



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
INSTITUTO DE ECOLOGÍA

**IDENTIFICACIÓN COMPUTACIONAL DE BLANCOS
FARMACOLÓGICOS ASOCIADOS A VÍAS ESPECÍFICAS PARA EL
TRATAMIENTO DE SUBTIPOS MOLECULARES DE CÁNCER DE
MAMA.**

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

PRESENTA:
RAÚL ALEJANDRO MEJÍA PEDROZA

DIRECTOR DE TESIS
Dr. Enrique Hernández Lemus
Instituto de Ecología

COMITÉ TUTOR
Dr. Mauricio Rodríguez Dorantes
Facultad de Medicina

Dr. Carlos Guadalupe Pérez Plascencia,
Facultad de Estudios Superiores Iztacala

CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO, SEPTIEMBRE, 2019



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.

JURADO ASIGNADO:

Presidente: Dr. Pedro Eduardo Miramontes Vidal

Secretario: Dr. Enrique Hernández Lemus

Vocal: Dr. Marco Antonio José Valenzuela

Vocal: Dr. Jesús Espinal Enríquez

Vocal: Dra. Marcela Lizano Soberón

La tesis se realizó en: Instituto Nacional de Medicina Genómica (INMEGEN).

TUTOR DE TESIS:

Dr. Enrique Hernández Lemus
Instituto de Ecología



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

**IDENTIFICACIÓN COMPUTACIONAL DE
BLANCOS FARMACOLÓGICOS ASOCIADOS A
VÍAS ESPECÍFICAS PARA EL TRATAMIENTO DE
SUBTIPOS MOLECULARES DE CÁNCER DE
MAMA.**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

Raúl Alejandro Mejía Pedroza

TUTOR:

Dr. Enrique Hernández Lemus

Instituto de Ecología



Ciudad Universitaria, Ciudad de México, Septiembre, 2019

*A mi familia:
y a todas las personas que lo hicieron posible;*

Agradecimientos

Este documento, representa no sólo los resultados y lo aprendido en mi proyecto de doctorado, sino también la culminación de un esfuerzo y una etapa importante de mi vida. Por lo anterior y sin la intención de dejar fuera a nadie que haya sido parte de este ciclo, quiero agradecer profundamente a:

Mi mamá por su afecto y apoyo incondicional en cada etapa de mi vida, a mi papá que aunque ya no está con nosotros ha sido una guía para mí así como una fuente de inspiración. A mis hermanos y toda mi familia por su cariño y apoyo.

Aunque muchas personas han influido e inspirado mi vida académica y profesional, deseo mencionar a aquellos que en ciertas etapas fueron de trascendencia singular: A el Dr. Rogelio Cárdenas por haber sido quien me ayudó a despertar mi interés por las “ciencias duras” y la lógica. Dr. Antonio Mendoza Suárez, Dra. María Luisa Pérez Segui en la facultad de Física y Matemáticas de la UMNSH. Dr. Francisco Jaramillo González y Dr. Alejandro Rosas Cabral profesores de la facultad de medicina de la Universidad Autónoma de Aguascalientes. Mis compañeros de la misma facultad Jesús Antonio Viana Rojas, Juan Pablo Acero , Francisco Javier Pozos Varela y Jesús Mandujano por su excelente desempeño, quienes me motivaban a esforzarme y alcanzarlos. Emerson Armando Collazo Gúzman, Sergio Antonio Alcalá Corona, Rodrigo García Herrera de quienes aprendí mucho y he recibido apoyo sobremano. Agradezco también a todos los compañeros y profesores que han aportado para mi formación académica y humana, una disculpa por las omisiones.

Muy en especial, a **Enrique Hernández Lemus** (mi tutor), por todo su apoyo, por creer una y otra vez en mí. Quien estuvo siempre ahí para darme guía en mi proyecto, en mi carrera académica y en la vida. consejo concienzudo, presionarme cuando fue necesario y también por darme la libertad para desarrollar mis ideas. Sin duda, no sería la persona y el científico que soy ahora sin su tutoría. Además por su amistad, por las pláticas y lecciones de vida que me ha dado. En verdad estaré siempre muy agradecido.

A mi comité tutorial: **Carlos Guadalupe Pérez Plascencia** y **Mauricio Rodríguez Dorantes**, por sus enseñanzas, sus observaciones y sus palabras que me han hecho ser cada vez mejor persona y científico.

Resumen

El presente trabajo recoge y resume los principales alcances del proyecto de doctorado intitulado: “Identificación computacional de blancos farmacológicos asociados a vías específicas para el tratamiento de subtipos moleculares de cáncer de mama.”.

El proyecto propone una metodología computacional, sistemática y robusta cuyo objetivo es detectar las vías desreguladas en cada subtipo tumoral de cáncer de mama las cuales resultan particularmente útiles en términos del reposicionamiento de fármacos. Esos fármacos candidatos han sido inferidos a partir de datos experimentales. Particularmente, de datos transcriptómicos de biopsias de cáncer de mama obtenidos de bases de datos públicas.

Las principales ventajas de esta metodología se centran en el poder tomar datos de tecnologías de alto rendimiento directamente de pacientes. Esto logra que nuestras predicciones estén ancladas a la realidad. Así también se explotan los recursos de conocimiento biológico disponible que nos ayudan a darle sentido a estos datos, tales como los repositorios que almacenan las definiciones de las vías moleculares conocidas hasta el momento, así como las bases de datos farmacológicas que nos hablan de cuales dianas cuentan ya con un fármaco que actúa sobre ellas, si el fármaco esta acualmente disponible en el mercado, el tipo de interacción que guarda con su blanco, entre otros recursos valiosos al momento de ponderar la utilidad de un medicamento para una nueva enfermedad o siendo el caso, reposicionar fármacos para una patología conocida.

Otra ventaja es que por cuanto este trabajo esta circunscrito a un abordaje de reposicionamiento de fármacos, puede tener una aplicación directa a la calidad de vida de los pacientes con cáncer de mama. Así también, de tener éxito el proceso de desarrollo de un fármaco se vería beneficiado, debido a que no iniciarían desde cero sino que tomaría ventaja de los años de inversión y cuantiosos recursos que fueron destinados a el fármaco para sacarlo al mercado. Así pues podríamos estar hablando en un escenario ideal de candidatos farmacológicos a reposicionamiento especialmente interesantes.

Por otro lado, una de las limitaciones de este enfoque es que proporciona inferencias a partir de datos experimentales. Sin embargo la realidad biológica puede ser diferente

por lo cual hay que realizar experimentos para poder corroborar dichas predicciones. De igual manera, debido a que estas tecnologías tienen un vertiginoso avance seguramente en un futuro no muy lejano habrá acceso a mejores datos para llevar un análisis de este tipo. Así también las bases de datos de anotación están limitadas al conocimiento existente hasta el momento. Todos estos son factores considerados en el diseño de este proyecto.

Comenzaremos con una reseña del marco teórico respecto al cáncer de mama, su clasificación, el aporte de las tecnologías de alto rendimiento y los enfoques propuestos para manejarlos, así como el marco conceptual necesario. Asimismo, se mostrarán los recientes aportes de la hoy llamada *farmacogenómica* mediante reposicionamiento de fármacos a la medicina.

Así entonces, se presentará nuestra propuesta metodológica para la identificación de blancos farmacológicos potencialmente útiles para subtipos moleculares (intrínsecos) de cáncer de mama. Dicha propuesta consiste en implementar un abordaje centrado en blancos farmacológicos y su contexto. Ya sea la vía molecular a la que pertenece, su estado de expresión diferencial en el fenotipo o la interacción de esta diana con su fármaco. Al obtener dichas asociaciones entre blancos y subtipos moleculares realiza una priorización de los mismos de acuerdo a los datos disponibles.

Por último, se presentarán los resultados de aplicar esta metodología a los subtipos intrínsecos del cáncer de mama, así como vislumbrar su extensión a la medicina personalizada donde a partir de los datos biológicos particulares de un paciente se pueda establecer su tratamiento específico.

En general, el resultado más importante de nuestro estudio, es poder asociar blancos farmacológicos a propiedades y funciones bien definidas, en distintos niveles de organización biológica dentro del contexto de una enfermedad posiblemente muy diferente a la que fueron concebidos en un inicio.

Índice general

Índice de figuras	XI
Índice de tablas	XIII
1. Marco Teórico.	1
1.1. Cáncer de mama.	1
1.1.1. Clasificación y terapia farmacológica	2
1.2. Tecnologías genómicas.	3
1.2.1. Transcriptómica.	5
1.2.1.1. Microarreglos.	5
1.2.1.2. Secuenciación de RNA	5
1.2.2. Bioinformática y biología de sistemas.	8
1.3. Vías moleculares	9
1.3.1. Repositorios de vías moleculares.	10
1.3.2. Análisis de vías moléculares.	11
1.3.2.1. Pathifier	12
1.4. Farmacogenómica	12
1.4.1. Proceso de desarrollo de un fármaco	13
1.4.2. Reposicionamiento de fármacos	13
1.4.3. Trabajos previos en reposicionamiento de fármacos	14
2. Objetivo y planteamiento del proyecto de investigación.	17
2.0.1. Pregunta de investigación y objetivos.	17
3. Resultados.	19
3.1. Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. <i>Frontiers in pharmacology</i> 9 (2018)	19
4. Resultados adicionales	33
4.1. Extensiones a medicina personalizada	33
4.1.0.1. Expresión genética individual	33
4.1.0.2. Activación de vías por individuo	34
4.2. Lista de blancos subtipo basal	34

ÍNDICE GENERAL

4.3. Lista de blancos por subtipo	38
5. Métodos	41
5.1. Expresión Diferencial de genes	41
5.1.1. Implementación	41
5.2. Pathifier	42
5.3. PAM50	44
6. Discusión y conclusiones	47
7. Perspectivas	51
7.1. Integración de la base de datos con otros recursos	51
7.2. validación	52
7.2.1. validación biológica	52
7.2.2. validación sintética	52
Bibliografía	53

Índice de figuras

1.1. <i>Hallmarks of Cáncer: Los signos distintivos del cáncer</i> conjunto de fenómenos biológicos que exhiben lo compleja que es esta enfermedad.	1
1.2. La <i>Biología de Sistemas</i> estudia a los sistemas biológicos en diferentes escalas, a partir de las grandes cantidades de datos <i>omicos</i> disponibles hoy en día.	4
1.3. Microarreglo de expresión.	6
1.4. Esquema representativo que muestra el carácter integrativo, interdisciplinario y de retroalimentación de la <i>Biología de Sistemas</i>	8
1.5. a) Se obtienen las diferencias de expresión génica identificadas para enfermos (en azul) y para controles (en amarillo). b) Se ordenan y comparan los perfiles de expresión génica para ser comparados contra un compendio de firmas moleculares conocidas. c) Mediante los resultados se puede establecer una asociación entre la acción de los compuesto sy los efectos conocidos. Proyecto “Connectivity Map” Lamb J. et al, 2006.	15
4.1. En la imagen vemos los niveles de expresión para los blancos farmacológicos ejemplificados para los pacientes con subtipo basal, podemos	34
4.2. Aunque las distribuciones estan más diferenciadas para los valores de desregulación de vías en comparación a los niveles de expresión. Aún así hay pacientes atípicos cuyos valores individuales bien podrían pasar por controles. Especial diferenciación muestra la vía molecular de la apoptosis cuyos valores grupales son bastante diferentes entre los fenotipos.	35
5.1. En la imagen se observa un "gráfico de volcán" (<i>volcano-plot</i>). Ésta es una visualización típica para evaluar la expresión diferencial. En el eje de las abscisas tenemos el logaritmo de la razón de cambio (LFC) entre dos condiciones. En el eje de las ordenadas un parámetro que nos habla que tan significativo es ese cambio. Para ambas variables se suelen asignar puntos de corte de acuerdo a los objetivos del estudio. En rojo los genes sobre expresados de manera significativa y en azul los significativamente subexpresados.	43

5.2.	En el cuadro superior, la nube de puntos con su curva de ajuste donde cada eje corresponde a un elemento de la vía. El "Pathway Dereglulation Score" (PDS) es la distancia sobre la curva entre la proyección del punto y el origen. Pathifier repite dicho algoritmo para cada vía. En el cuadro inferior, los renglones corresponden a vías y las columnas a pacientes el color de cada celda esta relacionado con el PDS asignado a esa vía para ese paciente (muestra). El código de colores va en un rango de cero a uno, donde el cero corresponde con el origen de la curva (azul) y uno al punto más distante (rojo).	45
5.3.	En esta visualización se muestra mediante un análisis de componentes principales los subtipos intrínsecos de cáncer de mama definidos por el algoritmo PAM50	46

Índice de tablas

4.1. Lista de blancos del subtipo basal	36
4.2. Lista de blancos por subtipo	38

Marco Teórico.

1.1. Cáncer de mama.

El cáncer de mama se caracteriza por una proliferación descontrolada de células provenientes del tejido mamario. Presentando dicho fenómeno además uno o más de los llamados "signos distintivos del cáncer" (hallmarks of cancer en inglés) [1].

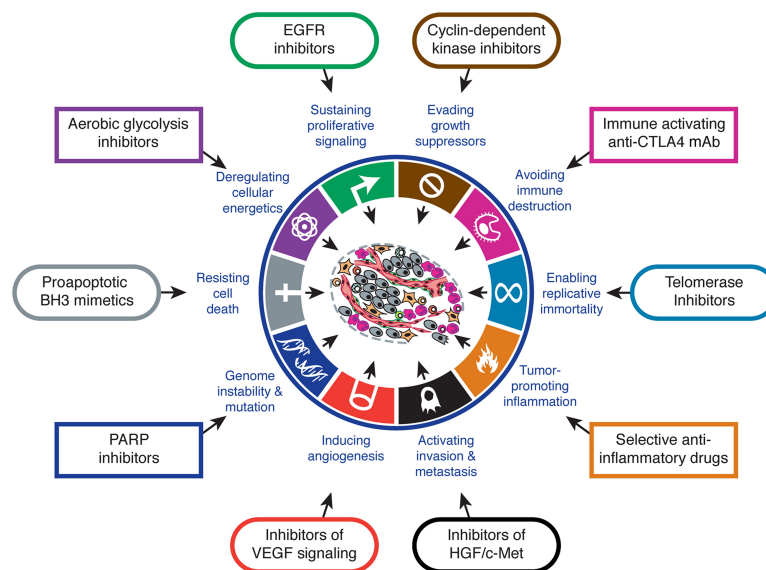


Figura 1.1: *Hallmarks of Cáncer*: Los *signos distintivos del cáncer* conjunto de fenómenos biológicos que exhiben lo compleja que es esta enfermedad.

El cáncer de mama es la neoplasia maligna con mayor incidencia y mortalidad entre las mujeres del mundo [2]. Uno de los principales desafíos para su tratamiento es su naturaleza heterogénea ya que esta conlleva a un amplio espectro de manifestaciones clínicas, fisiopatológicas, y de supervivencia dando lugar a diferentes opciones de esque-

mas terapéuticos [3].

Reconocer la importancia de comprender los elementos a nivel genómico y sus interrelaciones podría ser disruptivo en los paradigmas terapéuticos de las personas que padecen esta enfermedad, mejorándolos de una manera considerable. En estos términos, acceder a tecnologías de alto rendimiento disponibles para estudios genómicos puede ayudar en este empeño. Dado que aún existe un déficit de opciones terapéuticas efectivas y seguras especialmente para ciertos subtipos de cáncer de mama, un enfoque sistemático que evalúe de manera global en cada subtipo de cáncer de mama los elementos genéticos alterados y los procesos biológicos a los que estos pertenecen podría ofrecer luz guiando el proceso de identificar vías moleculares y/o opciones farmacológicas potencialmente útiles o particularmente interesantes para el tratamiento del cáncer de mama.

1.1.1. Clasificación y terapia farmacológica

Para manejar la amplia heterogeneidad que presenta esta enfermedad se han desarrollado diferentes métodos de clasificación. La subtipificación o clasificación molecular basada en patrones de expresión genética ha demostrado que al implementarlos proveen un diagnóstico más preciso así como un pronóstico más certero [4]. Motivos por los cuales no solo ha sido incluida en las guías de práctica clínica internacionales para esta enfermedad sino que es una pieza central en la toma de decisiones terapéuticas [5]. Entre estas clasificaciones basadas en los perfiles de expresión genética la que tiene más aceptación hasta el momento es aquella propuesta por Perou et al. [6] la cual a partir del perfil de expresión génica establece cuatro subtipos moleculares principales: *Luminal A*, *Luminal B*, *HER2+* y *basal*.

Luminal A. Alrededor de la mitad del total de los casos de cáncer de mama corresponden a tumores *luminales A* [7]. Estos tumores, regularmente son positivos para los receptores de estrógeno (*ER*) y progesterona (*PR*) mientras que usualmente son negativos para el *factor de crecimiento epidérmico 2* o *ERBB2* (*HER2-*), también presentan sobreexpresión de genes regulados por el receptor de estrógeno. Este subtipo detenta el mejor pronóstico [8], las menores tasas de recurrencia [9, 10] y generalmente hay buena respuesta al tratamiento hormonal aunque no la totalidad de dichos pacientes responden a ese tratamiento.

Luminal B. Este subtipo tiene un patrón de expresión y otras similitudes respecto a luminal A, se diferencia del anterior por una mayor variabilidad en la expresión de el *ER* y una mayor expresión de genes proliferativos, también se han encontrado en este subtipo, mutaciones asociadas con *TP53* e inestabilidad genómica. Alrededor del 20% del total de tumores de cáncer de mama se corresponden a este fenotipo [11]. Tiene un peor pronóstico que el subtipo *luminal A* [12]. Para el tratamiento de este subtipo

puede disponerse de terapia hormonal, en ocasiones esquemas de quimioterapia y/o inhibidores de *ERBB2* dependiendo de la expresión de este último.

HER2 enriquecido. Este subtipo intrínseco en su mayoría presenta una sobreexpresión del receptor *ERBB2* (HER2+), así como fenómenos de amplificación de dicho gen [13]. La terapéutica se cimienta en la inhibición de *ERBB2*, pues suelen ser negativos para los receptores de estrógeno y progesterona. De no recibir tratamiento específico tienen un pronóstico más precario que los de los subtipos luminales [14].

Basal. Los tumores basales tienen un perfil de expresión génica similar al *epitelio mamario basal*, correspondiente aproximadamente el 20% de los tumores de mama. Aproximadamente un 80 por ciento [15] de los tumores *triple negativo* (es decir aquellos que por inmunohistoquímica no expresan receptores de estrógeno, progesterona ni HER2) corresponden al subtipo basal. La subexpresión de estas moléculas suele presentarse también a nivel transcripcional. Los tumores basales están asociados con una mayor inestabilidad genómica, son más agresivos y presentan el peor pronóstico. La mayoría de los tumores mamarios relacionados a mutaciones de *BRCA-1* pertenecen a este subtipo [9, 12, 16, 17, 18]. Es importante resaltar que para el subtipo basal actualmente no se cuenta con una terapia dirigida (no hay terapia hormonal ni biológica disponible) [6] por ello la única opción hasta el momento es la quimioterapia la cual presenta mayores riesgos, complicaciones y múltiples efectos indeseables. En consecuencia son especialmente requeridas nuevas opciones farmacológicas menos agresivas y más eficaces para el tratamiento de este subtipo tumoral.

