



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE MEDICINA
BIOMEDICINA

MODULACIÓN DE LAS VÍAS DE SEÑALIZACIÓN POR FACTORES DE TRANSCRIPCIÓN
EN CÁNCER DE MAMA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

M. EN C. TAPIA CARRILLO DIANA ELVIRA

TUTOR PRINCIPAL DE TESIS: DR. ENRIQUE HERNÁNDEZ LEMUS
FACULTAD DE MEDICINA, UNAM

COMITÉ TUTOR: DR. LUIS ANTONIO MENDOZA SIERRA
INSTITUTO DE INVESTIGACIONES BIOMEDICAS, UNAM
DR. PEDRO EDUARDO MIRAMONTES VIDAL
FACULTAD DE CIENCIAS, UNAM



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE MEDICINA
BIOMEDICINA

MODULACIÓN DE LAS VÍAS DE SEÑALIZACIÓN POR FACTORES DE TRANSCRIPCIÓN
EN CÁNCER DE MAMA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

M. EN C. TAPIA CARRILLO DIANA ELVIRA

TUTOR PRINCIPAL DE TESIS: DR. ENRIQUE HERNÁNDEZ LEMUS
FACULTAD DE MEDICINA, UNAM

COMITÉ TUTOR: DR. LUIS ANTONIO MENDOZA SIERRA
INSTITUTO DE INVESTIGACIONES BIOMEDICAS, UNAM
DR. PEDRO EDUARDO MIRAMONTES VIDAL
FACULTAD DE CIENCIAS, UNAM

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE MEDICINA
OFICIO CPCB/047/2021
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión virtual del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **26 de octubre de 2020** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **TAPIA CARRILLO DIANA ELVIRA** con número de cuenta **99306010** con la tesis titulada **“MODULACIÓN DE LAS VÍAS DE SEÑALIZACIÓN POR FACTORES DE TRANSCRIPCIÓN EN CÁNCER DE MAMA”**, realizada bajo la dirección del **DR. ENRIQUE HERNÁNDEZ LEMUS**, quedando integrado de la siguiente manera:

Presidente: DR. ARTURO CARLOS II BECERRA BRACHO
Vocal: DRA. VILMA ARACELI MALDONADO LAGUNAS
Secretario: DR. LUIS ANTONIO MENDOZA SIERRA
Suplente: DR. HUGO ANTONIO TOVAR ROMERO
Suplente: DR. GUSTAVO CARLOS MARTÍNEZ MEKLER

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

ATENTAMENTE
“POR MI RAZA HABLARÁ EL ESPÍRITU”
Cd. Universitaria, Cd. Mx., a 20 de enero de 2021

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA



Agradecimientos institucionales

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México (UNAM) a la cual le estoy profundamente agradecida por mi formación académica.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada (No de beca 449453) y los apoyos recibidos para el desarrollo de este proyecto (No de proyecto Conacyt 2115/2018 y 285544/2016)

A mi tutor principal el Dr. Enrique Hernandez Lemus por todo su apoyo, paciencia, buen animo y guía en los momentos difíciles. Por ser un verdadero tutor.

Así mismo agradezco infinitamente a los miembros de mi comité tutor: el Dr. Pedro Eduardo Miramontes Vidal y el Dr. Luis Antonio Mendoza Sierra por enriquecer este trabajo constantemente con sus valiosos comentarios e ideas.

Agradecimientos

Al Instituto Nacional de Medicina Genómica por los recursos federales asignados y a la Cátedra Marcos Moshinsky 2016 en Ciencias Físicas otorgada al Dr. Enrique Hernández Lemus

A diferencia de la licenciatura, pienso que la culminación de un Doctorado requiere de la suma de los esfuerzos de muchas personas, a las cuales estoy profundamente agradecida.

A mi mamá por su apoyo incondicional, por estar a mi lado siempre.

A mi Tío Alfonso. Gracias por enseñarme a nunca darme por vencida y mostrarme que los límites están dentro de nosotros, a siempre a intentar y nunca darme por vencida.

A mis hermanas Erika y Mariana, por siempre estar dispuestas a ayudarme.

A Julián y Valentina, que son el amor en mi vida, por las tardes que les debo y la alegría y energía que irradian y contagian a su paso; por mostrarme lo bueno en mí y lo que me falta por crecer.

A Tadeo por todos los té, las risas, las largas noches estudiando, por estar y dejarme ser, para llegar hasta donde yo quiera.

A David, por todas las noches que perseguimos el último metro para volver a casa.

A Natalia, gracias por todas las noches que me consolaste, por tu amistad incondicional, tu sinceridad y honestidad, porque sé que siempre me has ayudado aún cuando no tenías la posibilidad o no era fácil. Tú definitivamente eres de otro planeta.

El Doctorado requiere de la habilidad de poder generar ideas nuevas, de poder analizar y discutir. En el doctorado tuve grandes maestros por ello quiero agradecer:

A mi tutor, el Dr. Enrique Hernández, por su pasión para enseñar a sus alumnos y mostrarme los límites de cuando terminar una investigación, estar siempre dispuesto a explicar, escuchar y sobretodo a ayudar.

Al Dr. Hugo Tovar, le agradezco su infinita paciencia, por adoptarme cuando recién llegué al laboratorio y no sabía absolutamente nada de la biología de sistemas, por escucharme siempre, ser mi amigo y devolverme el rumbo cuando me he perdido.

A la Dra. Eréndira Avendaño, por compartirme su visión de lo que es la ciencia, por confiar en mí, retarme y mostrarme las cosas que soy capaz de hacer; por su

amistad y guía.

A Rocío Chávez del posgrado de ciencias biológicas, que sin ser parte de su trabajo me ha escuchado y aconsejado en estos 4 años, por siempre buscar la forma de ayudarme y tener siempre una palabra amable y positiva. Rocío ¿Qué habría sido de mi sin ti?

Finalmente, quiero agradecer a mis amigas Yedith, Melissa, Erandi y Paulette por todos aquellos cafés que devolvían el alma cuando las cosas iban mal.

Al leer mis agradecimientos y pensar en ustedes me doy cuenta que estoy rodeada de gente extraordinaria. Soy muy afortunada.

A todos gracias.

Dedicatoria

A la memoria de

mi Tio el Arq. Ignacio Alfonso Carrillo Flores.

Estas en todo y siempre...

“Como no estás experimentado en las cosas del mundo, todas las cosas que tienen algo de dificultad te parecen imposibles.

**- Miguel de Cervantes de Saavedra
"Don Quijote de la Mancha"**

Índice general

Índice de figuras

Índice de tablas

Resumen	1
Abstract	2
1 Introducción	3
2 Objetivos	6
2.1 Objetivo general	6
2.2 Objetivos particulares	7
3 Antecedentes	8
3.1 Cáncer	8
3.2 Vías de señalización en cáncer	9
3.3 Regulación de las vías de señalización	11
3.4 Factores de transcripción	11
3.4.1 Regulación de la actividad de los factores de transcripción . .	12
3.4.2 Factores de transcripción en cáncer	13
3.4.3 Mecanismos de desregulación de los factores de transcripción en cáncer	14
3.4.4 Importancia del estudio de los factores de transcripción en cáncer	15
3.5 Factores transcripcionales maestros	16
3.6 Biología de sistemas y cáncer	17
4 Metodología	20
4.1 Obtención y pre-procesamiento de datos	20
4.2 Análisis de desregulación de vías.	20
4.3 Inferencia de reguladores transcripcionales maestros	22
4.4 Generación de la red de regulones	22
4.5 Generación de la firma molecular de las vías de transducción de señales	24
4.6 Generación del modelo nulo	25
4.7 Inferencia de los reguladores transcripcionales maestros de las vías de transducción de señales.	25

4.8	Enriquecimiento funcional de la red de los 10 primeros regulones con las vías de KEGG	28
4.9	Enriquecimiento de los regulones con procesos biológicos de Gene Ontology	29
5	Resultados y discusión	30
5.1	Análisis de desregulación de vías	30
5.2	Reguladores transcripcionales maestros de las vías de transducción de señales	32
5.3	Características de los 10 primeros reguladores transcripcionales maestros	37
5.4	Enriquecimiento de las redes de regulones en las vías de KEGG	40
5.5	Enriquecimiento de cada regulón en procesos de Gene Ontology	43
5.6	Familia HOXA enriquecida en los regulones	45
6	Conclusiones	48
7	Referencias bibliográficas	50
	Anexo: artículo científico para la obtención del grado	66

Índice de figuras

3.1	Características distintivas del cáncer	9
3.2	Vías de transducción de señales implicadas en cáncer.	10
4.1	Representación gráfica del algoritmo Pathifier.	21
4.2	Cálculo de información mutua.	23
4.3	Flujo de trabajo para obtener los reguladores transcripcionales de las vías de transducción de señales.	27
5.1	Mapa de calor del puntaje de desregulación de vías de las vías de transducción de señales.	31
5.2	Porcentaje de la firma molecular regulada por los reguladores transcripcionales maestros de las vías de señalización.	33
5.3	Los 10 principales reguladores maestros de las vías de transducción de señales.	35
5.4	Red de los 10 primeros regulones de los factores transcripcionales maestros de las vías de señalización.	36
5.5	Genes de la familia HOXA presentes en los 10 primeros Reguladores Maestros de las vías de señalización.	46

Índice de tablas

5.1	Primeros 10 Reguladores Maestros de las vías de transduccion de señales.	34
5.2	Análisis de enriquecimiento sobre las vías de KEGG	42
5.3	Principales enriquecimientos significativos de los procesos biológicos de Gene Ontology por regulón.	44

Resumen

El cáncer es una enfermedad compleja que se caracteriza por la desregulación de un amplio conjunto de genes. La deconvolución de redes de regulación genética han conducido a la identificación de Reguladores Transcripcionales Maestros los cuales controlan los genes más importantes para el establecimiento de un fenotipo específico. En este trabajo identificamos los reguladores transcripcionales maestros que controlan los genes de las vías de transducción de señales. Este nuevo enfoque en el cual nos centramos en un conjunto específico de genes, constituye una adaptación del “Análisis de Reguladores Maestros“ original. Para ello, partimos de un conjunto de genes exclusivos de las vías de señalización, catalogadas como “Signal Transduction Pathway“ en KEGG. Los 10 principales Reguladores Transcripcionales de las vías de señalización en cáncer de mama obtenidos fueron: TSHZ2, HOXA2, MEIS2, HOXA3, HAND2, HOXA5, TBX18, PEG3, GLI2 y CLOCK. El enriquecimiento funcional del conjunto de blancos asociado a cada uno de estos factores de transcripción, mostró una proporción importante de procesos relacionados con la morfogénesis. Los resultados sugieren que, como parte de la regulación aberrante de las vías de señalización en cáncer de mama, vías como las de diferenciación celular, desarrollo del sistema cardiovascular y desarrollo de la vasculatura, podrían estar siendo desreguladas a través de la acción de estos factores de transcripción promoviendo el desarrollo tumoral.

Abstract

Cancer is a complex disease characterized by the dysregulation of a large set of genes. The deconvolution of gene regulatory networks has led to the identification of Transcriptional Master Regulators which control the most important genes for the establishment of a specific phenotype. In this work we identify the Transcriptional Master Regulators that control the genes of the signal transduction pathways. This new approach in which we focus on a specific set of genes constitutes an adaptation of the original “Master Regulator Analysis”. To do this, we use a set of genes unique to signaling pathways, classified as "Signal Transduction Pathways" in KEGG. The main 10 Transcriptional Regulators of the signaling pathways in breast cancer obtained were: TSHZ2, HOXA2, MEIS2, HOXA3, HAND2, HOXA5, TBX18, PEG3, GLI2 and CLOCK. The functional enrichment of the set of targets associated with each of these transcription factors and their regulons, showed an important proportion of processes related to morphogenesis. The results suggest that, as part of the aberrant regulation of the signaling pathways in breast cancer, pathways such as cell differentiation, development of the cardiovascular system and development of the vasculature, could be being deregulated promoting tumor development through the action of these transcription factors.

Capítulo 1

Introducción

El cáncer se origina a partir de células sanas que son reprogramadas y que adquieren características como proliferación ilimitada, renovación, resistencia a la muerte celular, angiogénesis, entre otras, que en su conjunto son conocidas como “Hallmarks del cáncer” ([Hanahan and Weinberg, 2011](#)). Estos procesos resultan de la combinación de interacciones moleculares altamente específicas cuyo estudio puede ser simplificado en forma de vías de señalización. En este contexto es razonable pensar que las células cancerosas hacen uso de vías ya existentes modulándolas de forma aberrante ([Hanahan and Weinberg, 2011](#)). La regulación transcripcional puede jugar un papel importante en la alteración de dichos mecanismos ([Kolch et al., 2015](#)).

Las vías de transducción de señales son mecanismos moleculares intrincados que conducen a las células a detectar señales específicas, y tienen un papel importante en la integración de la información ([Dhanasekaran, 1998](#)). Entre estas acciones, la activación de los factores de transcripción puede modificar, la expresión de genes dentro de estas vías dando lugar a diferentes estados fenotípicos.

Las vías de transducción de señales en sí mismas pueden ser reguladas por medio de la acción de factores de transcripción, estos modulan la transcripción de conjun-

tos de genes que participan dentro de estas (Carroll and Brown, 2006; Laurent et al., 2015; Morgan and Pandha, 2017). Dado que las vías de señalización intervienen en procesos celulares claves en esta enfermedad, su regulación es de interés farmacológico, por lo que entender los mecanismo de regulación de estas vías y sus componentes puede ayudar en la búsqueda de blancos terapéuticos (Martin, 2003).

Una forma de identificar a los factores de transcripción que son importantes para la regulación de las vías de señalización, es a través del análisis de la estructura de una red regulatoria de genes que contiene factores de transcripción y sus blancos, junto con información de sus valores de expresión diferencial; este abordaje permite la identificación de los factores de transcripción que controlan a los genes más diferencialmente expresados y que presumiblemente tienen la mayor influencia sobre la expresión de un fenotipo (Lefebvre et al., 2010). Los factores de transcripción identificados son denominados Factores Transcripcionales Maestros (FTM).

El Algoritmo para identificar reguladores maestros: “Master Regulator Analysis” (MRA) (Lefebvre et al., 2010), encuentra a aquellos factores de transcripción con mayor influencia en la transición entre dos condiciones, como es el caso entre fenotipo normal y de cáncer de mama (Tovar et al., 2015; Lefebvre et al., 2010). En este trabajo usamos una versión modificada de este algoritmo para encontrar los factores de transcripción más importantes, no para todo el fenotipo molecular, sino para algunas vías de especial interés. En este trabajo, analizamos su efecto en las vías de transducción de señales. Los conjuntos de genes en las vías que se utilizaron fueron tomados de la categoría “Signal transduction” de la base de datos Kyoto Encyclopedia of Genes and Genomes, KEGG (Kanehisa and Goto, 2000).

En el cáncer de mama, donde múltiples factores de transcripción que pueden tener cientos o miles de blancos que se encuentran simultáneamente desregulados

un enfoque integrativo puede ayudar a entender la biología que subyace a esta enfermedad. A partir de las bases de datos KEGG (Kanehisa and Goto, 2000) y Gene Ontology (GO) (Ashburner et al., 2000), el flujo de trabajo modificado del MRA es complementado con análisis de enriquecimiento estadístico para identificar las posibles funciones biológicas que estarían siendo reguladas por los factores transcripcionales maestros en sus conjuntos de genes blanco (regulones) (García-Campos et al., 2015). Dada la amplitud de la expresión de genes en la célula, el centrarnos en un conjunto específico de procesos o vías tiene como ventaja reducir la complejidad de los resultados y nos permite centrarnos en la regulación y procesos relevantes para el fenotipo canceroso.

Capítulo 2

Objetivos

2.1. Objetivo general

Para comprender la desregulación de las vías de transducción de señales buscamos inferir los Reguladores Transcripcionales Maestros que modulan estas vías en cáncer de mama.

2.2. Objetivos particulares

1. A partir de las matrices de expresión obtenidas. Inferir redes de coexpresión esto con la finalidad de conocer cuales son los genes que se encuentran coregulados en cáncer de mama.
2. Determinar el grado de desregulación de las vías de transducción de señales, en el fenotipo de cáncer de mama con respecto al tejido normal para conocer las diferencias entre ambos fenotipos.
3. Inferir los Reguladores Transcripcionales Maestros de las vías de señalización en cáncer de mama. Esto como un acercamiento a entender la regulación de las vías de transducción de señales por factores de transcripción.

Capítulo 3

Antecedentes

3.1. Cáncer

El cáncer es una enfermedad compleja en donde las células adquieren mutaciones, que conducen a la inestabilidad genética, esto a su vez lleva a la desregulación de genes que forman parte de vías de señalización que son claves para la adquisición de las características distintivas de esta enfermedad o “Hallmarks del cáncer“ ([Hanahan and Weinberg, 2011](#); [Stratton et al., 2009](#)). Los tumores malignos van progresivamente adquiriendo proliferación sostenida, resistencia a la muerte celular, inmortalidad replicativa, evasión de las señales de crecimiento, angiogénesis e invasión y metástasis [Figura 3.1.](#) ([Hanahan and Weinberg, 2011](#)). Otra característica importante de esta enfermedad es que las células cancerosas generan un ambiente inflamatorio crónico, lo que provoca un desgaste de la respuesta inmunitaria y la posterior evasión de la respuesta inmune. ([Lussier and Schreiber, 2016](#); [Ward and Thompson, 2012](#)). Adicional a esto el establecimiento del cáncer no depende de un único tipo celular y los tumores están formados de varios tipos celulares por lo que éstos pueden ser vistos como tejidos complejos en los que coexisten distintas células alteradas interactuando entre sí en favor de la supervivencia tumoral ([Hanahan and Weinberg, 2011](#)).

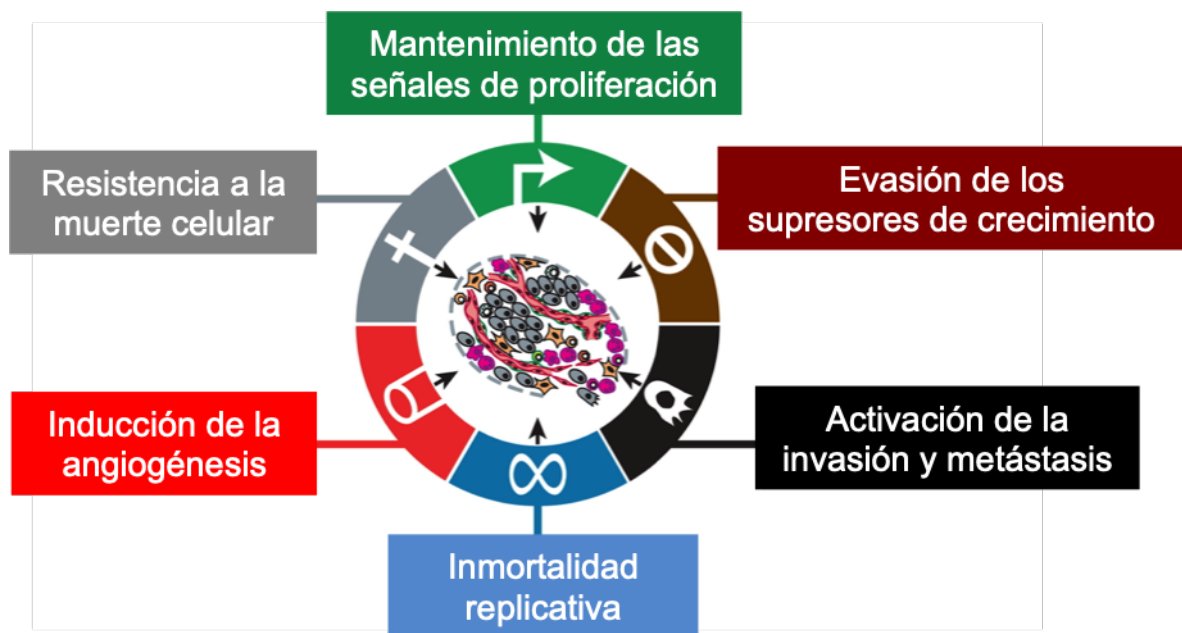


Figura 3.1: Características distintivas del cáncer. (Tomado de Hanahan and Weinberg, 2011)

3.2. Vías de señalización en cáncer

En un contexto normal, las vías de transducción de señales actúan como mediadores importantes, entre los estímulos que ocurren dentro y fuera de las células, estas vías se caracterizan por tener como elementos claves de su señalización un receptor, un segundo mensajero y un efector (Alberts, 2018). La adquisición de los “Hallmarks del cancer” esta relacionada, en gran medida, con la desregulación de estas vías de señalización (Hanahan and Weinberg, 2011).

