



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Caracterización de la heterogeneidad poblacional en cáncer a través de RNA-seq de célula única.

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestra en Ciencias

PRESENTA:

Thelma Escobedo Tapia

TUTOR PRINCIPAL:

Dr. Osbaldo Resendis Antonio

[Red de Apoyo a la Investigación-UNAM](#)

[Instituto Nacional de Medicina Genómica \(INMEGEN\)](#)

MIEMBROS DEL COMITÉ TUTOR:

Dra. Ayari Fuentes Hernández

[Centro de Ciencias Genómicas \(CCG\)](#)

Dr. Sergio Manuel Encarnación Guevara

[Centro de Ciencias Genómicas \(CCG\)](#)

Ciudad de México. Octubre, 2022



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.

*Este proyecto se realizó con la ayuda del financiamiento
por CONACyT beca 1043695 en el
Laboratorio de Biología de Sistemas y
Enfermedades Humanas del Instituto Nacional de
Medicina Genómica (INMEGEN) bajo la dirección
del Dr. Osbaldo Resendis Antonio.*

Índice general

Índice de figuras	5
Índice de tablas	5
Abreviaturas	6
Resumen	7
Abstract	8
Introducción	9
1. Cáncer de mama	9
2. Esferoides como modelos de sistemas biológicos en cáncer de mama	13
3. Heterogeneidad en cáncer de mama	15
4. Datos de célula única	17
4.1 Datos de célula única revelan heterogeneidad en cáncer	17
4.2 Reconstrucción de trayectorias celulares con datos de célula única	19
5. Redes de regulación	20
5.1 Redes Booleanas	20
Antecedentes	21
Justificación	24
Hipótesis	25
Objetivos	25
Objetivo general	25
Objetivos específicos	25
Métodos	26
Datos	26
Procesamiento de los datos de secuenciación de RNA de célula única	26
Reconstrucción de trayectorias	27
Identificación de genes de divergencia	28
Análisis de enriquecimiento	28
Reconstrucción de red de regulación	28
Traducción de conexiones regulatorias a funciones Booleanas	29
Modelado Booleano	30
Simulación y análisis de atractores con el modelo Booleano	31
Perturbaciones <i>in silico</i>	32
Resultados	33

Los datos de célula única de los MCTS reflejan tres trayectorias celulares	34
Reconstrucción de red de regulación transcripcional	41
La simulación Booleana de la red predice tres conjuntos de atractores	45
Modulación de la heterogeneidad a través de perturbaciones <i>in silico</i>	46
Discusión	49
Conclusión	55
Perspectivas	56
Material suplementario	58
Referencias	61

Índice de figuras

Figura 1. Panorama global del cáncer.....	9
Figura 2. Carcinoma mamario invasivo.....	11
Figura 3. Formación de modelos esféricos.....	14
Figura 4. Heterogeneidad en cáncer de mama.....	15
Figura 5. Diferencia entre la secuenciación de RNA de célula única y la secuenciación de RNA en masa.....	18
Figura 6. Subpoblaciones en esferoides tumorales multicelulares (MCTS) de la línea celular MCF7.....	23
Figura 7. Descripción metodológica del análisis RNA-seq de célula única.....	27
Figura 8. Estados de encendido y apagado de los genes para los atractores dentro de los clusters.....	31
Figura 9. Comparación de datos antes y después de la normalización.....	33
Figura 10. Reconstrucción de trayectorias de los esferoides de la línea celular MCF7 con datos de scRNA-seq.....	34
Figura 11. Genes de divergencia entre la trayectoria azul vs la trayectoria verde.....	37
Figura 12. Genes expresados diferencialmente en la trayectoria verde y azul.....	38
Figura 13. Análisis de enriquecimiento para el conjunto de genes de cada subpoblación...	40
Figura 14. Red de regulación transcripcional.....	42
Figura 15. Red de regulación transcripcional reducida.....	43
Figura 16. Red de regulación transcripcional propuesta.....	44
Figura 17. Clustering de los atractores encontrados a través del modelado Booleano de la red.....	46
Figura 18. Perturbaciones <i>in silico</i> de la red de regulación.....	47

Índice de tablas

Tabla 1. Tipos de tumor en cáncer de mama y subtipos histológicos.....	11
Tabla 2. Establecimiento de las reglas lógicas para las conexiones de la red de regulación.....	29
Tabla 3. Proporciones celulares normalizadas para las células tanto del día 6 como del día 19 para cada una de las trayectorias celulares o clusters.....	36

Abreviaturas

DNA: Ácido desoxirribonucleico (ADN)

CSC: Células troncales cancerosas (Cancer stem cells)

DE: Genes diferencialmente expresados

D2HG: D-2-hidroxiglutarato

ERBB2 (HER2): Receptor 2 del factor de crecimiento epidérmico humano

ER α : Receptor a estrógenos alfa

IDH2: Isocitrato deshidrogenasa 2

ISG: Genes estimulados por interferón

IPA: Análisis de vías de ingenio (Ingenuity pathway analysis)