Es importante mencionar que no solo hay una carencia de *blancos* farmacológicos para terapia dirigida especialmente en ciertos subtipos de cáncer de mama, sino que la presencia de un receptor específico por ejemplo *ERBB2* no es garantía absoluta que habrá respuesta a su tratamiento modulador respectivo. Más aún un gran porcentaje de los pacientes con subtipos de buen pronóstico los cuales ya cuentan con alguna terapia específica eventualmente desarrollan *resistencia* al tratamiento. Por ejemplo hay reportes de pacientes que reciben terapia adyuvante con tamoxifen donde hasta un 40 por ciento desarrollan resistencia al tratamiento. [19].

1.2. Tecnologías genómicas.

La secuencia unidimensional de nucleótidos en su forma de *Ácido desoxirribonucleico* (ADN o DNA en inglés) también conocida como *genoma* es una pieza fundamental para entender la composición de un individuo y poder hacer comparaciones entre ellos. En 2001, el Consorcio Internacional para la Secuenciación del Genoma Humano publicó la primer versión del genoma humano [20]; en 2004, hizo público la segunda fase del mismo

1. MARCO TEÓRICO.

[21]. A partir de la conclusión del proyecto del genoma humano ha devenido una panoplia de tecnologías que nos hacen accesible la medición masiva y detallada de moléculas de DNA así como moléculas relacionadas. Esto incluye secuencias de RNA mensajero, sitios de metilación, dominios de proteínas, entre otras. A estas técnicas también se les llama usualmente *tecnologías de alto rendimiento*.

Algunas características genómicas (en el sentido amplio de la palabra) han sido relacionadas a rasgos fenotípicos diferentes entre los individuos tales como color de ojos, estatura, capacidades metabólicas, enfermedades complejas como el cáncer, la diabetes, entre otras. La genómica propiamente dicha refiere al estudio del *Ácido desoxirribonucleico* (ADN o DNA en inglés) aunque no es infrecuente el uso del término para referirse a este tipo de tecnologías en su conjunto. La proteómica pretende estudiar la totalidad de proteínas que constituyen un ente biológico. La transcriptómica examina los transcritos presentes en la célula, la metabolómica el cúmulo de metabolitos, entre otras. Si bien los fenómenos biológicos pueden estudiarse desde la perspectiva de múltiples tecnologías genómicas o simplemente "ómicas", cada una de éstas nos ofrece información diferente.

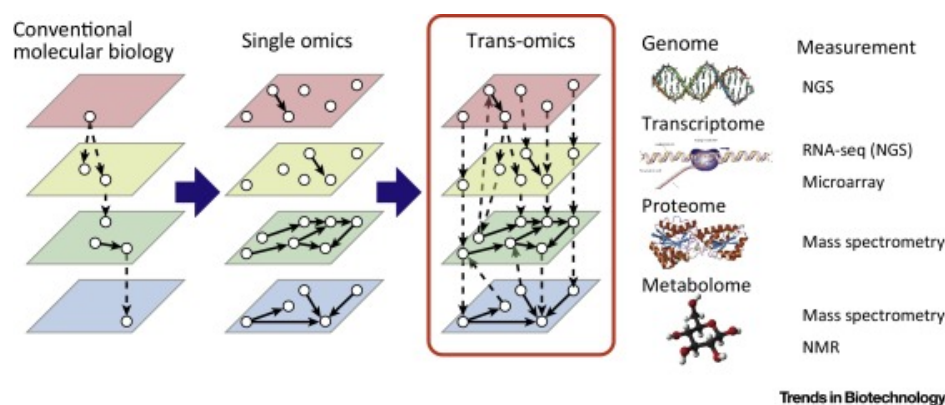


Figura 1.2: La *Biología de Sistemas* estudia a los sistemas biológicos en diferentes escalas, a partir de las grandes cantidades de datos *ómicos* disponibles hoy en día.

Nuestro marco teórico actual en la investigación *ómica* se desarrollado en torno a la aceptación que el análisis de datos biológicos provenientes de tecnologías de alto rendimiento nos ayudan a entender los mecanismos subyacentes que determinan los procesos de salud y enfermedad, especialmente aquellas de etiología compleja. Constituyendo uno de los más interesantes desafíos de la medicina hoy en día es dilucidar estas relaciones. Consecuentemente este conocimiento podría en un futuro dar frutos en forma de nuevas opciones de diagnósticas y terapéuticas para una multitud de enfermedades que el día de hoy afligen al género humano.

Por otra parte la disposición de una cada vez creciente cantidad de información biológica ha incrementado la necesidad de recursos computacionales así como de modelación matemática apropiada para manejar y darle inteligibilidad a este enorme cúmulo de datos. Se habla [22] de cifras astronómicas al tomar por poner un ejemplo los aproximadamente 25000 genes en el genoma humano, y modelando sus estados sólo como *encendido* o *apagado*, eso daría como resultado 5.6×10^{7525} de posibles perfiles de expresión genética.

1.2.1. Transcriptómica.

Entre las técnicas mencionadas la transcriptómica es de particular interés debido al dinamismo del proceso de transcripción genética. Otro punto a favor es que es de las tecnologías más intensamente estudiadas y con mayor volumen de datos disponibles. Dentro de las tecnologías de alto rendimiento para la aplicadas a la transcriptómica están los microarreglos y la secuenciación de RNA.

1.2.1.1. Microarreglos.

Un microarreglo de expresión de (o chip de ADN) es una superficie sólida (vidrio o plástico), a la cual se une un repertorio de secuencias de ADN o ARN (ADN complementario) de longitud definida (figura 1.3). De essta manera puede cuantificarse de manera simultánea los niveles expresión de miles de genes. **El nivel de expresión de un gen es una cantidad relacionada al número de ARN mensajero presente.**

Así el funcionamiento de los microarreglos consiste, básicamente, en medir la hibridación entre una secuencia específica de ADN de longitud determinada (*sonda* o *probe*), y una molécula problema (*target*). Estas últimas son secuencias de ARN mensajero extraídas del tejido en estudio. Además, las *sondas* tienen asociada una molécula fluorescente las cuales se hacen patentes después del proceso de lavado, ya que solo las *sondas* que lograron hibridar permanecerán. Así, es posible cuantificar el nivel de expresión de un gen mediante fluorescencia, a través de un análisis de imagen, que refleja el nivel de expresión del gen.

1.2.1.2. Secuenciación de RNA

La secuenciación de RNA se ha establecido en concierto con el rápido desarrollo de las tecnologías de alto rendimiento para la secuenciación de DNA. Antes de que los transcritos de RNA sean secuenciados, varios pasos son necesarios. La sensibilidad de un experimento de RNA-seq puede aumentarse enriqueciendo las clases de RNA de interés y depletando otros RNAs abundantes (por ejemplo RNA ribosomales). Posteriormente

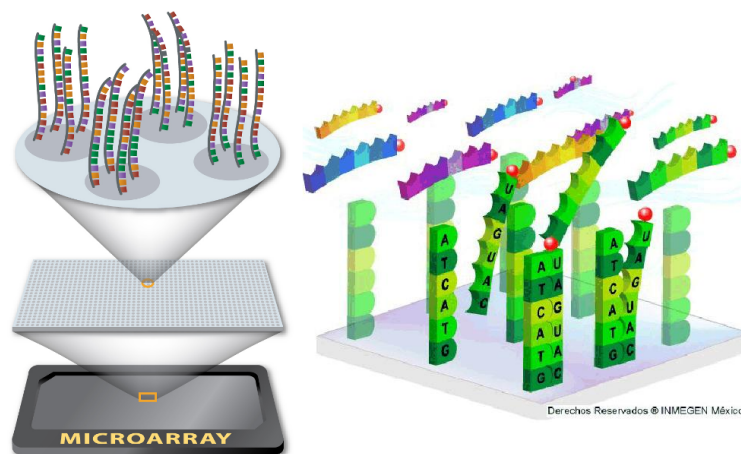


Figura 1.3: Microarreglo de expresión.

usualmente los transcritos son fragmentados, proceso que clave en la construcción de librerías [23]. Durante la preparación para la secuenciación, copias de DNA complementario a los transcritos pueden ser amplificadas mediante la reacción en cadena de la polimerasa (*PCR*) [24]. La amplificación también se usa para permitir la secuenciación de muy pequeñas cantidades de RNA. Los identificadores moleculares únicos (*UMI*) son secuencias cortas aleatorias que permiten marcar fragmentos de secuenciación durante la preparación de librerías lo cual da la oportunidad de corregir eventuales sesgos subsecuentes.

Una vez que los transcritos han sido preparados, pueden ser secuenciados en una dirección (*single-end*) o en ambas direcciones (*paired-end*). De ahí en adelante la secuenciación de RNA se sustenta en convertir las moléculas en DNA complementario (*cDNA*). De ahí que subsecuentemente las plataformas de secuenciación genómicas y transcriptómicas sean las mismas.

En cuanto a las tecnologías de secuenciación se basan en diferentes principios físico-químicos. Las técnicas empleadas con mayor frecuencia generan lecturas cortas (aproximadamente 100 pares de bases) con una tasa de error entre el 0.5% y 0.01%. Las tecnologías que generan lecturas cortas, se clasifican en tres tipos: El primero, secuenciación por ligación, se basa en la unión de una sonda con una o dos bases conocidas y el resto compuesto por bases degeneradas y el templado de DNA; al darse la ligación se libera un fluoróforo lo que permite la identificación de la o las bases sucesivas del fragmento de DNA que se está secuenciando. SOLiD y Complete Genomics emplean esta técnica. Segundo, secuenciación por síntesis en la categoría de terminación reversible por ciclo, a saber, en cada ciclo de secuenciación son agregados los cuatro nucleótidos marcados con un fluoróforo distinto cada uno y bloqueados en el extremo 3', la base

complementaria a el siguiente nucleótido en la secuencia blanco es añadida por la polimerasa a la secuencia naciente y el resto es lavado, en cada ciclo se toma una foto utilizando cuatro longitudes de onda específicas para identificar cuál nucleótido ha sido agregado; para ambas tecnologías, como un paso previo a la secuenciación se generan aglomerados de secuencias idénticas, donde cada aglomerado puede añadir una base distinta en cada ciclo, permitiendo que sean bastante escalables [25]. Este método es utilizado por Illumina y Quiagen. Tercero, secuenciación por síntesis en la categoría de adición de un sólo nucleótido, utiliza nucleótidos que no están bloqueados en el extremo 3'; en cada ciclo de secuenciación se agrega sólo un tipo de nucleótido y de acuerdo a la técnica empleada la adición de dicha base se registra como presencia de luz o como un cambio en el pH; una desventaja de esta técnica es que varios nucleótidos del mismo tipo podrían añadirse en un ciclo, tomando en consideración una restringida capacidad de detección, la longitud de algunos homopolímeros podría ser establecida de manera incorrecta [26]. Como ejemplos 454 e Ion Torrent.

Recientemente se ha logrado incrementar el tamaño de las lecturas pero obteniendo una tasas de errores hasta de un 12 %. Por otro lado, se pueden agrupar en dos tipos las tecnologías que generan lecturas largas: La primera, secuenciación en tiempo real de una sola molécula, las cuales pueden producir lecturas de hasta 20000 nucleótidos. Emplean este principio PacBio y Oxford Nanopore Technology (ONT). PacBio emplea este principio fijando en el asiento de un pozo de fondo transparente la polimerasa, un lector óptico registra las irradiaciones de fluorescencia cuando diferentes nucleótidos son añadidos a la molécula de DNA que se está sintetizando. En el caso de ONT, se secuencia el DNA basándose en la alteración de voltaje características de cada nucleótido mientras pasa por un poro proteico. Segunda, secuenciación por enfoques sintéticos; de hasta 200 mil nucleótidos pueden ser las lecturas. Inicialmente se fragmenta el DNA y el mismo código (barcode, en inglés) es agregado, experimentalmente, a moléculas de DNA que corresponden a regiones genómicas contiguas; a continuación, las moléculas de DNA son secuenciadas con tecnologías que generan lecturas cortas, incluyendo en la muestra moléculas de DNA con códigos distintos; por último, las lecturas con el mismo código son ensambladas originando lecturas sintéticas largas. la plataforma de Illumina para la secuenciación de lecturas largas sintéticas, y el sistema basado en emulsión 10X genomics son las tecnologías que usan este sistema [27].

1.2.2. Bioinformática y biología de sistemas.

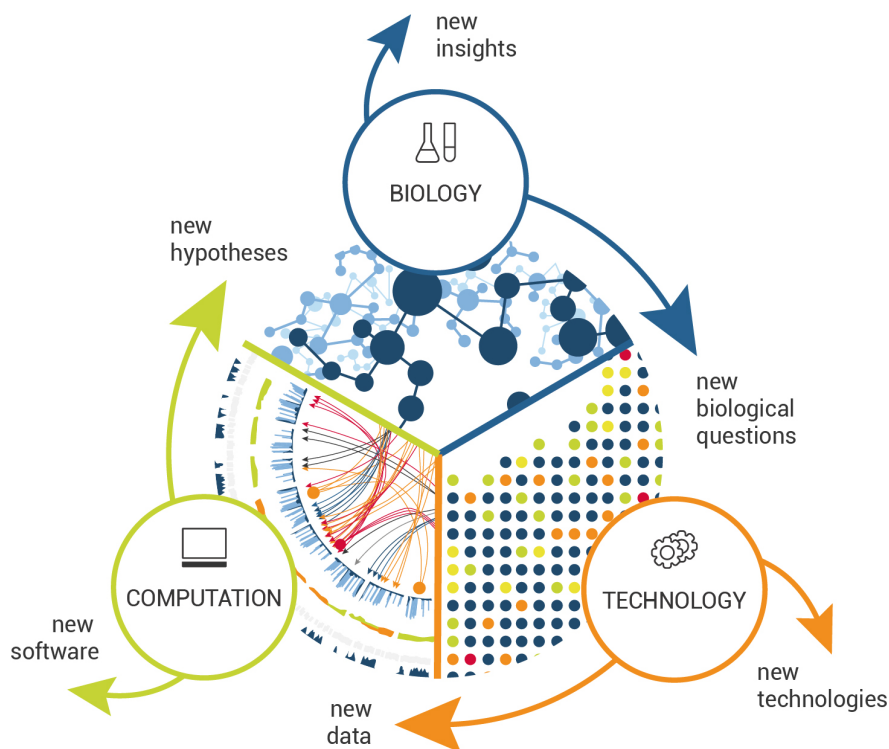


Figura 1.4: Esquema representativo que muestra el carácter integrativo, interdisciplinario y de retroalimentación de la *Biología de Sistemas*.

La aparición de tecnologías de alto rendimiento y la explosión en el volumen de datos han motivado el desarrollo de infraestructura para el almacenamiento y procesamiento de la información, datos derivados del estudio del genoma y compuestos relacionados por ejemplo Metaboloma, Metiloma, entre otros. De ahí surge la bioinformática para la gestión y análisis de datos biológicos mediante métodos computacionales.

Estas condiciones abren la posibilidad a cambiar de enfoque. En lugar de enfocarse en los genes particulares y sus efectos, podemos plantear atacar una pregunta simple, pero que requiere un enorme volumen de información para su respuesta: ¿cómo trabajan los genes dentro de una célula que son capaces de gobernar todas las reacciones que ocurren en ella? De este modo, a partir de la vasta cantidad de datos y la avance en la modelación matemática de sistemas complejos, recientemente ha surgido enfoque conocido como *Biología de Sistemas*, el cual tiene como objetivo responder preguntas como la anterior a partir técnicas matemáticas y computacionales.

Así, la biología de sistemas puede pensarse como el modelado computacional y matemático de sistemas biológicos complejos. Un campo de estudio interdisciplinario basado en las interacciones entre los componentes de los sistemas biológicos, y como éstas dan lugar al comportamiento y función de el sistema. Emplea un enfoque holístico (en lugar del tradicional reduccionismo) para entender la complejidad de los sistemas biológicos partiendo de la idea de que los organismos vivos son más que la suma de sus partes. Asimismo, es multidisciplinaria integrando desde la biología, ingeniería, física, informática, bioinformática entre otras. Uno de las metas de la biología de sistemas es modelar y descubrir propiedades emergentes, en los organismos, tejidos y células (por ejemplo, las enzimas y metabolitos en una vía metabólica o la frecuencia cardíaca), así como estudiar el comportamiento y organización de los procesos biológicos a través de sus componentes moleculares.

De esta manera la capacidad de integrar y analizar conjuntos de datos heterogéneos de múltiples experimentos empleando herramientas interdisciplinarias, es parte fundamental en la *Biología de Sistemas*; algunos de sus fuentes de información relevantes son: la epigenómica¹, genómica, transcriptómica, proteómica², entre otras (figura 1.4).

No obstante, no sólo los datos edifican la *Biología de Sistemas*, la construcción de un marco teórico para poder analizar y generar modelos a partir de los datos es un elemento central. Esto conlleva a desarrollo de modelos mecanísticos, como los de los sistemas dinámicos a partir de las propiedades cuantitativas de bloques de construcción básicos. Por ejemplo, una red celular se puede modelar matemáticamente empleando técnicas que provienen de la cinética química y la teoría de control; usando conjuntos de ecuaciones diferenciales. Empero, debido a la gran cantidad de parámetros, variables y restricciones en las redes celulares, a menudo se utilizan métodos numéricos y computacionales para dichos modelos.

1.3. Vías moleculares

De manera general, rara vez en la naturaleza un proceso biológico entero puede ser llevado a cabo por una sola molécula o gen [28]. Desde la década de los 50's [29] con los primeros mapas genéticos la idea de que un conjunto de genes llevan a cabo una función biológica específica ha estado rondando. Por ejemplo, al observar que las funciones biosintéticas en *Neurospora* y la asociación de éstas con los genes no era al azar [30]. O

¹factores reguladores transcriptómicos no codificados en la secuencia genómica como metilación de ADN, acetilación de histonas y desacetilación, etc.

²mediciones de proteínas y péptidos de un ente biológico, técnicas como la electroforesis en gel bidimensional, espectrometría de masas o técnicas de identificación de proteínas multidimensionales (sistemas avanzados de HPLC acoplados con espectrometría de masas) se usan para esta tarea

en los genes en las etapas de desarrollo temprano en *Drosophila* [31].

En este sentido varios modelos conceptuales se han propuesto para llegar a entender lo complicado de éstas interacciones. De los primeros en establecerse, es el reconocimiento de que puede darse un cambio fenotípico debido a una sola alteración génica. Tal es el caso de la anemia de células falciformes una enfermedad hematológica crónica hereditaria donde una mutación en el gen que codifica para la Beta hemoglobina [32] es la causante. Aunque este abordaje ha dado frutos y es parte de nuestros preceptos biológicos, no es el idóneo para el análisis de datos masivos de tecnologías de alto rendimiento.

Se ha propuesto el reconocimiento de módulos funcionales como un nivel crítico de organización biológica. Donde un módulo es una entidad cuya función emerge a partir de la interacción entre sus elementos y es separable de otros módulos [28]. A través de estas unidades separables de estudio propuestas por Hartwell, de aquí en adelante llamadas vías moleculares (o pathways) están en realidad interconectadas en una gran red conformada por todas las posibles interacciones entre los componentes biológicos. Cada elemento de una vía molecular puede pertenecer a diferentes vías moleculares al mismo tiempo, y su función individual varía de acuerdo al contexto celular, por ejemplo genes como Ras, Myc y NF-kB pueden tanto estimular el crecimiento y supervivencia celular o inducir apoptosis [33].

Las diferentes definiciones de "vía moleculares" pueden ser divididas en líneas generales en dos tipos diferentes: Como conjuntos de genes y aquellas que consideran la topología de las vías moleculares. Los conjuntos de genes (o Gene-sets) son listas de componentes biológicos a las cuales se les atribuye una función específica. En contraste las vías moleculares topológicas no solo enlistan los elementos si no que también describen las interacciones entre ellos, describiendo quién está interactuando con quién y que tipo de interacción es, de ahí se obtiene una descripción de como los componentes moleculares se desempeñan dentro de un proceso biológico.