En el cáncer se han identificado vías de señalización importantes para el desarrollo de esta enfermedad, particularmente aquellas que poseen receptores transmembranales y que forman parte de la categoría de “vías de transducción de señales” han sido ampliamente estudiadas en el contexto tumoral (Bianco et al., 2006; Hanahan and Weinberg, 2011; Rosen et al., 2010). Algunos receptores transmembranales que son parte de esta categoría y que son importantes para el desarrollo de esta enfermedad son: el receptor de crecimiento epidérmico (EGFR), el receptor de crecimiento

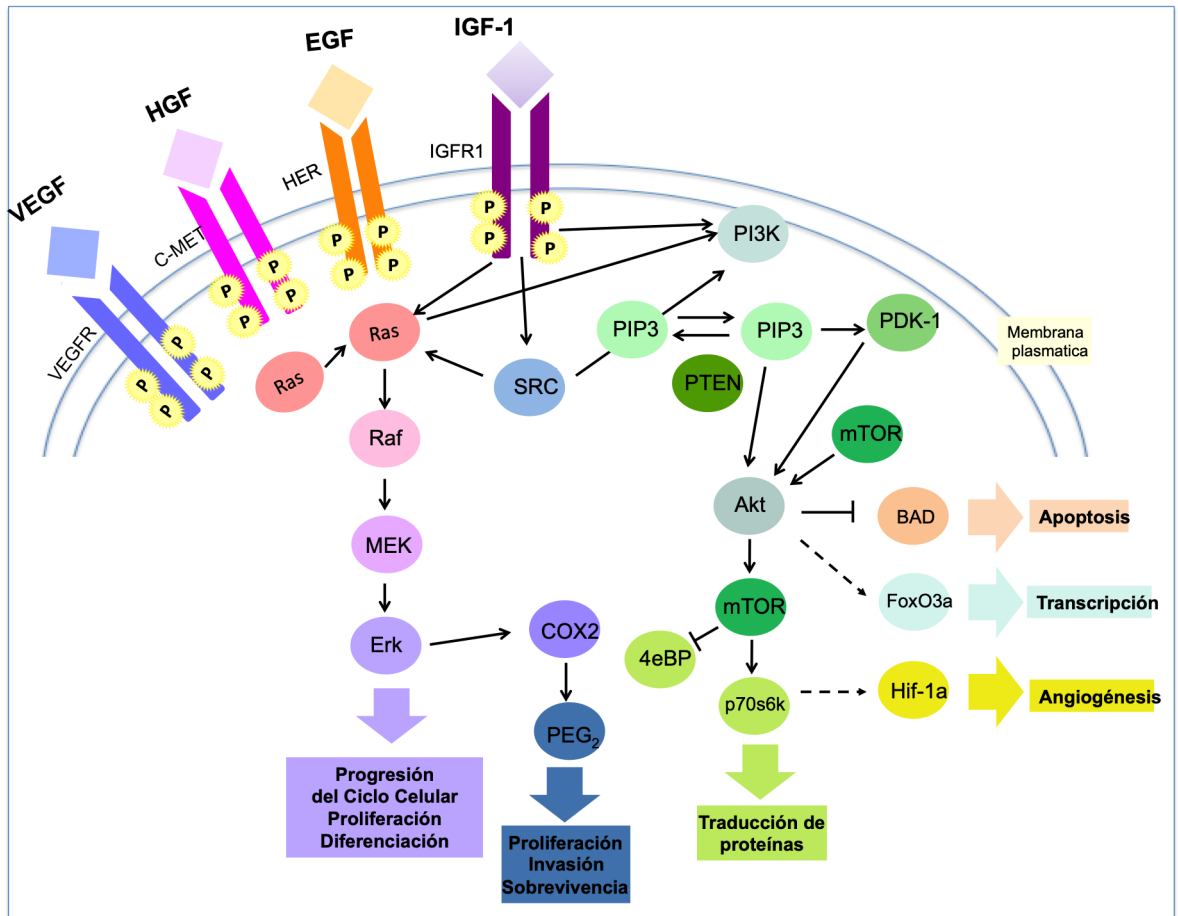


Figura 3.2: Vías de transducción de señales implicadas en cáncer.
(Modificado de Siena 2009)

insulínico (IGFR), el receptor de crecimiento vascular y endotelial (VEGFR), el receptor para el factor de crecimiento transformante beta $TGF\beta$ y el receptor 2 del factor de crecimiento epidérmico humano (HER2); éstos a través de sus receptores y la activación de cascadas de señalización controlan procesos celulares importantes para el establecimiento tumoral, tales como: la diferenciación, el crecimiento, la angiogénesis y la inhibición de la muerte celular (Figura 3.2); todas ellas características distintivas del cáncer. (Di Domenico and Giordano, 2017; Siena et al., 2009)

3.3. Regulación de las vías de señalización

Una de las características principales que presentan estas vías de transducción en un contexto normal es que la señal que emiten dentro de la célula ocurre de manera transitoria. Dentro del contexto normal el apagado de la señal puede ocurrir por la internalización de los receptores dentro de la célula o por la disponibilidad de receptores o ligando es decir de la molécula que inicia la señal (Alberts, 2018; Sampieri-Cabrera, 2020). En el cáncer la desregulación de estas vías de señalización está dada a diferentes niveles. Uno de los principales mecanismos estudiados es la presencia de mutaciones que inciden directamente sobre la función del receptor o sobre las proteínas que forman parte de la vía o que regulan la señal de apagado de la vía (Tvorogov et al., 2009; Hanahan and Weinberg, 2011; Kandoth et al., 2013). Por otro lado, la alteración de la actividad de los factores de transcripción que transcriben tanto los receptores como las proteínas señalizadoras influye en la abundancia de transcrito tanto de los elementos río abajo de la vía como de los receptores (Nebert, 2002).

Numerosas señales de crecimiento, proliferación y metástasis son captadas desde el exterior de la célula a través de receptores que forman parte de las vías de transducción de señales, fungiendo como elementos claves para el desarrollo tumoral, dado que los factores de transcripción pueden controlar la expresión de diversas moléculas dentro de una o entre diferentes vías, por lo que existe un gran interés por su estudio y por el desarrollo de fármacos que regulen su función (Bushweller, 2019).

3.4. Factores de transcripción

Los factores de transcripción son proteínas que poseen sitios de unión al ADN altamente conservados que modulan la transcripción de genes dentro del genoma.

Éstos regulan, en gran medida, el fenotipo celular así como la capacidad de la célula para responder a cambios en el medio externo a través de la activación o represión de la expresión de genes (Alberts, 2018; Allen and Taatjes, 2015; Latchman, 1997).

Los factores de transcripción que participan en la regulación positiva de la transcripción actúan primordialmente como andamios a los que se unen moléculas necesarias para formar el complejo de iniciación de la transcripción. Ejemplos de estas moléculas incluyen a la RNA polimerasa II y los factores de transcripción TFIIB y TFIID (Allen and Taatjes, 2015). En el caso de la regulación negativa de la transcripción, algunos factores de transcripción compiten directamente con las secuencias promotoras de otros factores de transcripción positivos impidiendo su unión. (Diamond et al., 1990). Un ejemplo de este mecanismo está representado por Bach1 el cual compite por la región promotora de NRF2 impidiendo la transcripción de sus genes blanco (Zhou et al., 2016).

3.4.1. Regulación de la actividad de los factores de transcripción

Los factores de transcripción pueden ser regulados a diferentes niveles. En primer lugar por un aumento de la cantidad de su transcrito y subsecuente aumento de su abundancia como proteína funcional, lo cual puede ocurrir como consecuencia de la activación de vías de señalización, así como por la actividad de otros factores de transcripción (Hunter and Karin, 1992). En segundo lugar las modificaciones post traduccionales, que son cambios covalentes dados por la adición de grupos funcionales a los aminoácidos o bien por rompimiento proteolítico, pueden determinar el estado de actividad, localización, recambio e interacción con otras proteínas (Hunter and Karin, 1992).

Si bien existen muchos tipos de modificaciones post traduccionales, la fosforilación juega un papel muy importante en la regulación de los factores de transcripción (Mann and Jensen, 2003) Esta modificación adiciona grupos fosfato preferentemente en tirosinas, serinas y treoninas a través de cinasas. (Hunter and Karin, 1992; Whitmarsh and Davis, 2000).

La fosforilación induce cambios alostéricos cuyo efecto depende del sitio en el que ocurre, activando o inhibiendo la actividad del factor de transcripción. Así, por ejemplo, la fosforilación puede promover la translocación hacia el núcleo, aumentar la afinidad por proteínas adaptadoras que favorezcan la transcripción o bien, por el contrario, conducir a la exposición de sitios que sean susceptibles a ubiquitinación induciendo su degradación (Cohen, 2000; Whitmarsh and Davis, 2000). La fosforilación también puede promover la formación de homo o heterodímeros para modular su actividad y afinidad (Hunter and Karin, 1992).

Las regiones promotoras en el DNA deben estar expuestas para que los FT puedan unirse y favorecer la transcripción. por lo que la regulación epigenética también juega un papel muy importante en la regulación de la actividad de los factores de transcripción. Típicamente las metilaciones en islas CpG reprimen la transcripción mientras que las acetilaciones en la cromatina conducen a la relajación de la hebra de ADN permitiendo la transcripción (Berger, 2007).

3.4.2. Factores de transcripción en cáncer

El cáncer también se caracteriza por una reprogramación transcripcional esta es debida a la desregulación de la función de los Factores de Transcripción, que puede estar dada por mutaciones como amplificaciones, deleciones o mutaciones puntuales en los genes que codifican a estos FT y que pueden dar como resultado una pérdida o

ganancia de función (Bhagwat and Vakoc, 2015). Algunos factores de transcripción que se han encontrado desregulados y que son relevantes para el establecimiento de esta enfermedad son: GABP, GA binding protein; GATA3, GATA binding factor 3; KLF8, Krüppel-like factor 8; RAR, retinoic acid receptor ; RUNX, runt-related transcription factor; SIX1, sine oculis homeobox homologue 1; STAT1, signal transducer and activator of transcription 1; ER, estrogen receptor entre muchos otros(Bhagwat and Vakoc, 2015; Bushweller, 2019)

En algunos tipos de cáncer el cambio en la expresión genética es tan característico que se utilizan estos patrones específicos en los niveles de expresión de genes como firmas moleculares para discriminar entre subtipos de cáncer; tal es el caso del cáncer de mama, el cual clínicamente puede ser separado a partir de los niveles de expresión de 50 genes en los subtipos Luminal A, Luminal B, Her2 y Basal (Parker et al., 2009). Por lo que la identificación de factores transcripción es relevante para entender la desregulación de la expresión genética global en esta enfermedad.

3.4.3. Mecanismos de desregulación de los factores de transcripción en cáncer

La desregulación de los factores de transcripción es mediada por diversos mecanismos: mutaciones, sobreexpresión, ausencia de represores, aumento de la tasa de fosforilación o alteraciones en la maquinaria de degradación en el proteosoma (Darnell Jr, 2002; Kansanen et al., 2013). Todos estos pueden dar origen a la función aberrante de los factores de transcripción conduciendo al establecimiento de los patrones de expresión patogénicos que son cruciales para el establecimiento de esta enfermedad (Hanahan and Weinberg, 2011).

Dentro de los primeros factores de transcripción que se descubrieron como parte importante en el desarrollo del cáncer se encontraron los receptores de esteroides.

Éstos incluyen al receptor de estrógenos, andrógenos, progesterona, glucocorticoides y mineralocorticoides, todos ellos son receptores intracelulares cuya activación induce su translocación hacia el núcleo (Darnell Jr, 2002). De éstos, los receptores para estrógenos han sido de los más estudiados ya que, aunque son necesarios para la función normal del sistema reproductivo femenino promoviendo la proliferación y diferenciación del epitelio mamario sano, los niveles endógenos de estrógeno correlacionan con el desarrollo de cáncer de mama (Ahmad and Kumar, 2011), el cual es uno de los más prevalentes en mujeres (Siegel et al., 2015). Dos de los mecanismos por los que este factor de transcripción promueve el desarrollo del cáncer de mama son por aumento de la transcripción de c-Myc y por unión a p53 impidiendo la transcripción de sus genes blanco teniendo como resultado general un aumento de la proliferación y sobrevivencia (Mohibi et al., 2011).

3.4.4. Importancia del estudio de los factores de transcripción en cáncer

Debido a que los factores de transcripción tienen la capacidad de modular la transcripción de genes, estos son de enorme interés en el estudio de enfermedades complejas como el cáncer. El conocimiento de la regulación de los programas transcripcionales es clave para el entendimiento de esta enfermedad de ahí la importancia de organizar y recopilar la información existente sobre ellos y los genes que regulan. Esto ha llevado a la creación de bases de datos tales como TRANSFAC (Matys et al., 2006) y JASPAR (Khan et al., 2017) que reúnen información sobre la evidencia experimental, el sitio de unión en el DNA y los genes que regula un factor de transcripción. Estas bases de datos también poseen herramientas que permiten predecir nuevos sitios de unión en el DNA (Stormo, 2013). Las recopilaciones de las diferentes bases de datos estiman un total de 980 factores de transcripción humanos con evidencia experimental y se han predicho bioinformáticamente con base a su

secuencia y homología 3400 (Chawla et al., 2013).

3.5. Factores transcripcionales maestros

Los factores transcripcionales maestros fueron definidos en un inicio como aquellos factores de transcripción que se encontraban en la posición de más rango en una jerarquía regulatoria, esto es, que no se encuentran bajo la jerarquía de otro gen y que llevan a la transición entre dos fenotipos (Chan and Kyba, 2013). Esta idea estaba ligada a la diferenciación celular durante el desarrollo embrionario, sin embargo este concepto ha ido evolucionando de tal manera que, en la actualidad, el término es usado para denominar aquellos factores de transcripción que son claves para llevar a cabo un proceso celular, establecer un fenotipo celular o que son fundamentales durante la progresión de una enfermedad (Su et al., 2012). En el caso del cáncer existen, por un lado, factores transcripcionales maestros asociados a un único proceso celular y que son nombrados como maestros, por ser claves para la regulación de este proceso único (Lu and Kang, 2010) y por otro lado están aquellos más generales que regulan varios procesos celulares modulando la actividad de una vasta cantidad de genes (Chan and Kyba, 2013). Como señala Chan y Kyba (2013), la ambigüedad que hay en este concepto nos lleva a pensar que, en general, un factor transcripcional maestro es aquel factor de transcripción que regula un determinado conjunto de genes que son fundamentales para el mantenimiento de un proceso o fenotipo de interés.

En la actualidad, continúa la búsqueda por identificar nuevos factores transcripcionales maestros que determinen el fenotipo canceroso. Debido a que la mayoría de los estudios en cáncer se centra en la activación de genes específicos, sería interesante ampliar este panorama y evaluar el alcance transcripcional de estos factores trans-

cripcionales maestros en la desregulación de los patrones de expresión dentro de los diferentes tipos de cáncer, este es el propósito del presente trabajo de investigación.

3.6. Biología de sistemas y cáncer

Entender fenómenos biológicos como el cáncer requiere tomar en consideración numerosos niveles de complejidad y la interacción entre una gran cantidad de elementos, lo que incluye desde aquellas interacciones moleculares que integran vías de señalización individuales, hasta las complicadas interacciones entre tipos celulares distintos e incluso tejidos distintos. La biología de sistemas busca integrar estos diferentes niveles tomando grandes volúmenes de datos para explorar y comprender las características dinámicas de los sistemas biológicos (Green et al., 2018).

Un factor crucial en el camino para elaborar estas explicaciones consiste en identificar aquellas variables de mayor importancia en el fenómeno, dejando de lado otras menos importantes por su magnitud, lo que denominamos como reducción de la dimensionalidad. Esto se puede ejemplificar con la identificación de conjuntos de genes más relevantes, o vías más importantes, de entre muchas, idealmente todas las opciones posibles. Esta tarea se torna más difícil en cuanto consideramos el gran volumen de información y conocimiento biológico actual, que ha conducido a la creación de bases de datos con la intención de recopilar información de forma exhaustiva, conteniendo catálogos de todos los genes, todas las proteínas o todas las vías. Mismas que se hallan frecuentemente disponibles para su consulta (Kanehisa and Goto, 2000; Khan et al., 2017; Matys et al., 2006; Pontén et al., 2008; Sussman et al., 1998).

Este gran volumen de datos disponible y el reto de compaginarlo con la nueva información adquirida de forma experimental lleva a la necesidad de desarrollar nue-

vas metodologías para el análisis e interpretación de resultados que a su vez pueden informar sobre el curso apropiado de nuevos experimentos; En particular, el análisis de los patrones globales de expresión de genes nos proporciona información sobre los patrones de regulación génica que se presentan en un fenotipo particular ([Salem et al., 2017](#)), por ejemplo en el caso de tipos particulares de cáncer de los que se tienen bases de datos de alta calidad como "The Cancer Genome Atlas"(TCGA) ([Weinstein et al., 2013](#)). Una forma de modelar estos patrones es por medio de redes de regulación genética inferidas a partir de estas bases de datos ([Alcalá-Corona et al., 2018](#); [Lu et al., 2017](#)).

Entre las opciones disponibles para la reconstrucción de redes de regulación transcripcional, las redes de Información Mutua permiten obtener patrones de asociación globales de carácter general entre genes. Esto es debido a que la Información Mutua es una medida de la dependencia estadística entre los niveles de expresión de genes que requiere conocer las distribuciones de probabilidad conjuntas y marginales de cada par de genes, lo que puede inferirse con un conjunto de datos suficientemente grande, al mismo tiempo que está exenta de aplicar algún modelo particular de correlación ([Hernández-Lemus and Siqueiros-García, 2013](#)). Para el cálculo de interacciones por pares, es decir cómo se relacionan dos genes entre sí, es necesario contar al menos con alrededor de un centenar de muestras, lo que es posible para ciertos tipos de cáncer con los conjuntos de datos disponibles en la actualidad ([Margolin et al., 2006](#))

Una vez que contamos con nuestra red, podemos hacer preguntas biológicas relacionadas con la estructura de la red y su conectividad, lo que equivale a realizar experimentos en la computadora sacando provecho de datos experimentales y complementando los resultados por medio de conocimiento biológico almacenado en otras bases de datos. Un ejemplo lo tenemos con la identificación de grupos de genes aso-

ciados en la red, que son comparados con conjuntos de genes asociados a funciones biológicas.([Alcalá-Corona et al., 2016](#); [Velazquez-Caldelas et al., 2019](#)).

Capítulo 4

Metodología

4.1. Obtención y pre-procesamiento de datos

La matriz de expresión se generó a partir de datos de nivel 3 de la plataforma IlluminaHiSeq RNA-seq obtenidos de TCGA (The Cáncer Genome Atlas) correspondientes a 780 muestras de carcinoma invasivo de mama y 101 muestras de tejido sano adyacente los cuales fueron pre-procesados utilizando las bibliotecas de R NOI-seq ([Tarazona et al., 2011](#)) y EDASeq ([Risso et al., 2011](#)) la primera se utilizó para el preprocesamiento general de los datos, mientras que la segunda para el eliminar el efecto de lote, el cual fue eliminado mediante la biblioteca de R EDASeq. Para mayores detalles consultar [Espinal-Enriquez et al., \(2017\)](#)

4.2. Análisis de desregulación de vías.

Para determinar si las vías de transducción de señales, se encuentran desreguladas a nivel de expresión de genes en nuestro conjunto de datos de cáncer de mama, estimamos su grado de desregulación utilizando el algoritmo Pathifier ([Drier et al., 2013](#)). Pathifier evalúa cada vía de forma independiente y le asigna un puntaje denominado puntaje de desregulación de vías (PDS) para cada vía en una muestra.

Pathifier, para una vía dada, define un espacio multidimensional. En este espacio cada dimensión representa el nivel de expresión de un gen en una vía y cada punto representa una muestra cuya posición esta dada por los niveles de expresión del conjunto de genes para esa vía. (Figura 4.1 panel A).

Una vez definida la nube de puntos proveniente de todas las muestras se ajusta una línea llamada curva principal, que minimiza la distancia entre todos ellos, la posición de cada muestra es proyectada sobre esta curva; el PDS corresponde a la medida de la distancia a lo largo de la curva principal entre la proyección de la muestra del tejido enfermo y la proyección de las muestras del tejido normal (Figura 4.1 panel B (Drier et al., 2013)).

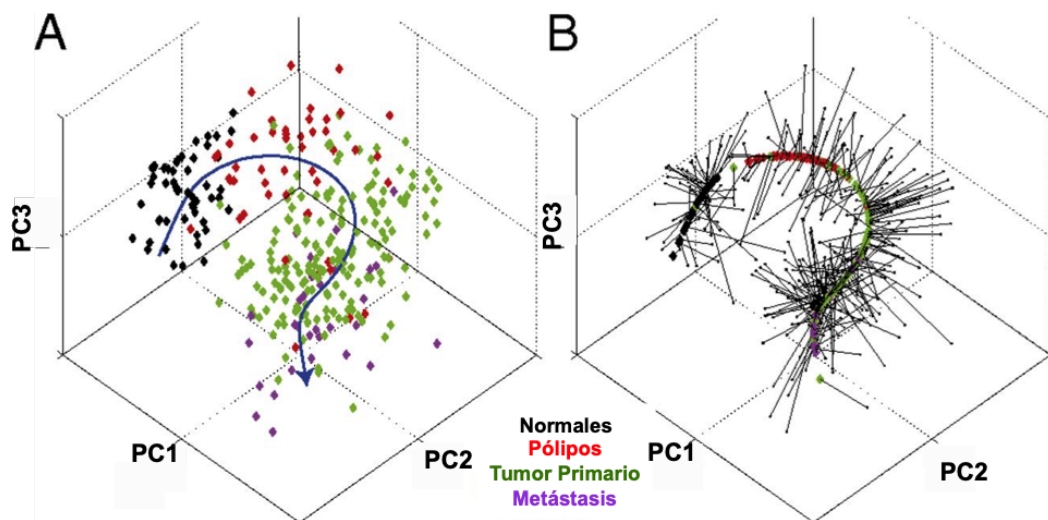


Figura 4.1: Representación gráfica del algoritmo Pathifier. Cada muestra es ubicada dentro del espacio multidimensional definido por la vía de interés. A) Determinación de la curva principal que atraviesa la nube de puntos. B) Proyección de cada muestra sobre la curva principal. (Tomado de Drier et al., 2013)

4.3. Inferencia de reguladores transcripcionales maestros

Los reguladores transcripcionales maestros se infirieron con MARINa (Master Regulator Inference Algorithm; [Lefebvre et al. \(2010\)](#) que es una implementación del MRA. MARINa identifica a los Factores de Transcripción mas importantes para un fenotipo para ello considera la diferencia de expresión entre dos fenotipos de estudio en este caso cáncer de mama y tejido normal. La inferencia de los reguladores maestros con MARINa requiere una red de regulones, una firma molecular de expresión de genes y un modelo nulo ([Lefebvre et al., 2010](#)) (Figura 4.3). La construcción de estos elementos se describe a continuación.

4.4. Generación de la red de regulones

Un regulón es definido en este trabajo como una red dirigida donde las aristas describen interacciones regulatorias desde un factor de transcripción y sus blancos transcripcionales (TF Target). Una red de regulones esta conformada por la unión de muchos regulones.

Para obtener el conjunto de regulones, usamos una lista de factores de transcripción. Esta lista fue tomada de la base de datos TFCheckpoint ([Tripathi et al., 2013](#)) seleccionando aquellos factores de transcripción que tenían evidencia experimental de unión a secuencias específicas en el DNA. Para generar las redes de regulones se consideraron en total 771 ya que estos fueron los que encontraron dentro de la matriz de expresión de cáncer de mama.

En primer lugar se generó una red de información mutua la cual asoció a los

factores de transcripción con sus blancos. Para ello usamos el “Algorithm for the Reconstruction of Accurate Cellular Networks” (ARACNe) (Margolin et al., 2006); ARACNe parte de la idea de que la expresión de los factores de transcripción y sus blancos son estadísticamente dependientes y por lo tanto se pueden predecir sus blancos transcripcionales en un contexto biológico particular a partir de la medición de estas dependencias estadísticas. La forma en que ARACNe encuentra estas asociaciones, es por medio del cálculo de información mutua entre pares de genes (en este caso FT-Gen). Para ello calcula las distribuciones de probabilidad marginales (que son las distribuciones de probabilidad para un solo gen) y conjuntas es decir la distribución de probabilidad de los niveles de expresión en un par de genes. Figura 4.2 (Hernández-Lemus and Siqueiros-García, 2013; Margolin et al., 2006).

