciales o no ser aun identificados como oncogenes o genes antitumorales en cáncer de mama, suponen una oportunidad de investigación excelente para subsecuente validación experimental. Ya sea determinando su papel como reguladores positivos o negativos de la apoptosis, o como promotores o supresores del fenotipo tumoral, ya que muchos de estos candidatos aún no han sido reportados de forma experimental con roles relevantes al cáncer de mama o los procesos de apoptosis y autofagia.

Finalmente la validación con otros conjuntos de datos, evidencian la importancia de realizar este tipo de análisis para obtener resultados reproducibles en torno a los procesos de apoptosis y autofagia y su relevancia en cáncer de mama y fungir como guía para experimentación prospectiva. Los resultados muestran que es posible una refinación e integración de todos los conjuntos de datos disponibles para realizar una determinación definitiva de los mejores candidatos como genes de intersección, además de lograr caracterizar a un grupo de muestras que presenta patrones distintivos de desregulación en los procesos de apoptosis y autofagia que no es identificado como un subtipo molecular aún.

LITERATURA CITADA

- Abedin, M.J., Wang, D., McDonnell, M.A., Lehmann, U. and Kelekar, A., 2007. Autophagy delays apoptotic death in breast cancer cells following DNA damage. *Cell Death & Differentiation*, 14(3), pp.500-510.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. and Harris, M.A., 2000. Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), pp.25-29.
- Baca-López, K., Mayorga, M., Hidalgo-Miranda, A., Gutiérrez-Nájera, N. and Hernández-Lemus, E., 2012. The role of master regulators in the metabolic/transcriptional coupling in breast carcinomas. *PLoS one*, 7(8), p.e42678.
- Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R. and Califano, A., 2005. Reverse engineering of regulatory networks in human B cells. *Nature genetics*, 37(4), pp.382-390.
- Cancer Genome Atlas Network, 2012. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), pp.61-70.
- Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S. and Web Presence Working Group, 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics*, 25(2), pp.288-289.
- Chen, C.Y., Fang, H.Y., Chiou, S.H., Yi, S.E., Huang, C.Y., Chiang, S.F., Chang, H.W., Lin, T.Y., Chiang, I. and Chow, K.C., 2011. Sumoylation of eukaryotic elongation factor 2 is vital for protein stability and anti-apoptotic activity in lung adenocarcinoma cells. *Cancer science*, 102(8), pp.1582-1589.
- Choi, J., Jung, W. and Koo, J.S., 2013. Expression of autophagy-related markers beclin-1, light chain 3A, light chain 3B and p62 according to the molecular subtype of breast cancer. *Histopathology*, 62(2), pp.275-286.
- Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C. and Bowlby, R., 2015. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*, 163(2), pp.506-519.
- Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y. and Gräf, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403), pp.346-352.
- Delettre, C., Lenaers, G., Griffoin, J.M., Gigarel, N., Lorenzo, C., Belenguer, P., Pelloquin, L., Grosgeorge, J., Turc-Carel, C., Perret, E. and Astarie-Dequeker, C., 2000. Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nature genetics*, 26(2), pp.207-210.
- Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M.S. and Bergh, J., 2007. 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), pp.3207-3214.

- Drier, Y., Sheffer, M. and Domany, E., 2013. Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences*, 110(16), pp.6388-6393. <http://www.weizmann.ac.il/complex/compphys/software/yotam/pathifier/installation.html>
- Easton, D.F., Ford, D. and Bishop, D.T., 1995. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Breast Cancer Linkage Consortium. American journal of human genetics*, 56(1), p.265.
- Economou, A.D., Ohazama, A., Porntaveetus, T., Sharpe, P.T., Kondo, S., Basson, M.A., Gritli-Linde, A., Cobourne, M.T. and Green, J.B., 2012. Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate. *Nature genetics*, 44(3), pp.348-351.
- Eisenberg-Lerner, A. et al., 2009. Life and death partners: apoptosis, autophagy and the cross-talk between them. *Cell death and differentiation*, 16(7), pp.966-975
- Elmore, S., 2007. Apoptosis: a review of programmed cell death. *Toxicologic pathology*, 35(4), pp.495-516.
- Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., MacGrogan, G., Bergh, J., Cameron, D., Goldstein, D. and Duss, S., 2005. Identification of molecular apocrine breast tumours by microarray analysis. *Breast Cancer Research*, 7(2), p.1.
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F., 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. Disponible en: <http://globocan.iarc.fr>
- Ford, D., Easton, D.F., Bishop, D.T., Narod, S.A. and Goldgar, D.E., 1994. Risks of cancer in BRCA1-mutation carriers. *The Lancet*, 343(8899), pp.692-695.
- Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J. and Sobol, H., 1998. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics*, 62(3), pp.676-689.
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. and Cerami, E., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, 6(269), p.p11.
- Gao, J., Barzel, B. and Barabási, A.L., 2016. Universal resilience patterns in complex networks. *Nature*, 530(7590), pp.307-312.
- García-Campos, M.A., 2016. Pa3BC: Pathway Analyses of Apoptosis and Autophagy on Breast Cancer, First release, DOI: 10.5281/zenodo.166107
- García-Campos, M.A., Espinal-Enríquez, J. and Hernández-Lemus, E., 2015. Pathway analysis: state of the art. *Frontiers in physiology*, 6.
- Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A., 2004. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3), pp.307-315.

Gene Ontology Consortium, 2008. The gene ontology project in 2008. *Nucleic acids research*, 36(suppl 1), pp.D440-D444.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. and Hornik, K., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10), p.R80.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. and Bloomfield, C.D., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439), pp.531-537.

Gorrini, C., Harris, I.S. and Mak, T.W., 2013. Modulation of oxidative stress as an anticancer strategy. *Nature reviews Drug discovery*, 12(12), pp.931-947.

Hanahan, D. & Weinberg, R. a., 2011. Hallmarks of cancer: The next generation. *Cell*, 144(5), pp.646-674.

Hanahan, D. & Weinberg, R.A., 2000. The Hallmarks of Cancer. *Cell*, 100(1), pp.57–70.

Harris, V.K., Coticchia, C.M., List, H.J., Wellstein, A. and Riegel, A.T., 2000. Mitogen-induced expression of the fibroblast growth factor-binding protein is transcriptionally repressed through a non-canonical E-box element. *Journal of Biological Chemistry*, 275(37), pp.28539-28548.

Hernández-Lemus, E. and Rangel-Escareño, C., 2011. The role of information theory in gene regulatory network inference. *Information Theory: New Research*, pp.109-144.

Hillmer, R.A., 2015. Systems Biology for Biologists. *PLoS Pathog*, 11(5), p.e1004786.

Hiratsuka, M., Inoue, T., Toda, T., Kimura, N., Shirayoshi, Y., Kamitani, H., Watanabe, T., Ohama, E., Tahimic, C.G., Kurimasa, A. and Oshimura, M., 2003. Proteomics-based identification of differentially expressed genes in human gliomas: down-regulation of SIRT2 gene. *Biochemical and biophysical research communications*, 309(3), pp.558-566.

Howlett, C.J., Bisson, S.A., Resek, M.E., Tigley, A.W. and Robbins, S.M., 1999. The proto-oncogene p120 Cbl is a downstream substrate of the Hck protein-tyrosine kinase. *Biochemical and biophysical research communications*, 257(1), pp.129-138.

Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H. and Wong, J.E., 2006. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer research*, 66(21), pp.10292-10301.

James, D.I., Parone, P.A., Mattenberger, Y. and Martinou, J.C., 2003. hFis1, a novel component of the mammalian mitochondrial fission machinery. *Journal of Biological Chemistry*, 278(38), pp.36373-36379.

Johnstone, R.W., Ruefli, A.A. and Lowe, S.W., 2002. Apoptosis: a link between cancer genetics and chemotherapy. *Cell*, 108(2), pp.153-164.

Kerr, J.F., Wyllie, A.H. and Currie, A.R., 1972. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer*, 26(4), p.239.

Kim, I., Rodriguez-Enriquez, S. and Lemasters, J.J., 2007. Selective degradation of mitochondria by mitophagy. *Archives of biochemistry and biophysics*, 462(2), pp.245-253.

Kirsammer, G., Strizzi, L., Margaryan, N.V., Gilgur, A., Hyser, M., Atkinson, J., Kirschmann, D.A., Seftor, E.A. and Hendrix, M.J., 2014, December. Nodal signaling promotes a tumorigenic phenotype in human breast cancer. In *Seminars in cancer biology* (Vol. 29, pp. 40-50). Academic Press.