1.3.1. Repositorios de vías moleculares.

Los esfuerzos para reunir y estructurar el conocimiento en las vías moléculares han dado como fruto la generación de **Bases de datos** de vías moleculares o (PDBs por sus siglas en inglés Pathway Databases), las cuales acopian el conocimiento de las interacciones biológicas en repertorios de vías moleculares. Estas bases de datos usualmente extraen la información de diferentes fuentes. Habitualmente, evidencia experimental proveniente de la literatura, la cual se somete a un proceso de revisión por expertos en el área, así como la predicción computacional de funciones biológicas a partir de biomoléculas similares. Además, el cruce de información entre bases de datos similares es una tarea usual.

La oferta de bases de datos es amplia donde cada una puede tener un enfoque diferente desde aquellas que se concentran en alguna especie, proceso de curación, tipo

de vías e interacciones cubiertas, entre otros. Por ejemplo las anotaciones dentro de la base de datos de Reactome [34] son manualmente curadas a partir de la literatura por expertos en colaboración con su equipo editorial y luego contrastada tanto con bases de datos relacionadas así como literatura primaria. RegulonDB [35] Se especializa en integrar el conocimiento en torno a la regulación genética bacteriana del organismo *mejor caracterizado* Escherichia coli K-12. Las correcciones y adiciones a las bases de datos de vías moleculares deben hacerse de manera frecuente para incrementar así su calidad y cobertura. Algunas son capaces de actualizar su información de manera periódica para tratar de mantenerse al ritmo de los nuevos descubrimientos. Por ejemplo, para el caso del humano la base de datos de Gene Ontology [36] se actualiza cada mes, mientras que KEGG [37] hace actualizaciones de manera semanal más aún esta última nos ofrece información topológica detallada de las vías moleculares.

1.3.2. Análisis de vías moléculares.

El advenimiento de cantidades masivas de datos a partir de las tecnológicas de alto rendimiento ha permitido medir de manera simultánea miles de componentes biológicos. Para extraer información de ellos nos valemos entre otras cosas del poder computacional, las herramientas matemáticas y el marco teórico de las vías moleculares. Lo que permite indentificar cuales son las vías y por ende, los procesos biológicos relevantes en la condición biológica estudiada. De tal manera que han sido necesarios métodos para identificar clases de genes (se hablará de genes pero estos principios pueden ser aplicados a otros términos ómicos) que estan sobre representados dentro de un gran conjunto de genes. Estas técnicas usan abordajes estadísticos para identificar vías moleculares que estan depletadas o enriquecidas de una manera significativa. A estos métodos de análisis de vías moleculares también se les conoce como **enriquecimiento de vías** (o pathway enrichment). Los cuales se pueden clasificar en tres categorías [22]: **sobre representación** (over-representation), **enriquecimiento funcional** (functional class scoring o functional enrichment analysis) y **análisis de vías basados en topología** (pathway topology-based).

Métodos por sobrerepresentación, la hipótesis básica es que las vías pueden ser detectadas en proporción de los genes diferencialmente expresados si dentro de una vía dada se excede la proporción de genes que podrían esperarse al azar. En los métodos de enriquecimiento funcional la hipótesis central es que no solo los grandes cambios en la expresión genética tienen un efecto significativo sobre una vía, si no que cambios más discretos pero coordinados pueden tener un impacto global en el estado de la vía molecular. Para los análisis de vías basados en topología, pueden ser vistos como una extensión de los métodos de sobrerepresentación y enriquecimiento funcional ya que comparten pasos generales, pero añaden la información topológica de la redes construidas en torno a la vías moleculares.

1.3.2.1. Pathifier

Pathifier es un relevante algoritmo de enriquecimiento funcional, el cual busca integrar datos de tecnologías de alto rendimiento con información de vías moleculares en una representación biológica compacta. La representación de las muestras centradas en una perspectiva de vías moleculares puede ser útil para facilitar los hallazgos de fenómenos biológicos.

Haciendo una breve recapitulación de su funcionamiento, Pathifier evalúa los datos provenientes de experimentos de alto rendimiento para una vía molecular a la vez mediante un análisis de componentes principales. A continuación genera una nube de puntos al localizar cada muestra en un sistema de coordenadas donde cada eje representa un elemento de la vía molecular. Seguidamente, se calcula una curva de ajuste mediante el algoritmo de Hastie y Stuetzle donde se asigna el centroide del grupo control como el origen de la curva. Finalmente se obtiene la proyección de los puntos sobre la curva y esa distancia sobre la curva entre la proyección del punto y el origen es el "Pathway Deregulation Score" (**PDS**) asignado a esa muestra para esa vía.

Una de las mayores diferencias entre Pathifier y otros métodos de enriquecimiento funcional es que los resultados no son una lista de vías relevantes sino un puntaje para cada pareja muestra-vía respecto a un grupo control.

1.4. Farmacogenómica

La farmacogenética estudia el rol de una variación hereditaria individual en la respuesta a un fármaco, desde fenotipos que responden con efectos secundarios que ponen en riesgo la vida, hasta ausencia de reacción o eficacia terapéutica. La Farmacogenética ha sido motor de una de las mayores aplicaciones clínicas que en la ciencia médica [38], esta revolución en la terapia farmacológica ha tomado lugar en el siglo XX y continua en el XXI. Enfermedades como la leucemia infantil o infecciones virales han sido curadas o controladas por primera vez. Sin embargo la base conceptual de la Farmacogenética tiene más de 50 años [39]. En paralelo con el desarrollo de las tecnologías de alto rendimiento ha evolucionado el concepto de farmacogenética a la farmacogenómica. La transición del estudio de rasgos monogénicos a la integración de los estudios de genoma completo han sido pilares para este proceso de evolución. La habilidad para considerar todos los factores celulares relacionados con la respuesta a un fármaco pueden ayudarnos a entender mejor los mecanismos subyacentes involucrados y ofrecer mejores tratamientos a los pacientes, con el objetivo en mente de una verdadera medicina personalizada [40].

1.4.1. Proceso de desarrollo de un fármaco

El programa tradicional para el desarrollo de un fármaco es un proceso largo y difícil. En éste, los compuestos de probable utilidad se ensayan en animales para observar si generan los efectos deseados o si producen toxicidad. Aquellos compuestos seguros y eficaces se vuelven candidatos para subsecuentes estudios necesarios en seres humanos. Entonces la FDA y un comité de revisión institucional deben aprobar un protocolo en el que se detalle el ensayo clínico, posteriormente la FDA emite un permiso para el empleo del compuesto con fines de investigación. El nuevo medicamento no puede comercializarse hasta la aprobación

Fase 1, la seguridad y toxicidad es evaluada para seres humanos. en la fase 2 se corrobora la eficacia del fármaco para la enfermedad a la que esta destinado. Fase 3, se evalúa en poblaciones más numerosas y heterogéneas el desempeño del medicamento, con el objetivo de verificar su eficacia para la pretendida aplicación clínica. Así mismo durante esta fase se contrasta el efecto del fármaco con otros tratamientos ya existentes, con placebos o ambos. Finalmente, la fase 4 se da cuando el fármaco ya ha sido aprobado y comercializado, constituyendo la farmacovigilancia o vigilancia poscomercialización. Permitiendo detectar los efectos adversos infrecuentes o de lento desarrollo los cuales son poco probables y difíciles de detectar en los estudios más cortos. Además el uso de fármacos en el mundo real no se restringe a pacientes seleccionados con estrictos criterios de inclusión sino que tienden a ser utilizados por pacientes con comorbilidades y riesgo mayor de efectos adversos.

Se estima que el costo total para el desarrollo de un fármaco hasta su aprobación oscila en los 2.8 mil millones de dólares (mdd), más aún para los fármacos para el tratamiento del Alzheimer puede ascender hasta 5.6 mil millones de dólares. Un proceso que toma 13 años desde los estudios preclínicos hasta la aprobación por la FDA [41].

1.4.2. Reposicionamiento de fármacos

El reposicionamiento de fármacos es una estrategia que busca la reutilización de fármacos actualmente disponibles en el mercado para nuevas indicaciones médicas, diferentes a las que el fármaco fue concebido inicialmente. Este enfoque presenta varias ventajas. La primera es que el riesgo de fallo es menor ya que el fármaco ha probado con anterioridad su seguridad en estudios preclínicos y en humanos. Segunda ventaja, el tiempo para la nueva aplicación de este fármaco es menor. Tercera, se requiere menos costo económico para el desarrollo del candidato a reposición [42].

1.4.3. Trabajos previos en reposicionamiento de fármacos

Múltiples trabajos sobre reposicionamiento de fármacos se han realizado. Desde aquellos que buscan computacionalmente similaridad en la estructura molecular de los compuestos, bajo la hipótesis de que compuestos parecidos tendrán funciones similares. Sin embargo estos abordajes no tienen acceso a el desempeño del fármaco dentro del contexto celular y tisular real. Por otro lado, hay abordajes que se soportan en datos experimentales reales para tal búsqueda.

En 2006, el proyecto **Connectivity Map** [43] estableció un abordaje sistemático para el descubrimiento de conexiones funcionales entre enfermedades y acción farmacológica. Crearon una colección de perfiles de expresión de células humanas tratadas con moléculas bioactivas. Uno de los objetivos era encontrar mecanismos de acción compartidos entre los compuestos químicos y los procesos fisiológicos, y de esa manera establecer relaciones entre enfermedades y fármacos. Las líneas celulares permiten un fácil manejo y probar una gran cantidad de fármacos. Sin embargo, hay que considerar sus limitaciones ya que las líneas celulares han sido modificadas para hacer posible su manejo y supervivencia. Sometidas a un proceso para evitar su senescencia (inmortalización) después del cual en muchas ocasiones muestran alteraciones nucleares. Éstas están separadas del contexto tisular en el que normalmente se nutren, comunican y cumplen sus funciones celulares; sometidas a antibióticos y un entorno estrictamente controlado. De ahí que dichos modelos biológicos no necesariamente reflejen de manera fiel la dinámica de las células que pretenden modelar.

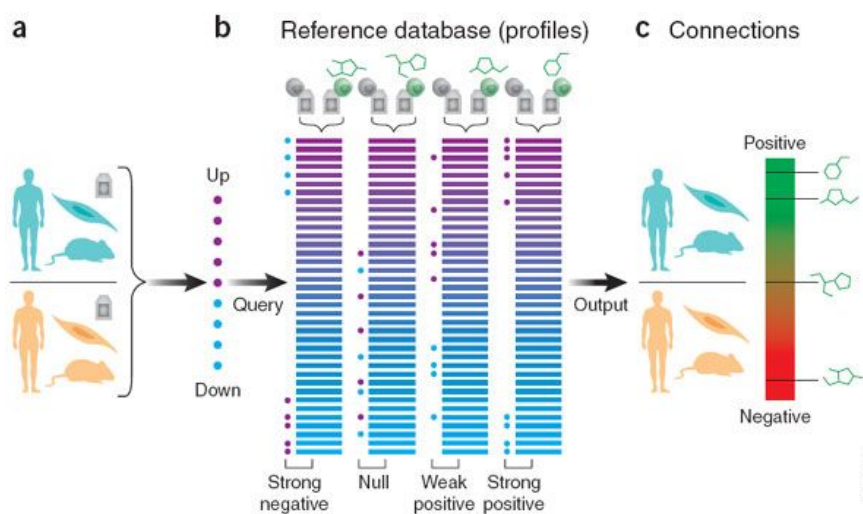


Figura 1.5: a) Se obtienen las diferencias de expresión génica identificadas para enfermos (en azul) y para controles (en amarillo). b) Se ordenan y comparan los perfiles de expresión génica para ser comparados contra un compendio de firmas moleculares conocidas. c) Mediante los resultados se puede establecer una asociación entre la acción de los compuesto sy los efectos conocidos. Proyecto “Connectivity Map” Lamb J. et al, 2006.

Draghici et al, en 2018 [44] construyó una red de fármacos y enfermedades a partir de los datos de *Connectivity Map*, Microarreglos de expresión públicos obtenidos de “Gene Expression Omnibus” (GEO) y la base de datos de *KEGG*. Construye una red uniendo todas las vías moleculares en KEGG, anota los genes que tienen un fármaco conocido, así como aquellos relacionados a una enfermedad. Posteriormente a partir de la red global extrae un subgrafo que consiste de todos los caminos más cortos que conectan a los genes (y sus blancos) asociados a una enfermedad, para posteriormente calcular un score entre fármacos y enfermedades a partir de estimaciones de perturbaciones en los subgrafos. Aunque novedoso, dicho abordaje no explota más niveles de información biológica contenida dentro de los resultados de experimentos de alto rendimiento, solo se nutre del análisis de las diferencias a nivel de genes y montando éstos sobre las rutas moleculares. Perdiendo así la información de las vías desreguladas concretamente en esa enfermedad. También vale la pena mencionar que pueden usarse tecnologías de alto rendimiento recientes que los microarreglos.

También existen antecedentes de trabajos que aportan para el reposicionamiento de fármacos basados específicamente para el caso de cáncer de mama, los cuales se basan en datos del alto rendimiento. Tal es el caso de Segura-Cabrera en 2015 [45] donde propone un algoritmo de caminante aleatorio *Netwalker* [46] sobre una red. Dicha red

1. MARCO TEÓRICO.

está construida a partir de diversas bases de datos vías moleculares y posteriormente anotadas con información farmacológica para luego incorporar datos de tecnologías de alto rendimiento así como datos de perturbaciones. No obstante estos datos de perturbaciones están basados líneas celulares que fueron sometidas a pruebas de senescencia. Además al crear la gran red no dirigida pierde la información de las bases de datos que trabajan con redes dirigidas así como aquellas que guardan el tipo de interacción entre sus elementos, olvidando igualmente el papel del blanco farmacológico. Y así como en el caso anterior no hay un proceso enriquecimiento de vías para los datos. La clasificación del cáncer de mama en dicho estudio es la tradicional por inmunohistoquímica.

Tomando en cuenta lo anterior proponemos un proyecto apoyado en herramientas bioinformáticas para el reposicionamiento de fármacos en subtipos de cáncer de mama con los siguientes objetivos y perspectivas.

Objetivo y planteamiento del proyecto de investigación.

Una vez expuesto el marco en el que se desenvuelve el trabajo doctoral y el problema a abordar el cual es la falta de opciones terapéuticas para el tratamiento del cáncer de mama especialmente para algunos subtipos, procederemos a plantear el proyecto y sus objetivos.

2.0.1. Pregunta de investigación y objetivos.

La pregunta de investigación en el presente proyecto es la siguiente: **¿Cuáles son las vías moleculares más activas por subtipo de cáncer de mama que resultan relevantes en términos de la biología del cáncer para la elección de un tratamiento farmacológico?**

Como objetivo principal del proyecto se pretende: **Identificar mediante métodos computacionales, el reposicionamiento de fármacos potencialmente útiles para los respectivos subtipos moleculares en cáncer de mama.**

Enunciamos como objetivos particulares:

- Encontrar las vías más activas por subtipo de cáncer de mama.
- Identificar cuales de estas vías moleculares son susceptibles de intervención mediante tratamiento farmacológico.
- Clasificar los fármacos de las vías susceptibles a intervención en cada subtipo tumoral de acuerdo a la evidencia existente.

Resultados.

En este trabajo se demuestra que basados en datos de elementos genómicos de un grupo de individuos que comparten cierta enfermedad es posible identificar patrones en los distintos niveles de organización biológica (en nuestro caso genes, vías moleculares y tejidos) y las bases de datos disponibles para proponer blancos farmacológicos potencialmente útiles (que posteriormente necesitan ser validados) para esa enfermedad. Más aún estas ideas pueden ser extendidas a un nivel de medicina personalizada. Los resultados principales de este trabajo se describen en el artículo "Pathway-Based Drug Repositioning for Breast Cancer Molecular Subtypes" publicado en la revista "frontiers in Pharmacology".^{el} 15 de Agosto de 2018. El artículo completo se puede consultar en la presente sección.

3.1. Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. Frontiers in pharmacology 9 (2018)



Pathway-Based Drug Repositioning for Breast Cancer Molecular Subtypes

Raúl A. Mejía-Pedroza¹, Jesús Espinal-Enríquez^{1,2} and Enrique Hernández-Lemus^{1,2*}

¹ Computational Genomics Division, National Institute of Genomic Medicine, Mexico City, Mexico, ² Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

OPEN ACCESS

Edited by:

Zhi Sheng,
Virginia Tech, United States

Reviewed by:

Vijay Pandey,
Cancer Science Institute of Singapore,
National University of Singapore,
Singapore

Robin T. Varghese,
Edward via College of Osteopathic
Medicine, United States

*Correspondence:

Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

Specialty section:

This article was submitted to
Cancer Molecular Targets and
Therapeutics,
a section of the journal
Frontiers in Pharmacology

Received: 27 April 2018

Accepted: 24 July 2018

Published: 15 August 2018

Citation:

Mejía-Pedroza RA, Espinal-Enríquez J
and Hernández-Lemus E (2018)
Pathway-Based Drug Repositioning
for Breast Cancer Molecular
Subtypes. *Front. Pharmacol.* 9:905.
doi: 10.3389/fphar.2018.00905

Breast cancer is a major public health problem which treatment needs new pharmacological options. In the last decades, during the postgenomic era new theoretical and technological tools that give us novel and promising ways to address these problems have emerged. In this work, we integrate several tools that exploit disease-specific experimental transcriptomic results in addition to information from biological and pharmacological data bases obtaining a contextual prioritization of pathways and drugs in breast cancer subtypes. The usefulness of these results should be evaluated in terms of drug repurposing in each breast cancer molecular subtype therapy. In favor of breast cancer patients, this methodology could be further developed to provide personalized treatment schemes. The latter are particularly needed in those breast cancer subtypes with limited therapeutic options or those who have developed resistance to the current pharmacological schemes.

Keywords: drug repositioning, breast neoplasms, systems biology, databases, genetic, pathway analysis, personalized medicine

1. INTRODUCTION

Breast cancer is a major public health concern and a main cause of death in young women worldwide (Ferlay et al., 2015). Breast cancer is characterized by an heterogeneous nature at the histological, molecular, and systemic levels. To cope with this heterogeneity, multiple prognostic and therapeutic approaches have been developed to handle this disease. From classifications based on clinical parameters or histopathologic markers (for example estrogen receptor[ER], progesterone receptor [PR], and epidermic human growth factor receptor [HER2]) (Prat et al., 2015) to molecular classifications based on gene expression values of the samples, such is the case of PAM50 classification (Parker et al., 2009).

The four breast cancer intrinsic subtypes (Luminal A, Luminal B, Basal, and Her2-enriched) defined by PAM50 offer more accurate clinical information than the classifications based on histopathologic parameters (Prat et al., 2015). The predictive power of PAM50 regarding patient prognosis is an example of this increase in accuracy (Prat et al., 2014, 2015). Furthermore, the new breast cancer international guidelines (Coates et al., 2015; Senkus et al., 2015) not only mention the utility of PAM50 intrinsic subtypes, but the latter states that they are indeed a central piece of the therapeutic and classification algorithms.

3.1 Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. *Frontiers in pharmacology* 9 (2018)

Mejía-Pedroza et al.

Drug Repositioning in Breast Cancer Subtypes

It is worth mentioning that the scarcity of effective therapeutic options, in particular for certain subtypes—such as basal, or triple-negative tumors—still present a challenge for clinicians who often have to resort to highly unspecific cytotoxic therapies.