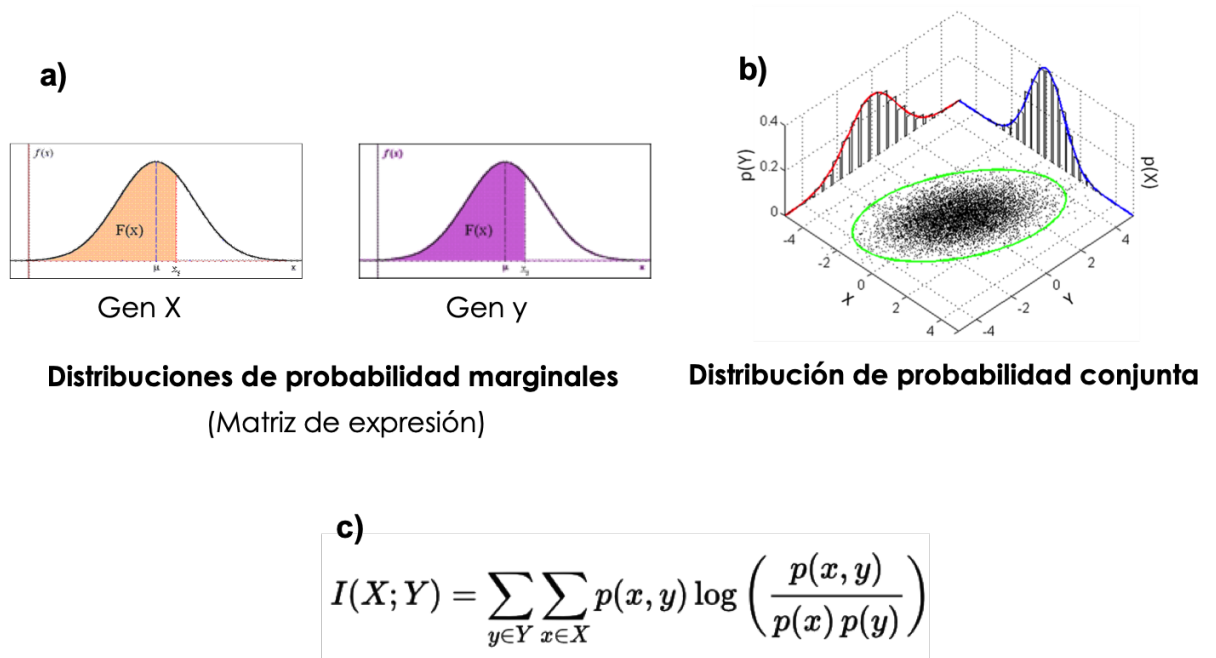


Figura 4.2: Cálculo de información mutua. A partir de los valores de expresión para un par de genes a)ARACNe calcula las distribuciones de probabilidad marginales y b) conjuntas. c) Formula de la información mutua. I:Informacion, p(x) y p(y): Distribución de probabilidad marginal de x y y, p(x,y): Probabilidad conjunta de (x,y) (Tomado de.

)

Para esta red se tomaron todas la posibles interacciones entre factores de transcripción y genes en la matriz de expresión y fueron consideradas si su valor de p

estaba por debajo de 0.005. Dado que la información mutua puede detectar tanto interacciones directas como indirectas, ARACNe elimina las interacciones indirectas aplicando el teorema “Data Processing Inequality” (DPI) el cual considera que en un triángulo de interacciones la mas débil tiene una mayor probabilidad de ser indirecta, si esta diferencia es mucho mayor con respecto a las otras dos interacciones del triángulo (Dembo et al., 1991). Aplicamos un valor de DPI de 0.2 (Margolin et al., 2006), lo cual significa que las interacciones mas débiles de los triángulos en la red fueron eliminados para no introducir un excesivo numero de falsos positivos si su valor de MI era menor de 0.2.

El tipo de asociación (activación o represión) del factor de transcripción es determinado a partir de la correlación de Spearman entre el factor de transcripción con el nivel de expresión de todos sus genes blancos. (Lefebvre et al., 2010). Este cálculo fue generado por la función `aracne2regulon` en el paquete `viper` (Alvarez et al., 2016) de R (Team et al., 2013). Esta función transforma la red de información mutua no dirigida generada con ARACNe en una red de regulones dirigida (FT->Gen).

4.5. Generación de la firma molecular de las vías de transducción de señales

En el flujo de trabajo estándar de MAR, la firma molecular es construida comparando las distribuciones de los niveles de expresión del total de genes entre dos condiciones como por ejemplo sano y enfermo. Para este trabajo construimos una firma molecular usando solo aquellos genes anotados dentro de la base de datos de KEGG de la categoría “Signal Transduction” (Kanehisa and Goto, 2000). Para la especie humana, esta categoría contiene 25 vías. El propósito de este filtrado fue centrarnos en la búsqueda de los factores de transcripción que regulan la actividad

aberrante de estas vías de transducción de señales en cáncer de mama. El total de genes presentes en esta categoría fueron 1700, de los cuales 1395 se encontraron en la matriz de expresión. Los resultados de esta prueba fueron normalizados con una prueba z para permitir que estos fueran comparables (Lefebvre et al., 2010).

4.6. Generación del modelo nulo

Para estimar la probabilidad de que un conjunto de genes enriquecidos dependa del contexto biológico y no del azar, se generó un modelo nulo permutando entre casos y controles las muestras y recalculando los valores de expresión diferencial (Lefebvre et al., 2010).

4.7. Inferencia de los reguladores transcripcionales maestros de las vías de transducción de señales.

Con la firma molecular, la red de regulones y el modelo nulo, se estimaron los principales regulones que enriquecieron por poseer los genes más diferencialmente expresados en la firma molecular a través de un análisis de enriquecimiento de genes (GSEA; (Subramanian et al., 2005)) considerando solo aquellos factores de transcripción que poseyeran 20 o más blancos (Lefebvre et al., 2010). A continuación fue estimado un valor de p para cada regulón a partir de la evaluación del score de enriquecimiento (ES) con referencia a la distribución de los puntajes del modelo nulo (Lefebvre et al., 2010). Para la inferencia de los reguladores transcripcionales maestros usamos el paquete viper (Alvarez et al., 2016) para R (Team et al., 2013) disponible a través de bioconductor (Gentleman et al., 2004), Huber et al. (2015)

La diferencia de este trabajo con respecto a la forma estándar del análisis de

MARINa radica en el uso de un conjunto específico de genes (firma de las vías de señalización) lo cual da como resultado un ordenamiento jerárquico de los regulones para este grupo específico de genes. Es importante destacar que las redes de regulones utilizadas como entrada son las mismas que las que utilizaría el MAR estándar pero la jerarquía de los regulones se centra en los genes de una firma específica. El conjunto de genes que constituyen cada regulón puede incluir genes que no son parte de la firma molecular y que pueden ser parte de una gama más diversa de funciones biológicas que las vías de transducción de señales. Por esta razón realizamos posteriormente análisis de enriquecimiento de los regulones sobre todas las vías de KEGG.

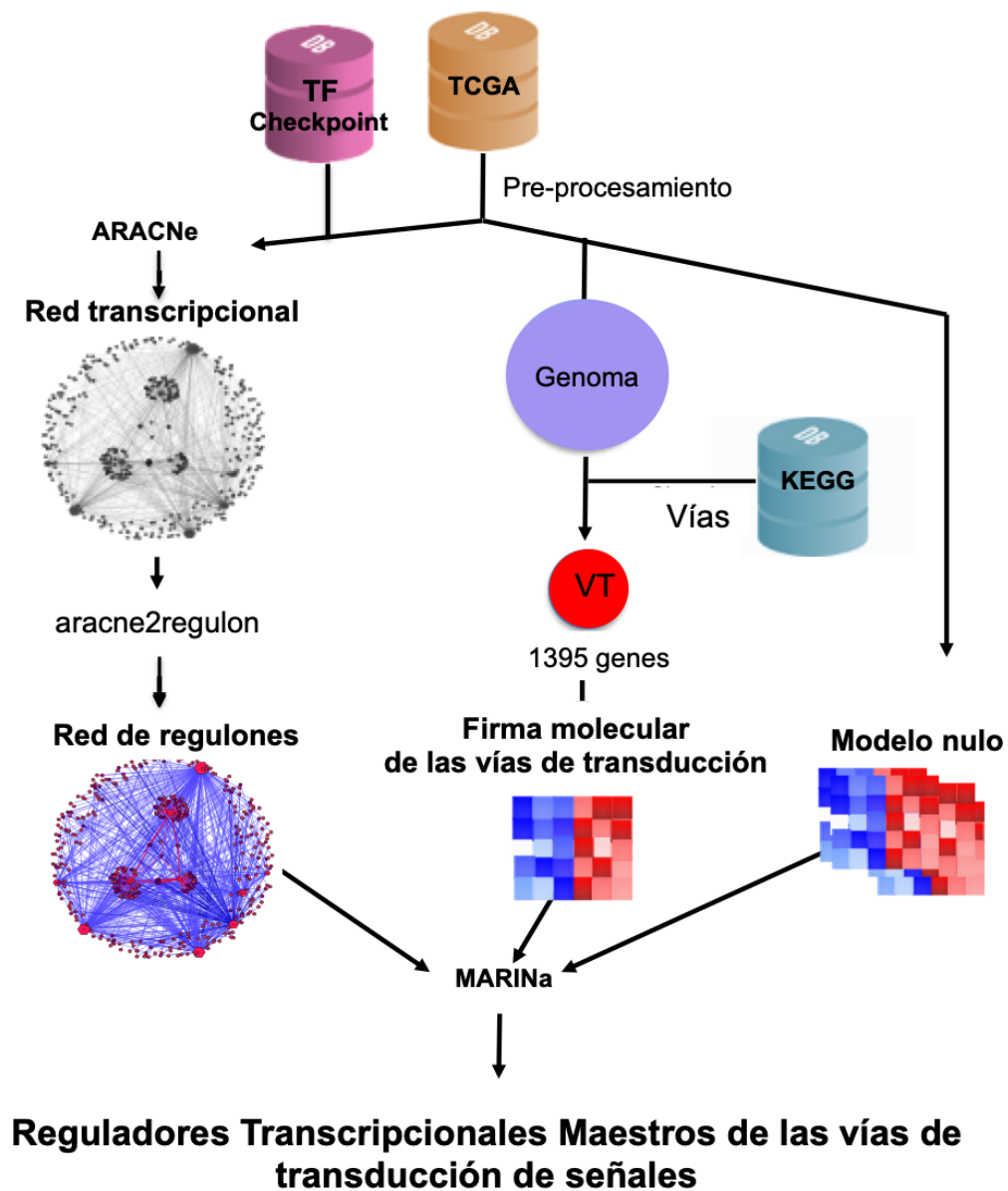


Figura 4.3: Flujo de trabajo para obtener los reguladores transcripcionales de las vías de transducción de señales. Los datos de RNAseq de los 780 carcinomas mamaros invasivos y 101 muestras de tejido adyacentes tomados de TCGA se procesaron para obtener una matriz de expresión (cilindro naranja). La matriz de expresión y una lista de factores de transcripción de la base de datos TFCheckpoint (cilindro rosa) sirvieron como entrada para inferir una red de regulación transcripcional con ARACNe. Utilizando la función *aracne2regulon* de *viper* se obtuvo una red de regulones que asocia el nivel de expresión de los blancos con su factor de transcripción (lado izquierdo). Para la generación de la firma molecular, consideramos los genes en la matriz de expresión en la categoría de “Signal Transduccion” de KEGG (cilindro azul). Finalmente, se generó un modelo nulo permutando las etiquetas de las muestra y recalculando la firma molecular (derecha). Estos tres elementos son la entrada a MARINa para la inferencia de los reguladores maestros transcripcionales de las vías de transducción de señales.

4.8. Enriquecimiento funcional de la red de los 10 primeros regulones con las vías de KEGG

Una vía sobre-representada en un conjunto de genes está definida como aquella para la cual encontramos estadísticamente más genes de los que se esperaría al tomar una muestra al azar del mismo tamaño ([García-Campos et al., 2015](#)) cuando esto ocurre decimos que este conjunto de datos está enriquecido con genes de una vía y a su vez esto puede sugerir relevancia biológica. La significancia estadística de un enriquecimiento puede ser evaluado por la media de una prueba hipergeométrica. Para saber si la combinación de los regulones de los 10 primeros factores transcripcionales maestros estaban enriquecidos para vías biológicas, utilizamos WebGestalt a través de su plataforma web para realizar un análisis de sobre-representación (ORA)([Zhang et al., 2005](#)) con KEGG como base de datos funcional como referencia ([Kanehisa and Goto, 2000](#)). El corte de significancia estadística después de la corrección hecha por la tasa de descubrimiento falso se estableció en ≤ 0.05 ($FDR \leq 0.05$).

Este análisis de sobre-representación de vías realizado a la red nos permite conocer cuales de las vías de transducción de señales originales están siendo reguladas por los reguladores maestros.

La composición genética de los regulones está determinada por las dependencias estadísticas entre los niveles de expresión de los genes de los factores de transcripción y el resto de los genes en el tejido. A pesar de que conocemos la identidad de los factores de transcripción, esto no es garantía de que todos los genes de los factores de transcripción y la firma molecular estén presentes en la red. Esto significa que el agrupamiento de los genes no es conocido a priori o impuesto desde una base de datos conocida como KEGG o GO. La co-expresión de genes pertenecientes a una función o vía en diferentes módulos de una red se han observado previamente

(Alcalá-Corona et al., 2018).

4.9. Enriquecimiento de los regulones con procesos biológicos de Gene Ontology

Para conocer cómo los 10 primeros reguladores transcripcionales maestros de las vías de transducción de señales, estaban contribuyendo al fenotipo canceroso, realizamos un ORA individual para cada uno de los regulones sobre la categoría de procesos biológicos de Gene Ontology (Ashburner et al., 2000). Utilizamos GO debido a que esta base de datos toma en consideración procesos biológicos con diferentes grados de especificidad, desde procesos muy generales que serían altamente esperados en todas las células hasta subconjuntos muy específicos de un proceso, adicional a esto, esta base de datos nos proporciona una referencia que es independiente de las listas de genes KEGG que utilizamos originalmente. Los enriquecimientos de GO se calcularon con WebGestalt (Zhang et al., 2005). El umbral de significancia estadística después de la corrección FDR se estableció en $FDR \leq 0.05$.

Capítulo 5

Resultados y discusión

5.1. Análisis de desregulación de vías

Para determinar si las vías de transducción de señales elegidas como parte de la firma molecular eran importantes para nuestro fenotipo de estudio evaluamos si éstas se encontraban desreguladas a nivel de la expresión de sus genes, en cáncer de mama, para ello usamos el algoritmo Pathifier ([Drier et al., 2013](#)).

Pathifier calcula el puntaje de la desregulación de vías (PDS) el cual es una medida de la diferencia global en los niveles de expresión de un conjunto de genes comparados con una referencia (casos vs control). El PDS es calculado para cada muestra en cada vía. Los resultados del análisis de desregulación mostraron que las 25 vías de transducción de señales que forman parte de la categoría de KEGG “Signal Transduction” tienen un patrón distintivo en los tumores de mama con respecto a los tejidos de mama normal. Esto puede observarse en el agrupamiento no supervisado del dendograma que se observa en la parte superior (Figura 5.1): las dos principales ramas de este dendograma separan claramente dos grupos (tejido normal y canceroso de mama)

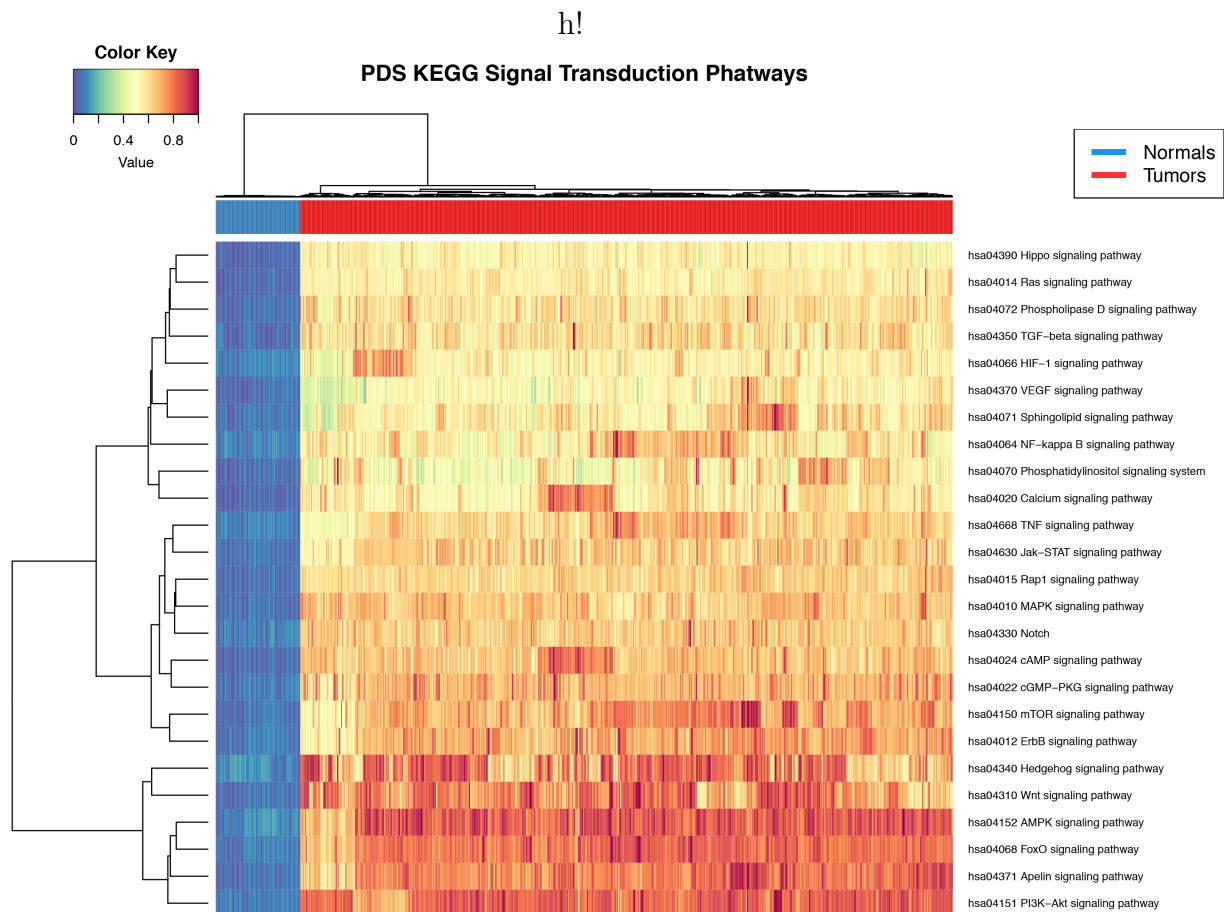


Figura 5.1: Mapa de calor del puntaje de desregulación de vías de las vías de transducción de señales. En la barra superior, las muestras control se muestran en azul mientras que las muestras de cáncer de mama en color rojo. La escala de color en el mapa representan el grado de desregulación por vía en referencia al tejido sano. En el lado superior izquierdo se observa la clave de color de los valores de PDS calculado para cada muestra y vía.

5.2. Reguladores transcripcionales maestros de las vías de transducción de señales

La red de regulones esta conformada por 765 factores de transcripción. De estos 338 cumplían con el requisito de tener al menos 20 blancos transcripcionales formando parte de la firma molecular. Del MRA se obtiene una lista jerarquizada de los regulones basada en la integración de estructura de la red de regulones y la expresión diferencial de los tumores con respecto al tejido normal. De la lista de los 338 factores transcripcionales maestros los 10 primeros regulan aproximadamente el 30% de los genes que forman parte de la firma molecular (Figura 5.2) Los 10 primeros reguladores Transcripcionales maestros de las vías de transducción de señales son: GLI Family Zinc Finger 2 (GLI2), Paternally Expressed 3 (PEG3), T Box 18 (TBX18), Homeobox A5 (HOXA5), Heart and Neural Crest Derivatives Expressed 2 (HAND2), Homeobox A3 (HOXA3), Meis Homeobox 2 (MEIS2), Homeobox A2 (HOXA2), Teashirt Zinc Finger Homeobox 2 (TSHZ2) y Clock Circadian Regulator (CLOCK) (Tabla 5.1).

Los resultados del MRA mostraron que, con excepción de CLOCK, la actividad de estos factores de transcripción sobre sus blancos es de represión (Columna Act del lado derecho de la Figura 4.3, valores negativos de NES de la Tabla 5.1 y los links rojos de la red en la Figura 5.4). Las interacciones regulatorias en los regulones están definidas como activación si los blancos de ese factor de transcripción se encuentran sobreexpresados o de inhibición si los blancos están sub expresados. La red de regulones de los 10 primeros reguladores transcripcionales de las vías de señalización mostraron una alta proporción de interacciones negativas (represión) sobre sus genes blanco (Figura 5.4). En esta red, GLI2 es el único regulador maestro interactuando con mas de un regulador maestro, interactuando con PEG3, TBX18, HAND2, HOXA3 HOXA2, y HOXA5. Estos genes junto con TSHZ2 y MEIS2 están

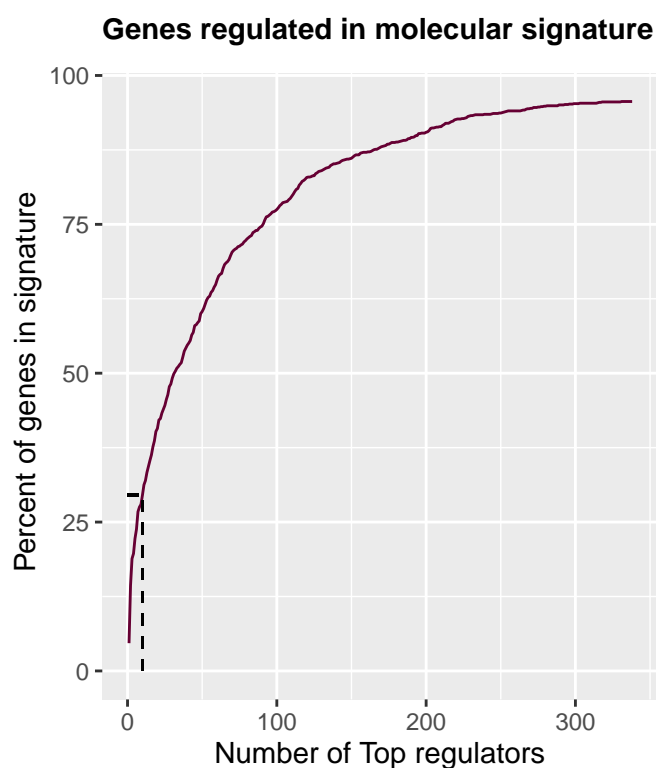


Figura 5.2: Porcentaje de la firma molecular regulada por los reguladores transcripcionales maestros de las vías de señalización. Los primeros 10 reguladores maestros controlan el 30% de los genes pertenecientes a la firma molecular. Si quisiéramos cubrir el doble de este porcentaje tendríamos que tomar 5 veces mas reguladores transcripcionales maestros; por lo que tomando 10 reguladores optimizamos el análisis.

reportados como factores de transcripción involucrados en procesos de embriogénesis, desarrollo embrionario y remodelación de tejido adulto (Kuroiwa et al., 1996) Srivastava (1999) Ruiz i Altaba et al. (2007) Melvin et al. (2013) Takeichi et al. (2013) Chojnowski et al. (2014) Amin et al. (2015) Machon et al. (2015) Jeannotte et al. (2016).