Kroemer, G. et al., 2009. Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. *Cell Death Differ*, 16(1), pp.3–11.

Langle, O., 2016. *Complex_Networks* First Release, DOI: 10.5281/zenodo.49699

Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M.F., 2007. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *New England Journal of Medicine*, 356(3), pp.217-226.

Livshits, A., Git, A., Fuks, G., Caldas, C. and Domany, E., 2015. Pathway-based personalized analysis of breast cancer expression data. *Molecular oncology*, 9(7), pp.1471-1483.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R.D. and Califano, A., 2006. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics*, 7(Suppl 1), p.S7.

Martin, K.J., Patrick, D.R., Bissell, M.J. and Fournier, M.V., 2008. Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets. *PloS one*, 3(8), p.e2994.

Merico, D., Isserlin, R., Stueker, O., Emili, A. and Bader, G.D., 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one*, 5(11), p.e13984.

Meyer, P.E., Lafitte, F. and Bontempi, G., 2008. Minet: an open source R/Bioconductor package for mutual information based network inference. *BMC bioinformatics*, 9, p.461.

Miki Y, Swensen J, Shattuck-Eidens D, et al. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 266: 66-71.

Milioli, H.H., Vimieiro, R., Tishchenko, I., Riveros, C., Berretta, R. and Moscato, P., 2016. Iteratively refining breast cancer intrinsic subtypes in the METABRIC dataset. *BioData mining*, 9(1), p.1.

Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T. and Bergh, J., 2005. 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), pp.13550-13555.

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. and Massagué, J., 2005. Genes that mediate breast cancer metastasis to lung. *Nature*, 436(7050), pp.518-524.

Molina, T.J., Kishihara, K., Siderovskid, D.P., Van Ewijk, W., Narendran, A., Timms, E., Wakeham, A., Paige, C.J., Hartmann, K.U., Veillette, A. and Davidson, D., 1992. Profound block in thymocyte development in mice lacking p56lck.

Morone, F. and Makse, H.A., 2015. Influence maximization in complex networks through optimal percolation. *Nature*.

Murtagh, F. and Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *Journal of Classification*, 31(3), pp.274-295.

Newman, M.E., 2003. The structure and function of complex networks. *SIAM review*, 45(2), pp.167-256.

Orvedahl, A., Sumpter Jr, R., Xiao, G., Ng, A., Zou, Z., Tang, Y., Narimatsu, M., Gilpin, C., Sun, Q., Roth, M. and Forst, C.V., 2011. Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature*, 480(7375), pp.113-117.

Otsubo, T., Akiyama, Y., Yanagihara, K. and Yuasa, Y., 2008. SOX2 is frequently downregulated in gastric cancers and inhibits cell growth through cell-cycle arrest and apoptosis. *British journal of cancer*, 98(4), pp.824-831.

Paglin, S., Hollister, T., Delohery, T., Hackett, N., McMahill, M., Sphicas, E., Domingo, D. and Yahalom, J., 2001. A novel response of cancer cells to radiation involves autophagy and formation of acidic vesicles. *Cancer research*, 61(2), pp.439-444.

Paliwal, P., Radha, V. and Swarup, G., 2007. Regulation of p73 by Hck through kinase-dependent and independent mechanisms. *BMC molecular biology*, 8(1), p.1.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z. and Quackenbush, J.F., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology*, 27(8), pp.1160-1167.

Pawitan, Y., Bjöhle, J., Amler, L., Borg, A.L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S. and Liu, E.T., 2005. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Research*, 7(6), p.1.

Pletscher-Frankild, S., Pallegà, A., Tsafou, K., Binder, J.X. and Jensen, L.J., 2015. DISEASES: Text mining and data integration of disease-gene associations. *Methods*, 74, pp.83-89.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J. and Lancet, D., 1997. GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics*, 13(4), p.163.

Rhodes, D.R. and Chinnaiyan, A.M., 2005. Integrative analysis of the cancer transcriptome. *Nature genetics*, 37, pp.S31-S37.

Samraj, A.K., Stroh, C., Fischer, U. and Schulze-Osthoff, K., 2006. The tyrosine kinase Lck is a positive regulator of the mitochondrial apoptosis pathway by controlling Bak expression. *Oncogene*, 25(2), pp.186-197.

Senkus, E., Kyriakides, S., Penault-Llorca, F., Poortmans, P., Thompson, A., Zackrisson, S. and Cardoso, F., 2013. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, p.mdt284.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498-2504.

Smyth, G.K., 2005. Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor* (pp. 397-420). Springer New York.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B. and Desmedt, C., 2006. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *Journal of the National Cancer Institute*, 98(4), pp.262-272.

Team, R.C., 2013. R: A language and environment for statistical computing.

Thorburn, A., 2008. Apoptosis and autophagy: regulatory connections between two supposedly different processes. *Apoptosis*, 13(1), pp.1-9.

Thorlacius, S., Struewing, J.P., Hartage, P., Olafsdottir, G.H., Sigvaldason, H., Tryggvadottir, L., Wacholder, S., Tulinius, H. and Eyfjörd, J.E., 1998. Population-based study of risk of breast cancer in carriers of BRCA2 mutation. *The Lancet*, 352(9137), pp.1337-1339.

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. and Gerry, N.P., 2008. Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *International journal of cancer*, 122(7), pp.1557-1566.

Turing, A.M., 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 237(641), pp.37-72.

Ullman, E., Fan, Y., Stawowczyk, M., Chen, H.M., Yue, Z. and Zong, W.X., 2008. Autophagy promotes necrosis in apoptosis-deficient cells in response to ER stress. *Cell death and differentiation*, 15(2), p.422.

UniProt Consortium, 2014. UniProt: a hub for protein information. *Nucleic acids research*, p.gku989.

Van't Veer, L.J., Dai, H., Van De Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. and Schreiber, G.J., 2002. Gene expression profiling predicts clinical outcome of breast cancer. *nature*, 415(6871), pp.530-536.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S. and Schwartz, M., 2009. gplots: Various R programming tools for plotting data. *R package version*, 2(4).

Wickham, H. and Francois, R., 2014. dplyr: A grammar of data manipulation. *R package version 0.2*.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G. and Barfoot, R., 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature*, 378(6559), pp.789-792.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W., 1997. Gene expression profiles in normal and cancer cells. *Science*, 276(5316), pp.1268-1272.

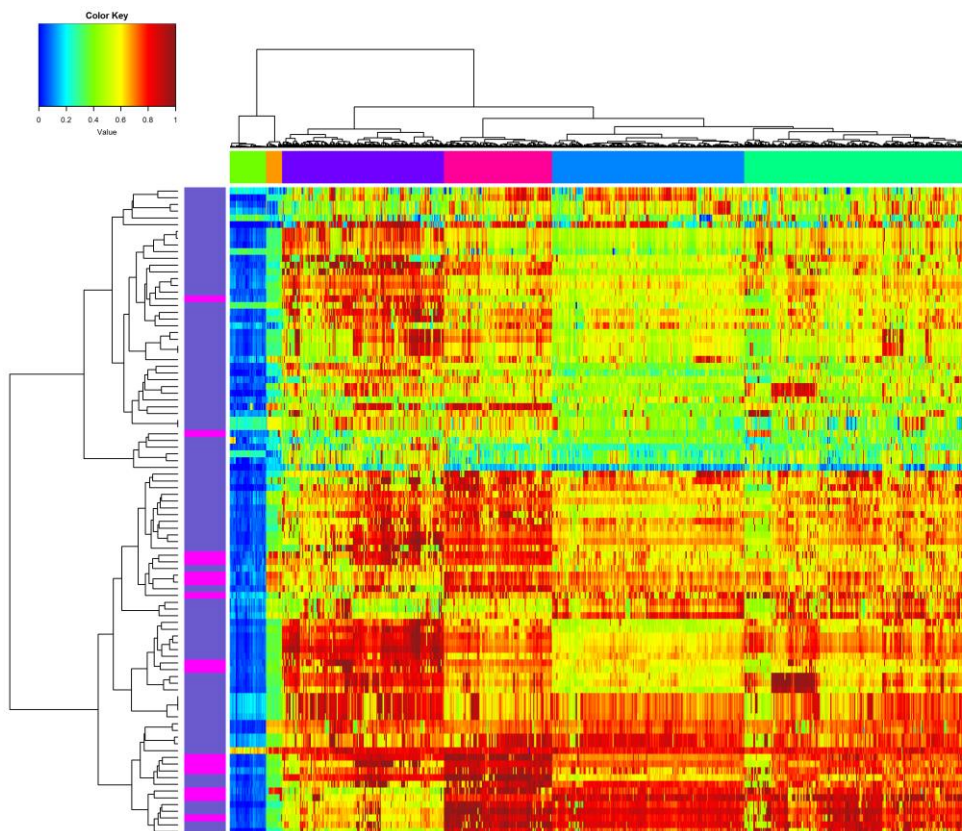
Zhao, Y., Yang, J., Liao, W., Liu, X., Zhang, H., Wang, S., Wang, D., Feng, J., Yu, L. and Zhu, W.G., 2010. Cytosolic FoxO1 is essential for the induction of autophagy and tumour suppressor activity. *Nature cell biology*, 12(7), pp.665-675.