In order to help solve this problem, theoretical and methodological advances are aimed at the development of improved therapeutic options firmly based on a deeper biological understanding of the disease. Three of these advances are particularly important in our view.

First, high throughput technologies capable of retrieve a huge amount of biological data; for example, expression microarrays or RNAseq which allow us to approach the study of the transcriptome, i.e., the set of RNA molecules present in biological entities. Second, The conception that generally speaking, a biological function of the living cell is a result of many interacting molecules; It can not be attributed to just a single gene or a single molecule. The above can be represented as a set of genes in the genome linked through a network of interacting molecules in the cell, such as a “Pathway,” representing a higher order biological function (Kanehisa and Goto, 2000). Third, the development of areas such as pharmacogenomics, which try to analyze by means of a global approach all the genes involved in the response to a certain drug in a variety of conditions (Pirmohamed, 2001). Such pharmacogenomics approach has evolved from a view centered on isolated genes to one based on pathways (García-Campos et al., 2015). Pharmacogenomics has progressively adopt a variety of high throughput technologies including genomics, transcriptomics, proteomics, and others, to enhance its capacity for generating and testing hypotheses and transfer these hypotheses to the clinical practice (Wang, 2010).

Despite these advances, for many breast cancer patients prognosis remains poor for both survival and quality of life (which is further affected by side effects of the pharmacological treatment itself). New, more effective and less harmful therapeutic options are needed, especially for certain subgroups of breast cancer patients, for example, those who suffer from basal subtype or from a tumor which develop resistance to usual therapeutic schemes (Prat et al., 2014; Yu and Jones, 2016; Friese et al., 2017). Some causes of this deficit of therapeutic options are the high economic and logistic costs involved, not only in traditional drug development (i.e., the path from concept to drug's approval), but also in the clinical evaluation of its usefulness in diseases for which the drug was not initially conceived, or drug repurposing assessment. In both cases, a more accurate selection of pharmacological targets is crucial to achieve viability and high rate of success. For all of the aforementioned reasons, a main goal of the present study is to answer the question: is it possible to identify new associations between pharmacological targets and their respective pathways to each breast cancer molecular subtype? In order to answer this, we propose a computational analysis based on the systematic inquiry of all relevant deregulated pathways specific for each breast cancer subtype, and the assessment of target genes (belonging to the aforementioned pathways), which are susceptible to pharmacological modulation. This approach relies on large, well-curated gene expression datasets from high throughput technologies coming from the two flagship projects (TCGA and

METABRIC) of transcriptomic characterization of mammary tumors. Such projects are hence the gold standard reference in terms of quality and quantity of samples. Adding to this, the use of state of the art computational methods will allow us to develop a reliable, trustworthy study that supports the results that will be presented.

2. MATERIALS AND METHODS

2.1. Outline

The analysis pipeline followed in the present work is depicted in **Figure 1**. Briefly, we obtained two data sets of gene expression experiments, one data set came from METABRIC (Curtis et al., 2012) repository and the other one from TCGA (Ciriello et al., 2015). Each data set was build up trough different technologies, namely Microarrays (METABRIC) and RNAseq (TCGA). Then, the samples were classified in its respective molecular breast cancer subtype according the PAM50 algorithm (Parker et al., 2009). After that, we identified the most deregulated pathways in each breast cancer subtype through Pathifier algorithm (Drier et al., 2013). Each deregulated molecular pathway was associated to its known pharmacological targets according to information from pharmacological databases. Finally, these drug-subtype relations were classified according to information available in both the pharmacological databases as well as information from the gene expression data of the samples themselves. Supported by these analyses we develop a Database Based Drug Repurposing (DBDR) method that allows for the implementation of precision medicine approaches, that may be applied even at the individual (personalized medicine) level.

2.2. Data Sources

As already discussed, our analysis is supported with information coming from the largest, more reliable whole-genome breast cancer gene expression datasets available, the ones provided by the METABRIC and TCGA consortia, respectively. We will briefly describe those datasets—at the level used here—. For more detailed information, please consult the original sources as referenced.

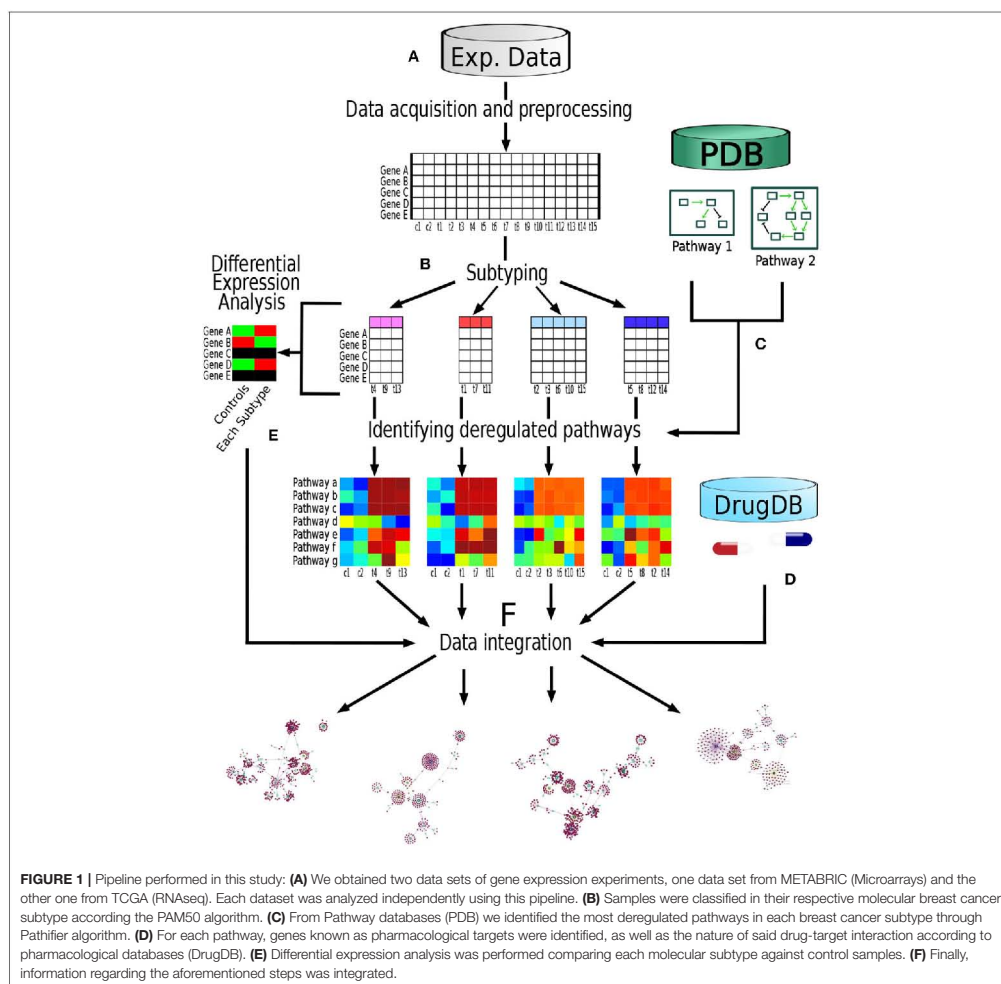
METABRIC (Molecular Taxonomy of Breast International Consortium) is a collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumor banks in the UK and Canada, more details about those data can be found in Curtis et al. (2012). To minimize batch effects, we only use the “discovery set” consisting in expression data from 997 samples of women with breast cancer and the “control set” a set of expression data from 144 samples from adjacent normal breast tissue.

Data were provided under request through the following download platform:

https://www.ebi.ac.uk/ega/about/your_EGA_account/download_streaming_client.

TCGA (The Cancer Genome Atlas) is a specialized database that stores high throughput technology data of at least 33 types of breast cancer obtained from more than eleven thousand patients. This big database is a reference point in terms of quality and

3. RESULTADOS.



sample size. We used the data set generated and published by Ciriello et al. (2015) which come from 808 biopsies of women with breast cancer and 112 from adjacent normal breast tissue.

Data were downloaded with their respective clinical information from the following internet address:

http://www.cbioportal.org/study?id=brca_tcga_pub2015#summary.

2.3. Preprocessing

In METABRIC Each RNA transcriptional profiling was extracted through the Illumina HT-12 v3 microarray

platform. The consortium provides the data in the form of preprocessed expression matrices. The detailed description of the preprocessing performed in the data can be found in the original publication (Curtis et al., 2012).

Similarly to the previous case, the data used for TCGA are provided as preprocessed expression matrices. Details about data preprocessing could be found in the original publication (Wilkerson et al., 2014). Briefly (As reported by the author), RNA sequencing was performed on Illumina HiSeq. Resulting sequencing reads were aligned to the human hg19 genome assembly using MapSlice (Wang et al., 2010). Gene expression was quantified for the transcript models corresponding to the

TABLE 1 | Table Distribution of subtypes by database.

Database	Control	Basal	Her2	Lumina A	Luminal B	Sum
METABRIC	144	118	87	466	268	1,083
TCGA	112	136	65	411	171	895
Sum	256	254	152	877	439	1,978

TCGA GAF 2.13 using RSEM4 and normalized within samples to a fixed upper quartile. Upper quartile normalized RSEM data were log2 transformed.

2.4. Subtyping

Samples were classified through PAM50 (Parker et al., 2009) algorithm by METABRIC and TCGA consortiums themselves. The proportions of samples per subtype inside each database are shown in **Table 1**.

2.5. Functional Enrichment Analysis

To identify the most deregulated molecular pathways we use Pathifier (Drier et al., 2013). Pathifier is an algorithm that integrates experimental data from high throughput technologies and pathway information from biological databases (**Supplementary Material**).

Pathifier provides as a result a deregulation value for each molecular pathway in each sample. This deregulation value is called “Pathway Deregulation Score” or PDS. This methodology has been used previously to observe deregulated pathways in breast cancer samples from TCGA (Drago-García et al., 2017). Importantly, Pathifier assigns a PDS for each individual sample, which in our case is a fundamental feature for choosing this method to assess the deregulation level for pathways of each breast cancer subtype.

In our case we use the molecular pathways defined by KEGG (Kanehisa and Goto, 2000) and the analysis was performed independently in each database (METABRIC and TCGA) for each tumor subgroup of breast cancer comparing it against its corresponding control group. **Figure 2** shows the result of the calculation of Pathifier for the basal subtype of METABRIC against its respective controls.

In order to identify the most deregulated pathways by subtype we performed a z-score transformation over the PDS that belonged to each pathway (row). This transformation allow us to compare the deregulation scores (called “PDSz” here after) between different molecular pathways as described by the author of the method (Livshits et al., 2015). We looked for those pathways whose patients exhibited the highest PDSz consistently in both databases. This was done as follows: the algorithm returns “n” PDSz values (one PDSz per each patient) for a given pathway. We used the median thus obtaining only one deregulation value per molecular pathway in each database. then we sort the values and got the intersection in METABRIC and TCGA data of top 30 most deregulated pathways. Results are shown in **Table 2**.

2.6. Drug Target Associations

The aim of this procedure was to identify which of the studied *genes* (in the broad sense of the word) could be considered as pharmacological targets by virtue of a drug acting upon it, according to pharmacological databases. Drug-target associations were identified through a tool known as *The Drug-Gene Interaction Database* (www.dgidb.org). This platform integrates information of at least 15 pharmacological databases which include information about drugs, pharmacological targets, type of drug-target interaction, data sources, and other characteristics. That tool was used through its implementation as bioconductor package which name is RDGIdb (Wagner et al., 2015).

2.7. Prioritization of Pharmacological Targets

Drugs were classified according to the following criteria. First, if there was a molecular pathway with which the drug *interacts*—i.e., if the drug has any kind of physicochemical effect in at least one of the pathway’s elements, having a noticeable effect in the pathway itself—. Second, if that molecular pathway resulted highly and consistently deregulated in a breast cancer molecular subtype (as it is found by applying pathway deregulation calculations in both gene expression datasets under study). Third, if the pharmacological target gene was differentially expressed in that breast cancer subtype in relation to healthy controls. Fourth, depending the type of interaction of the drug with the pharmacological target, in particular whether the effect of the drug (a) leads to a return of the target gene basal expression level (for presentation purposes, we will call *homeostasis*), (b) given an overexpressed/underexpressed target, leads to a higher overexpression/underexpression (anti-homeostasis), or (c) leads to an undetermined behavior of the target.

2.8. Differential Expression Analysis

Differential expression analysis is a procedure which allows to identify if the same gene holds different expression levels with respect to two conditions (in our case a breast cancer subtype compared to adjacent normal breast tissue). In this study, differential expression analysis was made through the *limma* (Ritchie et al., 2015) bioconductor (Gentleman et al., 2004) package for the case of METABRIC data and with *DESeq2* (Love et al., 2014) in the case of TCGA data. Threshold parameters used for those analysis were a absolute log fold change (base 2) >1, a Benjamini-Hochberg adjusted *p*-value lesser than 0.001 and a FDR lesser than 0.1. All the code was developed using the Emacs speak statistics package (Rossini et al., 2004)

2.9. Individualized Proof-Of-Concept

Since Pathifier provides a PDS per sample, the deregulation of any given pathway is different for each patient. The final section is composed by two examples of deregulated pathways in two patients with basal subtype. There we show the PDS for their most deregulated pathways, the genes which are susceptible to pharmacological modulation, and also the drugs that modulate these targets and concomitantly, the associated pathways, paying special attention to those drugs that have a mechanism of action

3.1 Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. *Frontiers in pharmacology* 9 (2018)

TABLE 2 | Concordance between the most deregulated pathways.

Molecular subtype	Deregulated pathways
Basal	Olfactory transduction, Hedgehog signaling pathway, Cell cycle, Base excision repair, ErbB signaling pathway, Neurotrophin signaling pathway, Metabolic pathways, Apoptosis, Oocyte meiosis, Drug metabolism - other enzymes
Her2	Toil-like receptor signaling pathway, Apoptosis, Vasopressin-regulated water reabsorption, Aldosterone-regulated sodium reabsorption, Glycine, serine and threonine metabolism, DNA replication, ErbB signaling pathway, Melanogenesis
LumA	Jak-STAT signaling pathway, NF-kappa B signaling pathway, Glycerolipid metabolism, Fatty acid degradation, TNF signaling pathway, Fc epsilon RI signaling pathway, Leukocyte transendothelial migration, Osteoclast differentiation, ECM-receptor interaction, Ascorbate and aldarate metabolism, FoxO signaling pathway
LumB	Steroid biosynthesis, Retinol metabolism, cAMP signaling pathway, Vasopressin-regulated water reabsorption, Adrenergic signaling in cardiomyocytes, Progesterone-mediated oocyte maturation, Thyroid hormone synthesis, GnRH signaling pathway, Glutamatergic synapse

of direct and selective interaction with the identified target, thus reducing the undesired side effects.

3. RESULTS

3.1. Deregulated Pathways by Breast Cancer Subtype

3.1.1. Deregulated Pathways Were Consistent Across Different Technologies

The deregulated pathways by subtype were consistent across different technologies. Unsupervised hierarchical clustering was performed to see if the set of pathway deregulation levels related to one particular breast cancer subtype was consistent with the subtype itself or instead it was consistent with the technology where the data come from. Before carrying out this procedure, were calculated the PDSz (the level of deregulation associated to each pathway in each tumoral subtype) as described in section Materials and Methods. Thus, there were eight sets of PDSz, considering four tumoral subtypes and two different data sources. After that, unsupervised hierarchical clustering was applied. The result of this analysis is shown in **Figure 3**, where it can be seen that data do not group according to the technology by which they were obtained, instead they group according to the biological condition from which they come.

3.1.2. Agreement Between the Most Deregulated Pathways

For each tumoral subtype the 30 pathways with higher deregulation score were taken and then we selected the ones which coincided in both databases (from now on called the most deregulated pathways by breast cancer subtype). These are mentioned in **Table 2**.

Finally, information regarding the deregulation value of the pathway associated to the pharmacological target for the breast

cancer subtype, the differential gene expression status of the pharmacological target, type of interaction between drug and pharmacological target, the relationship between the type of drug-target interaction and the differential expression status of the latter, is presented in **Supplementary Material**.

A brief example of the above mentioned database is provided in **Table 3**. Columns show several criteria by means of which the pharmacological targets were classified: *Drug* associated to the pharmacological target, *Gene* associated to the target, *Pathway* to the molecular pathway which pharmacological target belongs to. *logFC* refers to the logarithmic differential expression ratio of the pharmacological target gene in the breast cancer subtype relative to its controls. *InteractionType* to the type of interaction between the drug and the pharmacological target. *Homeostasis* is a term determined according to the differential expression status of the pharmacological target gene and the type of drug-target interaction. Finally, *Source* refers to the database from which the information was obtained.

3.2. Pharmacological Targets Are Associated to Specific Molecular Subtypes

Pharmacological target were selected from the pathways that are the most deregulated for a specific subtype of breast cancer and also coincide in both databases. Additionally, the target has to be differentially expressed.

The selected pathways and the number of associated pharmacological targets are shown in **Table 4**. There, rows are the most deregulated molecular pathways in both databases by breast cancer subtype as shown in **Table 2**. Columns are PAM50 breast cancer subtypes and number of pharmacological targets that are also differentially expressed in each subtype of breast cancer.

We can notice that there are many more pharmacological targets associated to each molecular subtype than those currently used for their clinical treatment. In **Figure 4** we can see a multipartite network of pathways-targets-drugs for basal tumors. A number of potential therapies (not currently used) is evident, as is the redundancy of the network, which could be indicative for flexible therapeutic approaches.

Seventy-nine drugs leading to homeostasis over eight highly deregulated pathways have been found for the Basal breast cancer subtype. This acquires particular relevance since this subtype is closely associated to triple negative tumors, for which directed (i.e., non-purely cytotoxic) pharmacological therapies are scarce.

The fact that a high number of drugs already in the market (i.e., after passing phase I, II, and III clinical trials and receiving legal approval) are able to target directly some of the most deregulated pathways in this hard-to-treat tumors points out to the actual scope of pathway-based drug repurposing approaches. A further application of these findings on the road to personalized medicine is shown in the Discussion section.