La mayoría de los regulones de estos RTMs están enriquecidos con genes que forman parte de la vía de señalización Hedgehog. Esta vía esta relacionada con procesos biológicos como la morfogénesis temprana y, junto con Wnt, participan en la autorrenovación de las células madre (Reya et al., 2001). Ambas vías se encuentran desreguladas en el cáncer (Reya et al. (2001); Taipale and Beachy (2001). El fac-

Tabla 5.1: Primeros 10 Reguladores Maestros de las vías de transduccion de señales.

Regulon	Tamaño	NES	p Value	FDR
CLOCK	59	3.68	2.31E-04	0.00592
TSHZ2	35	-3.7	2.13E-04	0.00592
HOXA2	54	-3.71	2.11E-04	0.00592
MEIS2	128	-3.73	1.89E-04	0.00592
HOXA3	67	-3.84	1.23E-04	0.00592
HAND2	93	-4.05	5.06E-05	0.00324
HOXA5	30	-4.07	4.66E-05	0.00324
TBX18	132	-4.12	4.66E-05	0.00324
PEG3	162	-4.18	2.80E-05	0.00324
GLI2	64	-4.36	1.30E-05	0.00324

Estos factores de transcripción controlan los genes de las vías de transducción de señales expresadas de manera más diferencial en el tejido tumoral. Con la excepción de CLOCK, estos reguladores se describen comúnmente en el contexto del desarrollo embrionario, y todos ellos se han informado en asociación con el cáncer. El número total de genes controlados por estos regulones es 412, lo que representa casi un tercio de los genes totales en la firma molecular.

tor de transcripción TSHZ2 forma un complejo con GLI1, que funciona de manera coordinada con GLI2 y GLI3 dentro de la vía Hedgehog ([Riku et al., 2016](#)). Esto se recuperó de la red de regulones en forma de asociaciones de expresión génica. Además, se informó previamente una relación con TBX18 y la vía Hedgehog en experimentos Knockout de TBX18 donde mostró una marcada disminución en la expresión de genes de la vía Hedgehog. [Wu et al. \(2013\)](#).

GLI2 es el unico RTM que contiene múltiples interacciones con otros RTMs (seis en total; [Figura 5.4](#)). GLI2, junto con GLI1, GLI3, y TSHZ2 (otro de los 10 primeros RMTs), son moléculas efectoras activadas dentro de la vía Hedgehog que modulan los procesos de desdiferenciación y diferenciación durante morfogénesis temprana ([Ruiz i Altaba et al., 2007](#); [Scales and de Sauvage, 2009](#)). Aquí estos RTMs también los encontramos asociados a un contexto donde se encuentra representada la vía de Hedgehog.

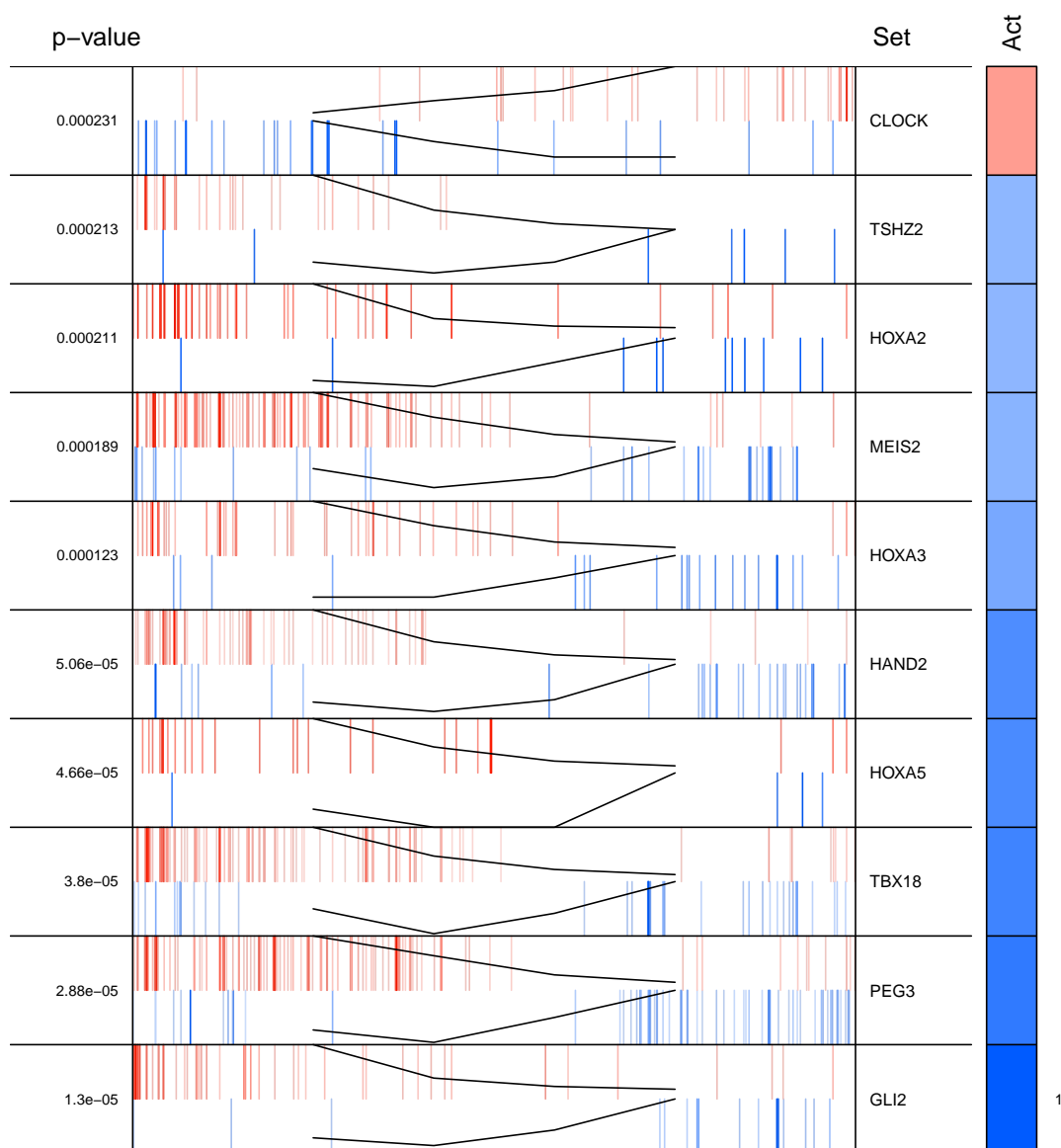


Figura 5.3: Los 10 principales reguladores maestros de las vías de transducción de señales. Estos factores de transcripción controlan los genes de las vías de transducción de señales expresadas de manera más diferencial en el tejido tumoral. El número total de genes de las vías de transducción de señales controladas por estos regulones es 421, lo que representa casi un tercio de los genes totales en la firma molecular. La primera columna muestra el valor de enriquecimiento p para cada regulón. Los nombres de los reguladores maestros están en el lado derecho. La columna “Act” indica la actividad del regulador maestro sobre sus genes blanco. El color rojo representa la sobreexpresión y el color azul representa la subexpresión con respecto al tejido normal. La columna de código de barras muestra cada gen en la firma molecular como una barra vertical. La línea negra en las líneas verticales azul y roja muestra el recorrido por la lista rankeada de genes del análisis de enriquecimiento del conjunto de genes en los genes de firma molecular.

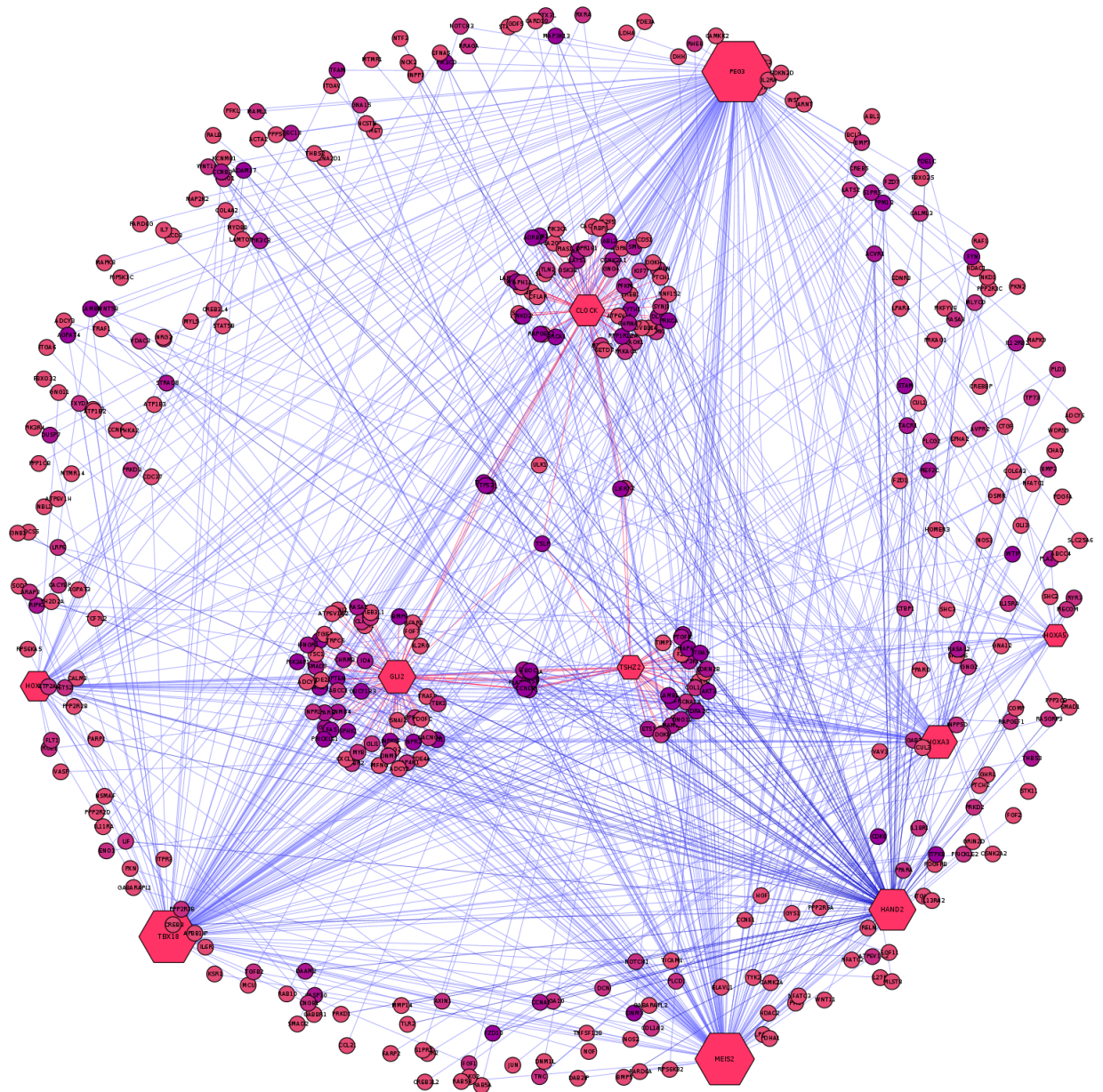


Figura 5.4: Red los 10 primeros regulones de los factores transcripcionales maestros de las vías de señalización. La red de los factores de transcripcionales maestros (hexágonos grandes) y sus blancos (círculos) muestra una mayor cantidad de interacciones de represión (líneas azules) en comparación con las interacciones de sobre-expresión (líneas rojas). GLI2 es el factor transcripcional maestro con el puntaje de enriquecimiento mas alto ES de los 10 primeros. GLI2 mantiene interacción de activación con algunos de sus blancos aunque la mayoría son de represión. CLOCK es el único de los reguladores que mantiene una mayor proporción de interacción de activación imagen generada con Cytoscape (Shannon et al., 2003).

5.3. Características de los 10 primeros reguladores transcripcionales maestros

Los genes homeobox se encuentran dentro de 4 clusters de genes denominados de la letra A- D. El cluster A se encuentran dentro del cromosoma 7 (Apiou et al., 1996) y regulan la expresión de genes que llevan a la morfogénesis y diferenciación, su especificidad de unión al ADN depende de su unión a cofactores para seleccionar sus blancos transcripcionales (Merabet and Lohmann, 2015).

HOXA2 puede a través de su heterodimerización con otros factores de transcripción como Pdx y MEIS2 participar dentro del desarrollo del segundo arco branquial y se ha visto que mutaciones en HOXA2 están asociadas a defectos congénitos del desarrollo de la oreja (microtia)(Piceci et al., 2017; Alasti et al., 2008)

En cáncer HOXA2 actúa como un represor transcripcional ya que se ha reportado que la hipermetilación de su promotor conlleva a la baja expresión de ARNm, por ejemplo, reprime la expresión de la metaloproteínasa 9 MMP-9 involucrada en la metástasis, al bloquear su secuencia promotora. En estudios de cáncer de mama los resultados no son concluyentes ya que se ha encontrado tanto subexpresado (Makiyama et al., 2005) como sobreexpresado. (Hur et al., 2014)

HOXA3 participa en la organogénesis del timo y la paratiroides (Chojnowski et al., 2014) y está involucrado en la migración de las células epibranquiales (Watari-Goshima and Chisaka, 2011) Por otro lado, este factor de transcripción se encuentra diferencialmente expresado en diferentes líneas celulares de cáncer de mama y piel (Svingen and Tonissen, 2003). Algunos estudios en cáncer de colon mostraron que

la sobre expresión de HOXA3 está asociada a un pobre pronóstico en pacientes, esto mediado por la activación de la vía EGFR/Ras/Raf/MEK/ERK promoviendo la sobrevivencia, celular por lo que se propuso como un oncogen (Zhang et al., 2018); también participa en la migración de células endoteliales y epiteliales por incremento de la metaloproteinasa-14 y u-Par e induce la angiogénesis (Mace et al., 2005).

En cáncer de mama la pérdida de HOXA5 está asociado a pérdida de las funciones de p53 (Raman et al., 2000) pérdida de rasgos epiteliales, aumento de la troncalidad, plasticidad celular, y la adquisición de fenotipos más agresivos por regulación de CD24 y E-Cadherina (Teo et al., 2016) promoviendo la tumorigenesis, en cáncer colorectal donde se encuentra subexpresado. Se encontró que es controlado de manera indirecta por la vía de WNT a través de c-Myc el cual reprime la expresión de HOXA5 promoviendo el mantenimiento de la troncalidad intestinal (Ordóñez-Morán et al., 2015). Por otro lado en carcinoma de células escamosas su subexpresión reduce la tumorigénesis, por inhibición de b-catenina ciclina-D1 y c-Myc disminuyendo la proliferación y la metástasis (Zhang et al., 2018).

Durante el desarrollo embrionario TBX18 participa en la organogénesis del corazón, columna vertebral y formación de los ureteros (Vivante et al., 2015) actúa como represor transcripcional al competir con otros miembros de la misma familia por las mismas regiones de unión al DNA (Farin et al., 2007) durante el desarrollo cardiaco la disminución de TBX18 inhibe la transición epitelio mesénquima epicárdica (Takeichi et al., 2013) análisis de expresión genética en ratones Knockout de TBX18 mostraron una disminución de genes que participan en la vasculatura coronaria durante el desarrollo, tales como VEGF-A, VEGF-B, Endoglin, y Angiopietin-1 y una marcada sub expresión de genes de la vía de Hedgehog como Ptch1, Smo y Gli2 (Wu et al., 2013).

En células de neuroblastoma MEIS2 mantiene la expresión de genes del ciclo celular tardío promoviendo la promoción hacia a la fase M del ciclo celular por transcripción de genes necesarios para la replicación del DNA regulando la supervivencia y la proliferación (Zha et al., 2014). Su sub-expresión conduce a las células de cáncer colorectal hacia la apoptosis, y se ha encontrado sobre expresado en cánceres de ovario y páncreas (Hong et al., 2010).

HAND2 interactúa directamente con Era/b formando un complejo que reprime la actividad transcripcional del receptor a estrógenos (ER)(Fukuda et al., 2015) se ha visto que es inducido por progestina y en cáncer endotelial de tipo 1 se ha encontrado metilado dentro de islas CPG (Srivastava, 1999) durante el desarrollo embrionario del corazón HAND2 modula varios componentes de la vía de VEGF permitiendo la vascularización coronaria y su expresión está controlada río debajo de la señalización de NOCT (VanDusen et al., 2014).

Experimentos de ChiP-seq mostraron que PEG3 se une al DNA en regiones promotoras de Per1 y Cry2 (Lee et al., 2015) regiones típicamente reguladas por CLOCK (Brown, 2014). En glioma se ha reportado que puede actuar como un gen supresor de tumores al inducir la apoptosis mediada por p53 (Kohda et al., 2001) experimentos de sobre expresión de PEG3 indicaron una disminución de β -catenina e inhibición de la vía de WNT, lo cual condujo a la inhibición de la proliferación (Jiang et al., 2010). Por otro lado PEG3 se ha propuesto como un marcador de riesgo ya que su metilación fue asociada a un elevado riesgo de cáncer de mama y cervical (Harrison et al., 2015). Este factor de transcripción también se expresa durante el desarrollo embrionario en somitas tempranas, arcos branquiales, hipotálamo y otras regiones mesodermicas (Kuroiwa et al., 1996)

GLI2 puede actuar como represor o activador de la transcripción de genes. Junto

con GLI1 y GLI3 son importantes moléculas efectoras activadas dentro de la vía de Hedgehog que modulan la diferenciación celular durante el desarrollo embrionario. (Ruiz i Altaba et al., 2007; Scales and de Sauvage, 2009). La desregulación de la vía de Hedgehog está asociada a varios tipos de cáncer como glioma, leucemia, páncreas, mama y próstata (Clement et al., 2007; Zhao et al., 2009; Thayer et al., 2003; Zardawi et al., 2009; Karhadkar et al., 2004). GLI2 en cáncer ha sido asociado a procesos de proliferación y metástasis (Javelaud et al., 2011; Gupta et al., 2015).

Finalmente está reportado que TSHZ2 actúa como un represor transcripcional formando un complejo con GLI1 y CtBP2 impidiendo la transcripción de genes regulados por GLI1 tales como AEBP1 y CXCR4. TSHZ2 se ha encontrado subexpresado en análisis de microarreglos de cáncer de mama esta disminución de la proteína correlaciona con un aumento de la expresión de los genes de GLI1 (Riku et al., 2016) estudios en líneas celulares de cáncer de mama y próstata donde se evaluó el estado de metilación del promotor de TSHZ2 éste se encontró no metilado, sin embargo la expresión del gen se encuentra disminuida lo que sugiere otros mecanismos de regulación transcripcionales (Yamamoto et al., 2011).

5.4. Enriquecimiento de las redes de regulones en las vías de KEGG

La diferencia entre este trabajo con respecto al MRA tradicional radica en el uso de un conjunto específico de genes para generar la firma molecular lo cual da como resultado una jerarquización de los regulones en función de esta firma específica, es decir de los genes de las vías de señalización. La red de regulones utilizada como entrada para el análisis es la misma que se usaría para el MRA tradicional, pero el ordenamiento es priorizado por la expresión de los genes de las vías de señalización

por el uso específico de la firma molecular.

El conjunto de genes que constituyen cada regulón puede variar en forma y tamaño desde unas decenas a cientos de genes que pueden o no incluir genes de la firma molecular pero son parte de funciones biológicas u otras vías de señalización. El análisis estadístico de sobre-representación conduce a la reducción de la dimensionalidad de una larga lista de genes individuales ó unos pocos y bien distinguibles procesos biológicos, lo cual es necesario dado el gran numero de genes incluidos en la red y la posibilidad de múltiples anotaciones para cada gen.

Los 10 primeros reguladores transcripcionales maestros de las vías de señalización asociaron por dependencia estadística 4052 genes. Para saber cuales vías estaban siendo posiblemente más reguladas, realizamos un análisis de sobrerrepresentación para todas las vías de KEGG humanas (Tabla 5.2). Esto nos ayudo a saber cuáles de las vías de transducción de señales estaban representadas, así como otras vías correguladas. La vía con el enriquecimiento mas significativo fue Pathways in cancer (hsa05200) con una coincidencia de 121 genes. Esta vía no fue considerada en la construcción de la firma molecular, aunque es de esperarse este enriquecimiento ya que es una vía muy grande compuesta de muchos genes de otras vías de señalización y porque inicialmente restringimos nuestro análisis a regulones que contenían al menos 20 genes de la firma molecular (ver Materiales y métodos).

Otras vías como Cell cycle (hsa04110) y Focal adhesion (hsa04510), se encuentran dentro del Top 3 de enriquecimientos. También encontramos enriquecidas vías presentes en nuestra firma molecular y que son conocidas por ser importantes en el desarrollo del cáncer, tales como PI3K-AKT signaling pathway (hsa04151), Phospholipase D signaling pathway (hsa04072), and Hedgehog signaling pathway (hsa04340) (Tabla 5.2). En general, estas vías parecen sugerir una señal coordinada hacia la

supervivencia, la proliferación y la diferenciación, que son consistentes con el conocimiento actual de la biología del cáncer (Hanahan and Weinberg, 2011). Se sabe que algunos de los genes que forman parte de estos regulones participan en procesos en los que tiene lugar el crecimiento y la diferenciación celular, es decir, procesos morfológicos (Li et al., 2019). Las funciones y la posible relevancia de estos genes en el contexto del cáncer de mama se comentan en las siguientes secciones.

Tabla 5.2: Análisis de enriquecimiento sobre las vías de KEGG

Vías de KEGG enriquecidas	KEGG ID	FDR	Genes	ER
Pathways in cancer	hsa05200	0.000347	121	1.48
Cell cycle	hsa04110	0.00226	46	1.8
Focal adhesion	hsa04510	0.00386	66	1.58
Glioma	hsa05214	0.00993	27	1.98
Prostate cancer	hsa05215	0.0143	33	1.8
Huntington's disease	hsa05016	0.0148	60	1.51
PI3K-Akt signaling pathway	hsa04151	0.0148	96	1.37
Phospholipase D signaling pathway	hsa04072	0.0148	47	1.58
EGFR tyrosine kinase inhibitor resistance	hsa01521	0.0148	30	1.8
Hedgehog signaling pathway	hsa04340	0.0148	20	2.06

El análisis de sobre representación estadística sobre las vías de KEGG para la red de los 10 regulones principales se realizó con Web Gestalt. El umbral de significación estadística se estableció en $p \leq 0.05$ después de la corrección con FDR. ER: Enrichment Ratio.

Durante los procesos asociados con la remodelación de los tejidos, las señales como los morfógenos, las citocinas y los factores de crecimiento están presentes en el entorno de la célula. Éstos activan vías de transducción de señales que a su vez inducen cambios dentro de la célula (Christian, 2000). En el organismo adulto, la remodelación del tejido ocurre después del daño, o como parte de procesos muy específicos como la lactancia donde la estructura de la glándula mamaria cambia dramáticamente (Hennighausen and Robinson, 2001; Macias and Hinck, 2012).

Este fenómeno comparte una serie de características, entre las que se encuentran la proliferación celular, la migración y la formación de estructuras tisulares como nuevos vasos sanguíneos y linfáticos o epitelios. Sin embargo, no es imposible que algunos componentes de esta maquinaria molecular sean similares entre esos procesos y el desarrollo del tumor donde se forman estructuras similares, aunque con resultados anormales.