ANEXOS

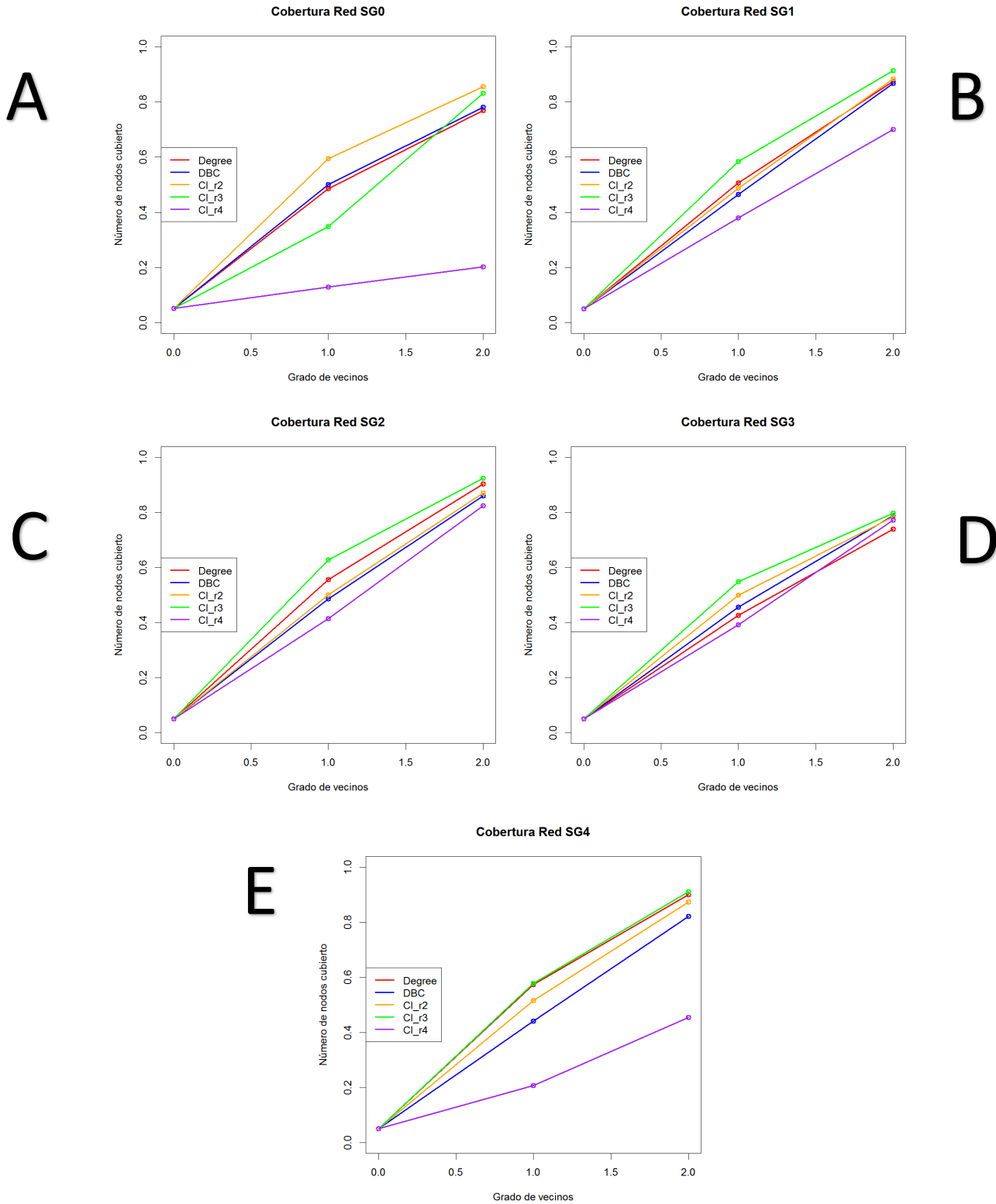
Anexo 1. Código SQL para consulta genérica de términos GeneOntology en GOOSE:

```
SELECT DISTINCT descendant.acc, descendant.name, descendant.term_type
FROM
  term
  INNER JOIN graph_path ON (term.id=graph_path.term1_id)
  INNER JOIN term AS descendant ON (descendant.id=graph_path.term2_id)
WHERE term.name='nombre de término GO' AND distance <> 0 ;
```

Anexo 2. *Heatmap* mostrando los resultados del agrupamiento jerárquico a un punto de corte de 6 subgrupos.



Anexo 3. Gráficos lineales en el que se representa la cobertura por los nodos más influyentes seleccionados por las distintas centralidades, a primero y segundo grado de vecinos. A) Gráfico para red de subgrupo 0, B) Gráfico para red de subgrupo 1, C) Gráfico para red de subgrupo 2, D) Gráfico para red de subgrupo 3, E) Gráfico para red de subgrupo 4.



Anexo 4. Tablas que contienen los genes más influyentes para cada red seleccionados por tener 5% mayor *Collective Influence (CI)*, ordenados de forma descendente por su calificación de CI. Se incluye para cada gen su calificación de Collective Influence y el proceso biológico al que pertenece. En los casos de los subgrupos tumorales se incluye la información de log-Fold-Change, expresión promedio y estadístico B.

Red Subgrupo 0

GENESYMBOL	CI	BIOPROCESS
SOX2	500688	Apoptosis
TFDP1	460000	Apoptosis
DAP3	454274	Apoptosis
PSMA3	449361	Apoptosis
BIRC2	446960	Apoptosis
TBX5	441168	Apoptosis
GAPDH	438949	Apoptosis
RHOT1	438438	Apoptosis
TARDBP	438048	Apoptosis
RASA1	437952	Apoptosis
RAPGEF2	437544	Apoptosis
DUSP6	437088	Apoptosis
RBM5	433566	Apoptosis
DNAJA1	432828	Apoptosis
FGA	432486	Apoptosis
RPS3	432130	Apoptosis
DNM2	429606	Apoptosis
DDX3X	428479	Apoptosis
BCL2L13	426610	Apoptosis
MRPS10	425340	Autophagy
IRS2	425060	Apoptosis
FGD1	424266	Apoptosis

Red Subgrupo 1

GENESYMBOL	LOGFC	AVEEXPR	B	CI	BIOPROCESS
IL13	0.768994	6.703037	102.7947	330675	Apoptosis
PNMA3	0.345131	7.230953	13.77379	327807	Apoptosis
FGFBP1	0.726246	7.323976	82.86645	322655	Autophagy
CCL5	0	8.167313	0	308611	Apoptosis
ADAMTS7	0.492074	6.637283	37.40814	300914	Autophagy
AGAP2	0.341859	7.258834	19.21781	300768	Apoptosis
LCK	0	7.741567	0	297264	Apoptosis
PSMB5	-1.21203	10.30259	123.5885	295020	Apoptosis
DCC	1.575055	5.737118	323.0634	292340	Apoptosis
STK17A	0	8.186288	0	290852	Apoptosis
GDF2	0.272532	6.982793	6.65878	286990	Apoptosis
PSMD10	-0.68984	9.565847	26.48389	286032	Apoptosis
SSTR3	-0.40914	7.696271	43.05493	282436	Apoptosis
PDCD1	0	7.393632	0	274560	Apoptosis
BMX	1.640537	6.22408	260.3279	263366	Apoptosis
NODAL	1.219388	7.17189	208.488	263012	Apoptosis
AVP	0	6.554695	0	261215	Apoptosis
BMP5	1.110495	6.74733	172.9156	255633	Apoptosis
TRIAP1	-0.75589	9.262917	38.43337	253470	Apoptosis