4. DISCUSSION

Breast cancer is a public health problem which treatment require new pharmacological options. In the present study, For each

3. RESULTADOS.

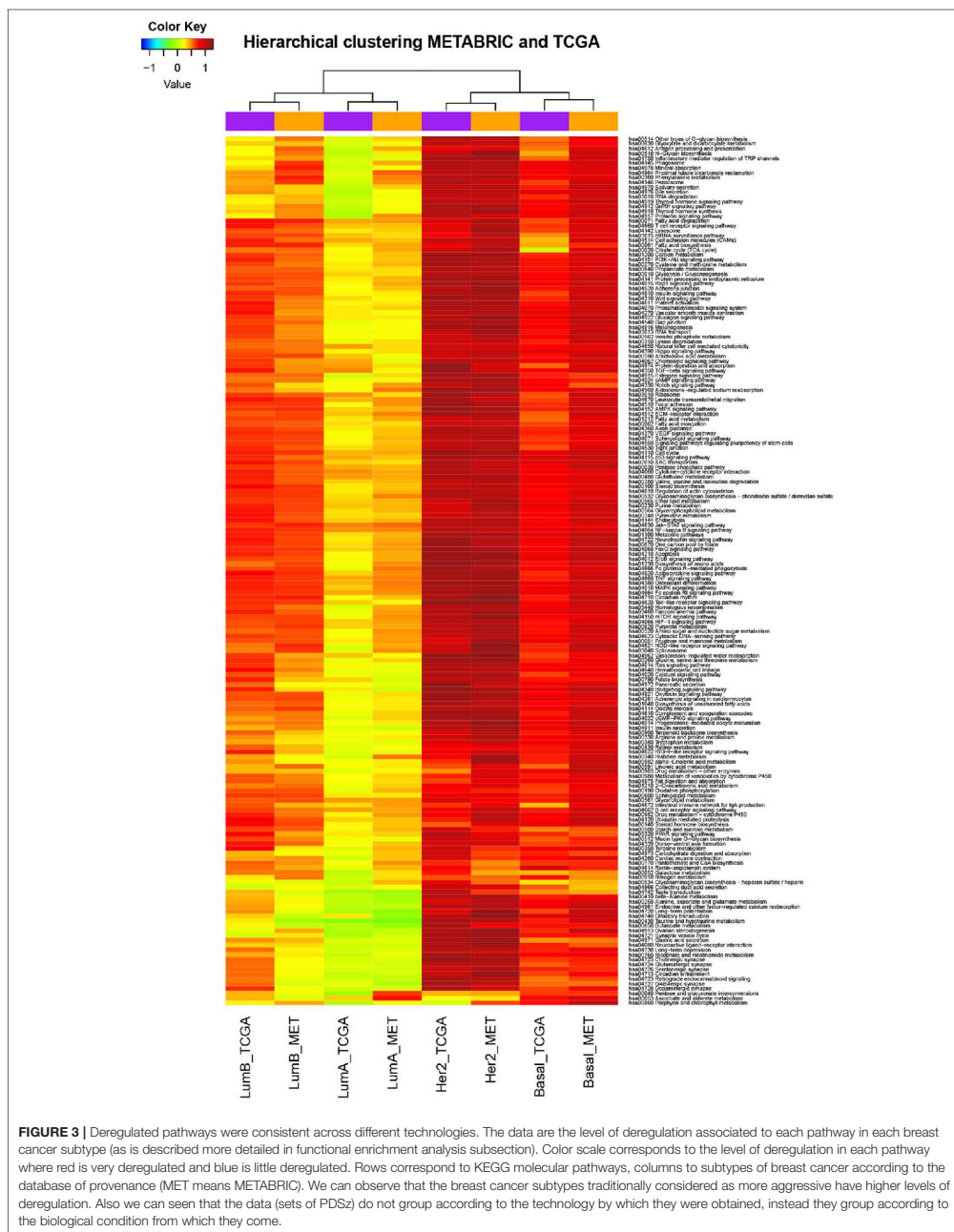


FIGURE 3 | Deregulated pathways were consistent across different technologies. The data are the level of deregulation associated to each pathway in each breast cancer subtype (as is described more detailed in functional enrichment analysis subsection). Color scale corresponds to the level of deregulation in each pathway where red is very deregulated and blue is little deregulated. Rows correspond to KEGG molecular pathways, columns to subtypes of breast cancer according to the database of provenance (MET means METABRIC). We can observe that the breast cancer subtypes traditionally considered as more aggressive have higher levels of deregulation. Also we can see that the data (sets of PDSz) do not group according to the technology by which they were obtained, instead they group according to the biological condition from which they come.

3.1 Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. *Frontiers in pharmacology* 9 (2018)

TABLE 3 | Example of the information contained in the database.

Pathway	Gene	Drug	Interaction type	Source	logFC	Effect
hsa00230 Purine metabolism	HPRT1	AZATHIOPRINE	Inhibitor	DrugBank	1.623	Homeostasis
hsa00230 Purine metabolism	PNP	CLADRINE	Inducer	DrugBank	1.628	Anti-homeostasis
hsa00230 Purine metabolism	PDE2A	TOFISPAM	Inhibitor	DrugBank	-3.962	Anti-homeostasis
hsa00230 Purine metabolism	PDE8B	KETOTIFEN	Inhibitor	DrugBank	-2.509	Anti-homeostasis
hsa00310 Lysine degradation	ALDH2	DISULFIRAM	Inhibitor	DrugBank	-2.363	Anti-homeostasis
hsa00100 Steroid biosynthesis	SQLE	NAFTIFINE	Inhibitor	DrugBank	2.703	Homeostasis
hsa00100 Steroid biosynthesis	SQLE	TERBINAFINE	Inhibitor	DrugBank	2.703	Homeostasis
hsa00100 Steroid biosynthesis	SQLE	BUTENAFINE	Inhibitor	DrugBank	2.703	Homeostasis
hsa00100 Steroid biosynthesis	SQLE	ELLAGIC ACID	Inhibitor	DrugBank	2.703	Homeostasis

TABLE 4 | Pharmacological targets by tumor subtype.

Deregulated pathway	Tumor subtype	Number of druggable targets
Adipocytokine signaling pathway	Luminal B	6
Aldosterone-regulated sodium reabsorption	Her2+	6
Apoptosis	Basal	6
Apoptosis	Her2+	3
Ascorbate and aldarate metabolism	Luminal A	1
Cell cycle	Basal	7
Cell cycle	Luminal B	4
Drug metabolism - other enzymes	Basal	3
ECM receptor interaction	Luminal A	4
ErbB signaling pathway	Basal	4
ErbB signaling pathway	Her2+	5
Fatty acid degradation	Luminal A	5
Fatty acid degradation	Luminal B	5
Fc epsilon RI signaling pathway	Luminal A	1
FoxO signaling pathway	Luminal A	6
Glycerolipid metabolism	Luminal A	3
Glycerophospholipid metabolism	Luminal B	4
Glycine, serine and threonine metabolism	Her2+	6
Jak-STAT signaling pathway	Luminal A	5
Leukocyte transendothelial migration	Luminal A	2
Leukocyte transendothelial migration	Luminal B	4
Melanogenesis	Her2+	2
Metabolic pathways	Basal	46
Neurotrophin signaling pathway	Basal	7
NF-κ signaling pathway	Luminal A	1
Olfactory transduction	Basal	1
Oocyte meiosis	Basal	5
Osteoclast differentiation	Luminal A	3
Osteoclast differentiation	Luminal B	7
Steroid Biosynthesis	Luminal B	1
TNF signaling pathway	Luminal A	5
Toll-like receptor signaling pathway	Her2+	5

Notice that despite there are subtypes that present the same deregulated pathway, the number of targets between subtypes is different.

breast cancer molecular subtype we looked for deregulated pathways that are pharmacologically modulable and are not currently used in the corresponding breast cancer subtype therapy. We have identified new pharmacological targets (and their respective pathways) that are potentially useful in breast cancer subtypes treatment.

Significant amount of work has been dedicated to the discovery of novel drug-disease associations. Previous efforts have been made to associate drug response with gene signatures (Lamb et al., 2006; Lee, 2017). However, those tools were designed to be used in cell lines, their performance in real human cells has not been explored. Furthermore, those approaches were built up with a relatively small number of drugs, and they were not built specifically in the context of breast cancer.

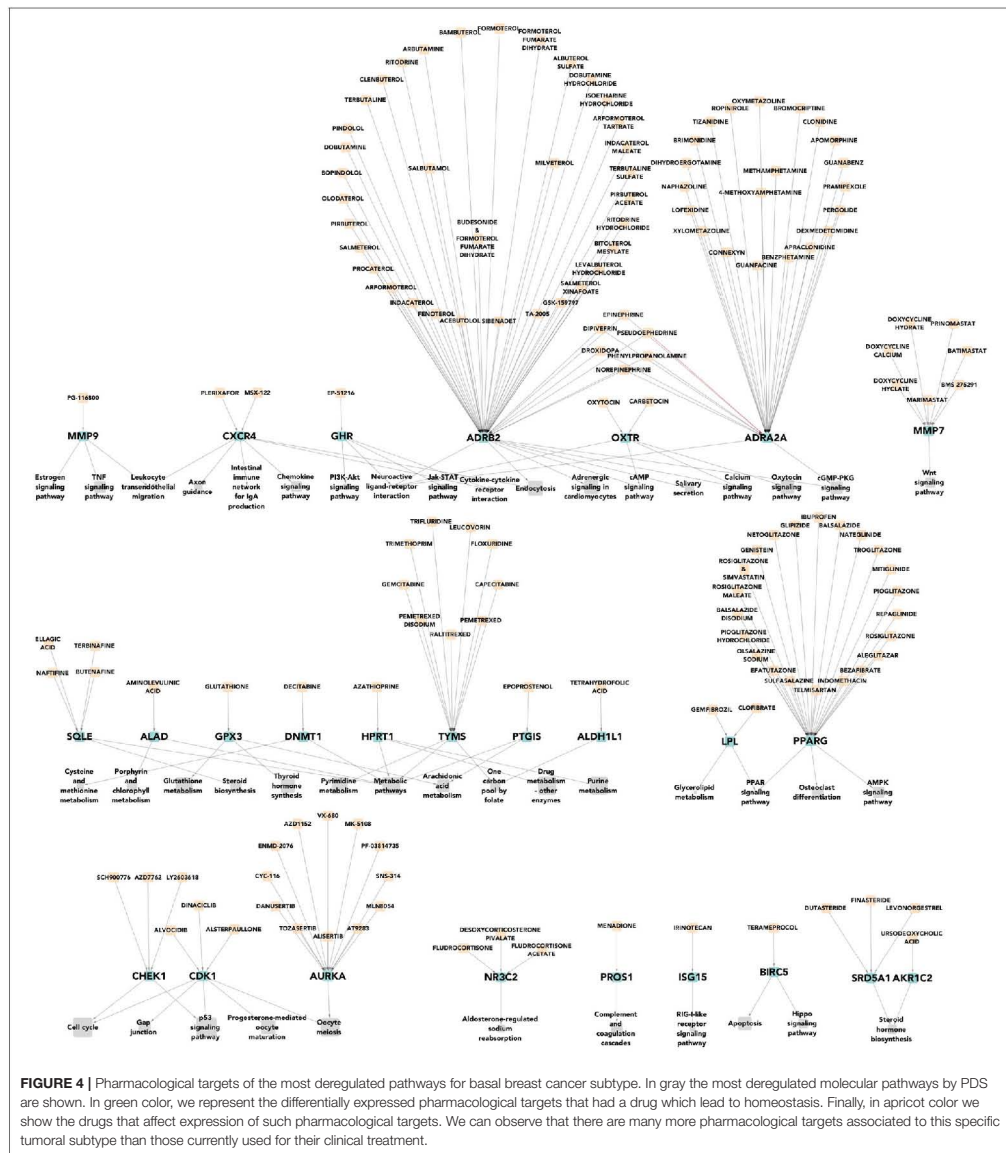
With this in mind, we performed a data-driven approach to the actual phenomenology of the different breast cancer subtypes as represented by their distinctive gene expression patterns, to look up for sets of patient-specific deregulated pathways, then by resorting to molecular and drug databases. We designed a methodology to find out druggable targets whose associated drugs were already in the market, i.e., we built a drug repositioning strategy.

4.1. New Relations Between Pathways and Subtypes

Table 2 shows the set of the highly deregulated molecular pathways associated to each subtype of breast cancer. The description of how the associations were made is discussed in the section Materials and Methods, as well as in the Results section. Some of the associations found in the present study between breast cancer subtypes and molecular pathways are already known and intensively studied, in fact, there are pharmacological targets belonging to these molecular pathways which are currently employed for the treatment of the corresponding breast cancer subtype. Some of these relationships will be mentioned below.

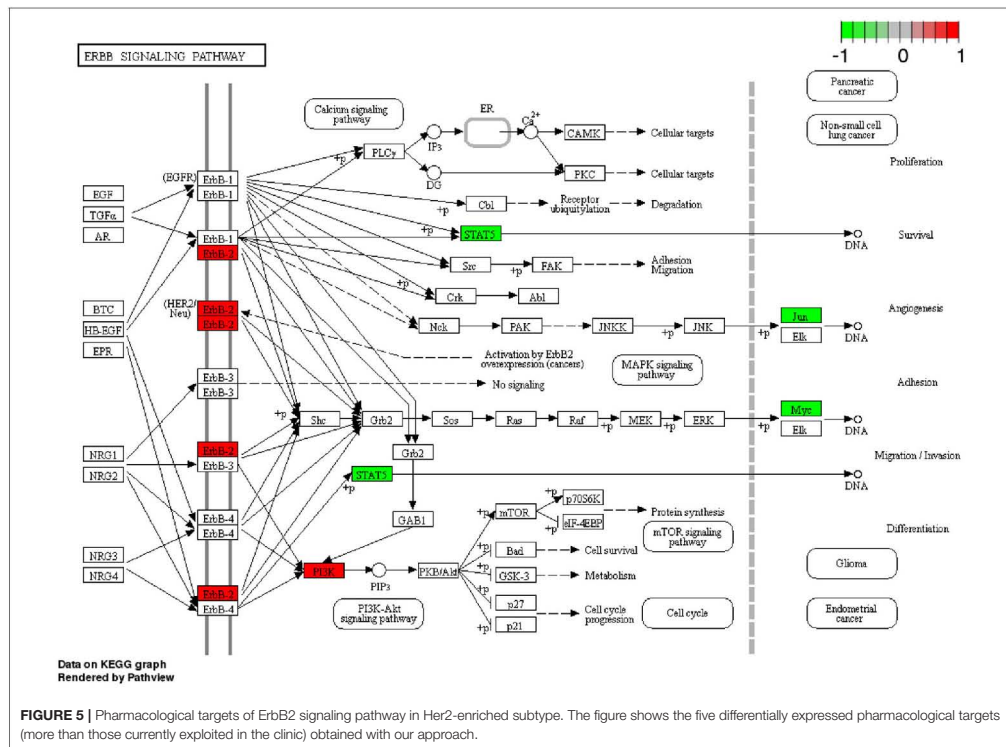
With regard to Her2-enriched subtype, in our study, it is associated to the ErbB signaling pathway one of the most studied pathways for this subtype (the name of this subtype of breast cancer is due to the relationship with this pathway). The ErbB2

3. RESULTADOS.



receptor is the mainstay of pharmacological treatment in patients with the Her2-enriched breast cancer subtype (Senkus et al., 2015). In the present study we also associate the Her2-enriched subtype with the molecular pathway of apoptosis and DNA

replication, where it has already been described in the literature to detail how the overexpression of Her2 leads to uncontrolled cell proliferation and suppression of apoptosis (Carpenter and Lo, 2013).



Regarding the basal subtype, its association with the molecular pathways of apoptosis and cell cycle is not new; the increase in cell proliferation and apoptosis is a well-known characteristic of this subtype (Choo and Nielsen, 2010). Moreover, lapatinib (a Her2 inhibitor) has been shown to be useful in treatment when combined with other agents in breast cancer models of the basal phenotype (Liu et al., 2011), therefore the association of this molecular subtype with Her2 signaling pathway is not entirely new.

Although several associations found in the present study between molecular pathways and subtypes of breast cancer are under intense study, here we propose novel associations that do not have been previously reported, based on the present study. This strategy results promising to evaluate their usefulness in generating knowledge regarding breast cancer treatment.

4.2. New Relations Between Pharmacological Targets and Subtypes

We found different pharmacological targets associated to each breast cancer subtype. The number of pharmacological targets associated to each breast cancer subtype is shown in Table 4

and how they were associated is described in the Results section. In that table we can note that even in the cases where two breast cancer subtypes share the same deregulated molecular pathway, the number of differentially expressed pharmacological target genes is different. By observing Table 4 we see that in the molecular subtype Her2-enriched the molecular pathway "ErbB signaling" has five deregulated pharmacological targets. These are shown in detail in Figure 5. However only one of these five pharmacological targets is currently indicated in the clinical practice guidelines for breast cancer: ErbB2.

Currently, 1–5 pharmacological targets are typically used in clinic, depending on the molecular subtype of the patient. In this work we show that more than a dozen pharmacological targets are associated to each breast cancer subtype. Therefore, a subsequent evaluation of these new pharmacological targets may lead to some new therapeutic options, needed in the clinic environment. Finally, it is worth to mention that each pharmacological target may have several (up to dozens) drugs or chemical compounds that exert actions over it. Such condition amplifies the ways to modulate the therapeutic effect.

3. RESULTADOS.

4.3. DBDR in the Age of Personalized Medicine

In what follows we will present, as a *proof-of-concept*, an application of the DBDR approach to two individual breast cancer patients (within our datasets) with known gene expression profiles. Based on such expression profile, specific molecular subtype and main deregulated pathways can be known.

We decided to focus these examples in the basal subtype, since the scarcity of non-cytotoxic therapeutic options and the poor prognosis of patients having this phenotype. In the two following boxes we show the deregulated pathway with their respective PDS, their associated target genes, and the drugs that lead to homeostasis of said pathways. Each proposed drug set was checked so that all of the drug pairs do not have any known drug interaction. Finally we discussed a personalized set of drugs for each individual.

4.3.1. Individual Therapy Designed for Patient [MB.3058]

This is a patient classified as having a basal tumor with the following deregulated pathways:

Pathway	PDS	Target	Drug
Steroid biosynthesis	0.937	SQLE	TERBINAFINE
Pyrimidine metabolism	0.466	TYMS	PEMETREXED or Capecitabine
Sphingolipid metabolism	0.628	SPHK1	SKI-I
Olfactory transduction	0.736	NA	NA
Apoptosis	0.398	BIRC5	ALVOCIDIB or Paclitaxel
Oocyte meiosis	0.406	AURKA, CDK1	ALISERTIB, DINACICLIB
Cell cycle	0.328	CDK1	DINACICLIB
Neurotrophin signaling pathway	0.765	NA	NA
ErbB signaling pathway	0.783	NA	NA
Drug metabolism—other enzymes	0.712	NA	NA

In view of this particular set of deregulated pathways and gene expression patterns, *individualized therapy* for this patient may include the following drugs: **TERBINAFINE**, **PEMETREXED**, or **CAPECITABINE**, and **SKI-I**. Alvocidib, paclitaxel, alisertib y dinaciclib are not considered since the low PDS in their respective pathways.

4.3.2. Individual Therapy Designed for Patient [MB.5387]

This is a patient classified as having a basal tumor with the following deregulated pathways:

Pathway	PDS	Target	Drug
Steroid biosynthesis	0.597	SQLE	TERBINAFINE
Pyrimidine metabolism	0.605	TYMS	PEMETREXED or Capecitabine
Sphingolipid metabolism	0.590	SPHK1	SKI-I
Olfactory transduction	0.500	NA	NA
Apoptosis	0.712	BIRC5	ALVOCIDIB or Paclitaxel
Oocyte meiosis	1	AURKA, CDK1	ALISERTIB, DINACICLIB
Cell cycle	0.718	CDK1	DINACICLIB
Neurotrophin signaling pathway	0.839	NA	NA
ErbB signaling pathway	0.616	NA	NA
Drug metabolism—other enzymes	0.846	NA	NA

In view of this particular set of deregulated pathways and gene expression patterns, *individualized therapy* for this patient may include the following drugs: **TERBINAFINE**, **PEMETREXED**, or **CAPECITABINE**, **SKI-I**, **ALVOCIDIB**, or **PACLITAXEL**, **ALISERTIB**, and **DINACICLIB**.

4.4. Scope and Limitations

A systems biology approach such as the one presented here is aimed at primary discovery and hypotheses generation. As such, there is a need to test the findings obtained here—in the context of the pharmacogenomics and lifestyles of actual patients—before they can be fully beneficial in a general clinical practice.

Considering the previous comments, we cannot stress enough that this theoretical approach is still far from constituting an actual aid to the clinician. For this reason, no further considerations on dose, drug combinations and patient features (other than tumor subtype, gene expression profile and deregulated pathways) are being considered.