5.5. Enriquecimiento de cada regulón en procesos de Gene Ontology

Los procesos más significativamente enriquecidos de cada regulón TMR se presentan en la Tabla 5.3. En las células de cáncer de mama, se observó que HOXA5 activaba al promotor del gen supresor de tumores p53 (Raman et al., 2000). Esto concuerda con la observación de que la expresión de HOXA5 en células de cáncer de mama que expresan p53 de tipo wild type condujo a la apoptosis, mientras que las que carecen del gen p53 no lo hicieron (Raman et al., 2000; Chen et al., 2004). Además, la región promotora HOXA5 fue metilada en el 80 % de las muestras de cáncer de mama p53 negativo.

Esta regulación aberrante de los genes HOX en el cáncer de mama sugiere que los genes HOX pueden ser componentes importantes en una red de mecanismos reguladores de genes relacionados con la homeostasis del tejido adulto (Bhatlekar et al., 2014).

Nueve de nuestros 10 TMR son reconocidos por participar en procesos morfogénicos (Kuroiwa et al., 1996; Srivastava, 1999; Ruiz i Altaba et al., 2007; Melvin et al., 2013; Takeichi et al., 2013; Chojnowski et al., 2014; Amin et al., 2015; Ma-

Tabla 5.3: Principales enriquecimientos significativos de los procesos biológicos de Gene Ontology por regulón.

Regulon	Enriched GO processes	ID	FDR	# Genes	ER
CLOCK	Mitotic cell cycle	GO:0000278	1.39E-02	64	1.84
GLI2	Regulation of cell differentiation	GO:0045595	1.22E-05	73	2.07
HAND2	Cardiovascular system development	GO:0072358	4.31E-06	53	2.52
HAND2	Vasculature development	GO:0001944	4.31E-06	52	2.51
HOXA3	Tube development	GO:0035295	8.94E-05	46	2.5
HOXA5	Proximal/distal pattern formation	GO:0009954	1.69E-02	6	1.98
HOXA5	Anterior/posterior pattern specification	GO:0009952	1.69E-02	12	4.98
HOXA5	Skeletal system development	GO:0001501	1.69E-02	19	3.28
MEIS2	Animal organ morphogenesis	GO:0009887	5.29E-08	107	1.97
PEG3	Cell cycle	GO:0007049	8.72E-08	225	1.53
TBX18	Tissue development	GO:0009888	3.59E-05	139	1.61
TBX18	Blood vessel development	GO:0001568	3.59E-05	62	2.16
TSHZ2	Regulation of cell proliferation	GO:0042127	4.91E-02	45	1.92
TSHZ2	Extracellular matrix organization	GO:0030198	4.91E-02	17	3.32
TSHZ2	Extracellular structure organization	GO:0043062	4.91E-02	17	3.31

Los regulones principales enriquecen principalmente procesos biológicos relacionados con el desarrollo embrionario. El umbral de significancia estadística se estableció arriba de $p \leq 0.05$ después de la corrección FDR. ER: Enrichment ratio.

chon et al., 2015; Jeannotte et al., 2016). Los enriquecimientos de los regulones de los TMR individuales en procesos biológicos GO recuperaron asociaciones entre los regulones de los TMR con procesos morfogénéticos generales.

Los procesos biológicos de GO enriquecidos se obtuvieron a partir de la firma molecular de las vías de transducción de señales. Los primeros lugares están ocupados por procesos asociados con la morfogénesis. Estos resultados concuerdan con la idea de que los tumores tienen aberraciones en procesos como crecimiento, diferenciación y organización de poblaciones celulares. Estos son procesos básicos que están estrechamente coordinados y controlados durante la embriogénesis, así como en los tejidos adultos. (Vinnitsky, 1993) propuso previamente una idea similar, con el nombre de "teoría oncogerminativa del desarrollo del cáncer"(Vinnitsky, 1993). Sugiere que el cáncer surge debido a la expresión aberrante de los genes del desarrollo, y donde la formación del tumor es un proceso dinámico de autoorganización

que produce efectivamente tejido nuevo incluso si está en una forma anormal. La transformación maligna de las células somáticas, que puede comenzar con mutaciones genéticas combinadas con desregulación epigenética, en última instancia puede producir células somáticas reprogramadas en células inmortales que imitan las células de la línea germinal. Estas se denominan células madre cancerosas o células oncogerminalivas (Vinnitsky, 1993; Bhatlekar et al., 2014).

Los enriquecimientos de los procesos biológicos GO para cada regulón individual en el top 10 muestran un tema recurrente en los procesos asociados con la morfogénesis (Kuroiwa et al., 1996; Srivastava, 1999; Ruiz i Altaba et al., 2007; Melvin et al., 2013; Takeichi et al., 2013; Chojnowski et al., 2014; Amin et al., 2015; Machon et al., 2015; Jeannotte et al., 2016). Estos resultados están en concordancia con la idea de que los tumores tienen aberraciones de crecimiento, diferenciación y organización de poblaciones celulares. Aunque nuestros resultados no pueden asegurar la actividad de los procesos morfogenéticos en los tumores, existe una asociación a nivel de los patrones de expresión génica de las moléculas representadas canónicamente en ellos.

5.6. Familia HOXA enriquecida en los regulones

Además, nuestros resultados muestran que 10 de los 11 miembros de la familia Homeobox A están incluidos dentro de los 10 principales regulones RTM (Figura 5.5). En humanos, HOXA consta de 11 genes (HOXA1, HOXA2, HOXA3, HOXA4, HOXA5, HOXA6, HOXA7, HOXA9, HOXA10, HOXA11 y HOXA13). Aunque los genes HOXA codifican proteínas con actividad de factor de transcripción, estos no suelen considerarse componentes de las vías de transducción de señales (Figura 5.5).

Los factores de transcripción HOXA actúan no sólo como activadores transcrip-

	CLOCK	GLI2	HAND2	HOXA2	HOXA3	HOXA5	MEIS2	PEG3	TBX18	TSHZ2	Total
HOXA1		•	•	•	•	•	•	•	•		8
HOXA2	•	•		•	•	•	•	•			7
HOXA3	•	•	•	•	•	•	•	•	•	•	10
HOXA4	•	•	•	•	•	•	•	•	•	•	10
HOXA5		•		•	•		•	•	•		6
HOXA6		•						•			2
HOXA7	•	•	•	•	•	•	•	•	•		9
HOXA9		•		•	•	•	•		•	•	7
HOXA10	•	•		•	•	•	•	•	•	•	9
HOXA11			•	•	•	•		•	•		6
Total	5	9	5	9	9	8	8	9	8	4	
pvalue	8.19E-05	9.78E-13	6.48E-05	8.00E-13	1.64E-11	4.72E-13	1.22E-07	1.69E-07	5.13E-08	5.25E-05	

Figura 5.5: Genes de la familia HOXA presentes en los 10 primeros Reguladores Maestros de las vías de señalización. Numerosos miembros de la familia HOXA son parte de los principales regulones maestros y se encontró un valor p significativo en todos los casos. Los parámetros de la prueba hipergeométrica son: tamaño de la población, $N = 15,802$ genes en la matriz de expresión; número de éxitos en la población, $M = 10$ de 11 genes HOXA humanos en la matriz de expresión; tamaño de la muestra, s es el tamaño del regulón y el número de éxitos y k es el número de genes HOXA presentes en el regulón. Es de destacar que el único gen HOXA humano que no se encuentra en los regulones de todos los factores transcripcionales maestros detectados por MARINa (338) fue HOXA13

cionales en los cánceres sino también como represores transcripcionales (Ladam y Sagerström, 2014); por lo tanto, tanto la regulación positiva como la regulación negativa de los miembros de esta familia pueden ser relevantes en la carcinogénesis. Se ha demostrado que muchos genes HOXA (HOXA1, A2, A3, A5 y A9) tienen niveles de expresión significativamente más bajos en los tejidos cancerosos en comparación con los no cancerosos.

Capítulo 6

Conclusiones

Con la generación de la red regulatoria de genes y la firma molecular centrada en las vías de señalización, presentamos un listado de factores de transcripción que podrían estar modulando las vías de señalización en cáncer de mama. Esta información podría ayudar al estudio de esta enfermedad en la cual las vías analizadas tomadas de la categoría de KEGG “signal transduction” se encontraron en su totalidad desreguladas a nivel de la expresión de genes en un amplio conjunto de muestras de cáncer de mama.

Identificamos a TSHZ2, HOXA2, MEIS2, HOXA3, HAND2, HOXA5, TBX18, PEG3, GLI2, y CLOCK como los 10 principales factores transcripcionales maestros que regulan las vías de transducción de señales en cáncer de mama. Estos genes regulan el 30% de los genes de estas vías. CLOCK es el único factor transcripcional maestro dentro de estos 10 primeros que mostró una regulación positiva hacia sus blancos predichos, el resto de los factores transcripcionales maestros muestran una regulación negativa, los mecanismos a través de los cuales estos factores de transcripción actúan podrían ser explorados en estudios posteriores

El análisis de enriquecimiento sobre la red de formada por los 10 primeros re-

reguladores maestros recuperó información sobre procesos que son bien reconocidos en cáncer como angiogénesis, organogénesis, proliferación, supervivencia etc. Estos resultados son razonables en el sentido de que estamos recuperando información biológica relevante del fenotipo de estudio en vez de asociaciones espurias. Mas aún, el análisis individual de los regulones llevo a la identificación de moléculas específicas que podrían estar jugando papeles claves en esta enfermedad. Es el caso de la familia de los genes HOXA, los cuales se encuentran regulados por los 10 primeros reguladores maestros como parte de sus regulones. Esta familia de genes son conocidos por su papel en procesos de morfogénesis así como en el mantenimiento del tejido adulto y por encontrarse su expresión alterada en los tumores.

Presentamos una versión modificada del algoritmo MRA que utiliza una firma de genes específica centrada en los genes de la categoría de KEGG “signal transduction” . El motivo de enfocarnos en un conjunto reducido de genes fue para encontrar los principales reguladores transcripcionales maestros de las vías de transducción de señales. Sin embargo este enfoque no está limitado a una vía o proceso en particular y la firma molecular puede ser modificada para responder a preguntas relacionadas con otras vías o procesos de interés.

Las vías de transducción de señales tienen un importante rol como integradores de la información celular por lo que sus componentes e interacciones son de gran interés y sujetos de un exhaustivo estudio. Más aún las vías de señalización son susceptibles de modulación farmacológica, por lo que un entendimiento de la regulación de las vías podría ser de utilidad en la búsqueda de blancos terapéuticos.

Capítulo 7

Referencias bibliográficas

Referencias bibliográficas

- Ahmad, N. and Kumar, R. (2011). Steroid hormone receptors in cancer development: a target for cancer therapeutics. *Cancer letters*, 300(1):1–9.
- Alasti, F., Sadeghi, A., Sanati, M. H., Farhadi, M., Stollar, E., Somers, T., and Van Camp, G. (2008). A mutation in *hoxa2* is responsible for autosomal-recessive microtia in an iranian family. *American journal of human genetics*, 82:982–991.
- Alberts, B. (2018). *Molecular biology of the cell*.
- Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enriquez, J., Tovar, H., and Hernández-Lemus, E. (2018). Network modularity and hierarchical structure in breast cancer molecular subtypes. In *International Conference on Complex Systems*, pages 352–358. Springer.
- Alcalá-Corona, S. A., Velázquez-Caldelas, T. E., Espinal-Enríquez, J., and Hernández-Lemus, E. (2016). Community structure reveals biologically functional modules in *mef2c* transcriptional regulatory network. *Frontiers in physiology*, 7:184.
- Allen, B. L. and Taatjes, D. J. (2015). The mediator complex: a central integrator of transcription. *Nature reviews Molecular cell biology*, 16(3):155–166.
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature genetics*, 48(8):838–47.

- Amin, S., Donaldson, I. J., Zannino, D. A., Hensman, J., Rattray, M., Losa, M., Spitz, F., Ladam, F., Sagerström, C., and Bobola, N. (2015). *Hoxa2* selectively enhances meis binding to change a branchial arch ground state. *Developmental cell*, 32:265–277.
- Apiou, F., Flagiello, D., Cillo, C., Malfoy, B., Poupon, M. F., and Dutrillaux, B. (1996). Fine mapping of human hox gene clusters. *Cytogenetics and cell genetics*, 73:114–115.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.
- Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature*, 447(7143):407.
- Bhagwat, A. S. and Vakoc, C. R. (2015). Targeting transcription factors in cancer. *Trends in cancer*, 1(1):53–65.
- Bhatlekar, S., Fields, J. Z., and Boman, B. M. (2014). Hox genes and their role in the development of human cancers. *Journal of molecular medicine*, 92(8):811–823.
- Bianco, R., Melisi, D., Ciardiello, F., and Tortora, G. (2006). Key cancer cell signal transduction pathways as therapeutic targets. *European journal of cancer*, 42(3):290–294.
- Brown, S. A. (2014). Circadian clock-mediated control of stem cell division and differentiation: beyond night and day. *Development (Cambridge, England)*, 141:3105–3111.

- Bushweller, J. H. (2019). Targeting transcription factors in cancer—from undruggable to reality. *Nature Reviews Cancer*, 19(11):611–624.
- Carroll, J. S. and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. *Molecular Endocrinology*, 20(8):1707–1714.
- Chan, S. S.-K. and Kyba, M. (2013). What is a master regulator? *Journal of stem cell research & therapy*, 3.
- Chawla, K., Tripathi, S., Thommesen, L., Læg Reid, A., and Kuiper, M. (2013). Tf-checkpoint: a curated compendium of specific dna-binding rna polymerase ii transcription factors. *Bioinformatics*, 29(19):2519–2520.
- Chen, H., Chung, S., and Sukumar, S. (2004). HOXA5-induced apoptosis in breast cancer cells is mediated by caspases 2 and 8. *Molecular and cellular biology*, 24(2):924–935.
- Chojnowski, J. L., Masuda, K., Trau, H. A., Thomas, K., Capecchi, M., and Manley, N. R. (2014). Multiple roles for *hoxa3* in regulating thymus and parathyroid differentiation and morphogenesis in mouse. *Development*, pages dev–110833.
- Christian, J. L. (2000). Bmp, wnt and hedgehog signals: how far can they go? *Current opinion in cell biology*, 12(2):244–249.
- Clement, V., Sanchez, P., de Tribolet, N., Radovanovic, I., and Ruiz i Altaba, A. (2007). Hedgehog-gli1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. *Current biology : CB*, 17:165–172.
- Cohen, P. (2000). The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends in biochemical sciences*, 25(12):596–601.
- Darnell Jr, J. E. (2002). Transcription factors as targets for cancer therapy. *Nature Reviews Cancer*, 2(10):740.

- Dembo, A., Cover, T. M., and Thomas, J. A. (1991). Information theoretic inequalities. *IEEE Transactions on Information theory*, 37(6):1501–1518.
- Dhanasekaran, N. (1998). Cell signaling: an overview. *Oncogene*, 17(11):1329.
- Di Domenico, M. and Giordano, A. (2017). Signal transduction growth factors: the effective governance of transcription and cellular adhesion in cancer invasion. *Oncotarget*, 8(22):36869.
- Diamond, M. I., Miner, J. N., Yoshinaga, S. K., and Yamamoto, K. R. (1990). Transcription factor interactions: selectors of positive or negative regulation from a single dna element. *Science*, 249(4974):1266–1272.
- Drier, Y., Sheffer, M., and Domany, E. (2013). Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 110:6388–6393.
- Farin, H. F., Bussen, M., Schmidt, M. K., Singh, M. K., Schuster-Gossler, K., and Kispert, A. (2007). Transcriptional repression by the t-box proteins tbx18 and tbx15 depends on groucho corepressors. *The Journal of biological chemistry*, 282:25748–25759.
- Fukuda, T., Shirane, A., Wada-Hiraike, O., Oda, K., Tanikawa, M., Sakuabashi, A., Hirano, M., Fu, H., Morita, Y., Miyamoto, Y., Inaba, K., Kawana, K., Osuga, Y., and Fujii, T. (2015). Hand2-mediated proteolysis negatively regulates the function of estrogen receptor . *Molecular medicine reports*, 12:5538–5544.
- García-Campos, M. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Pathway analysis: state of the art. *Frontiers in physiology*, 6:383.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open soft-

- ware development for computational biology and bioinformatics. *Genome biology*, 5(10):R80.
- Green, S., Şerban, M., Scholl, R., Jones, N., Brigandt, I., and Bechtel, W. (2018). Network analyses in systems biology: new strategies for dealing with biological complexity. *Synthese*, 195(4):1751–1777.
- Gupta, P., Ranjan, A., Fofaria, N. M., and Srivastava, S. K. (2015). Gli2 inhibition suppresses metastasis of her2+ breast cancer.
- Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, 144(5):646–674.
- Harrison, K., Hoad, G., Scott, P., Simpson, L., Horgan, G. W., Smyth, E., Heys, S. D., and Haggarty, P. (2015). Breast cancer risk and imprinting methylation in blood. *Clinical epigenetics*, 7:92.
- Hennighausen, L. and Robinson, G. W. (2001). Signaling pathways in mammary gland development. *Developmental cell*, 1(4):467–475.
- Hernández-Lemus, E. and Siqueiros-García, J. M. (2013). Information theoretical methods for complex network structure reconstruction. *Complex Adaptive Systems Modeling*, 1(1):1–22.
- Hong, C. S., Zhao, X.-F., Moon, S. M., Lee, H. W., Lee, S. R., and Park, Y. K. (2010). Down-regulation of meis2 leads to apoptosis in human colorectal cancer cells.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with bioconductor. *Nature methods*, 12(2):115–121.
- Hunter, T. and Karin, M. (1992). The regulation of transcription by phosphorylation. *Cell*, 70:375–387.

- Hur, H., Lee, J.-Y., Yun, H. J., Park, B. W., and Kim, M. H. (2014). Analysis of hox gene expression patterns in human breast cancer. *Molecular biotechnology*, 56(1):64–71.
- Javelaud, D., Alexaki, V. I., Dennler, S., Mohammad, K. S., Guise, T. A., and Mauviel, A. (2011). Tgf-/smad/gli2 signaling axis in cancer progression and metastasis. *Cancer research*, 71:5606–5610.
- Jeannotte, L., Gotti, F., and Landry-Truchon, K. (2016). Hoxa5: a key player in development and disease. *Journal of developmental biology*, 4(2):13.
- Jiang, X., Yu, Y., Yang, H. W., Agar, N. Y. R., Frado, L., and Johnson, M. D. (2010). The imprinted gene peg3 inhibits wnt signaling and regulates glioma growth. *The Journal of biological chemistry*, 285:8472–8480.
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333–339.
- Kanehisa, M. and Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28:27–30.
- Kansanen, E., Kuosmanen, S. M., Leinonen, H., and Levonen, A.-L. (2013). The keap1-nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox biology*, 1:45–49. Original DateCompleted: 20140624.
- Karhadkar, S. S., Bova, G. S., Abdallah, N., Dhara, S., Gardner, D., Maitra, A., Isaacs, J. T., Berman, D. M., and Beachy, P. A. (2004). Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature*, 431:707–712.
- Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S. R., Tan, G., et al. (2017). Jasp

2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research*, 46(D1):D260–D266.

Kohda, T., Asai, A., Kuroiwa, Y., Kobayashi, S., Aisaka, K., Nagashima, G., Yoshida, M. C., Kondo, Y., Kagiya, N., Kirino, T., Kaneko-Ishino, T., and Ishino, F. (2001). Tumour suppressor activity of human imprinted gene *peg3* in a glioma cell line. *Genes to cells : devoted to molecular & cellular mechanisms*, 6:237–247.

Kolch, W., Halasz, M., Granovskaya, M., and Kholodenko, B. N. (2015). The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer*, 15(9):515.

Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L. L., Tada, M., Suzuki, R., Yokoyama, M., Shiroishi, T., Wakana, S., Barton, S. C., Ishino, F., and Surani, M. A. (1996). *Peg3* imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. *Nature genetics*, 12:186–190.

Latchman, D. S. (1997). Transcription factors: an overview. *The international journal of biochemistry & cell biology*, 29:1305–1312.

Laurent, A., Calabrese, M., Warnatz, H.-J., Yaspo, M.-L., Tkachuk, V., Torres, M., Blasi, F., and Penkov, D. (2015). Chip-seq and rna-seq analyses identify components of the wnt and fgf signaling pathways as *prep1* target genes in mouse embryonic stem cells. *PloS one*, 10(4):e0122518.

Lee, S., Ye, A., and Kim, J. (2015). Dna-binding motif of the imprinted transcription factor *peg3*. *PloS one*, 10:e0145531.

Lefebvre, C., Rajbhandari, P., Alvarez, M. J., Bandaru, P., Lim, W. K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B. C., Basso, K., Beltrao, P., Krogan, N., Gautier, J., Dalla-Favera, R., and Califano, A. (2010). A human b-cell interactome identifies *myb* and *foxm1* as master regulators of proliferation in germinal centers. *Molecular systems biology*, 6:377.

- Li, B., Huang, Q., and Wei, G.-H. (2019). The role of hox transcription factors in cancer predisposition and progression. *Cancers*, 11(4):528.
- Lu, J., Wang, W., Xu, M., Li, Y., Chen, C., and Wang, X. (2017). A global view of regulatory networks in lung cancer: An approach to understand homogeneity and heterogeneity. In *Seminars in Cancer Biology*, volume 42, pages 31–38. Elsevier.
- Lu, X. and Kang, Y. (2010). Hypoxia and hypoxia-inducible factors (hifs): master regulators of metastasis. *Clinical cancer research*, pages clincanres–1360.
- Lussier, D. M. and Schreiber, R. (2016). Cancer immunosurveillance: immunoediting. In *Immunity to Pathogens and Tumors*, pages 396–405. Elsevier Inc.
- Mace, K. A., Hansen, S. L., Myers, C., Young, D. M., and Boudreau, N. (2005). Hoxa3 induces cell migration in endothelial and epithelial cells promoting angiogenesis and wound repair. *Journal of cell science*, 118:2567–2577.
- Machon, O., Masek, J., Machonova, O., Krauss, S., and Kozmik, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. *BMC developmental biology*, 15(1):40.
- Macias, H. and Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(4):533–557.
- Makiyama, K., Hamada, J.-i., Takada, M., Murakawa, K., Takahashi, Y., Tada, M., Tamoto, E., Shindo, G., Matsunaga, A., Teramoto, K.-i., Komuro, K., Kondo, S., Katoh, H., Koike, T., and Moriuchi, T. (2005). Aberrant expression of HOX genes in human invasive breast carcinoma. *Oncology Reports*, 13(4):673–679.
- Mann, M. and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature biotechnology*, 21(3):255.
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). Aracne: an algorithm for the reconstruction of

- gene regulatory networks in a mammalian cellular context. *BMC bioinformatics*, 7 Suppl 1:S7.
- Martin, G. S. (2003). Cell signaling and cancer. *Cancer cell*, 4(3):167–174.
- Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). Transfac® and its module transcompel®: transcriptional gene regulation in eukaryotes. *Nucleic acids research*, 34(suppl_1):D108–D110.
- Melvin, V. S., Feng, W., Hernandez-Lagunas, L., Artinger, K. B., and Williams, T. (2013). A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. *Developmental Dynamics*, 242(7):817–831.
- Merabet, S. and Lohmann, I. (2015). Toward a new twist in hox and tale dna-binding specificity. *Developmental cell*, 32(3):259–261.
- Mohibi, S., Mirza, S., Band, H., and Band, V. (2011). Mouse models of estrogen receptor-positive breast cancer. *Journal of carcinogenesis*, 10.
- Morgan, R. and Pandha, H. S. (2017). HOX transcription factors and the prostate tumor microenvironment. *Journal of Cancer Metastasis and Treatment*, 3(12):278.
- Nebert, D. W. (2002). Transcription factors and cancer: an overview. *Toxicology*, 181:131–141.
- Ordóñez-Morán, P., Dafflon, C., Imajo, M., Nishida, E., and Huelsken, J. (2015). Hoxa5 counteracts stem cell traits by inhibiting wnt signaling in colorectal cancer. *Cancer cell*, 28:815–829.
- Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology*, 27(8):1160.