XDH	0.891547	7.974035	110.3978	249834	Apoptosis
SIRT2	-0.50032	9.537627	77.13844	249568	Both
CCR7	0.647978	6.727309	8.085991	249432	Apoptosis
IFI16	-1.1732	9.458529	52.05173	248976	Both
SLCO1A2	1.42891	6.808908	283.1461	243164	Autophagy
PSMC6	-1.50147	8.684348	112.1897	243012	Apoptosis
HCLS1	0	8.396957	0	242190	Apoptosis
TNFSF18	1.154954	5.75659	301.0682	241445	Apoptosis
PSMC2	-0.63949	10.541	41.41937	240149	Apoptosis
SKI	0	7.054376	0	237705	Apoptosis
TGFB1	-1.382	8.256121	80.41832	236222	Apoptosis
CALR	-1.45938	10.80532	212.2674	234969	Apoptosis
ACTN4	-2.15848	10.20061	264.3298	233730	Apoptosis
PPP2R5C	0	8.437704	0	233155	Apoptosis
TEK	0.993204	8.545391	196.2839	231035	Apoptosis
LYN	0	8.551838	0	229530	Apoptosis
CD3E	1.017372	8.256524	77.65564	226224	Apoptosis
ATG10	0.552095	6.387425	111.8921	225050	Autophagy
PRKG1	0.937046	6.849431	160.9827	224718	Autophagy
TARDBP	-0.75333	10.07395	48.83197	224004	Apoptosis
HAND2	0	6.080642	0	223482	Apoptosis
CNTF	1.843108	6.34175	367.903	223405	Apoptosis

Red Subgrupo 2

GeneSymbol	logFC	AveExpr	B	CI	BioProcess
FYN	-1.35773	8.974963	87.71247	315456	Apoptosis
HTT	0.651636	8.285522	65.48574	285384	Apoptosis
HIPK1	0	8.550563	0	280551	Apoptosis
LYN	0	8.551838	0	277060	Apoptosis
AXL	0	8.907115	0	271912	Apoptosis
PSMC6	-1.26241	8.684348	77.5413	268787	Apoptosis
NODAL	1.128172	7.17189	178.0318	268136	Apoptosis
DLC1	0	8.807128	0	262794	Apoptosis
SET	-1.45852	11.79744	123.2675	259182	Apoptosis
GAS1	0	9.30669	0	255170	Apoptosis
CUL3	-1.86307	9.509775	180.5571	254112	Apoptosis
WNT10B	0.854304	8.265612	140.3738	253786	Apoptosis
ACTN4	-2.39911	10.20061	302.2355	249014	Apoptosis
PPP2R5C	0	8.437704	0	243376	Apoptosis
DICER1	-1.13096	9.768186	42.7952	242760	Apoptosis
IL11	0.656473	7.788886	133.1723	238644	Apoptosis
AGAP2	0.409456	7.258834	29.51012	237188	Apoptosis
PRKCI	0	9.722836	0	236000	Apoptosis
LCK	0	7.741567	0	234060	Apoptosis
FAP	-1.32281	9.160412	43.01602	233442	Apoptosis
SIRT2	-0.51776	9.537627	79.94613	230657	Both
RBM5	-0.58872	9.934217	23.88832	229950	Apoptosis
CAMK2A	1.050037	7.603003	209.9352	228696	Apoptosis
NCKAP1L	0	8.031793	0	226780	Apoptosis
IFI16	-1.39904	9.458529	72.79189	226512	Both
CCR7	0.776673	6.727309	14.386	223328	Apoptosis
PNMA3	0.420406	7.230953	23.12413	222314	Apoptosis
BTG1	-2.0394	10.58276	155.2077	222123	Apoptosis
BMP5	1.05202	6.74733	152.8918	220350	Apoptosis

MFF	0	8.123531	0	218808	Apoptosis
PDCD1	0	7.393632	0	218221	Apoptosis
PSMD10	0	9.565847	0	216846	Apoptosis
CXCL12	0	9.963552	0	213240	Apoptosis
CD74	-3.42208	12.0683	366.8526	212085	Apoptosis
PSMA2	-1.44781	10.54132	147.9	210210	Apoptosis
FBXW7	0	7.586793	0	210018	Apoptosis
IL13	0.849307	6.703037	120.1388	206880	Apoptosis
CA7	0	7.33909	0	204700	Autophagy
HCK	0	7.770962	0	202748	Apoptosis
CAPN3	0	7.530838	0	202410	Apoptosis
TMBIM6	-1.95147	11.33027	284.7481	202272	Apoptosis
AVP	0	6.554695	0	201318	Apoptosis