5. CONCLUSIONS

New molecular pathways and their respective pharmacological targets of potential therapeutic utility were associated to

3.1 Mejía-Pedroza, Raúl A., et al. Pathway-based drug repositioning for breast cancer molecular subtypes. *Frontiers in pharmacology* 9 (2018)

Mejía-Pedroza et al.

Drug Repositioning in Breast Cancer Subtypes

each of the different breast cancer molecular subtypes. This allows subsequent focused assessment of the pharmacological targets associated to each subtype in order to find new therapeutic options for the treatment of patients suffering from this disease. The knowledge derived from this approach will result particularly useful for those patients that have developed pharmacological resistance to canonical treatment and/or suffer from tumors with reduced therapeutic options, such as basal or triple-negative breast cancers.

Here we present an application of pathway-based analyses based on the detailed study of two of the largest and more trustworthy international cohorts, including whole genome high quality transcriptomic data, coming from either RNASeq or microarray technologies: The METABRIC and TCGA collaborations. By analyzing high throughput gene expression data for thousands of patients, using a robust probabilistic approach to find the most deregulated pathways for each subtype (even at the per-patient level) and assessing the findings by crossing-out the information with highly curated knowledge-based drug target databases, we have been able to develop a methodology that allows for a detailed drug-repurposing scheme to treat the different breast cancer tumor types and that, in principle (as an introductory proof-of concept) may even be extended up to the single patient design, thus paving the way to personalized medicine to treat such complex pathologies as breast cancer using a pragmatic drug re-use approach.

Hopefully, the work presented here will contribute to save considerable amounts of financial resources—by resorting to drug repurposing strategies—, but above all, may soon allow us to save more and more lives of people suffering from this excruciating disease.

REFERENCES

- Carpenter, R. L., and Lo, H.-W. (2013). Regulation of apoptosis by her2 in breast cancer. *J. Carcinog. Mutagen.* 2013(Suppl. 7):003. doi: 10.4172/2157-2518.S7-003
- Choo, J. R., and Nielsen, T. O. (2010). Biomarkers for basal-like breast cancer. *Cancers* 2, 1040–1065. doi: 10.3390/cancers2021040
- Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., et al. (2015). Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 163, 506–519. doi: 10.1016/j.cell.2015.09.033
- Coates, A. S., Winer, E. P., Goldhirsch, A., Gelber, R. D., Gnant, M., Piccart-Gebhart, M., et al. (2015). Tailoring therapies—improving the management of early breast cancer: St Gallen international expert consensus on the primary therapy of early breast cancer 2015. *Ann. Oncol.* 26, 1533–1546. doi: 10.1093/annonc/mdv221
- Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346–352. doi: 10.1038/nature10983
- Drago-García, D., Espinal-Enríquez, J., and Hernández-Lemus, E. (2017). Network analysis of emt and met micro-rna regulation in breast cancer. *Sci. Rep.* 7:13534. doi: 10.1038/s41598-017-13903-1

AUTHOR CONTRIBUTIONS

RM-P carried out analysis and calculations, devised code, and database management duties and contributed to the writing of the manuscript. JE-E contributed to the discussion, reviewed analysis and contributed to the writing of the manuscript. EH-L designed and coordinated the project, contributed to the discussion, supervised, and reviewed the writing of the manuscript. All authors read and approved the final manuscript.

FUNDING

RM-P is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 419866 from CONACYT. This work was supported by CONACYT (grant no. 285544/2016), as well as by federal funding from the National Institute of Genomic Medicine (Mexico). Additional support has been granted by the Laboratorio Nacional de Ciencias de la Complejidad (grant no. 232647/2014 CONACYT). EH-L is recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

ACKNOWLEDGMENTS

Authors also thank to Miguel Ángel García-Campos from Weizman Institute of Sciences and Guillermo de Anda-Jáuregui from University of North Dakota for their valuable support and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.00905/full#supplementary-material>

- Drier, Y., Sheffer, M., and Domany, E. (2013). Pathway-based personalized analysis of cancer. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6388–6393. doi: 10.1073/pnas.1219651110
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int. J. Cancer* 136, E359–E386. doi: 10.1002/ijc.29210
- Friese, C. R., Harrison, J. M., Janz, N. K., Jaggi, R., Morrow, M., Li, Y., et al. (2017). Treatment-associated toxicities reported by patients with early-stage invasive breast cancer. *Cancer* 123, 1925–1934. doi: 10.1002/cncr.30547
- García-Campos, M. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Pathway analysis: state of the art. *Front. Physiol.* 6:383. doi: 10.3389/fphys.2015.00383
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi: 10.1186/gb-2004-5-10-r80
- Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi: 10.1126/science.1132939

3. RESULTADOS.

- Lee, G. (2017). *Design and Fabrication of Multilayer Structures in Dynamic Sensing and Transparent Nanocomposite Scintillators for High Energy Detection*. Ph.D. thesis, Georgia Institute of Technology.
- Liu, T., Yacoub, R., Taliaferro-Smith, L. D., Sun, S.-Y., Graham, T. R., Dolan, R., et al. (2011). Combinatorial effects of lapatinib and rapamycin in triple-negative breast cancer cells. *Mol. Cancer Ther.* 10, 1460–1469. doi: 10.1158/1535-7163.MCT-10-0925
- Livshits, A., Git, A., Fuks, G., Caldas, C., and Domany, E. (2015). Pathway-based personalized analysis of breast cancer expression data. *Mol. Oncol.* 9, 1471–1483. doi: 10.1016/j.molonc.2015.04.006
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27, 1160–1167. doi: 10.1200/JCO.2008.18.1370
- Pirmohamed, M. (2001). Pharmacogenetics and pharmacogenomics. *Br. J. Clin. Pharmacol.* 52, 345–347. doi: 10.1046/j.0306-5251.2001.01498.x
- Prat, A., Carey, L. A., Adamo, B., Vidal, M., Taberero, J., Cortés, J., et al. (2014). Molecular features and survival outcomes of the intrinsic subtypes within her2-positive breast cancer. *J. Natl. Cancer Inst.* 106:dju152. doi: 10.1093/jnci/dju152
- Prat, A., Pineda, E., Adamo, B., Galván, P., Fernández, A., Gaba, L., et al. (2015). Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* 24, S26–S35. doi: 10.1016/j.breast.2015.07.008
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rossini, A. J., Heiberger, R. M., Sparapani, R. A., Maechler, M., and Hornik, K. (2004). Emac speaks statistics: a multiplatform, multipackage development environment for statistical analysis. *J. Comput. Graph. Stat.* 13, 247–261. doi: 10.1198/1061860042985
- Senkus, E., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rutgers, E., et al. (2015). Primary breast cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 26(Suppl. 5), v8–v30. doi: 10.1093/annonc/mdv298
- Wagner, A. H., Coffman, A. C., Ainscough, B. J., Spies, N. C., Skidmore, Z. L., Campbell, K. M., et al. et al. (2015). DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* 44, D1036–D1044. doi: 10.1093/nar/gkv1165
- Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., et al. (2010). Mappedsplice: accurate mapping of rna-seq reads for splice junction discovery. *Nucleic Acids Res.* 38:e178. doi: 10.1093/nar/gkq622
- Wang, L. (2010). Pharmacogenomics: a systems approach. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2, 3–22. doi: 10.1002/wsbm.42
- Wilkerson, M. D., Cabanski, C. R., Sun, W., Hoadley, K. A., Walter, V., Mose, L. E., et al. (2014). Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* 42:e107. doi: 10.1093/nar/gku489
- Yu, A. F., and Jones, L. W. (2016). Breast cancer treatment-associated cardiovascular toxicity and effects of exercise countermeasures. *Cardio Oncol.* 2:1. doi: 10.1186/s40959-016-0011-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Mejía-Pedroza, Espinal-Enríquez and Hernández-Lemus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Resultados adicionales

Los resultados descritos en los siguientes capítulos incorporan extensiones de los resultados principales y algunos resultados adicionales.

4.1. Extensiones a medicina personalizada

4.1.0.1. Expresión genética individual

La importancia de tener una métrica por paciente y no sólo por subtipo se pone de relieve en la figura 4.1. En ella podemos observar la dispersión de los valores de expresión en cáncer de mama del subtipo basal y tejido sano para los blancos farmacológicos elegidos como prueba de concepto para el subtipo basal en los dos pacientes MB.5387 y MB.3058. Cada uno de estos blancos farmacológicos pasaron los criterios de selección. Pertenecen a una vía molecular de las más desreguladas consistentemente en el grupo de pacientes con cáncer de mama del subtipo basal. Por sí mismos los genes de dichos blancos están diferencialmente expresados en relación al tejido sano. Además, el tipo de relación entre el fármaco y su blanco conlleva a un estado de homeostasis tal como fue definida en nuestra publicación [47].

En virtud de lo anterior se convierten en candidatos para la reposición de fármacos para el subtipo basal. Sin embargo, si observamos la distribución de valores de expresión podemos ver que para algunos pacientes su valor de expresión cabría muy bien dentro del rango del grupo normal. Más aún algunos pacientes tienen una expresión génica del blanco aún por debajo de la media del grupo control, tal es el caso de los genes **SPHK1** y **SQLE** en la figura 4.1. Donde tal vez los pacientes con esos valores de expresión atípico no se vean beneficiados con la terapia dirigida contra tales blancos, e incluso pudiendo aumentar el riesgo a efectos secundarios. De ahí la valía de perfilarnos hacia un enfoque personalizado.

4. RESULTADOS ADICIONALES

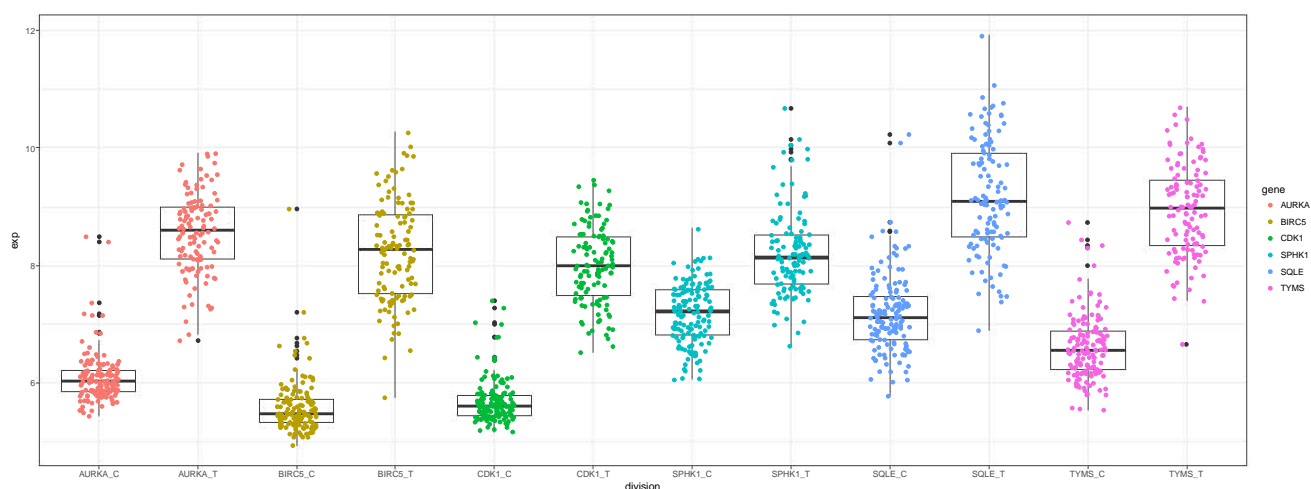


Figura 4.1: En la imagen vemos los niveles de expresión para los blancos farmacológicos ejemplificados para los pacientes con subtipo basal, podemos

4.1.0.2. Activación de vías por individuo

En el caso de las vías moleculares seleccionadas [47] para el subtipo basal, vale la pena reconocer que hay un menos solapamiento en las distribuciones (es decir, están más diferenciadas) en comparación a lo que veíamos en las distribuciones de expresión. Tal vez esto se debe a que la vía molecular en su conjunto tiene más robustez para resistir a las perturbaciones que los genes individuales, poniendo de relieve subyacentes mecanismos de compensación. No obstante siguen siendo importantes las diferencias entre los estados de activación entre el grupo del subtipo basal y controles ???. También podemos observar que existen un amplio rango en el estado de activación de la vía molecular entre cada uno de las muestras que pertenecen a algún fenotipo. Redundando así en la importancia de distinguir entre individuos.

4.2. Lista de blancos subtipo basal

Algunos de los blancos farmacológicos identificados fueron expuestos en el trabajo principal, no obstante la lista de medicamentos candidatos a reposición para el subtipo basal (de especial importancia debido a su mal pronóstico y déficit de opciones terapéuticas) no se expuso a detalle. De tal manera que los blancos farmacológicos que aquí se enuncian los compuestos que sobrepasaron los filtros y su detalle. Los filtros que debía pasar un blanco farmacológico para su selección fueron los siguientes, que el blanco éste diferencialmente expresado, que pertenezca a una vía molecular de las más desreguladas para ese subtipo en ambas bases de datos (METABRIC y TCGA). También en la tabla podemos ver el tipo de interacción entre el fármaco y su diana. Y si este tipo de



Figura 4.2: Aunque las distribuciones están más diferenciadas para los valores de desregulación de vías en comparación a los niveles de expresión. Aún así hay pacientes atípicos cuyos valores individuales bien podrían pasar por controles. Especial diferenciación muestra la vía molecular de la apoptosis cuyos valores grupales son bastante diferentes entre los fenotipos.

4. RESULTADOS ADICIONALES

interacción lleva a una relación de homeostasis o no en relación al fenotipo normal.

Tabla 4.1: Lista de blancos del subtipo basal

Blanco	Vía	LogFC	interacción	Homeostasis
ACSL1	hsa01100	-2.6399697772	n/a	0
ACSS2	hsa01100	-1.6103586762	n/a	0
ADH1B	hsa01100	-5.9702762301	n/a	0
AKR1C3	hsa01100	-1.8076181853	n/a	0
ALAD	hsa01100	-1.332487215	inducer	homeostasis
ALAD	hsa01100	-1.332487215	inhibitor	anti-homeostasis
ALAD	hsa01100	-1.332487215	n/a	0
ALDH1A1	hsa01100	-2.6847732817	n/a	0
ALDH2	hsa01100	-2.363885245	inhibitor	anti-homeostasis
ALDH2	hsa01100	-2.363885245	n/a	0
AOC3	hsa01100	-4.7780935025	inhibitor	anti-homeostasis
AOC3	hsa01100	-4.7780935025	n/a	0
AOX1	hsa01100	-3.1226576376	inhibitor	anti-homeostasis
AOX1	hsa01100	-3.1226576376	n/a	0
ASP	hsa01100	-4.7135988046	n/a	0
AURKA	hsa04114	3.9277072	inhibitor	homeostasis
AURKA	hsa04114	3.9277072	n/a	0
BHMT2	hsa01100	-3.2411154424	product of	0
BIRC5	hsa04210	4.712990326	inhibitor	homeostasis
BIRC5	hsa04210	4.712990326	n/a	0
CCNA2	hsa04110	3.9197713418	n/a	0
CCNE1	hsa04110	4.6613479126	n/a	0
CCNE1	hsa04114	4.6613479126	n/a	0
CDC25B	hsa04110	2.0561005719	n/a	0
CDK1	hsa04110	3.9382466189	binder	0
CDK1	hsa04110	3.9382466189	inhibitor	homeostasis
CDK1	hsa04110	3.9382466189	n/a	0
CDK1	hsa04114	3.9382466189	binder	0
CDK1	hsa04114	3.9382466189	inhibitor	homeostasis
CDK1	hsa04114	3.9382466189	n/a	0
CDKN2A	hsa04110	4.5553658821	n/a	0
CHEK1	hsa04110	2.525297321	inhibitor	homeostasis
CHEK1	hsa04110	2.525297321	n/a	0
CKMT2	hsa01100	-2.8101847205	ligand	0
DNMT1	hsa01100	1.6341976311	inhibitor	homeostasis
DNMT1	hsa01100	1.6341976311	n/a	0
DNMT1	hsa01100	1.6341976311	other/unknown	0
GMDS	hsa01100	1.5587723331	n/a	0
HPRT1	hsa00983	1.6234754998	inhibitor	homeostasis

4.2 Lista de blancos subtipo basal

HPRT1	hsa00983	1.6234754998	n/a	0
HPRT1	hsa01100	1.6234754998	inhibitor	homeostasis
HPRT1	hsa01100	1.6234754998	n/a	0
HSD17B6	hsa01100	2.8562771561	n/a	0
IDH2	hsa01100	1.5363130184	n/a	0
IDO1	hsa01100	4.0463570909	binder	0
IGF1	hsa04114	-2.9950161335	n/a	0
IL4I1	hsa01100	3.5054879038	n/a	0
IRS1	hsa04722	-1.9408248541	n/a	0
ITPR1	hsa04114	-2.7424945791	antagonist	anti-homeostasis
ITPR1	hsa04114	-2.7424945791	n/a	0
ITPR1	hsa04210	-2.7424945791	antagonist	anti-homeostasis
ITPR1	hsa04210	-2.7424945791	n/a	0
JUN	hsa04012	-1.6924933113	n/a	0
JUN	hsa04012	-1.6924933113	other/unknown	0
JUN	hsa04210	-1.6924933113	n/a	0
JUN	hsa04210	-1.6924933113	other/unknown	0
JUN	hsa04722	-1.6924933113	n/a	0
JUN	hsa04722	-1.6924933113	other/unknown	0
MAOA	hsa01100	-4.6610796109	antagonist	anti-homeostasis
MAOA	hsa01100	-4.6610796109	inhibitor	anti-homeostasis
MAOA	hsa01100	-4.6610796109	n/a	0
MAPK10	hsa04012	-2.0887390726	n/a	0
MAPK10	hsa04210	-2.0887390726	n/a	0
MAPK10	hsa04722	-2.0887390726	n/a	0
MGLL	hsa01100	-3.0047051321	n/a	0
NME1	hsa01100	1.9652960981	n/a	0
PAFAH1B3	hsa01100	2.2032848947	n/a	0
PCK1	hsa01100	-6.1673178228	n/a	0
PDE2A	hsa04740	-3.9629036985	inhibitor	anti-homeostasis
PDE2A	hsa04740	-3.9629036985	n/a	0
PIK3R1	hsa04012	-1.7138426025	inhibitor	anti-homeostasis
PIK3R1	hsa04012	-1.7138426025	n/a	0
PIK3R1	hsa04210	-1.7138426025	inhibitor	anti-homeostasis
PIK3R1	hsa04210	-1.7138426025	n/a	0
PIK3R1	hsa04722	-1.7138426025	inhibitor	anti-homeostasis
PIK3R1	hsa04722	-1.7138426025	n/a	0
PNP	hsa01100	1.6285268228	inducer	anti-homeostasis
PNP	hsa01100	1.6285268228	n/a	0
PTDSS1	hsa01100	1.2753789687	n/a	0
PTGIS	hsa01100	-1.7089122052	inducer	homeostasis
PTGIS	hsa01100	-1.7089122052	inhibitor	anti-homeostasis
PTGIS	hsa01100	-1.7089122052	n/a	0

4. RESULTADOS ADICIONALES

PYCR1	hsa01100	2.6457710984	n/a	0
SHMT2	hsa01100	1.5621290087	cofactor	anti-homeostasis
SHMT2	hsa01100	1.5621290087	n/a	0
SPHK1	hsa01100	1.7612685354	n/a	0
SQLE	hsa01100	2.7036726124	inhibitor	homeostasis
TK1	hsa00983	3.7151463326	n/a	0
TK1	hsa01100	3.7151463326	n/a	0
TYMS	hsa01100	3.2689937821	inhibitor	homeostasis
TYMS	hsa01100	3.2689937821	n/a	0
UCK2	hsa00983	1.9349024798	n/a	0
UCK2	hsa01100	1.9349024798	n/a	0

4.3. Lista de blancos por subtipo

En sintonía con los criterios de selección previamente expuestos, podemos seleccionar los blancos para cada uno de los subtipos moleculares. Mismos que podemos ver en la tabla 4.2.