- Piceci, F., Morlino, S., Castori, M., Buffone, E., De Luca, A., Grammatico, P., and Guida, V. (2017). Identification of a second *hoxa2* nonsense mutation in a family with autosomal dominant non-syndromic microtia and distinctive ear morphology. *Clinical genetics*, 91:774–779.
- Pontén, F., Jirstrom, K., and Uhlen, M. (2008). The human protein atlas—a tool for pathology. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 216(4):387–393.
- Raman, V., Martensen, S. A., Reisman, D., Evron, E., Odenwald, W. F., Jaffee, E., Marks, J., and Sukumar, S. (2000). Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature*, 405(6789):974–978.
- Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *nature*, 414(6859):105.
- Riku, M., Inaguma, S., Ito, H., Tsunoda, T., Ikeda, H., and Kasai, K. (2016). Down-regulation of the zinc-finger homeobox protein *tshz2* releases *gli1* from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget*, 7:5690–5701.
- Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). Gc-content normalization for rna-seq data. *BMC bioinformatics*, 12(1):480.
- Rosen, L. S., Ashurst, H. L., and Chap, L. (2010). Targeting signal transduction pathways in metastatic breast cancer: a comprehensive review. *The oncologist*, 15(3):216.
- Ruiz i Altaba, A., Mas, C., and Stecca, B. (2007). The gli code: an information nexus regulating cell fate, stemness and cancer. *Trends in cell biology*, 17:438–447.
- Salem, H., Attiya, G., and El-Fishawy, N. (2017). Classification of human cancer diseases by gene expression profiles. *Applied Soft Computing*, 50:124–134.

- Sampieri-Cabrera, R. (2020). Transducción de señales (signal transduction). *Available at SSRN 3559063*.
- Scales, S. J. and de Sauvage, F. J. (2009). Mechanisms of hedgehog pathway activation in cancer and implications for therapy. *Trends in pharmacological sciences*, 30:303–312.
- Siegel, R. L., Miller, K. D., and Jemal, A. (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1):5–29.
- Siena, S., Sartore-Bianchi, A., Di Nicolantonio, F., Balfour, J., and Bardelli, A. (2009). Biomarkers predicting clinical outcome of epidermal growth factor receptor–targeted therapy in metastatic colorectal cancer. *Journal of the National Cancer Institute*, 101(19):1308–1324.
- Srivastava, D. (1999). Hand proteins: molecular mediators of cardiac development and congenital heart disease. *Trends in cardiovascular medicine*, 9:11–18.
- Stormo, G. D. (2013). Modeling the specificity of protein-dna interactions. *Quantitative biology*, 1(2):115–130.
- Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239):719–724.
- Su, X., Gi, Y. J., Chakravarti, D., Chan, I. L., Zhang, A., Xia, X., Tsai, K. Y., and Flores, E. R. (2012). Tap63 is a master transcriptional regulator of lipid and glucose metabolism. *Cell metabolism*, 16:511–525.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., and Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102(43):15545–15550.

- Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., and Abola, E. E. (1998). Protein data bank (pdb): database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*, 54(6):1078–1084.
- Svingen, T. and Tonissen, K. F. (2003). Altered hox gene expression in human skin and breast cancer cells. *Cancer biology & therapy*, 2:518–523.
- Taipale, J. and Beachy, P. A. (2001). The Hedgehog and Wnt signalling pathways in cancer. *Nature*, 411(6835):349–354.
- Takeichi, M., Nimura, K., Mori, M., Nakagami, H., and Kaneda, Y. (2013). The transcription factors tbx18 and wt1 control the epicardial epithelial-mesenchymal transition through bi-directional regulation of slug in murine primary epicardial cells. *PLoS one*, 8:e57829.
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in rna-seq: a matter of depth. *Genome research*, 21:2213–2223.
- Team, R. C. et al. (2013). R: A language and environment for statistical computing.
- Teo, W. W., Merino, V. F., Cho, S., Korangath, P., Liang, X., Wu, R.-C., Neumann, N. M., Ewald, A. J., and Sukumar, S. (2016). Hoxa5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of e-cadherin and cd24. *Oncogene*, 35:5539–5551.
- Thayer, S. P., di Magliano, M. P., Heiser, P. W., Nielsen, C. M., Roberts, D. J., Lauwers, G. Y., Qi, Y. P., Gysin, S., Fernández-del Castillo, C., Yajnik, V., Antoniu, B., McMahon, M., Warshaw, A. L., and Hebrok, M. (2003). Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature*, 425:851–856.

- Tovar, H., García-Herrera, R., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Transcriptional master regulator analysis in breast cancer genetic networks. *Computational biology and chemistry*, 59 Pt B:67–77.
- Tripathi, S., Christie, K. R., Balakrishnan, R., Huntley, R., Hill, D. P., Thommesen, L., Blake, J. A., Kuiper, M., and Lægreid, A. (2013). Gene ontology annotation of sequence-specific dna binding transcription factors: setting the stage for a large-scale curation effort. *Database : the journal of biological databases and curation*, 2013:bat062.
- Tvorogov, D., Sundvall, M., Kurppa, K., Hollmén, M., Repo, S., Johnson, M. S., and Elenius, K. (2009). Somatic mutations of *erbb4* selective loss-of-function phenotype affecting signal transduction pathways in cancer. *Journal of Biological Chemistry*, 284(9):5582–5591.
- VanDusen, N. J., Casanovas, J., Vincentz, J. W., Firulli, B. A., Osterwalder, M., Lopez-Rios, J., Zeller, R., Zhou, B., Grego-Bessa, J., De La Pompa, J. L., Shou, W., and Firulli, A. B. (2014). *Hand2* is an essential regulator for two notch-dependent functions within the embryonic endocardium. *Cell reports*, 9:2071–2083.
- Velazquez-Caldelas, T. E., Alcalá-Corona, S. A., Espinal-Enríquez, J., and Hernandez-Lemus, E. (2019). Unveiling the link between inflammation and adaptive immunity in breast cancer. *Frontiers in immunology*, 10:56.
- Vinnitsky, V. B. (1993). Oncogerminative hypothesis of tumor formation. *Medical Hypotheses*, 40(1):19–27.
- Vivante, A., Kleppa, M.-J., Schulz, J., Kohl, S., Sharma, A., Chen, J., Shril, S., Hwang, D.-Y., Weiss, A.-C., Kaminski, M. M., Shukrun, R., Kemper, M. J., Lehnhardt, A., Beetz, R., Sanna-Cherchi, S., Verbitsky, M., Gharavi, A. G., Stuart, H. M., Feather, S. A., Goodship, J. A., Goodship, T. H. J., Woolf, A. S., Westra,

- S. J., Doody, D. P., Bauer, S. B., Lee, R. S., Adam, R. M., Lu, W., Reutter, H. M., Kehinde, E. O., Mancini, E. J., Lifton, R. P., Tasic, V., Lienkamp, S. S., Jüppner, H., Kispert, A., and Hildebrandt, F. (2015). Mutations in *tbx18* cause dominant urinary tract malformations via transcriptional dysregulation of ureter development. *American journal of human genetics*, 97:291–301.
- Ward, P. S. and Thompson, C. B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer cell*, 21(3):297–308.
- Watari-Goshima, N. and Chisaka, O. (2011). Chicken *hoxa3* gene: its expression pattern and role in branchial nerve precursor cell migration. *International journal of biological sciences*, 7:87–101.
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J. M., Network, C. G. A. R., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113.
- Whitmarsh, A. and Davis, R. (2000). Regulation of transcription factor function by phosphorylation. *Cellular and Molecular Life Sciences CMLS*, 57(8-9):1172–1183.
- Wu, S.-P., Dong, X.-R., Regan, J. N., Su, C., and Majesky, M. W. (2013). *Tbx18* regulates development of the epicardium and coronary vessels. *Developmental biology*, 383:307–320.
- Yamamoto, M., Cid, E., Bru, S., and Yamamoto, F. (2011). Rare and frequent promoter methylation, respectively, of *tshz2* and *3* genes that are both downregulated in expression in breast and prostate cancers. *PloS one*, 6:e17149.
- Zardawi, S. J., O’Toole, S. A., Sutherland, R. L., and Musgrove, E. A. (2009). Dysregulation of hedgehog, wnt and notch signalling pathways in breast cancer. *Histology and histopathology*, 24:385–398.

- Zha, Y., Xia, Y., Ding, J., Choi, J., Yang, L., Dong, Z., Yan, C., Huang, S., and Ding, H. (2014). Meis2 is essential for neuroblastoma cell survival and proliferation by transcriptional control of m-phase progression. *Cell death & disease*, 5(9):e1417.
- Zhang, B., Kirov, S., and Snoddy, J. (2005). Webgestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research*, 33:W741–W748.
- Zhang, X., Liu, G., Ding, L., Jiang, T., Shao, S., Gao, Y., and Lu, Y. (2018). Hoxa3 promotes tumor growth of human colon cancer through activating egfr/ras/raf/mek/erk signaling pathway. *Journal of cellular biochemistry*, 119:2864–2874.
- Zhao, C., Chen, A., Jamieson, C. H., Fereshteh, M., Abrahamsson, A., Blum, J., Kwon, H. Y., Kim, J., Chute, J. P., Rizzieri, D., Munchhof, M., VanArsdale, T., Beachy, P. A., and Reya, T. (2009). Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. *Nature*, 458:776–779.
- Zhou, Y., Wu, H., Zhao, M., Chang, C., and Lu, Q. (2016). The bach family of transcription factors: a comprehensive review. *Clinical reviews in allergy & immunology*, 50(3):345–356.

Anexo: artículo científico para la obtención del grado

Tapia-Carrillo, D., Tovar, H., Velázquez-Caldelas, T., Hernández-Lemus, E. (2019). Master regulators of signaling pathways: An application to the analysis of gene regulation in breast cancer. *Frontiers in genetics*, 10, 1180.



Master Regulators of Signaling Pathways: An Application to the Analysis of Gene Regulation in Breast Cancer

Diana Tapia-Carrillo^{1,2}, *Hugo Tovar*¹, *Tadeo Enrique Velazquez-Caldelas*¹ and *Enrique Hernandez-Lemus*^{1,3*}

¹ Computational Genomics Department, National Institute of Genomic Medicine (INMEGEN), Mexico City, Mexico, ² Graduate Program in Biological Sciences, National Autonomous University of Mexico (UNAM), Mexico City, Mexico, ³ Center for Complexity Sciences, National Autonomous University of Mexico (UNAM), Mexico City, Mexico

OPEN ACCESS

Edited by:

Matteo Barberis,
University of Surrey,
United Kingdom

Reviewed by:

Hiroyuki Kubota,
Kyushu University, Japan
Renee Van Amerongen,
University of Amsterdam,
Netherlands

*Correspondence:

Enrique Hernandez-Lemus
ehernandez@inmegen.gob.mx

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 27 May 2019

Accepted: 24 October 2019

Published: 03 December 2019

Citation:

Tapia-Carrillo D, Tovar H, Velazquez-Caldelas TE and Hernandez-Lemus E (2019) Master Regulators of Signaling Pathways: An Application to the Analysis of Gene Regulation in Breast Cancer. *Front. Genet.* 10:1180. doi: 10.3389/fgene.2019.01180

Analysis of gene regulatory networks allows the identification of master transcriptional factors that control specific groups of genes. In this work, we inferred a gene regulatory network from a large dataset of breast cancer samples to identify the master transcriptional factors that control the genes within signal transduction pathways. The focus in a particular subset of relevant genes constitutes an extension of the original Master Regulator Inference Algorithm (MARINA) analysis. This modified version of MARINA utilizes a restricted molecular signature containing genes from the 25 human pathways in KEGG's signal transduction category. Our breast cancer RNAseq expression dataset consists of 881 samples comprising tumors and normal mammary gland tissue. The top 10 master transcriptional factors found to regulate signal transduction pathways in breast cancer we identified are: TSHZ2, HOXA2, MEIS2, HOXA3, HAND2, HOXA5, TBX18, PEG3, GLI2, and CLOCK. The functional enrichment of the regulons of these master transcriptional factors showed an important proportion of processes related to morphogenesis. Our results suggest that, as part of the aberrant regulation of signaling pathways in breast cancer, pathways similar to the regulation of cell differentiation, cardiovascular system development, and vasculature development may be dysregulated and co-opted in favor of tumor development through the action of these transcription factors.

Keywords: breast cancer, master regulator, signaling pathways, transcription factors, development

INTRODUCTION

Breast cancers originate from healthy cells that are somehow reprogrammed to acquire unlimited proliferation and self-renewal capacity, among other properties, altogether referred to as "hallmarks of cancer" (Hanahan and Weinberg, 2011). These processes are the result of highly specific molecular interactions. In this context, it seems reasonable that cancerous cells make use of existing pathways through aberrant modulation mechanisms. Transcriptional regulation may play an important role in such altered mechanisms (Kolch et al., 2015).

Signal transduction pathways (STPs) are intricate molecular mechanisms that allow cells to sense specific signals, producing cellular actions in response and serve an important role in the integration of information (Dhanasekaran, 1998). Among these actions, the activation of transcription factors

through STPs can modify the expression of genes with varying degrees of phenotypical effect.

STPs themselves can be regulated through the action of transcription factors (TFs) that modulate the transcription of groups of genes participating in them (Carroll and Brown, 2006; Laurent et al., 2015; Morgan and Pandha, 2017). As STPs are susceptible to external pharmacological modulation, an understanding of the regulation of the pathways may be helpful in the search for therapeutic targets (Steven Martin., 2003).

The analysis of the structure of a gene regulatory network that contains TFs and their targets, together with information of differential expression values, allows the identification of TFs with the greatest influence over the differences in expression. Those TFs are denominated transcriptional master regulators (TMRs). The Master Regulator Inference Algorithm (MARINa) (Lefebvre et al., 2010) can infer the TFs with greater influence in the transition between two conditions, as is the case with normal breast and breast cancer phenotype (Lefebvre et al., 2010; Tovar et al., 2015). In this work, we used a modified version of this algorithm to find the most important transcription factors associated with genes of known and well-curated signal transduction pathways obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000).

In breast cancer, where multiple transcription factors with hundreds or even thousands of targets are simultaneously deregulated, an integrative approach can help us to understand the biology underlying this disease. Modified MARINa analysis is complemented with the use of statistical enrichment analysis to look at the possible biological functions that are affected by the regulons of the TMRs (García-Campos et al., 2015). The information of biological knowledge databases was used (KEGG (Kanehisa and Goto, 2000) and Gene Ontology (GO) (Ashburner et al., 2000)). Because the breadth of gene expression in the cell, the focus in a particular subset of processes or pathways has the advantage of giving us specific and less complex information about relevant processes and regulation in the cancerous phenotype.

MATERIALS AND METHODS

Obtaining and Preprocessing Data

The Expression matrix was obtained from Espinal-Enriquez et al. (Espinal-Enriquez et al., 2017). The data corresponds to The Cancer Genome Atlas (TCGA) level 3 available data of the Illumina HiSeq RNA-Seq platform, and it consists of 881 samples (**Supplementary Datasheet 1**), of which 780 are from breast cancer tissue and 101 from adjacent healthy mammary tissue. Quality control and batch effect removal were performed with *NOISeq* (Tarazona et al., 2011) and *EDASeq* (Risso et al., 2011) R libraries, respectively (Espinal-Enriquez et al., 2017).

Pathway Deregulation Analysis

To determine if signal transduction pathways are deregulated at the level of gene expression in our dataset of breast cancer, we estimated the degree of deregulation of KEGG Signal Transduction pathways by using the *Pathifier* algorithm (Drier et al., 2013).

Pathifier assigns a score, denominated pathway deregulation score (PDS) for each pathway in a sample. For this, the expression status of the genes in the pathway is evaluated with reference to its expression in normal tissues of the same origin. In brief, for a given pathway, a multidimensional space is defined where each dimension represents the expression level of a gene. All samples are positioned in this space according to the expression levels of all the genes in the pathway. Then, a principal curve (a smoothed curve of minimal distance to all points) is calculated and all samples are projected into it. The score corresponds to the distance of the sample projection measured over the principal curve with respect to the projection of the normal tissue samples (Drier et al., 2013).

The Master Regulator Inference Algorithm

TMRs were inferred using the MARINa (Lefebvre et al., 2010). MARINa identifies TMRs through an enrichment of TF regulons (a TF with its targets) with differentially expressed genes between the two phenotypes (breast cancer vs. adjacent healthy mammary tissue). TMR inference with MARINa requires as input a network of regulons, a gene expression, molecular signature, and a null model (Lefebvre et al., 2010) (**Figure 1**). The construction of these elements is described below.

Generation of the Regulons Network

A regulon is defined here as a directed network where interactions describe regulatory interactions from a transcription factor to its transcriptional targets (TF Target). A regulons network is a regulon set, which is formed by the union of many regulons.

To obtain a regulon set from the data, we used the expression matrix of the tumor samples and a list of transcription factors in the TFCheckpoint curated database (Tripathi et al., 2013). From this database we selected those TFs that had experimental evidence for TF activity. A total of 771 of these TFs were expressed in breast cancer samples and present in the expression matrix (**Supplementary Datasheet 2**).

As a first step, transcription factors are associated with other genes expressed in the tissue. We used the mutual information-based algorithm ARACNe (Margolin et al., 2006) which calculates the pairwise mutual information for a pair of genes using the empirical probability distributions of their expression levels. For this network all possible interactions between TFs and genes in the expression matrix were calculated and kept if its *p*-value was below 0.005.

Mutual information can detect both indirect and direct relationships. ARACNe constrains the number of indirect interactions applying the data processing inequality theorem (DPI), which considers that, in a triangle of interactions, the weakest one has a greater probability of being indirect if its difference is large with respect to the other two interactions (Hernández-Lemus and Siqueiros-García, 2013). We applied a DPI value of 0.2 as recommended in Margolin et al., 2006 (Margolin et al., 2006), which means that the weakest interactions of the triangles in the network were eliminated without introducing an excessive number of false positives.

The type of association (activation or repression) of the transcription factors is determined from the Spearman

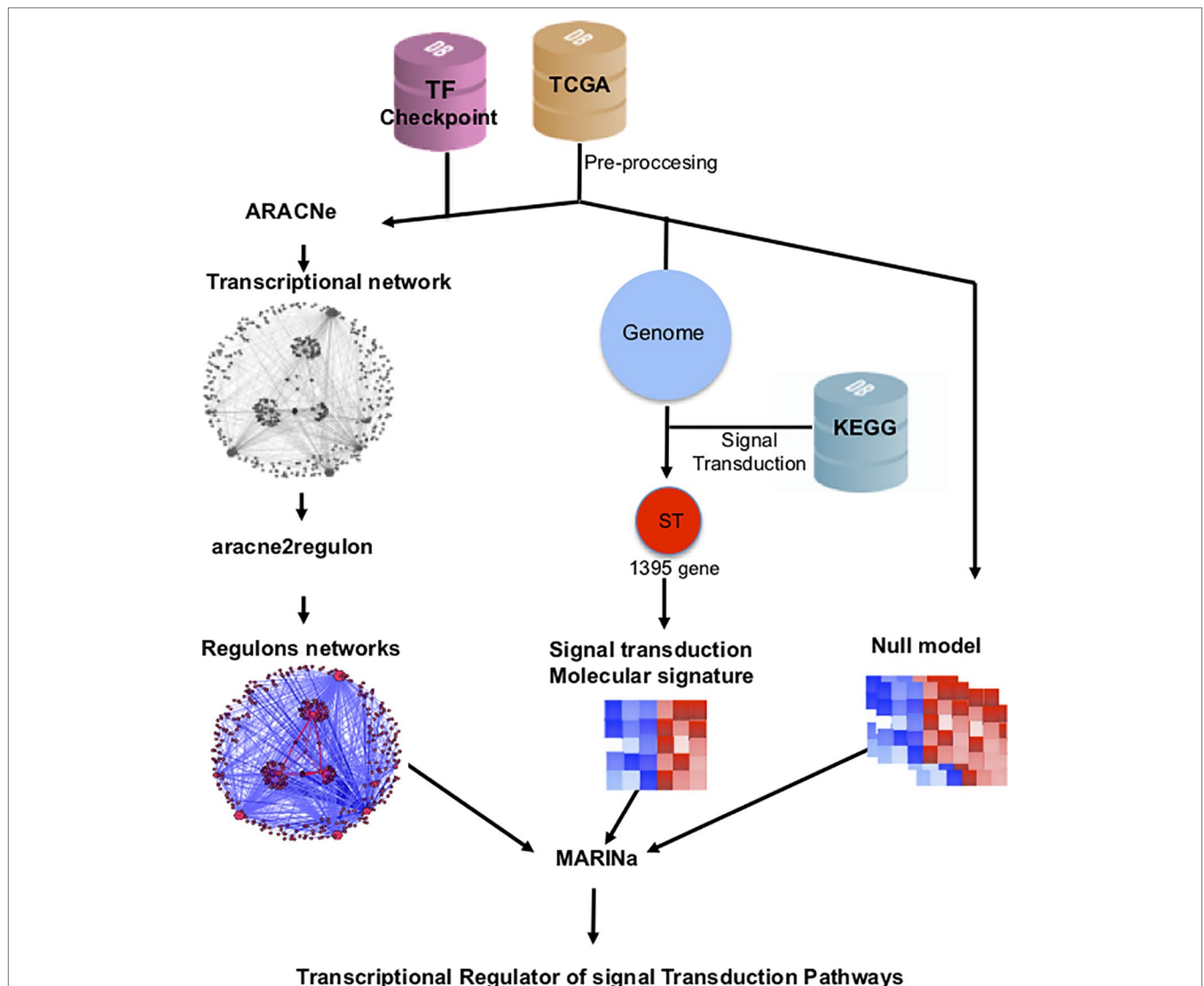


FIGURE 1 | Customized MARINa pipeline. RNAseq data from TCGA's 780 invasive mammary carcinomas and 101 adjacent tissue samples was processed to obtain an expression matrix (orange cylinder). The expression matrix and a list of transcription factors from the TFCheckpoint database (pink cylinder) served as input to infer a transcriptional regulatory network with ARACNe. A regulon network was obtained associating the expression level of the targets of all transcription factors using the *aracne2regulon* function from *viper* (left side). For the generation of the molecular signature, we considered genes in the expression matrix in KEGG's "signal transduction" category (blue cylinder). Finally, a null model was generated by permuting sample labels and recalculating the molecular signature (right). These three elements are the input to MARINa for the inference of the transcriptional master regulators (TMRs) of the signal transduction pathways.

correlation of the TF with the levels of expression of all its targets (Lefebvre et al., 2010). This calculation was performed by the *aracne2regulon* function in the *viper* R package (Alvarez et al., 2016). This function transforms the undirected MI network from ARACNE into a regulons network that is directed.