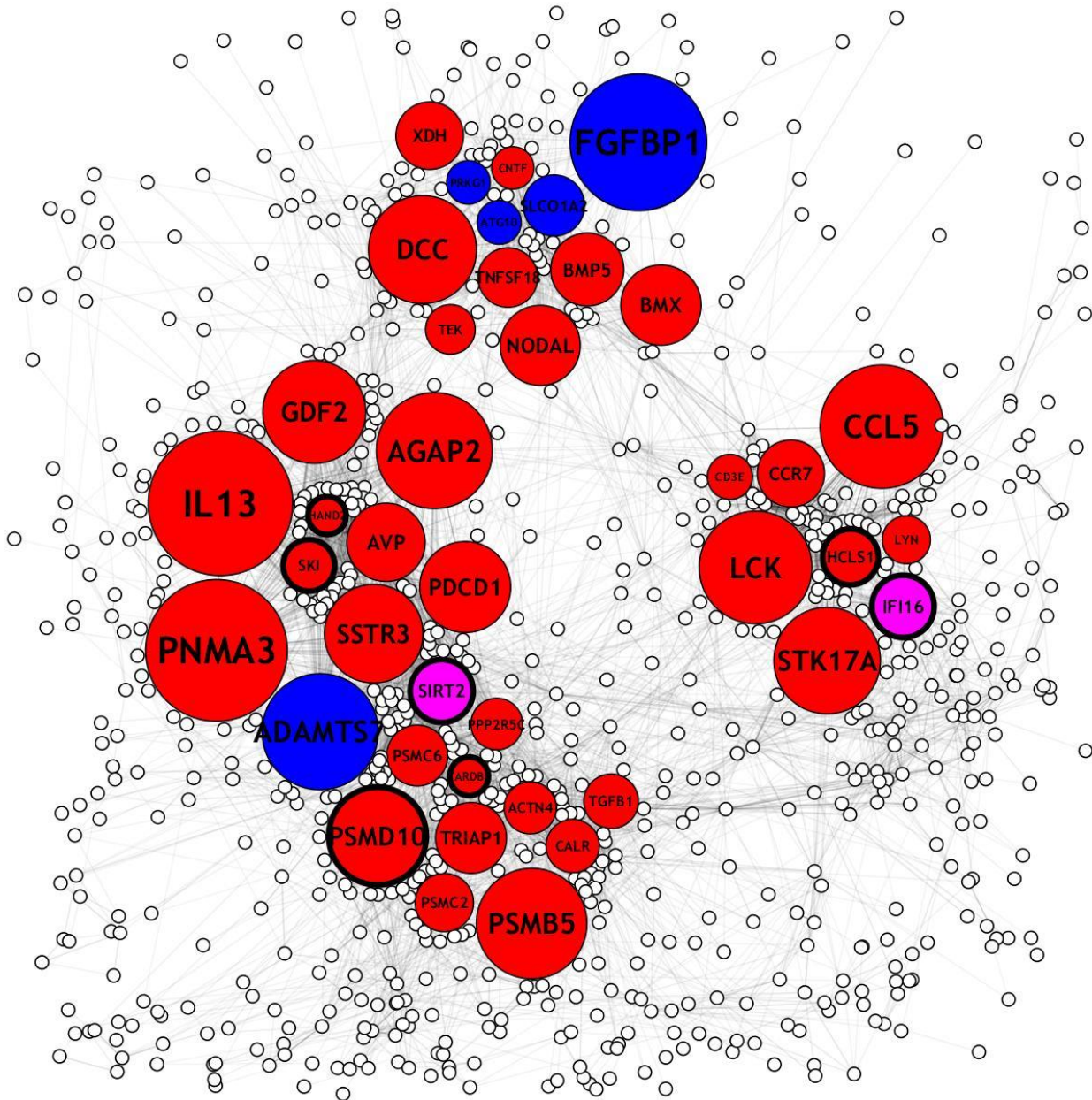
Red Subgrupo 3

GENESYMBOL	LOGFC	AVEEXPR	B	CI	BIOPROCESS
HCLS1	-0.74355	8.396957	17.12098	357048	Apoptosis
JAK3	0.263249	7.689568	16.58771	322136	Apoptosis
ATG10	0.718072	6.387425	168.466	320528	Autophagy
TEK	1.179179	8.545391	244.4767	312634	Apoptosis
LCK	0	7.741567	0	311212	Apoptosis
BMX	1.841147	6.22408	293.5608	287300	Apoptosis
GZMA	-0.68646	6.821563	7.168565	282658	Apoptosis
TNFAIP8	0	8.082893	0	267320	Apoptosis
CCR7	0	6.727309	0	260244	Apoptosis
PRKG1	1.162504	6.849431	217.8329	260064	Autophagy
IL2RB	0	8.696749	0	256310	Apoptosis
MNDA	0	7.111825	0	249343	Apoptosis
CD27	0	7.342939	0	245350	Apoptosis
CASP1	0	8.1365	0	231472	Apoptosis
CNTF	2.054101	6.34175	405.5337	230272	Apoptosis
GZMB	-0.91163	6.186539	25.74155	229835	Apoptosis
ACSL5	0.873675	7.253837	49.76206	227970	Apoptosis
AGAP2	0.309082	7.258834	12.77036	226750	Apoptosis
PSMB10	-2.36521	8.15068	226.5658	225302	Apoptosis
PAX8	0	9.3626	0	221169	Apoptosis
GPLD1	1.34602	6.280199	382.7169	220634	Apoptosis
SSTR3	-0.5291	7.696271	69.26066	219032	Apoptosis
SERPINB9	0.709367	7.86924	16.00401	213024	Apoptosis
TIAM2	1.147914	6.759245	207.82	198744	Apoptosis
PF4	0.591982	7.593534	94.60819	197291	Apoptosis
GDF2	0	6.982793	0	197234	Apoptosis
CYLD	-0.53641	8.293633	13.27638	196252	Apoptosis
SERPINB10	1.18051	6.715421	276.2264	194554	Autophagy
HCK	-0.57653	7.770962	7.735538	186172	Apoptosis
BIRC3	0	5.42525	0	184008	Apoptosis
DCC	1.687117	5.737118	338.4813	179975	Apoptosis
BIRC7	0	6.247062	0	178321	Apoptosis
HGF	0.23855	6.919414	7.512942	176954	Apoptosis
SLCO1A2	1.695183	6.808908	344.5373	176900	Autophagy
NODAL	1.381282	7.17189	240.4615	176792	Apoptosis

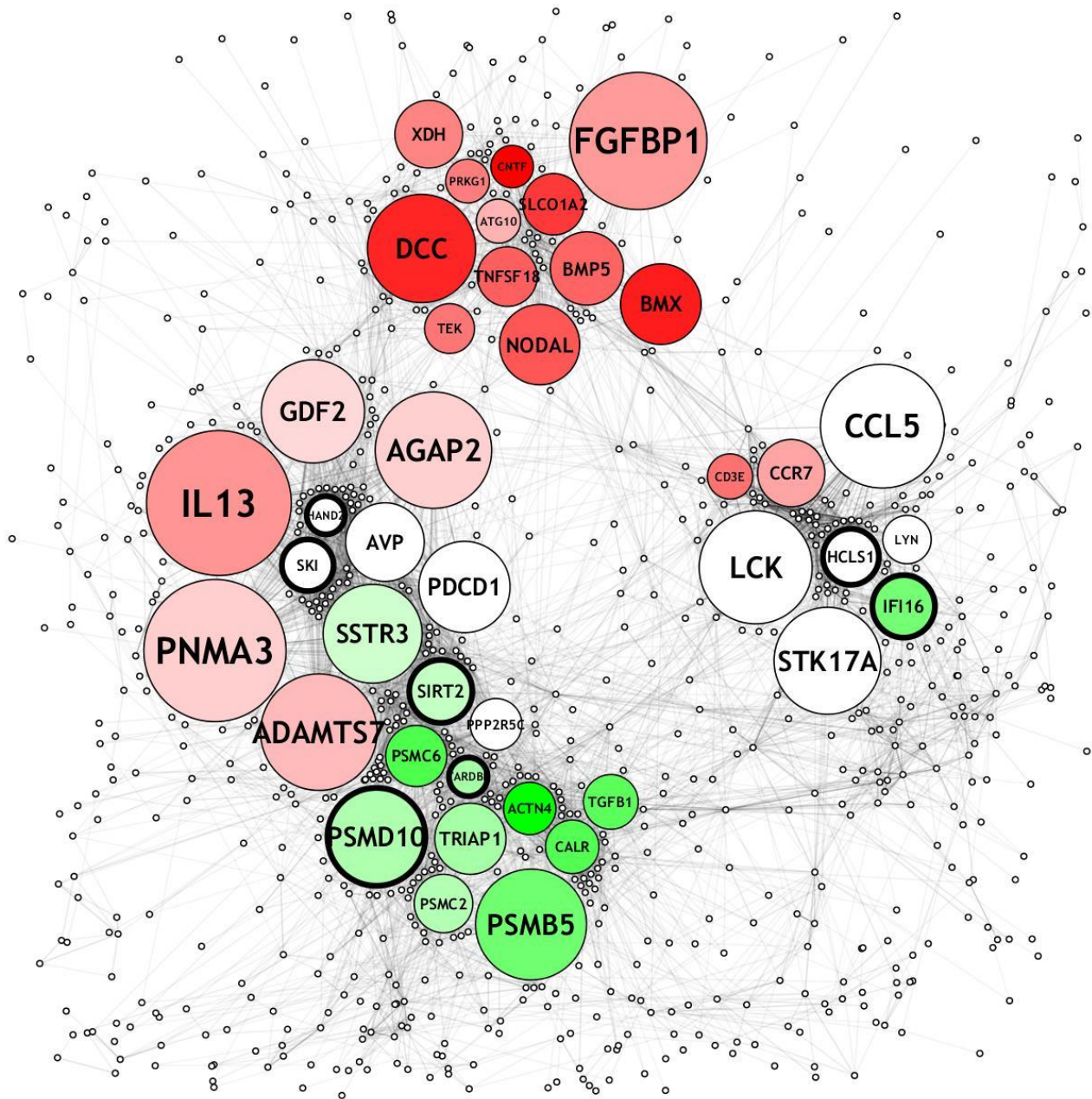
Red Subgrupo 4

GENESYMBOL	LOGFC	AVEEXPR	B	CI	BIOPROCESS
HCLS1	0	8.396957	0	326808	Apoptosis
LCK	0	7.741567	0	323616	Apoptosis
CCL5	0	8.167313	0	307440	Apoptosis
SIRT2	-0.50907	9.537627	66.10179	299341	Both
BCL2A1	0	5.46641	0	287492	Apoptosis
PSMB9	-1.44148	9.020792	41.44973	285675	Apoptosis
GZMA	0	6.821563	0	271600	Apoptosis
FIS1	-2.7054	9.545501	375.8681	256355	Both
NDUFS3	-1.65081	10.09662	296.6572	253456	Apoptosis
CAV1	0	9.35305	0	250039	Apoptosis
OPA1	1.559525	8.783411	170.908	243622	Apoptosis
NDUFA13	-3.61886	10.77512	526.233	243324	Apoptosis
PNMA3	0	7.230953	0	241906	Apoptosis
SRGN	-1.35364	9.523537	42.34573	229983	Apoptosis
MNDA	0	7.111825	0	228096	Apoptosis
BCLAF1	-0.85489	10.21915	38.95132	226734	Apoptosis
TNFAIP8	0	8.082893	0	226032	Apoptosis
VIM	-2.72054	12.21056	364.2105	223660	Apoptosis
CRYAB	0	8.371494	0	221724	Apoptosis
PHIP	0	9.514	0	221144	Apoptosis
LSM4	-1.97627	9.353827	225.6613	220662	Autophagy
AQP1	0	9.269862	0	220215	Apoptosis
NCKAP1L	0.495957	8.031793	12.07701	213869	Apoptosis
MYLK	-0.84953	9.832583	7.383976	210843	Autophagy
SLC1A3	0	7.842331	0	209280	Autophagy
GSN	-1.2018	10.84351	73.43173	208780	Apoptosis
LAMTOR5	-1.85799	10.19227	259.7635	208152	Apoptosis
PSMB5	-1.28032	10.30259	115.1396	207966	Apoptosis
GAS1	0	9.30669	0	207688	Apoptosis
HCK	0	7.770962	0	206832	Apoptosis
CASP1	0	8.1365	0	205535	Apoptosis
PSMA3	-2.80454	9.656437	372.022	203938	Apoptosis
CD27	0	7.342939	0	203825	Apoptosis
IL2RB	0.630864	8.696749	14.97414	203825	Apoptosis
PSMA5	-2.02413	9.756433	254.1502	201305	Apoptosis
LILRB1	0	8.340705	0	200928	Apoptosis
TCF7L2	0	9.743299	0	199872	Apoptosis
GDF2	0	6.982793	0	197287	Apoptosis
ACTN4	-1.72099	10.20061	156.1286	196716	Apoptosis
EPO	-0.43698	6.472571	31.78548	196633	Apoptosis
DCN	-2.07561	11.46201	138.7865	194320	Autophagy
CDK5	-1.1163	8.168353	124.0217	194060	Apoptosis
FOXO1	0.996433	8.149735	71.55866	193836	Both
SSTR3	-0.54202	7.696271	64.72564	192224	Apoptosis