Ahí podemos apreciar que para el subtipo Her2 se encuentra el blanco ERBB2, como era de esperarse pues es el pilar del tratamiento actual para dicho subtipo molecular y debido a su estrecha relación con esta molécula el subtipo recibe su nomenclatura. Para ese mismo subtipo podemos observar que hay más blancos propuestos esperando que alguno de ellos pudiera ser útil para estos pacientes. Le siguen los subtipos Luminal A y Luminal B, interesante notar que ambos comparten algunas de las moléculas farmacológicas propuestas de este estudio, no es una gran sorpresa ya que estos dos subtipos luminales guardan una relación más cercana en comparación al resto, no obstante hay compuestos que diferencian las listas. Por último pero no menos importante podemos ver que para el subtipo Basal resultó el mayor número de moléculas farmacológicas propuestas, lo cual es un dato alentador para aquellos que se encuentran en la búsqueda de nuevos posibles tratamientos para esta enfermedad. Podemos encontrar moléculas relacionadas al control del ciclo celular como **CDK1** o **CDKN2A**, a la síntesis de nucleótidos (**TYMS**), factores de transcripción como **JUN**, funciones metabólicas como la aldehído deshidrogenasa 2 **ALDH2**, entre otros.

Tabla 4.2: Lista de blancos por subtipo

Basal	Her2	LumA	LumB
MGLL	BIRC5	GHR	FYN
CKMT2	JUN	AOX1	JUN
ADH1B	MAP3K8	LEP	PIK3R1
DNMT1	CXCL10	PRLR	PPARG
ALAD	SPP1	S1PR1	MAPK10

ALDH1A1	EDNRB	EGFR	CCND1
ALDH2	KIT	MAPK10	CCNA2
ACSL1	ERBB2 ¹	SGK1	CDK1
BHMT2	MYC	PPARG	CDH5
IL4I1	MAOA	CDH5	MMP9
GMDS	MAOB	MMP9	PLA2G2A
AOX1	SHMT2	CXCL10	ETNK2
HPRT1	IGF1	PTGS2	ACSL1
IDH2	IRS1	COL1A1	LEP
IDO1	NR3C2	COMP	CD36
MAOA		SPP1	ADH1B
ASPA		CD36	ALDH2
NME1		MGLL	SQLE
PNP		ALDH2	
PAFAH1B3		LPL	
PCK1		ADH1B	
ACSS2		ACSL1	
PTGIS			
PYCR1			
SHMT2			
SQLE			
TK1			
TYMS			
UCK2			
HSD17B6			
AOC3			
AKR1C3			
SPHK1			
PTDSS1			
PDE2A			
BIRC5			
ITPR1			
JUN			
PIK3R1			
MAPK10			
IGF1			
AURKA			
CCNE1			
CDK1			
CDKN2A			

¹ERBB2 o HER2, es el receptor del factor de crecimiento epidérmico 2 pieza central en el tratamiento del subtipo molecular con el mismo nombre

4. RESULTADOS ADICIONALES

CHEK1
CCNA2
CDC25B
IRS1

5.1. Expresión Diferencial de genes

El nivel de expresión de un gen es una estimación de la cantidad de transcrito presente en la célula con algún fenotipo dado. Cuando se cuentan con datos provenientes de tecnologías de alto rendimiento usualmente se desea conocer si los transcritos de un gen (la expresión de un gen) se encuentran de manera significativa en número mayor o menor, en un grupo de muestras con respecto a otro; por ejemplo casos y controles). Para esta tarea existe la técnica de **análisis de expresión diferencial**.

A continuación profundizaremos en el funcionamiento de esta técnica. El abordaje más común en el análisis comparativo de datos transcriptómicos es evaluar la hipótesis nula de que el logaritmo de la razón de cambio (“logarithmic fold change” o **LFC**) del transcrito entre casos y controles es exactamente cero. Es decir que el la transcripción del gen no se ve afectada. A menudo el propósito de un análisis de expresión diferencial es obtener una lista de genes ordenada de acuerdo al “valor p” una vez ajustado por contrastes múltiples. Sin embargo, diferencias pequeñas incluso si son estadísticamente muy significativas pueden no ser los candidatos más interesantes para investigaciones subsecuentes. Por otra parte, ordenar por LFC es complicado por lo ruidoso que son los estimados de LFC en los genes que exhiben bajos conteos. Es importante recordar que el número de genes que se pueden encontrar de manera significativa diferencialmente expresados depende del número de muestras y otros aspectos del diseño experimental, así como de la tecnología utilizada para la medición.

5.1.1. Implementación

Actualmente ya están implementados de computacionalmente varios métodos para evaluar la expresión diferencial en los datos provenientes de tecnologías de alto rendimiento. Varias de estas técnicas para evaluar la expresión diferencial reúnen la infor-

mación entre genes para aprovechar la suposición de que la similaridad de las varianzas (dispersión) de genes diferentes medidos en el mismo experimento[48]. *edgeR* [49] modela el estimado de la dispersión para cada gen a través de un estimado común para todos los genes, o a través de un estimado local a partir de genes con un valor de expresión similar, usando una verosimilitud condicional ponderada. *DESeq* [50] detecta y corrige los estimados de dispersión que son muy bajos mediante la modelación de la dependencia de la dispersión respecto al nivel de expresión promedio del conjunto de muestras. *BBseq* [51] modela la dispersión sobre la media, donde se usa la desviación absoluta de la media de los estimados de dispersión para minimizar la influencia de los valores atípicos. *DSS*[52] usa un abordaje Bayesiano para proveer un estimado de la dispersión de genes individuales que aportan para la heterogeneidad de los valores de dispersión en otros genes. *BaySeq*[53] y *ShrinkBayes* [54] estiman las probabilidades apriori en un modelo Bayesiano en todos los genes, para luego proveer las probabilidades a posteriori o "false discovery rate" (**FDRs**) de la expresión diferencial. Para el caso de *limma* [55], la idea central es el ajuste de un modelo lineal para cada gen para evaluar su expresión diferencial. Para esto, se toma la matriz de expresión donde cada renglón corresponde a un gen (u otro ente medido) y cada columna a una muestra. Se aplica un modelo lineal para cada renglón, con los coeficientes de regresión y los errores estándar se estiman directamente a través de las comparaciones de interés o mediante contrastes.

5.2. Pathifier

Para identificar las vías más desreguladas usamos *Pathifier*[56]. *Pathifier* es un relevante algoritmo de enriquecimiento funcional. El cual integra datos experimentales de tecnologías de alto rendimiento con información de vías moleculares provenientes de bases de datos biológicas. *Pathifier* provee como resultado un valor de desregulación por vía para cada muestra. A este valor de desregulación se le llama "Pathway Dereglulation Score." (**PDS**). García Campos et al. menciona que de esta manera se obtiene una representación biológica compacta de la vía, donde esta representación de las muestras centradas en una perspectiva de vías moleculares puede ser útil para facilitar los hallazgos de fenómenos biológicos.

Haciendo una breve recapitulación de su funcionamiento, *Pathifier* evalúa los datos provenientes de experimentos de alto rendimiento una vía molecular a la vez mediante un análisis de componentes principales. A continuación genera una nube de puntos al localizar cada muestra en un sistema de coordenadas donde cada eje representa un elemento de la vía molecular. Seguidamente, se calcula una curva de ajuste mediante el algoritmo de Hastie y Stuetzle[57]. Donde se asigna el centroide del grupo control como el origen de la curva. Finalmente se obtiene la proyección de los puntos sobre la curva y esa distancia sobre la curva entre la proyección del punto y el origen es el "Patway Dereglulation Score" (**PDS**) asignado a esa muestra para esa vía[56].

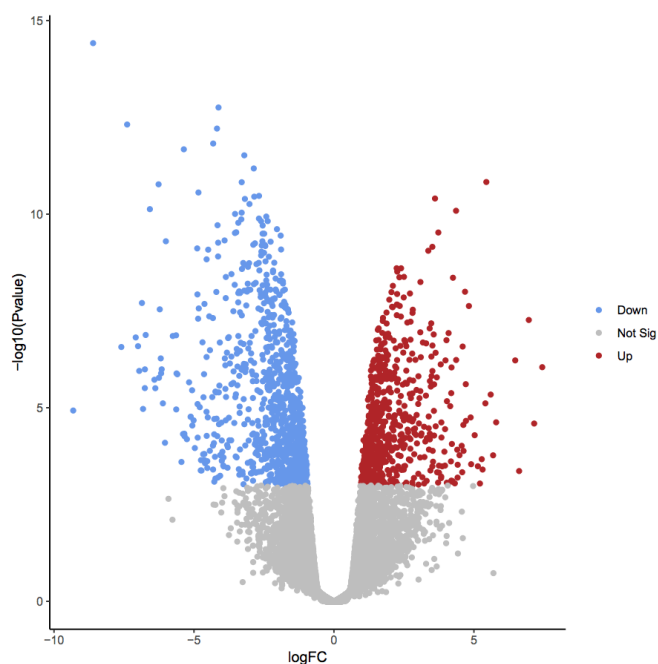


Figura 5.1: En la imagen se observa un "gráfico de volcán" (*volcano-plot*). Ésta es una visualización típica para evaluar la expresión diferencial. En el eje de las abscisas tenemos el logaritmo de la razón de cambio (**LFC**) entre dos condiciones. En el eje de las ordenadas un parámetro que nos habla que tan significativo es ese cambio. Para ambas variables se suelen asignar puntos de corte de acuerdo a los objetivos del estudio. En rojo los genes sobre expresados de manera significativa y en azul los significativamente subexpresados.

Este método se ha usado previamente para observar los valores de desregulación en muestras de cáncer de mama [58] [59]. Es importante recordar que Pathifier asigna un puntaje de desregulación por muestra, lo cual fue una característica fundamental al elegir este método. Ya que por un lado es posible evaluar el nivel de desregulación de vías por paciente en cada subtipo de cáncer de mama facilitando extender nuestra propuesta hacia la medicina personalizada. Por otro lado el mismo autor del método propone un ajuste para obtener un puntaje general del nivel de desregulación de la vía en un grupo de muestras completo [58], ambas características nos permitieron gran versatilidad.

5.3. PAM50

El algoritmo de PAM50 [60] (Partitioning around Medoids) está basado en el algoritmo de k-medias (k-medoids) el cual es un algoritmo de agrupamiento (relacionado al agrupamiento por k-means) para la partición de un conjunto de datos en k grupos (o clusters). Para luego considerar los niveles de expresión de 50 genes el algoritmo y compararlos mediante la medida de distancia Rho de Spearman contra cada uno de los cinco centroides definidos, los cuales representan los subtipos **Luminal A**, **Luminal B**, “**Basal-like**”, **Her2-enriquecido** así como un subtipo “**Normal**”.

Vale la pena mencionar que éste método de clasificación esta actualmente incorporado a las guías de práctica clínica internacionales [5] para el manejo de pacientes con cáncer de mama. Donde constituye una piedra angular para la decisión del esquema terapéutico.

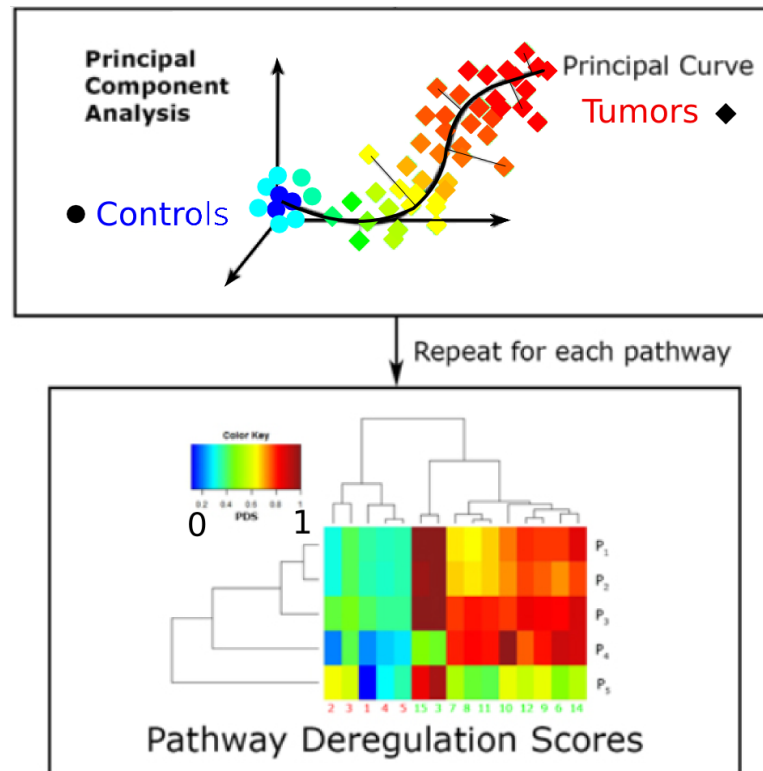


Figura 5.2: En el cuadro superior, la nube de puntos con su curva de ajuste donde cada eje corresponde a un elemento de la vía. El "Pathway Deregulation Score" (**PDS**) es la distancia sobre la curva entre la proyección del punto y el origen. Pathifier repite dicho algoritmo para cada vía. En el cuadro inferior, los renglones corresponden a vías y las columnas a pacientes el color de cada celda esta relacionado con el PDS asignado a esa vía para ese paciente (muestra). El código de colores va en un rango de cero a uno, donde el cero corresponde con el origen de la curva (azul) y uno al punto más distante (rojo).

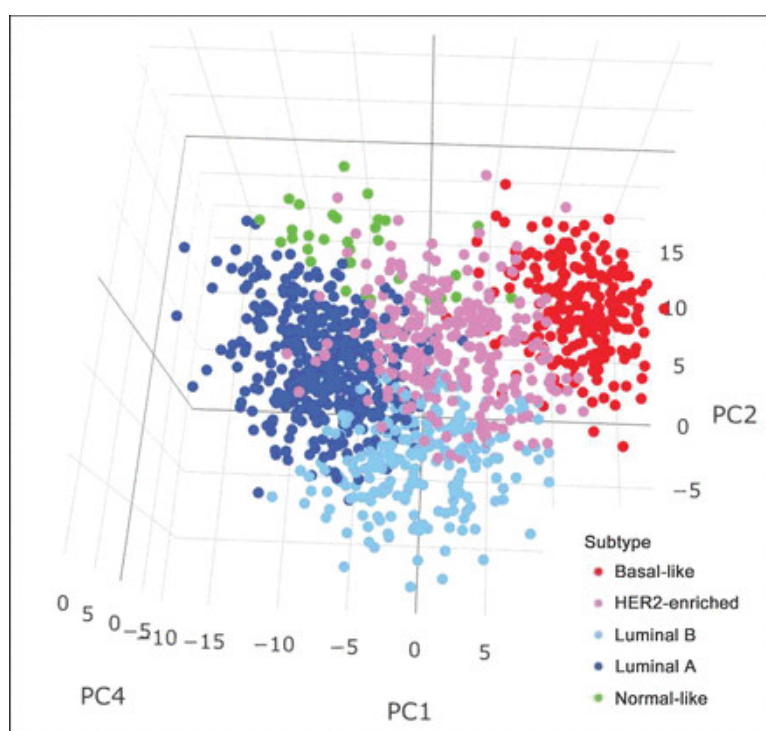


Figura 5.3: En esta visualización se muestra mediante un análisis de componentes principales los subtipos intrínsecos de cáncer de mama definidos por el algoritmo PAM50

Discusión y conclusiones

El proceso tradicional para el desarrollo de un fármaco es un proceso costoso y tardado. Y el caso del cáncer de mama no es la excepción. De ahí que el abordaje de reposicionamiento de fármacos sea sumamente atractivo para la industria y para el beneficio de los pacientes. Con el reposicionamiento de fármacos se ahorran años de desarrollo e inversión en etapas preclínicas y algunas fases clínicas, acortando así la brecha para traer un nuevo medicamento al ambiente clínico.

La presente metodología se circunscribe en el enfoque actual de la biología de sistemas que es capaz de integrar grandes datos provenientes de tecnologías de alto rendimiento con las que contamos hoy en día y con las cuales tenemos el acceso a enormes volúmenes de datos biológicos. Este volumen de datos representa un salto cualitativo que pretende ser un empuje disruptivo para los anteriores paradigmas y herramientas en los cuales estaba basada la ciencia médica y el desarrollo de fármacos.

En este trabajo hemos propuesto una metodología computacional con la cual es posible detectar propiedades relevantes en distintos niveles de organización biológica a partir de datos experimentales de alto rendimiento. Estas propiedades nos ayudaron a identificar blancos terapéuticos potencialmente útiles para enfermedades que actualmente sufren de una deficiencia de opciones farmacológicas. Vale la pena mencionar que los fármacos estudiados actualmente ya están aprobados y disponibles en el mercado por lo tanto se enmarca en un esfuerzo de reposicionamiento de fármacos.

Fuimos capaces de alcanzar los objetivos de investigación propuestos, encontramos las vías más activas en cada subtipo de cáncer de mama. Identificamos cuales de esas vías moleculares son susceptibles de intervención farmacológica ¹. Clasificamos y filtramos los medicamentos de las vías susceptibles de intervención farmacológica, de acuerdo a información disponible. Entre ella el estatus de expresión diferencial del blanco, el sta-

¹Aquellas vías a las que actualmente se conoce un fármaco que incide directamente sobre un elemento de ella

6. DISCUSIÓN Y CONCLUSIONES

tus regulatorio de de la vía, el tipo de interacción entre el fármaco y su blanco, así como el rol biológico de la diana.

Sin lugar a dudas, para el progreso de la biología de sistemas y la bioinformática ha tenido un aporte vital el estudio de sistemas biológicos, no obstante este avance no sería posible sin un marco matemático que sustente los algoritmos y más aún que no solo proporcione nuevas técnicas sino formas novedosas de comprender a los sistemas biológicos.

La estrategia presentada en este trabajo se se basa directamente en los datos experimentales de biopsias de pacientes con cáncer de mama. Otros algoritmos se sustentan en líneas celulares sin embargo éstas han sido modificadas para hacer posible su manejo y supervivencia. Sometidas a un proceso para evitar su senescencia (inmortalización) después del cual en muchas ocasiones muestran alteraciones nucleares; son separadas del contexto tisular en el que normalmente se nutren, comunican y cumplen sus funciones celulares; sometidas a antibióticos y un entorno estrictamente controlado. De ahí que estos modelos biológicos no necesariamente reflejen de manera fiel la dinámica de las células que pretenden modelar.

Sin embargo estas tecnologías donde provienen los datos de alto rendimiento tienen sus limitaciones desde que extraen el RNA de porciones de tejidos. De ahí que en la biopsia exista una heterogeneidad celular donde no todas las células son cancerígenas. Más aún, aún cuando todas lo fueran, distintas células cancerígenas presentan estados diferentes en sus ciclo celular o incluso en su cadena de DNA debido a su inestabilidad genómica. La técnica de secuenciación de una sola célula “Single cell sequencing” puede en principio superar este inconveniente, sin embargo al igual que las anteriores no esta exenta de ruido inherente a la medición y limitación en la sensibilidad.