Molecular Signature Generation of Signal Transduction Pathways

In the standard MARINa workflow, the molecular signature is built by comparing the expression level distributions of all genes between two conditions (e.g., healthy and diseased). For this work

we built a molecular signature using only those genes annotated within the *signal transduction pathways* category in the KEGG database (Kanehisa and Goto, 2000). For the human species, this category contains 25 curated pathways. The total number of genes present in this subset is 1,700, of which 1,395 are included in the expression matrix (Supplementary Datasheet 3). The purpose of this filtering is to focus our search on those transcription factors that regulate the activity of these signal transduction pathways in breast cancer. The molecular signature was built by applying a *t*-test to each gene of the expression matrix, between tumors and adjacent healthy mammary tissue. The results of this test were Z-score normalized to allow comparability (Lefebvre et al., 2010).

Null Model Generation

To estimate the probability that a gene set enrichment score depends on the biological context and thus is not merely random, a null model was generated by random permutation between cases and control samples and recalculating differential expression values (Lefebvre et al., 2010).

Inferring the Master Regulators of Signal Transduction Pathways

With the molecular signature, the regulon network and the null model, MARINa estimated the top regulons that enrich the most differentially expressed genes in the molecular signature through a gene set enrichment analysis (Subramanian et al., 2005). An additional constraint was to consider only TFs with 20 or more targets in the molecular signature (Lefebvre et al., 2010). A p value for each regulon was estimated by evaluating the enrichment score (ES) with reference to the distribution of scores of the null model (Lefebvre et al., 2010). For TMR inference we used *viper* package (Alvarez et al., 2016), an R implementation of MARINa available *via* Bioconductor.

The difference of this work with respect to MARINa lies in the use of a specific set of genes (signal transduction signature) which produces a ranking of the regulons for this specific subset. It is important to note that the regulons network used as input is the same as in regular MARINa, but the ranking is focused on the specific gene signature. The set of genes that constitute each regulon may include genes that are not in the molecular signature and can be part of a more diverse range of biological functions than signal transduction. This is the reason why we performed a subsequent enrichment analysis of the regulons with all KEGG human pathways.

Functional Enrichment of the Top 10 Regulons Network With KEGG Pathways

An overrepresented pathway is defined as one for which we found significantly more genes within a test set than the number expected from a random sampling (García-Campos et al., 2015) hence, we say this set is enriched with genes of the pathway—this may in turn suggest biological relevance. The statistical significance of an enrichment can be assessed by means of an hypergeometric test. In order to know if the combined regulons of our top 10 transcription master regulators are enriched for biological pathways, an overrepresentation enrichment analysis (ORA) was performed using the WebGestalt web platform (Zhang et al., 2005) with KEGG as the functional reference database (Kanehisa and Goto, 2000). Statistical significance threshold was set $top = 0.05$ after false discovery rate (FDR) correction.

The interrogation of the network for overrepresented pathways can evidence which of the original signal transduction pathways are being regulated. The genetic composition of the regulons is determined by the statistical dependencies between expression levels of the transcription factor genes and all other genes expressed in the tissue. Although we know the identity of the TFs, there is no guarantee that all transcription factor genes and the gene signature will be present in the network. This means that the clustering of the genes is not known *a priori* or imposed

from a biological knowledge database like KEGG or GO. The co-expression of genes belonging to a function or pathway in different network modules has been previously observed (Alcalá-Corona et al., 2018).

Regulon Enrichment of Gene Ontology Biological Processes

To gain insight on how our top 10 TMRs may contribute to this phenotype, we performed an ORA for each of the individual regulons with GO (Ashburner et al., 2000) biological process categories. GO was used because this database considers processes with various degrees of specificity, from very general processes expected in all cells to very specific subsets of a process (i.e., positive and negative regulation) and provides a reference that is independent from our original KEGG gene lists. GO biological process enrichments were calculated with WebGestalt (Zhang et al., 2005). Statistical significance threshold after FDR correction was set to $p \leq 0.05$.

RESULTS AND DISCUSSION

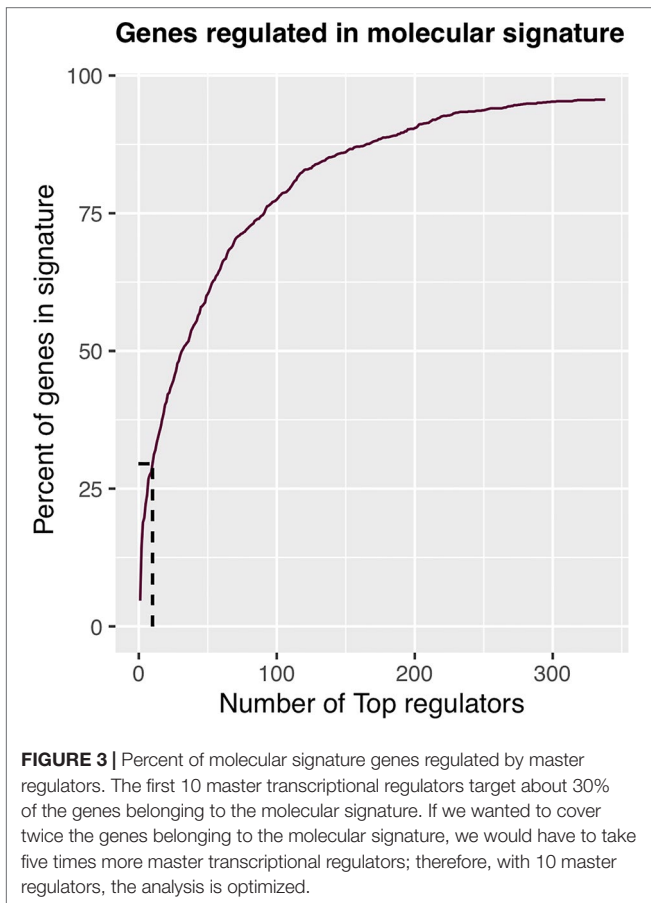
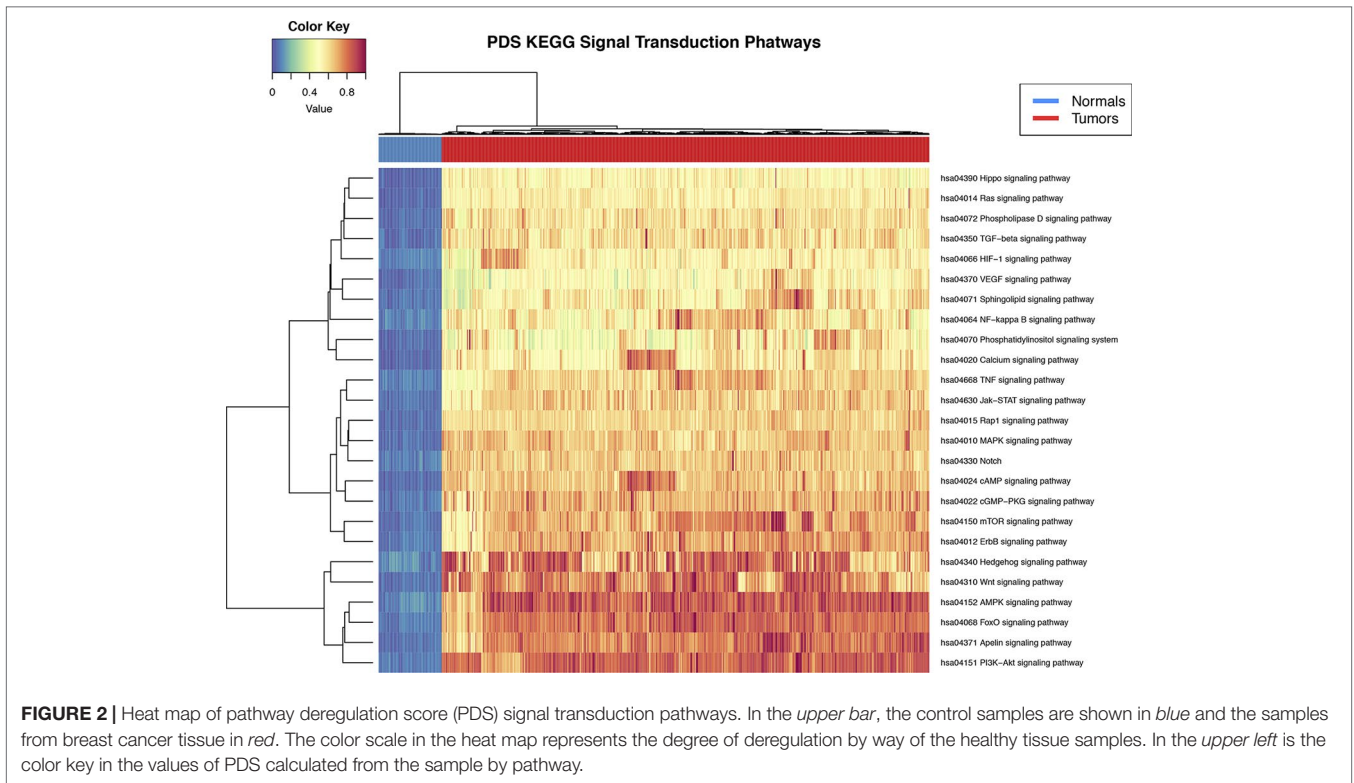
Pathway Deregulation Analysis

Pathifier's pathway deregulation score is a measure of global difference in the expression levels of a set of genes compared to a reference. To determine if signal transduction pathways are deregulated at the level of gene expression in the breast cancer phenotype, we used the *Pathifier* algorithm (Drier et al., 2013). PDS are calculated for each sample and for each pathway. The deregulation analysis results show that all 25 signal STPs have a distinctive pattern in breast tumors with respect to normal breast tissue. This can be seen in the non-supervised clustering dendrogram at the top of **Figure 2** in the two major branches that early separate between both groups.

Transcriptional Master Regulators of Signal Transduction Pathways

The regulons network contains 765 TFs. From the TFs in the network, 338 met the requirement of having at least 20 targets in the molecular signature (**Supplementary Datasheet 4**). The output from MARINa contains a ranking of these regulons based on the integration of the regulons network structure and the differential expression of tumors with respect to normal tissue. In the ranking of the 338 TMRs, the top 10 TMRs regulate approximately 30% of the genes that belong to the set of the molecular signature (**Figure 3**). The top 10 master regulators of signaling pathways in breast cancer are: GLI Family Zinc Finger 2 (GLI2), Paternally Expressed 3 (PEG3), T Box 18 (TBX18), Homeobox A5 (HOXA5), Heart and Neural Crest Derivatives Expressed 2 (HAND2), Homeobox A3 (HOXA3), Meis Homeobox 2 (MEIS2), Homeobox A2 (HOXA2), Teashirt Zinc Finger Homeobox 2 (TSHZ2), and Clock Circadian Regulator (CLOCK) (**Figure 4**).

The results of the MARINa analysis show that, with the exception of CLOCK, the activity of these transcription factors over their targets is repression (Act column on the right side of **Figure 4**, negative values of NES in **Table 1**, and red links in **Figure 5**). Regulatory



interactions in regulons are defined as activation if a target is overexpressed or inhibition if the target is unexpressed. The top 10 regulon network (**Supplementary Datasheet 5**) shows a higher proportion of repression interactions over their target genes (**Figure 5**). In this network, *GLI2* is the only TMR interacting with more than one TMR (*PEG3*, *TBX18*, *HAND2*, *HOXA3*, *HOXA2*, and *HOXA5*). These genes, together with *TSHZ2* and *MEIS2*, have been cited as transcription factors involved in morphogenetic processes like embryonic development and adult tissue remodeling like wound healing (Kuroiwa et al., 1996; Srivastava, 1999; Ruiz i Altaba et al., 2007; Melvin et al., 2013; Takeichi et al., 2013; Chojnowski et al., 2014; Amin et al., 2015; Machon et al., 2015; Jeannotte et al., 2016).

Most of the regulons of these TMRs are enriched with genes that are part of the Hedgehog Signaling pathway. Hedgehog is relevant during early morphogenesis and, in conjunction with Wnt, play a role in the self-renewal of stem cells (Reya et al., 2001). Both pathways have been previously described in cancer (Reya et al., 2001; Taipale and Beachy, 2001). Transcription factor *TSHZ2* forms a complex with *GLI1*, which functions in a coordinated manner with *GLI2* and *GLI3* within the Hedgehog pathway (Riku et al., 2016). This was recovered from the regulons network in the form of gene expression associations. Additionally, a relationship with *TBX18* and the Hedgehog pathway was previously reported in *TBX18* knockout experiments where it showed a marked decrease in the expression of Hedgehog pathway genes (Wu et al., 2013).

GLI2 is the only TMR that shows multiple interactions with other TMRs (six in total; **Figure 5**). *GLI2*, together with *GLI1*, *GLI3* (Ruiz i Altaba et al., 2007), and *TSHZ2* (another of our TMRs) (Riku et al., 2016), are effector molecules activated within the Hedgehog pathway that modulate dedifferentiation and

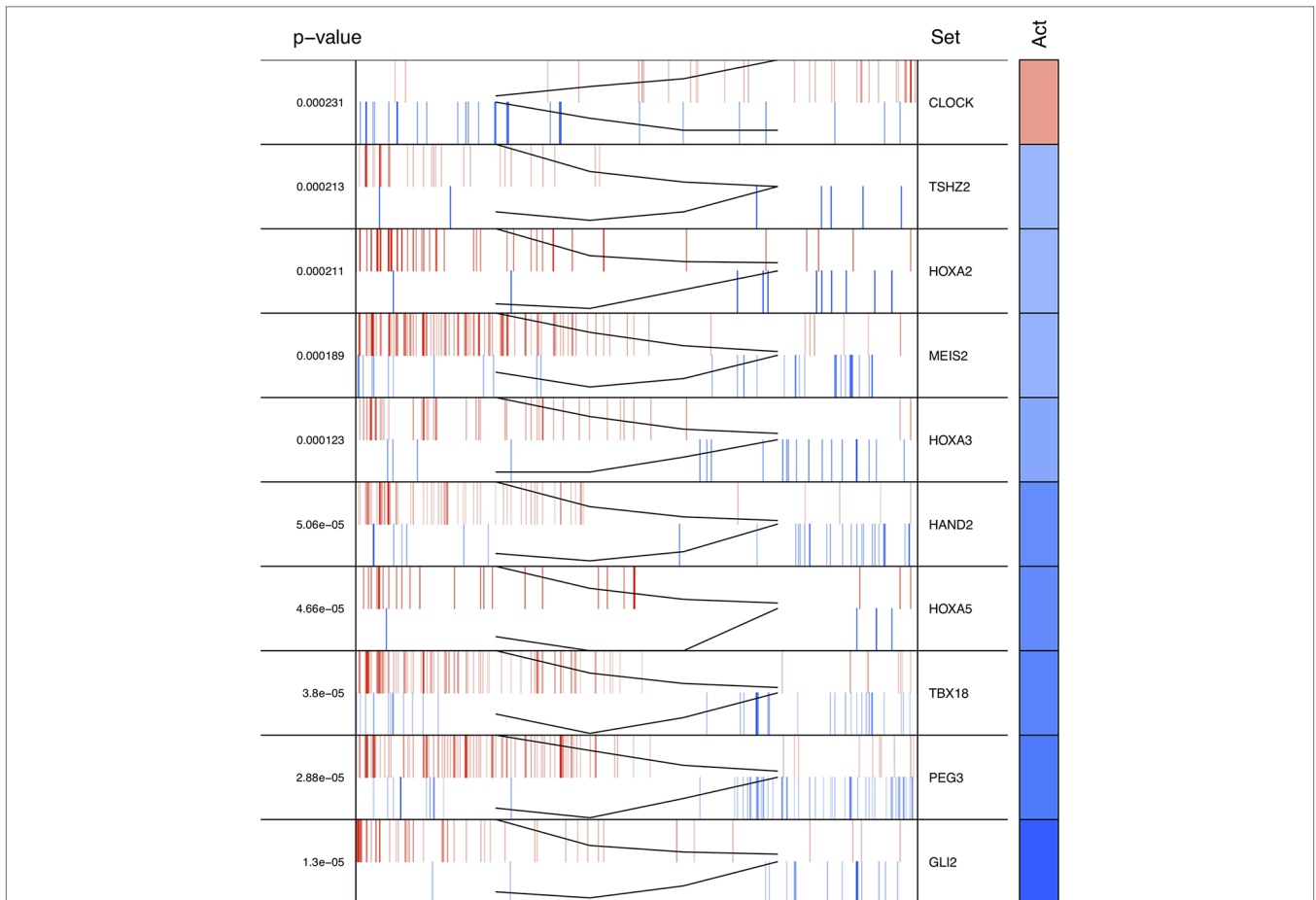


FIGURE 4 | Top 10 master regulators of signal transduction pathways. These transcription factors control the genes of signal transduction pathways more differentially expressed in the tumor tissue. The total number of genes of signal transduction pathways (STPs) controlled by these regulons is 421, representing almost one third of the total genes in the molecular signature. The first column shows the enrichment *p* value for each regulon. The names of the master regulators are on the right side. The "Act" column indicates the activity of the master regulator on its STPs targets. Red color represents the over expression and the blue color represents the sub-expression with respect to normal tissue. Barcode column shows each gene in the molecular signature as one vertical bar. The black line on the blue and red vertical lines shows the walking down the ranked list of genes of the gene set enrichment analysis on the molecular signature genes.

TABLE 1 | Top 10 master regulators of signal transduction pathways.

Regulon	Size	NES	<i>p</i> value	FDR
CLOCK	59	3.68	2.31E-04	0.00592
TSHZ2	35	-3.7	2.13E-04	0.00592
HOXA2	54	-3.71	2.11E-04	0.00592
MEIS2	128	-3.73	1.89E-04	0.00592
HOXA3	67	-3.84	1.23E-04	0.00592
HAND2	93	-4.05	5.06E-05	0.00342
HOXA5	30	-4.07	4.66E-05	0.00342
TBX18	132	-4.12	3.80E-05	0.00342
PEG3	162	-4.18	2.88E-05	0.00342
GLI2	64	-4.36	1.30E05	0.00342

These transcription factors control the genes of signal transduction pathways more differentially expressed in the tumor tissue. With the exception of CLOCK, these regulators are commonly described within the context of embryonic development, and all of them have been reported in association with cancer. The total number of genes controlled by these regulons is 412, representing almost one third of the total genes in the molecular signature.

differentiation processes during early morphogenesis (Ruiz i Altaba et al., 2007; Scales and de Sauvage, 2009). This TMR is associated here in a context where the Hedgehog pathway is also represented.

Regulons Network Enrichment of KEGG Pathways

The difference of this work with respect to MARINa lies in the use of a specific set of genes (signal transduction signature) which produces a ranking of the regulons for this specific subset of genes. The regulons network used as input is the same as in regular MARINa, but the ranking is constrained with the use of the specific gene signature. The sets of genes that constitute each regulon can vary in size from tens to several hundreds of genes and may include genes that are not in the molecular signature but are part of biological functions other than signal transduction. The statistical overrepresentation analysis allows the reduction

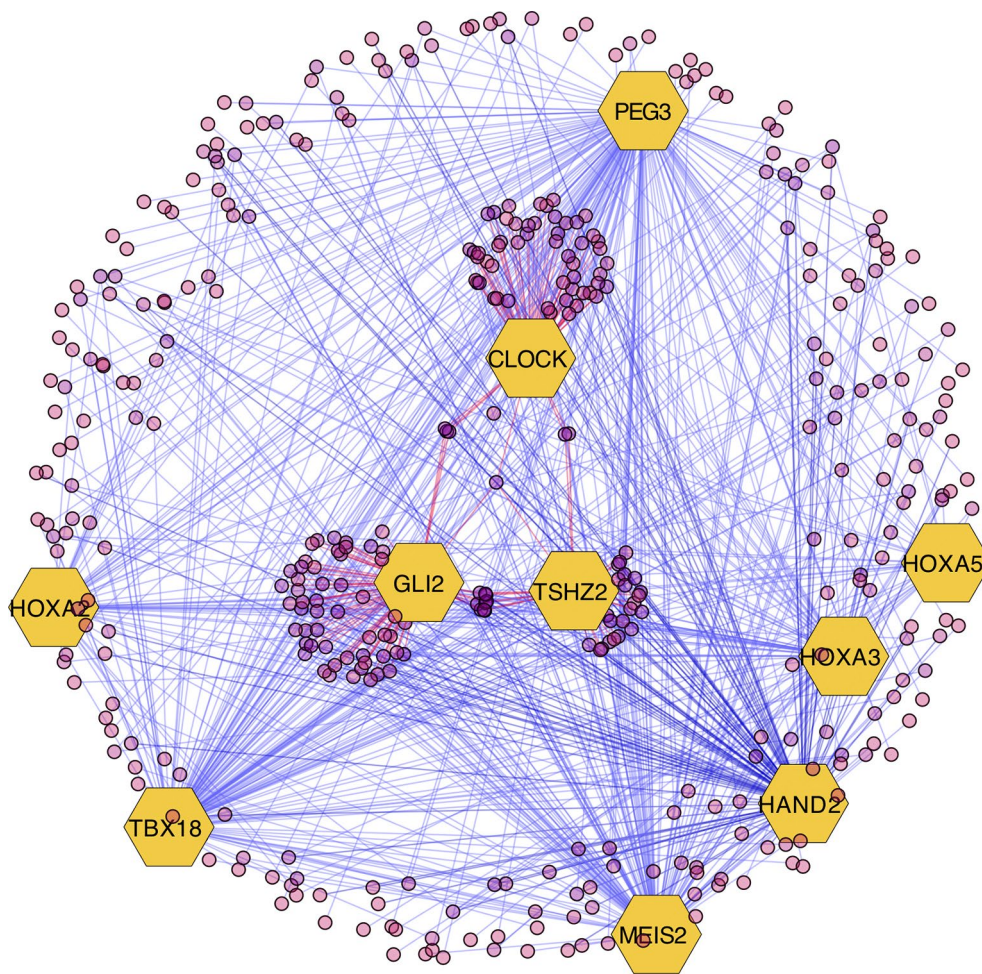


FIGURE 5 | Visualization of the top 10 transcriptional master regulators (TMRs) (big hexagons) and their targets (circles). TMRs show a majority of repression interactions of their targets (shown as blue links) and comparatively few activation interactions (shown as red links). GLI2 is the TMR with the highest enrichment score (ES) of the top 10. Although GLI2 maintains activation interactions with some of its targets, the majority of its interactions are inhibitory. CLOCK is the only TMR that maintains a greater proportion of activation interactions [image generated with Cytoscape (Shannon et al., 2003)].

of dimensionality from large lists of individual genes to fewer discernible biological functions, which is necessary given the large number of genes included in the network and the possibility of multiple annotations for each gene.