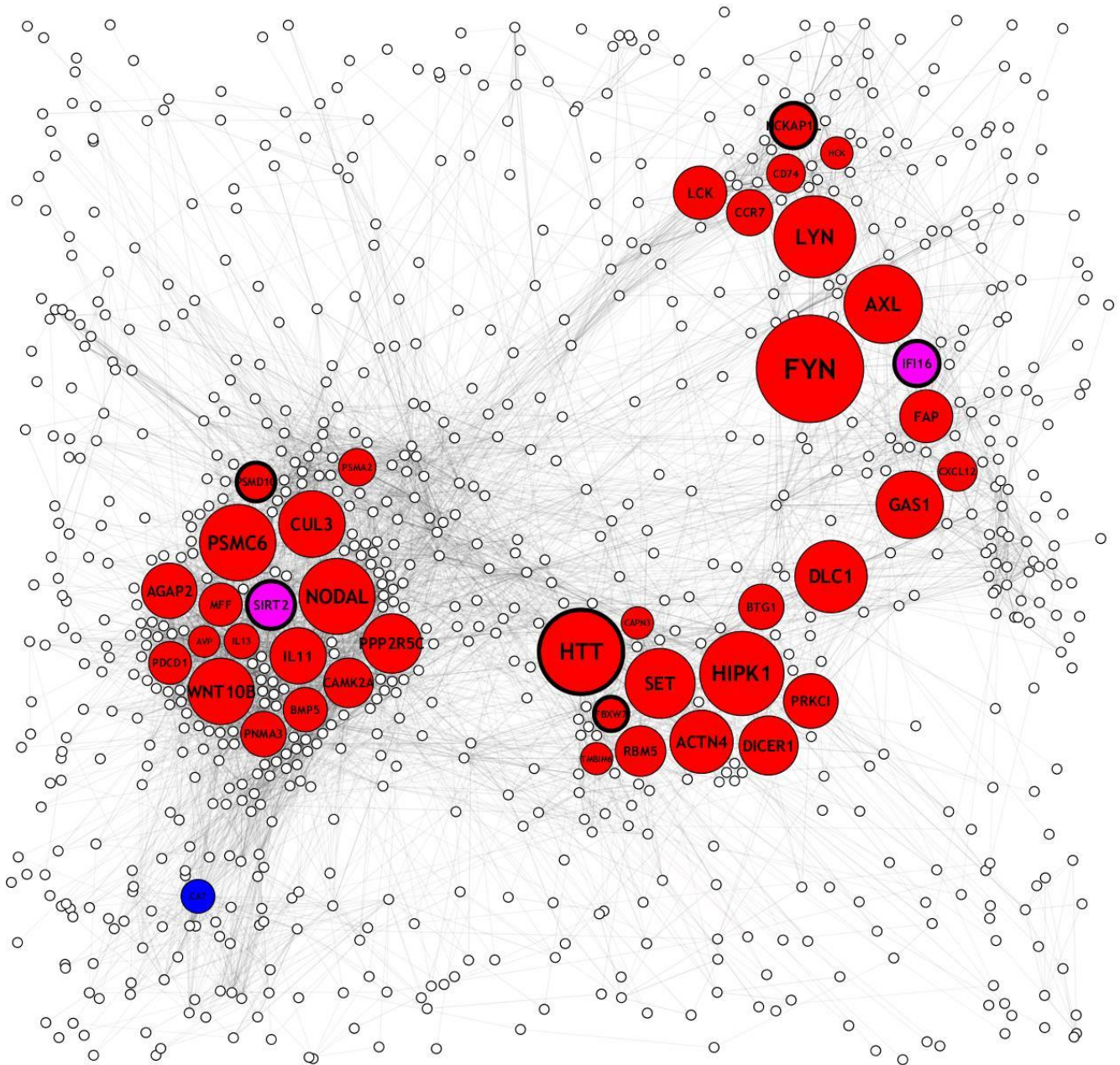
Anexo 5. Visualizaciones de las RRG en mayor tamaño.



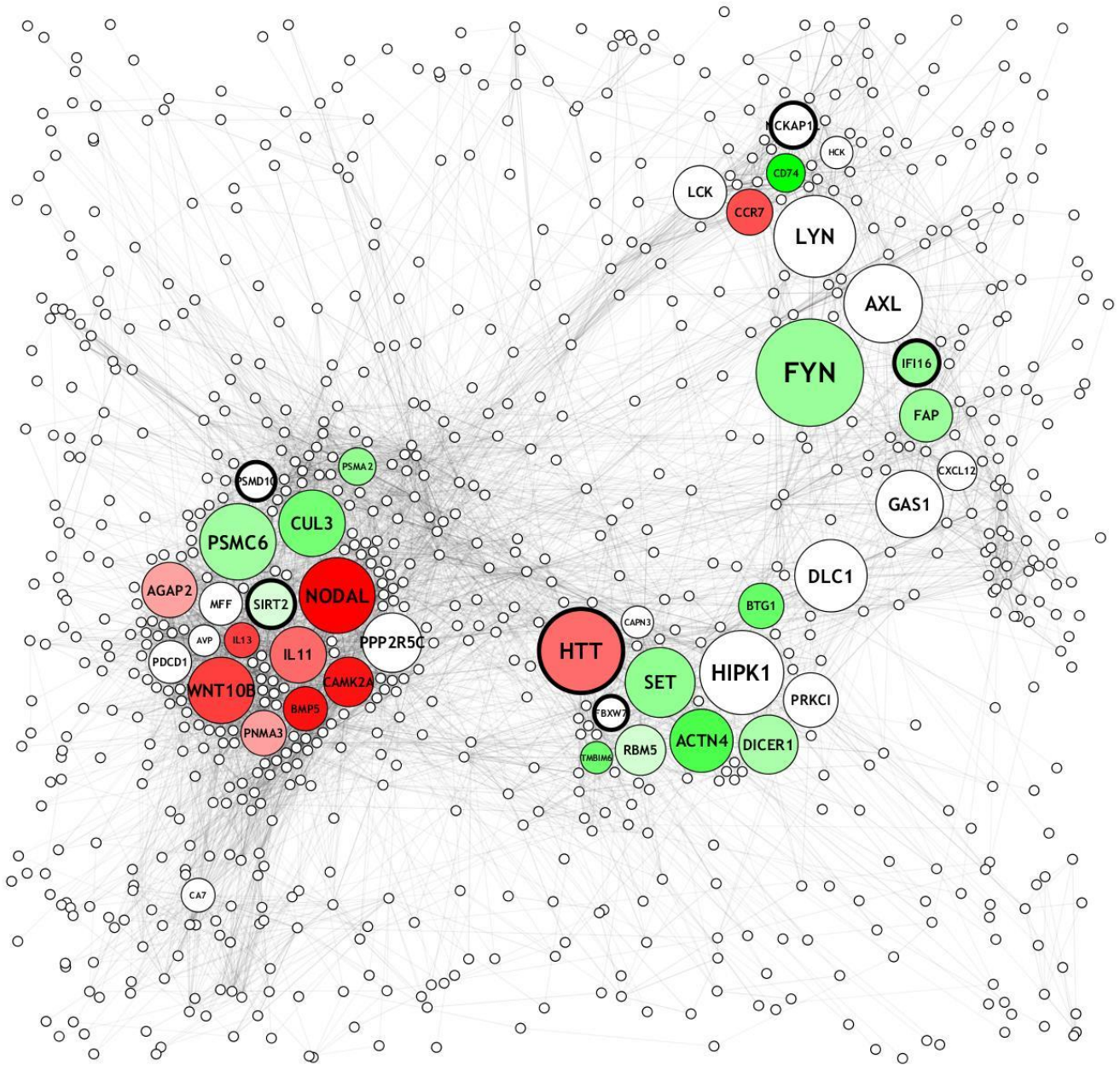
Red SG1-Basal - Procesos biológicos. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa el proceso al que pertenece: apoptosis en rojo, autofagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.