También concerniente al tema tecnológico una gran área de oportunidad es la integración de datos provenientes de más de una ómica (“multiomics”). De esta manera podríamos tener acceso a mediciones de diferentes niveles de organización celular, contrastar, complementar y robustecer nuestros hallazgos. Por ejemplo una vez que se ha identificado un gen subexpresado, sería complementario conocer su estado de metilación, y ¿Qué hay de su niveles de proteicos?, sin mencionar la diversidad de moléculas que pueden influir en su regulación (micro RNAs, RNA largos no codificantes, entre otros). ¿Y cómo se orquestan todas ellas para incidir en un fenotipo?. Así pues, el acceso a diferentes planos visuales un mismo fenómeno pueden brindarnos un mejor entendimiento de las enfermedades y como combatirlas.

Las bases de datos biológicas y farmacológicas como cualquier otro recurso de anotación, estan limitadas por el conocimiento disponible hasta el momento.

En cuanto al método de clasificación si bien es cierto que existen muchos, se optó por el más reciente incorporado en las guías de práctica clínica, ha mostrado su mejorar

la precisión en el pronóstico y es ampliamente abordado por la comunidad científica. también este trabajo sienta un precedente en el entendido de que el objetivo a largo plazo es acercarnos a la medicina personalizada y de acercarnos lo suficiente a tal meta, la clasificación de enfermedades en subgrupos quedarían en un segundo plano.

Este trabajo se distingue de los demás de los que le antecedieron en reposicionamiento de fármacos en cáncer de mama porque pretende usar cambalmente distintos niveles de organización biológica contenida dentro de los datos de alto rendimiento. Por ejemplo, usar la información del estado de las vías en las muestras no solo como definiciones en una base de datos. Emplea los datos experimentales más copiosos y de alta calidad disponibles hasta el momento de su realización.

Importante es recordar que al final del día en cuanto a que la farmacología es una ciencia experimental, de tal manera que solo el contraste con la realidad podrá definir cuales son las propuestas farmacológicas exitosas para un entorno concreto. Así pues, la validación en modelos biológicos o en el ambiente clínico es un paso necesario.

Perspectivas

7.1. Integración de la base de datos con otros recursos

La base de datos creada en el presente trabajo (con más de 80 mil relaciones vía-diana-fármaco) por sí misma es un recurso valioso de consulta para la búsqueda de posibles blancos para el reposicionamiento de fármacos, los criterios de filtrado pueden variar de acuerdo a las necesidades del quien la utilice. No obstante reste recursos se robustece y se espera que los hallazgos sean más precisos si se incorporan otras fuentes informativas. Por ejemplo:

- Datos sobre la topología de la red ¹ de interacciones local de cada una de las dianas, por ejemplo si la dina pertenece a un módulo identificado por topología y éste tiene alguna función atribuida.
- Información relativa al estado del fármaco en el mercado, su toxicidad, vía de administración, efetos adversos, incluso su costo puede ser información útil al momento de seleccionar.
- La incorporación de datos sobre el estado de las proteínas (proteoma), metilación, entre otras ómicas.
- Al incorporar nuevas ómicas pueden surgir propiedades que pueden ser determinadas con una ómica en particular.
- Cada propiedad en principio puede ser incluida como un criterio de priorización al momento de seleccionar los blancos farmacológicos idóneos.

¹construida sobre los datos de expresión génica

7.2. validación

Como ya se ha mencionado el proceso de validación de las dinas propuestas es un paso necesario para el ejercicio del fármaco en su ambiente terapéutico. He aquí algunas propuestas.

7.2.1. validación biológica

Un primer entorno más o menos accesible para validar los hallazgos sería mediante el uso de líneas celulares, para las cuales ya hay una asociada a cada subtipo de cáncer de mama. Otra posibilidad de mayor impacto sería la evaluación del desempeño del fármaco en un ensayo clínico, que por su naturaleza de reposicionamiento de fármacos se podría iniciar en principio con uno de Fase 2, ahorrando cuantiosos recursos y tiempo valioso.

7.2.2. validación sintética

La validación biológica tiene requerimientos como costos elevados, necesidad de infraestructura, consideraciones éticas, entre otros. Por ello en sintonía con la era de los grandes datos que hoy en día replantean la manera como se genera el conocimiento médico científico. Podemos imaginarnos el uso de grandes volúmenes de información de respuesta clínica en ocasiones llamada “del mundo real” para analizar retrospectivamente la respuesta de pacientes que recibieron el fármaco por algún motivo tangencial a su enfermedad principal. Por ejemplo, aquellos pacientes con cáncer de mama que recibieron Beta-agonistas para el tratamiento del asma, o antimicóticos por una eventual infección fúngica. Todos estos datos si son apropiados en volumen y precisión pueden analizarse para evaluar retrospectivamente el desempeño del fármaco. La desventaja es que se esta limitado a la información disponible y a la retrospectividad del estudio. Dentro de las ventajas esta su relativamente bajo costo, evaluación en un entorno de uso real con comorbilidades, condiciones especiales (embarazo, vejez), los cuales no suelen encontrarse en los estrictos criterios de inclusión en los que suelen desenvolverse los ensayos clínicos. Y explotando la ventaja que estos datos en un futuro, serán más abundantes.

Bibliografía

- [1] D. Hanahan and R. A. Weinberg, “Hallmarks of cancer: the next generation,” *cell*, vol. 144, no. 5, pp. 646–674, 2011. [1](#)
- [2] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, and F. Bray, “Globocan 2012 v1. 0, cancer incidence and mortality worldwide: Iarc cancerbase no. 11. 2013,” *Available from: globocan. iarc. fr*, 2014. [1](#)
- [3] K. Polyak, “Heterogeneity in breast cancer,” *The Journal of clinical investigation*, vol. 121, no. 10, pp. 3786–3788, 2011. [2](#)
- [4] A. Prat, L. A. Carey, B. Adamo, M. Vidal, J. Taberero, J. Cortés, J. S. Parker, C. M. Perou, and J. Baselga, “Molecular features and survival outcomes of the intrinsic subtypes within her2-positive breast cancer,” *Journal of the National Cancer Institute*, vol. 106, no. 8, p. dju152, 2014. [2](#)
- [5] E. Senkus, S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, F. Cardoso, A. Cervantes, G. Pentheroudakis, N. Pavlidis, R. A. Stahel, F. Cardoso, E. Senkus-Konefka, D. Arnold, A. Horwich, C. Sessa, N. Colombo, L. Licitra, C. Buske, M. Ladetto, S. Peters, J. Y. Blay, U. Keilholz, F. Strasser, A. A. Hassan, J. Herrstedt, K. H. McGregor, C. Bramley, J. Lamarre, M. Wallace, A. Galli, and S. Jezdic, “Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up,” *Ann. Oncol.*, vol. 26 Suppl 5, pp. 8–30, Sep 2015. [2](#), [44](#)
- [6] C. M. Perou, T. Sørli, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, *et al.*, “Molecular portraits of human breast tumours,” *Nature*, vol. 406, no. 6797, pp. 747–752, 2000. [2](#), [3](#)
- [7] C. Fan, D. S. Oh, L. Wessels, B. Weigelt, D. S. Nuyten, A. B. Nobel, L. J. Van’t Veer, and C. M. Perou, “Concordance among gene-expression-based predictors for breast cancer,” *New England Journal of Medicine*, vol. 355, no. 6, pp. 560–569, 2006. [2](#)

- [8] Z. Hu, C. Fan, D. S. Oh, J. S. Marron, X. He, B. F. Qaqish, C. Livasy, L. A. Carey, E. Reynolds, L. Dressler, *et al.*, “The molecular portraits of breast tumors are conserved across microarray platforms,” *BMC genomics*, vol. 7, no. 1, p. 96, 2006. [2](#)
- [9] O. Metzger-Filho, Z. Sun, G. Viale, K. N. Price, D. Crivellari, R. D. Snyder, R. D. Gelber, M. Castiglione-Gertsch, A. S. Coates, A. Goldhirsch, *et al.*, “Patterns of recurrence and outcome according to breast cancer subtypes in lymph node–negative disease: Results from international breast cancer study group trials viii and ix,” *Journal of Clinical Oncology*, pp. JCO–2012, 2013. [2](#), [3](#)
- [10] N. D. Arvold, A. G. Taghian, A. Niemierko, R. F. A. Raad, M. Sreedhara, P. L. Nguyen, J. R. Bellon, J. S. Wong, B. L. Smith, and J. R. Harris, “Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy,” *Journal of Clinical Oncology*, vol. 29, no. 29, pp. 3885–3891, 2011. [2](#)
- [11] T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. Van De Rijn, S. S. Jeffrey, *et al.*, “Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 19, pp. 10869–10874, 2001. [2](#)
- [12] R. Haque, S. A. Ahmed, G. Inzhakova, J. Shi, C. Avila, J. Polikoff, L. Bernstein, S. M. Enger, and M. F. Press, “Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades,” *Cancer Epidemiology Biomarkers & Prevention*, vol. 21, no. 10, pp. 1848–1855, 2012. [2](#), [3](#)
- [13] H. J. Burstein, “The distinctive nature of her2-positive breast cancers,” *New England Journal of Medicine*, vol. 353, no. 16, pp. 1652–1654, 2005. [3](#)
- [14] X. R. Yang, J. Chang-Claude, E. L. Goode, F. J. Couch, H. Nevanlinna, R. L. Milne, M. Gaudet, M. K. Schmidt, A. Broeks, A. Cox, *et al.*, “Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the breast cancer association consortium studies,” *Journal of the National Cancer Institute*, vol. 103, no. 3, pp. 250–263, 2011. [3](#)
- [15] E. Senkus, S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, and F. Cardoso, “Primary breast cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up,” *Annals of oncology*, vol. 26, no. suppl_5, pp. v8–v30, 2015. [3](#)
- [16] S. Bayraktar and S. Glück, “Molecularly targeted therapies for metastatic triple-negative breast cancer,” *Breast cancer research and treatment*, vol. 138, no. 1, pp. 21–35, 2013. [3](#)
- [17] K. D. Voduc, M. C. Cheang, S. Tyldesley, K. Gelmon, T. O. Nielsen, and H. Kennecke, “Breast cancer subtypes and the risk of local and regional relapse,” *Journal of Clinical Oncology*, vol. 28, no. 10, pp. 1684–1691, 2010. [3](#)

-
- [18] P. Singha, S. Pandeswara, M. Venkatachalam, and P. Saikumar, “Abstract p2-06-08: Interplay of smad2 and smad3 during tgf- β induced tmepai/pmepa1 mediated triple negative breast cancer cell growth,” *Cancer Research*, vol. 76, no. 4 Supplement, pp. P2–06, 2016. [3](#)
- [19] A. Ring and M. Dowsett, “Mechanisms of tamoxifen resistance,” *Endocrine-related cancer*, vol. 11, no. 4, pp. 643–658, 2004. [3](#)
- [20] I. H. G. S. Consortium *et al.*, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, p. 860, 2001. [3](#)
- [21] I. H. G. S. Consortium *et al.*, “Finishing the euchromatic sequence of the human genome,” *Nature*, vol. 431, no. 7011, p. 931, 2004. [4](#)
- [22] M. A. García-Campos, J. Espinal-Enríquez, and E. Hernández-Lemus, “Pathway analysis: State of the art,” *Frontiers in physiology*, vol. 6, 2015. [5](#), [11](#)
- [23] E. Knierim, B. Lucke, J. M. Schwarz, M. Schuelke, and D. Seelow, “Systematic comparison of three methods for fragmentation of long-range pcr products for next generation sequencing,” *PloS one*, vol. 6, no. 11, p. e28240, 2011. [6](#)
- [24] S. Parekh, C. Ziegenhain, B. Vieth, W. Enard, and I. Hellmann, “The impact of amplification on differential expression analyses by rna-seq,” *Scientific reports*, vol. 6, p. 25533, 2016. [6](#)
- [25] R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, *et al.*, “Human genome sequencing using unchained base reads on self-assembling dna nanoarrays,” *Science*, vol. 327, no. 5961, pp. 78–81, 2010. [7](#)
- [26] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, *et al.*, “An integrated semiconductor device enabling non-optical genome sequencing,” *Nature*, vol. 475, no. 7356, p. 348, 2011. [7](#)
- [27] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley, “Continuous base identification for single-molecule nanopore dna sequencing,” *Nature nanotechnology*, vol. 4, no. 4, p. 265, 2009. [7](#)
- [28] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, “From molecular to modular cell biology,” *Nature*, vol. 402, no. 6761supp, p. C47, 1999. [9](#), [10](#)
- [29] J. Lawrence, “Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes,” *Current opinion in genetics & development*, vol. 9, no. 6, pp. 642–648, 1999. [9](#)
- [30] R. W. Barratt, D. Newmeyer, D. D. Perkins, and L. Garnjobst, “Map construction in neurospora crassa,” in *Advances in genetics*, vol. 6, pp. 1–93, Elsevier, 1954. [9](#)

BIBLIOGRAFÍA

- [31] E. B. Lewis, “The pseudoallelism of white and apricot in *Drosophila melanogaster*,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38, no. 11, p. 953, 1952. [10](#)
- [32] S. K. Ballas, M. R. Kesen, M. F. Goldberg, G. A. Luty, C. Dampier, I. Osunkwo, W. C. Wang, C. Hoppe, W. Hagar, D. S. Darbari, *et al.*, “Beyond the definitions of the phenotypic complications of sickle cell disease: an update on management,” *The Scientific World Journal*, vol. 2012, 2012. [10](#)
- [33] S. Huang, “The practical problems of post-genomic biology,” *Nature biotechnology*, vol. 18, no. 5, p. 471, 2000. [10](#)
- [34] I. Vastrik, P. D’Eustachio, E. Schmidt, G. Joshi-Tope, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, *et al.*, “Reactome: a knowledge base of biologic pathways and processes,” *Genome biology*, vol. 8, no. 3, p. R39, 2007. [11](#)
- [35] S. Gama-Castro, H. Salgado, A. Santos-Zavaleta, D. Ledezma-Tejeida, L. Muniz-Rascado, J. S. García-Sotelo, K. Alquicira-Hernández, I. Martínez-Flores, L. Pannier, J. A. Castro-Mondragón, *et al.*, “RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond,” *Nucleic acids research*, vol. 44, no. D1, pp. D133–D143, 2015. [11](#)
- [36] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, *et al.*, “Gene ontology: tool for the unification of biology,” *Nature genetics*, vol. 25, no. 1, p. 25, 2000. [11](#)
- [37] M. Kanehisa, “A database for post-genome analysis,” *Trends in Genetics*, vol. 13, no. 9, pp. 375–376, 1997. [11](#)
- [38] A. E. Guttmacher and F. S. Collins, “Realizing the promise of genomics in biomedical research,” *JAMA*, vol. 294, pp. 1399–1402, Sep 2005. [12](#)
- [39] C. A. CLARKE, “Pharmacogenetics—a study of inherited variability in the response to drugs,” *J. Pharm. Pharmacol.*, vol. 14(Suppl), pp. 20T–30T, Dec 1962. [12](#)
- [40] L. Wang, “Pharmacogenomics: a systems approach,” *Wiley Interdiscip Rev Syst Biol Med*, vol. 2, no. 1, pp. 3–22, 2010. [12](#)
- [41] T. J. Scott, A. C. O’Connor, A. N. Link, and T. J. Beaulieu, “Economic analysis of opportunities to accelerate Alzheimer’s disease research and development,” *Ann. N. Y. Acad. Sci.*, vol. 1313, pp. 17–34, Apr 2014. [13](#)
- [42] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Williams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed, “Drug repurposing: progress, challenges and recommendations,” *Nat Rev Drug Discov*, Oct 2018. [13](#)

- [43] A. Musa, L. S. Ghoraie, S. D. Zhang, G. Glazko, O. Yli-Harja, M. Dehmer, B. Haibe-Kains, and F. Emmert-Streib, “A review of connectivity map and computational approaches in pharmacogenomics,” *Brief. Bioinformatics*, vol. 19, pp. 506–523, 05 2018. [14](#)
- [44] A. Peyvandipour, N. Saberian, A. Shafi, M. Donato, and S. Draghici, “A novel computational approach for drug repurposing using systems biology,” *Bioinformatics*, vol. 34, pp. 2817–2825, 08 2018. [15](#)
- [45] A. Segura-Cabrera, N. Singh, and K. Komurov, “An integrated network platform for contextual prioritization of drugs and pathways,” *Mol Biosyst*, vol. 11, pp. 2850–2859, Nov 2015. [15](#)
- [46] K. Komurov, M. A. White, and P. T. Ram, “Use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data,” *PLoS Comput. Biol.*, vol. 6, Aug 2010. [15](#)
- [47] R. A. Mejia-Pedroza, J. Espinal-Enriquez, and E. Hernandez-Lemus, “Pathway-Based Drug Repositioning for Breast Cancer Molecular Subtypes,” *Front Pharmacol*, vol. 9, p. 905, 2018. [33](#), [34](#)
- [48] I. Lönnstedt and T. Speed, “Replicated microarray data,” *Statistica sinica*, pp. 31–46, 2002. [42](#)
- [49] D. J. McCarthy, Y. Chen, and G. K. Smyth, “Differential expression analysis of multifactor rna-seq experiments with respect to biological variation,” *Nucleic acids research*, vol. 40, no. 10, pp. 4288–4297, 2012. [42](#)
- [50] S. Anders and W. Huber, “Differential expression analysis for sequence count data,” *Genome biology*, vol. 11, no. 10, p. R106, 2010. [42](#)
- [51] Y. H. Zhou, K. Xia, and F. A. Wright, “A powerful and flexible approach to the analysis of RNA sequence count data,” *Bioinformatics*, vol. 27, pp. 2672–2678, Oct 2011. [42](#)
- [52] H. Wu, C. Wang, and Z. Wu, “A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data,” *Biostatistics*, vol. 14, pp. 232–243, Apr 2013. [42](#)
- [53] T. J. Hardcastle and K. A. Kelly, “baySeq: empirical Bayesian methods for identifying differential expression in sequence count data,” *BMC Bioinformatics*, vol. 11, p. 422, Aug 2010. [42](#)
- [54] M. A. Van De Wiel, G. G. Leday, L. Pardo, H. Rue, A. W. Van Der Vaart, and W. N. Van Wieringen, “Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors,” *Biostatistics*, vol. 14, pp. 113–128, Jan 2013. [42](#)

- [55] G. K. Smyth, “Linear models and empirical bayes methods for assessing differential expression in microarray experiments,” *Stat Appl Genet Mol Biol*, vol. 3, p. Article3, 2004. [42](#)
- [56] Y. Drier and M. A. Yitzhaky, “Package ‘pathifier’,” 2013. [42](#)
- [57] T. Hastie and W. Stuetzle, “Principal curves,” *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989. [42](#)
- [58] A. Livshits, A. Git, G. Fuks, C. Caldas, and E. Domany, “Pathway-based personalized analysis of breast cancer expression data,” *Molecular oncology*, vol. 9, no. 7, pp. 1471–1483, 2015. [43](#)
- [59] D. Drago-García, J. Espinal-Enríquez, and E. Hernández-Lemus, “Network analysis of emt and met micro-rna regulation in breast cancer,” *Scientific reports*, vol. 7, no. 1, p. 13534, 2017. [43](#)
- [60] J. S. Parker, M. Mullins, M. C. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J. F. Quackenbush, I. J. Stijleman, J. Palazzo, J. S. Marron, A. B. Nobel, E. Mardis, T. O. Nielsen, M. J. Ellis, C. M. Perou, and P. S. Bernard, “Supervised risk predictor of breast cancer based on intrinsic subtypes,” *J. Clin. Oncol.*, vol. 27, pp. 1160–1167, Mar 2009. [44](#)