The top 10 regulons contain 4,052 genes associated by statistical dependencies. To know which pathways are most likely being regulated, we performed an overrepresentation analysis for all human KEGG pathways. This helped us to know which of the signal transduction pathways are represented, as well as other co-regulated pathways. The pathway with the most statistically significant enrichment was *Pathways in cancer* (hsa05200) with a coincidence of 121 genes. This pathway was not considered in the construction of the gene signature, although the enrichment is to be expected since it is a very large pathway composed of genes from many other signal transduction pathways and because we initially constrained our analysis to regulons that contained at least 20 genes of the molecular signature (see *Materials and Methods*).

Other pathways such as *Cell cycle* (hsa04110) and *Focal adhesion* (hsa04510) follow in the top 3 enrichments. Also enriched are

signaling pathways present within our molecular signature and that are known to be important in the development of cancer, such as *PI3K-AKT signaling pathway* (hsa04151), *Phospholipase D signaling pathway* (hsa04072), and *Hedgehog signaling pathway* (hsa04340) (Table 2). As a whole, these pathways seem suggestive of coordinated signaling towards survival, proliferation, and differentiation, which are consistent with current knowledge of cancer biology. Some of the genes that are part of these regulons are known to take part in processes where cell growth and differentiation take place (i.e., morphogenetic processes). The functions and the possible relevance of these genes in the context of breast cancer are commented in the following sections.

During the processes associated with tissue remodeling, signals such as morphogens, cytokines, and growth factors are present in the cell's environment. These activate signal transduction pathways that in turn induce changes within the cell (Christian, 2000). In the adult organism, tissue remodeling occurs after damage, or as part of very specific processes like lactation where the mammary gland structure changes dramatically (Hennighausen

TABLE 2 | Enrichment analysis.

Enriched KEGG pathway	ID	FDR	No. of genes	Enrichment ratio
Pathways in cancer	hsa05200	0.000347	121	1.48
Cell cycle	hsa04110	0.00226	46	1.8
Focal adhesion	hsa04510	0.00386	66	1.58
Glioma	hsa05214	0.00993	27	1.98
Prostate cancer	hsa05215	0.0143	33	1.8
Huntington's disease	hsa05016	0.0148	60	1.51
PI3K-Akt signaling pathway	hsa04151	0.0148	96	1.37
Phospholipase D signaling pathway	hsa04072	0.0148	47	1.58
EGFR tyrosine kinase inhibitor resistance	hsa01521	0.0148	30	1.8
Hedgehog signaling pathway	hsa04340	0.0148	20	2.06

Statistical over representation analysis of KEGG pathways for the top 10 TMR regulons network was performed with Web Gestalt. The statistical significance threshold was set top ≤ 0.05 after FDR correction.

and Robinson, 2001; Macias and Hinck, 2012). This phenomena share a number of features, among which are cell proliferation, migration, and the formation of tissue structures like new blood and lymphatic vessels or epithelia. It is not unreasonable, though, that some components of the molecular machinery are similar between those processes and tumor development where similar structures are formed, although with abnormal results.

Enrichment of Each Regulon in GO Processes

The most significantly enriched processes of each TMR regulon are presented in **Table 3**. In human breast cancer cells, HOXA5 was observed to activate the p53 tumor suppressor gene promoter (Raman et al., 2000). This agrees with the observation that the expression of HOXA5 in breast cancer cells expressing wild-type p53 led to apoptosis, while those lacking the p53 gene did not (Raman et al., 2000; Chen et al., 2004). Furthermore, the HOXA5 promoter region was methylated in 80% of p53-negative breast

cancer specimens. (Raman et al., 2000). This aberrant regulation of HOX genes in breast cancer suggests that HOX genes may be important components in a network of gene regulatory mechanisms related to adult tissue homeostasis (Bhatlekar et al., 2014).

Nine out of our 10 TMRs are recognized for taking part in morphogenetic processes (Kuroiwa et al., 1996; Srivastava, 1999; Ruiz i Altaba et al., 2007; Melvin et al., 2013; Takeichi et al., 2013; Chojnowski et al., 2014; Amin et al., 2015; Machon et al., 2015; Jeannotte et al., 2016). Enrichments of individual TMR regulons in GO biological processes recovered associations between TMR regulons with general morphogenetic processes.

Enriched GO biological processes are obtained from the molecular signature of the signal transduction pathways. The top places are occupied by processes associated with morphogenesis. These results are in line with the idea that tumors bear aberrations of growth, differentiation, and organization of cell populations. These are basic processes that are tightly coordinated and controlled during embryogenesis as well as in adult tissues. A similar idea has been previously proposed by Vinnitsky (1993), with the name of "oncogerminative theory of cancer development" (Vinnitsky, 1993). It suggests that cancer arises due to aberrant expression of developmental genes, and where the tumor formation is a dynamic self-organizing process that effectively produces new tissue even if in an abnormal form. The malignant transformation of somatic cells, which can start with gene mutations combined with epigenetic dysregulation, can ultimately produce somatic cells reprogrammed into immortal cells that mimic germline cells. These mimics are termed "cancer stem cells" or "oncogerminative cells" (Vinnitsky, 1993; Bhatlekar et al., 2014).

Enrichments of GO biological processes for each individual regulon in the top 10 show a recurrent theme in processes associated with morphogenesis (Kuroiwa et al., 1996; Srivastava, 1999; Ruiz i Altaba et al., 2007; Melvin et al., 2013; Takeichi et al., 2013; Chojnowski et al., 2014; Amin et al., 2015; Machon et al., 2015; Jeannotte et al., 2016). These results are in line with the idea that tumors bear aberrations of growth, differentiation, and organization of cell populations. Although our results cannot assure the activity of morphogenetic processes in tumors, there is

TABLE 3 | Top significant enrichments of Gene Ontology biological processes per regulon.

Regulon	Enriched GO processes	ID	FDR	No. of genes	Enrichment ratio
CLOCK	Mitotic cell cycle	GO:000278	1.39E-02	64	1.84
GLI2	Regulation of cell differentiation	GO:0045595	1.22E-05	73	2.07
HAND2	Cardiovascular system development	GO:0072358	4.31E-06	53	2.52
HAND2	Vasculature development	GO:0001944	4.31E-06	52	2.51
HOXA3	Tube development	GO:0035295	8.94E-05	46	2.5
HOXA5	Proximal/distal pattern formation	GO:0009954	1.69E-02	6	1.98
HOXA5	Anterior/posterior pattern specification	GO:0009952	1.69E-02	12	4.98
HOXA5	Skeletal system development	GO:0001501	1.69E-02	19	3.28
MEIS2	Animal organ morphogenesis	GO:0009887	5.29E-08	107	1.97
PEG3	Cell cycle	GO:0007049	8.72E-08	225	1.53
TBX18	Tissue development	GO:0009888	3.59E-05	139	1.61
TBX18	Blood vessel development	GO:0001568	3.59E-05	62	2.16
TSHZ2	Regulation of cell proliferation	GO:0042127	4.91E-02	45	1.92
TSHZ2	Extracellular matrix organization	GO:0030198	4.91E-02	17	3.32
TSHZ2	Extracellular structure organization	GO:0043062	4.91E-02	17	3.31

The top 10 regulons enrich more biological processes related to embryonic development. The statistical significance threshold was set top ≤ 0.05 after FDR correction.

an association at the level of gene expression patterns of molecules canonically represented in them.

HOXA Family Enriched in Regulons

In addition, our results show that 10 of the 11 members of the Homeobox A family are included within the top 10 TMR regulons (Figure 6). In humans, HOXA consists of 11 genes (HOXA1, HOXA2, HOXA3, HOXA4, HOXA5, HOXA6, HOXA7, HOXA9, HOXA10, HOXA11, and HOXA13). Although the HOXA genes code for proteins with transcription factor activity, these are not typically considered as components of signal transduction pathways. HOXA TFs act not only as transcriptional activators in cancers but also as transcriptional repressors (Ladam and Sagerström, 2014); thus, both upregulation and downregulation of the members of this family may be relevant in carcinogenesis. Many HOXA genes (HOXA1, A2, A3, A5, and A9) have been shown to have significantly lower expression levels in cancerous tissues compared to non-cancerous ones.

CONCLUSIONS

With the generation of gene regulatory networks and a molecular signature centered on signal transduction pathways,

we present a list of genes and their transcriptional regulators that may be modulated signal transduction pathways in breast cancer. This information may be helpful in the study of this disease where pathway analysis showed that all pathways from KEGG in this category are deregulated in a large dataset of breast cancer samples.

We identified TSHZ2, HOXA2, MEIS2, HOXA3, HAND2, HOXA5, TBX18, PEG3, GLI2, and CLOCK as the top 10 TMRs regulating signal transduction pathways in breast cancer. These genes regulate 30% of the genes in these pathways. CLOCK is the only TMR in the top 10 that shows a positive regulation of its predicted targets, while the other top TMRs show negative regulation relationships, although the molecular mechanisms by which those TMRs act should be explored in future studies.

The enrichment analysis of the top 10 TMR regulons recovers information about processes that are well recognized in cancer (angiogenesis, organogenesis, proliferation, survival, etc.). These results are reassuring in the sense that we know we are recovering relevant biological information from the phenotypes under study instead of random associations. Furthermore, the analysis of individual regulons allows the identification of specific molecules that may be playing key roles. This seems to be the case with genes in the HOXA family, which are within the top 10 TMRs and as part of regulons. Genes of this family are known for their role in morphogenetic processes as well as in the maintenance of adult tissues and with altered expression in tumors.

We present a modified version of MARINa that utilizes a specific gene signature with genes in KEGG's Signal Transduction category. The reason for this is to narrow the search of TMRs to those most relevant to signal transduction pathways. This approach, however, is not limited to a particular pathway or process and the molecular signature can be modified to reflect other research questions relative to specific pathways or processes.

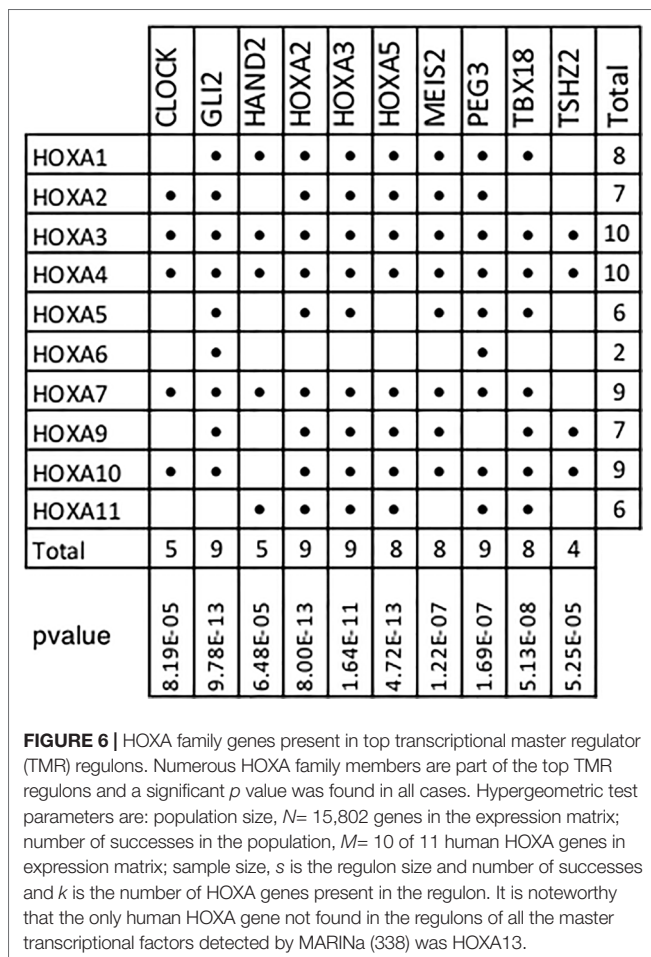
Signal transduction pathways serve an important role as information integrators in the cell. Components and their interactions are thus of great interest and subject of a thorough study. Furthermore, signal transduction pathways are susceptible to external pharmacological modulation. An understanding of the regulation of the pathways may be helpful in the search for therapeutic targets.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Data comes from TCGA/GDC public repositories. The list of GDC samples used in this work is in the **Supplementary Datasheet 1**. The list of transcription factors with experimental evidence obtained in TFCheckpoint.org is in the **Supplementary Datasheet 2**.

AUTHOR CONTRIBUTIONS

DT-C implemented methods, performed calculations, analyzed results, and drafted the paper. HT implemented methods and analyzed and discussed the results. TV-C



contributed to the discussion and drafting of the paper. EH-L designed and supervised the study, contributed to the analysis of the results and discussion, and revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

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 the recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

REFERENCES

- Alcalá-Corona, S. A., de Anda-Jáuregui, G., JEspinal-Enriquez, J., Tovar, H., and Hernández-Lemus, E. (2018). Network modularity and hierarchical structure in breast cancer molecular subtypes, in *In International Conference on Complex Systems* (Cham: Springer), 352–358. doi:10.1007/978-3-319-96661-8_36
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., et al. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48 (8), 838–847. doi:10.1038/ng.3593
- Amin, S., Donaldson, I. J., Zannino, D. A., Hensman, J., Rattray, M., Losa, M., et al. (2015). Hoxa2 selectively enhances meis binding to change a branchial arch ground state. *Dev. Cell* 32, 265–277. doi: 10.1016/j.devcel.2014.12.024
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556
- Bhatlekar, S., Fields, J. Z., and Boman, B. M. (2014). Hox genes and their role in the development of human cancers. *J. Mol. Med.* 92 (8), 811–823. doi: 10.1007/s00109-014-1181-y
- Carroll, J. S., and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. *Mol. Endocrinol.* 20 (8), 1707–1714. doi: 10.1210/me.2005-0334
- Chen, H., Chung, S., and Sukumar, S. (2004). HOXA5-induced apoptosis in breast cancer cells is mediated by caspases 2 and 8. *Mol. Cell. Biol.* 24 (2), 924–935. doi: 10.1128/MCB.24.2.924-935.2004
- Chojnowski, J. L., Masuda, K., Trau, H. A., Thomas, K., Capocchi, M., and Manley, N. R. (2014). Multiple roles for hoxa3 in regulating thymus and parathyroid differentiation and morphogenesis in mouse. *Development* 141, dev-110833. doi: 10.1242/dev.110833
- Christian, J. L. (2000). Bmp, wnt and hedgehog signals: how far can they go? *Curr. Opin. Cell Biol.* 12 (2), 244–249. doi: 10.1016/S0955-0674(99)00082-4
- Dhanasekaran, N. (1998). Cell signaling: an overview. *Oncogene* 17 (11), 1329. doi: 10.1038/sj.onc.1202170
- Drier, Y., Sheffer, M., and Domany, E. (2013). Pathway-based personalized analysis of cancer. *Proc. Natl. Acad. Sci. United States America* 110, 6388–6393. doi: 10.1073/pnas.1219651110
- Espinal-Enriquez, J., Fresno, C., Anda-Jáuregui, G., and Hernández-Lemus, E. (2017). Rna-seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Sci. Rep.* 7, 1760. doi: 10.1038/s41598-017-01314-1
- García-Campos, M. A., Espinal-Enriquez, J., and Hernández-Lemus, J. (2015). Pathway analysis: state of the art. *Front. Physiol.* 6, 383. doi: 10.3389/fphys.2015.00383
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674. doi: 10.1016/j.cell.2011.02.013
- Hennighausen, L., and Robinson, G. W. (2001). Signaling pathways in mammary gland development. *Dev. Cell* 1 (4), 467–475. doi: 10.1016/S1534-5807(01)00064-8

ACKNOWLEDGMENTS

This paper constitutes a partial fulfillment of the Graduate Program in Biological Sciences of the National Autonomous University of México (UNAM). DT-C acknowledges the scholarship and financial support provided by the National Council of Science and Technology (CONACyT) and UNAM. This manuscript has been released as a pre-print at BioRxiv (Tapia-Carrillo et al., 2018). DOI: 10.1101/425777.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01180/full#supplementary-material>

- Hernández-Lemus, E., and Siqueiros-García, J. M. (2013). Information theoretical methods for complex network structure reconstruction. *Complex Adaptive Syst. Modeling* 1 (1), 8. doi: 10.1186/2194-3206-1-8
- Jeanotte, L., Gotti, F., and Landry-Truchon, K. (2016). Hoxa5: a key player in development and disease. *J. Dev. Biol.* 4 (2), 13. doi: 10.3390/jdb4020013
- 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
- Kolch, W., Halasz, M., and Granovskaya, M. (2015). Kholodenko BN The dynamic control of signal transduction networks in cancer cells. *Nat. Rev. Cancer* 15 (9), 515. doi: 10.1038/nrc3983
- Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L. L., Tada, M., et al. (1996). Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. *Nat. Genet.* 12, 186–190. doi: 10.1038/ng0296-186
- Ladam, F., and Sagerström, C. G. (2014). Hox regulation of transcription: more complex (es). *Dev. Dynamics* 243 (1), 4–15. doi: 10.1002/dvdy.23997
- Laurent, A., Calabrese, M., Warnatz, H. J., Yaspo, M. L., Tkachuk, V., Torres, M., et al. (2015). Chip-seq and rna-seq analyses identify components of the wnt and fgf signaling pathways as prep1 target genes in mouse embryonic stem cells. *PLoS One* 10 (4), e0122518. doi: 10.1371/journal.pone.0122518
- Lefebvre, C., Rajbhandari, P., Alvarez, M. J., Bandaru, P., Lim, W. K., Sato, M., et al. (2010). A human b-cell interactome identifies myb and foxm1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377. doi: 10.1038/msb.2010.31
- Machon, O., Masek, J., Machonova, O., Krauss, S., and Kozmik, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. *BMC Dev. Biol.* 15 (1), 40. doi: 10.1186/s12861-015-0093-6
- Macias, H., and Hinck, L. (2012). Mammary gland development. *Wiley Interdiscip. Reviews: Dev. Biol.* 1 (4), 533–557. doi: 10.1002/wdev.35
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., and Dalla Favera, R. (2006). Califano A.: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinf.* 7 Suppl 1, S7. doi: 10.1186/1471-2105-7-S1-S7
- Melvin, V. S., Feng, W., Hernandez-Lagunas, L., Artinger, K. B., and Williams, T. (2013). A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. *Dev. Dynamics* 242 (7), 817–831. doi: 10.1002/dvdy.23969
- Morgan, R., and Pandha, H. S. (2017). HOX transcription factors and the prostate tumor microenvironment. *J. Cancer Metastasis Treat* 3 (12), 278. doi: 10.20517/2394-4722.2017.31
- Raman, V., Martensen, S. A., Reisman, D., Evron, E., Odenwald, W. F., Jaffee, E., et al. (2000). Compromised hoxa5 function can limit p53 expression in human breast tumours. *Nature* 405, 974–978. doi: 10.1038/35016125
- Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature* 414 (6859), 105. doi: 10.1038/35102167
- Riku, M., Inaguma, S., Ito, H., Tsunoda, T., Ikeda, H., and Kasai, K. (2016). Down-regulation of the zinc-finger homeobox protein tshz2 releases gli1

- from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget* 7, 5690–5701. doi: 10.18632/oncotarget.6788
- Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). Gc-content normalization for rna-seq data. *BMC Bioinf.* 12 (1), 480. doi: 10.1186/1471-2105-12-480
- Ruiz i Altaba, A., Mas, C., and Stecca, B. (2007). The gli code: an information nexus regulating cell fate, stemness and cancer. *Trends Cell Biol.* 17, 438–447. doi: 10.1016/j.tcb.2007.06.007
- Scales, S. J., and de Sauvage, F. J. (2009). Mechanisms of hedgehog pathway activation in cancer and implications for therapy. *Trends Pharmacol. Sci.* 30 (6), 303–312. doi: 10.1016/j.tips.2009.03.007
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Srivastava, D. (1999). Hand proteins: molecular mediators of cardiac development and congenital heart disease. *Trends Cardiovasc. Med.* 9, 11–18. doi: 10.1016/S1050-1738(98)00033-4
- Steven Martin, G. (2003). Cell signaling and cancer. *Cancer Cell* 4 (3), 167–174. doi: 10.1016/S1535-6108(03)00216-2
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102 (43), 15545–15550. doi: 10.1073/pnas.0506580102
- Taipale, J., and Beachy, P. A. (2001). The Hedgehog and Wnt signalling pathways in cancer. *Nature* 411 (6835), 349–354. doi: 10.1038/35077219
- Takeichi, M., Nimura, K., Mori, M., Nakagami, H., and Kaneda, Y. (2013). The transcription factors tbx18 and wt1 control the epicardial epithelial-mesenchymal transition through bi-directional regulation of slug in murine primary epicardial cells. *PLoS One* 8, e57829. doi: 10.1371/journal.pone.0057829
- Tapia-Carrillo, D., Tovar, H., Velázquez-Caldelas, T. E., and Hernandez-Lemus, S. (2018). Master regulators of signaling pathways coordinate key processes of embryonic development in breast cancer. *BioRxiv*, 425777. doi: 10.1101/425777
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in rna-seq: a matter of depth. *Genome Res.* 21, 2213–2223. doi: 10.1101/gr.124321.111
- Tovar, H., García-Herrera, R., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Transcriptional master regulator analysis in breast cancer genetic networks. *Comput. Biol. Chem.* 59 Pt B, 67–77. doi: 10.1016/j.compbiolchem.2015.08.007
- Tripathi, S., Christie, K. R., Balakrishnan, R., Huntley, R., Hill, D. P., Thommesen, L., et al. (2013). Gene ontology annotation of sequence-specific dna binding transcription factors: setting the stage for a large-scale curation effort. *Database: J. Biol. databases curation* 2013, bat062. doi: 10.1093/database/bat062
- Vinnitsky, V. B. (1993). Oncogerminative hypothesis of tumor formation. *Med. Hypotheses* 40 (1), 19–27. doi: 10.1016/0306-9877(93)90191-R
- Wu, S.-P., Dong, X.-R., Regan, J. N., Su, C., and Majesky, M. W. (2013). Tbx18 regulates development of the epicardium and coronary vessels. *Dev. Biol.* 383, 307–320. doi: 10.1016/j.ydbio.2013.08.019
- Zhang, B., Kirov, S., and Snoddy, J. (2005). Webgestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 33, W741–W748. doi: 10.1093/nar/gki475

Conflict of Interest: 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 © 2019 Tapia-Carrillo, Tovar, Velázquez-Caldelas and Hernandez-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.