RRG SG1-Basal - Log-Fold Change. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa su nivel de expresión diferencial en comparación con el grupo control. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.

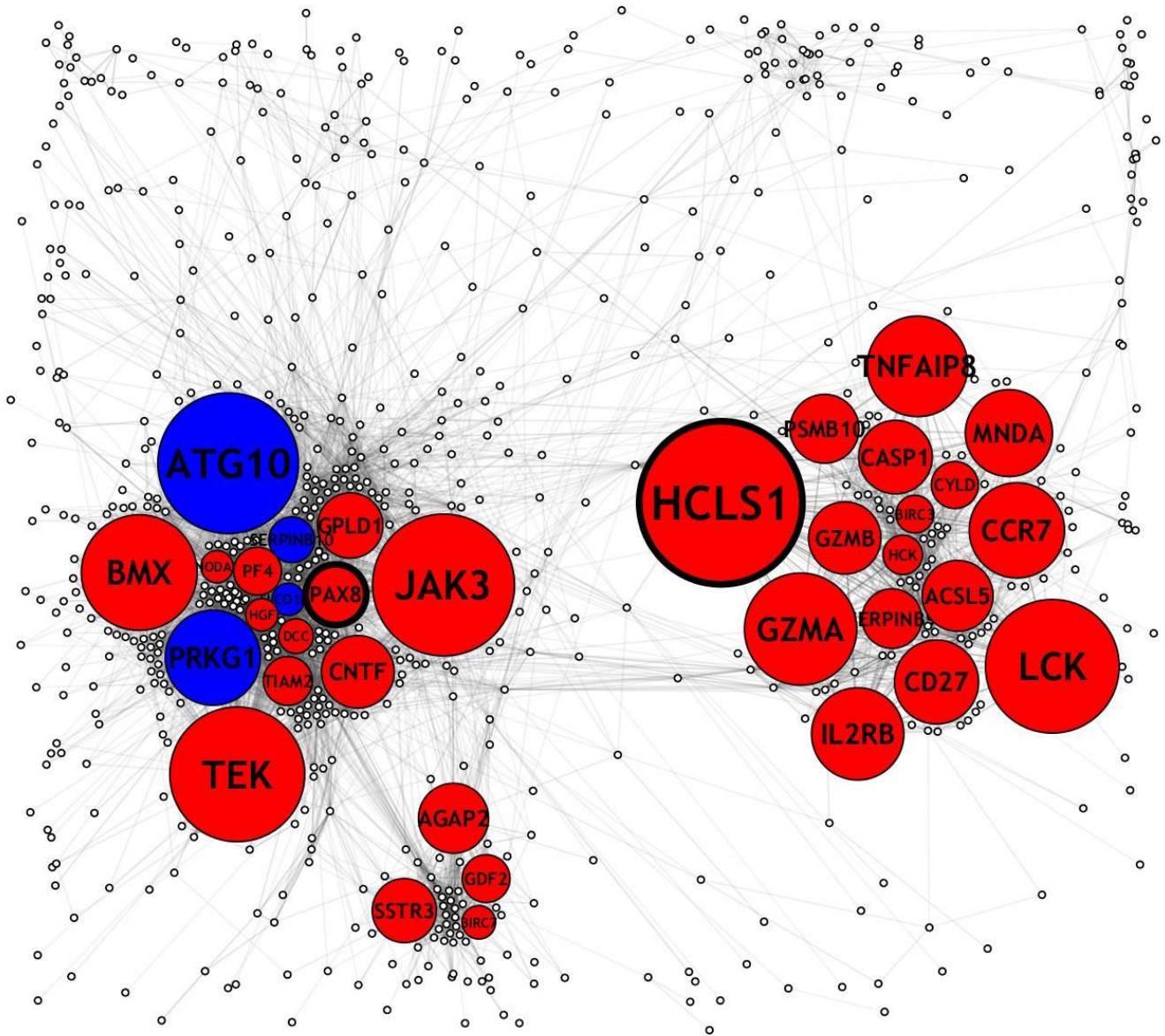


Red SG2 - Procesos biológicos. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa el proceso al que pertenece: apoptosis en rojo, autophagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.

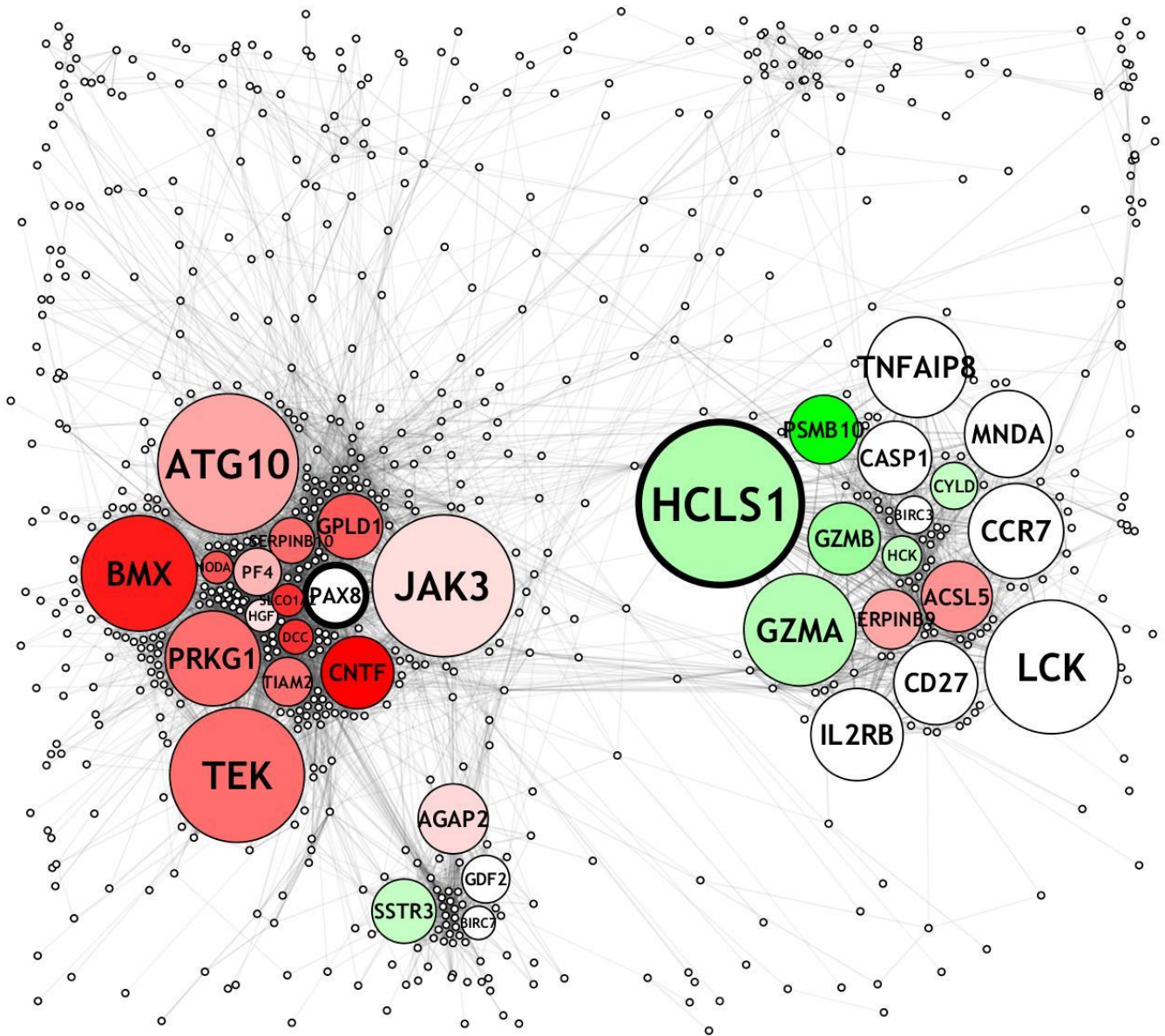


Red SG2- Log-Fold Change. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa su nivel de expresión diferencial en comparación con el grupo control.

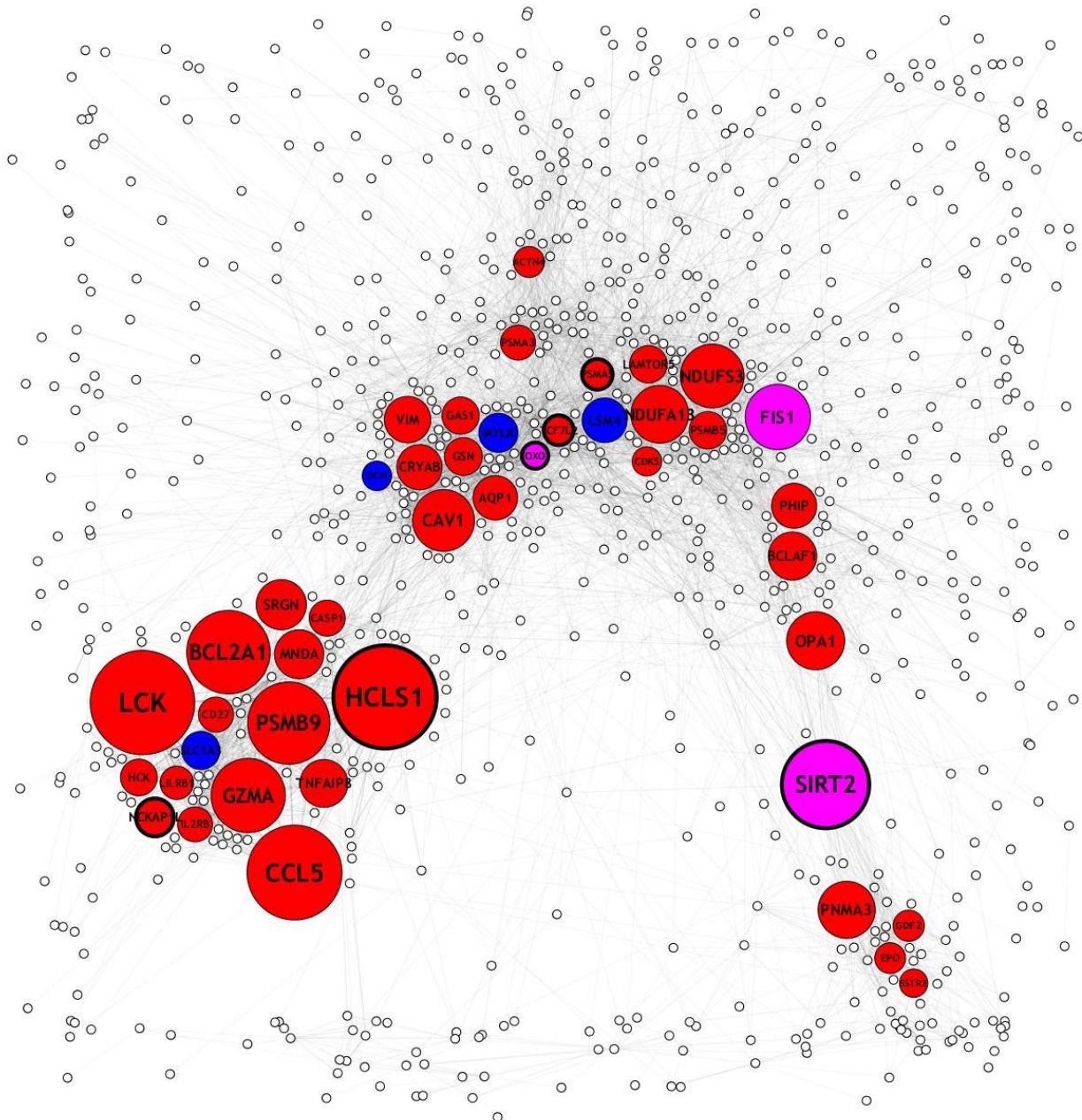
Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.



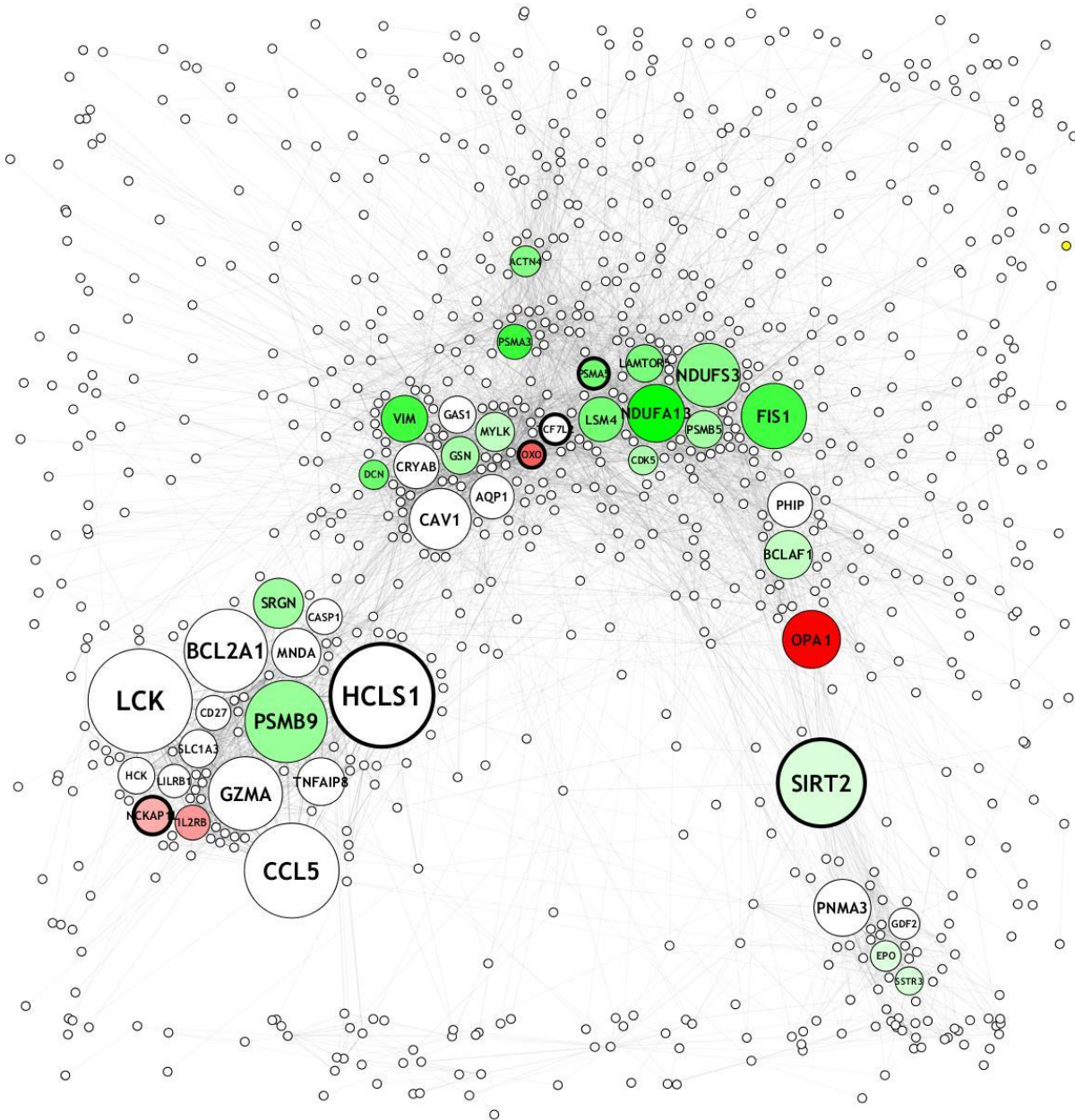
Red SG3-LumA - Procesos biológicos. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa el proceso al que pertenece: apoptosis en rojo, autofagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.



RRG SG3-LumA - Log-Fold Change. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa su nivel de expresión diferencial en comparación con el grupo control. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.



Red SG4-LumB - Procesos biológicos. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa el proceso al que pertenece: apoptosis en rojo, autofagia en azul, ambivalentes en rosa. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.



RRG SG4-LumB - Log-Fold Change. El tamaño de los nodos refleja su nivel de influencia sobre la red de acuerdo al CI. El color del nodo representa su nivel de expresión diferencial en comparación con el grupo control. Los genes que se conocen como factores de transcripción presentan un borde más ancho. La organización de los nodos refleja una mayor cantidad de relaciones entre genes agrupados.

Anexo 6. Agrupamiento jerárquico de genes de apoptosis y autofagia. Se tuvieron enriquecimientos significativos (valor de p corregido < 0.001) en el SG3, para el subtipo basal y en elSG2 para el Luminal A. No se aprecian diferencias evidentes entre subgrupos. El SG1 está conformado por muestras de tejido sano, sin embargo son agrupadas con las muestras tumorales.