LOESS: Suavizado de diagrama de dispersión estimado localmente (Locally estimated scatterplot smoothing)

MCTS: Esferoides tumorales multicelulares (Multicellular tumor spheroid)

MHCII: Complejo mayor de histocompatibilidad tipo II

MLLE: Incrustación lineal localmente modificada (Modified locally linear embedding)

mRNA: ARN mensajero

ORA: Análisis de sobre-representación (Over-representation analysis)

RE: Receptor de estrógenos

RNA: Ácido ribonucleico (ARN)

RNA-seq: Secuenciación de ARN

RP: Receptor de progesterona

scRNA-seq: Secuenciación de ARN de célula única

STREAM: Reconstrucción, exploración y mapeo de trayectorias de célula única (Single-cell trajectory reconstruction, exploration and mapping)

TNBC: Cáncer de mama triple negativo

UMAP: Aproximación y proyección de variedades uniformes (Uniform manifold approximation and projection)

Resumen

Actualmente, la implementación de la secuenciación de RNA de célula única (scRNA-seq) junto con diferentes análisis computacionales han brindado la oportunidad de explorar la heterogeneidad en el cáncer (Puram et al., 2017). Así, existen estudios que han resaltado la importancia de entender esta heterogeneidad, su relación con la progresión de la enfermedad, así como con el impacto en la eficiencia terapéutica (Gerlinger et al., 2012). En este proyecto de investigación planteamos un método computacional a nivel de biología de sistemas para develar la composición heterogénea de subpoblaciones intratumorales provenientes de esferoides tumorales multicelulares (MCTS) de la línea celular de cáncer de mama MCF7. Para ello, a partir del uso de datos transcriptómicos (scRNA-seq) generados por Muciño-Olmos y colaboradores (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020), se identificaron las subpoblaciones, sus fenotipos, y las posibles trayectorias que postulen sus transformaciones. Posteriormente, se reconstruyó una red de regulación transcripcional clave en la inducción de la heterogeneidad intratumoral. A partir de la red de regulación genética propuesta, con 23 nodos y 40 interacciones, se hizo la dinámica Booleana para identificar los potenciales fenotipos y su relación con las subpoblaciones en los esferoides de MCF7. Los resultados de la simulación arrojaron la generación de tres conjuntos de atractores con diferentes características. Uno de ellos presentó un fenotipo proliferativo, otro un fenotipo invasivo y otro un fenotipo combinado, es decir, tanto proliferativo como invasivo. Interesantemente, dichos atractores o estados estacionarios convergieron con los resultados en cuanto a las tres subpoblaciones determinadas a través de las trayectorias celulares. Finalmente, las perturbaciones *in silico* de los genes *ESR1*, *TP53*, *CCND1*, *FOXA1*, *BRCA1* y *STAT3* mostraron una modulación inherente de las subpoblaciones y, por tanto, de la heterogeneidad intratumoral. En conclusión, la red de regulación propuesta y su simulación dinámica permitieron inferir mecanismos de regulación entre las subpoblaciones a través de un análisis bioinformático de datos de scRNA-seq provenientes de MCTS.

Abstract

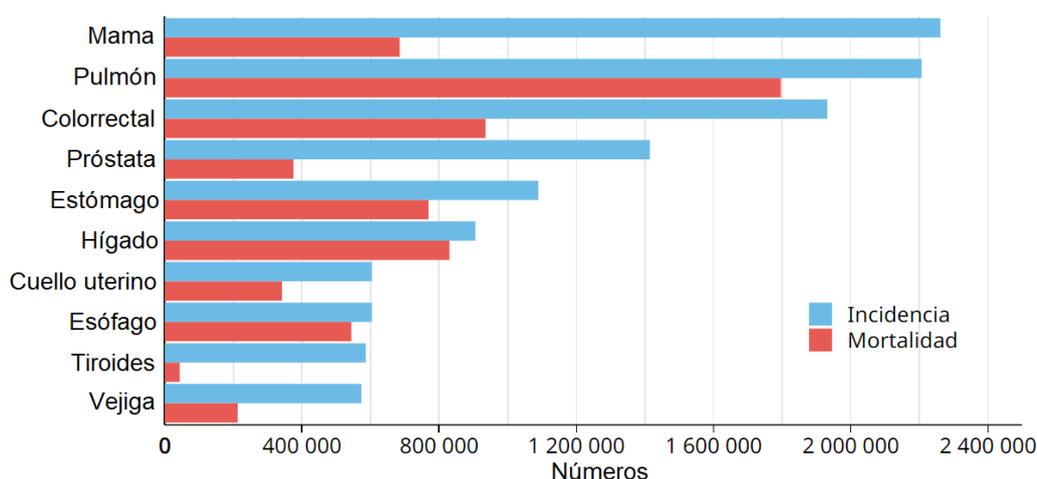
Currently, the implementation of single-cell RNA sequencing (scRNA-seq) along with different computational analysis have provided the opportunity to explore heterogeneity in the study of cancer (Puram et al. 2017). Thus, there are studies that have highlighted the importance of understanding this heterogeneity in cancer, its relationship with disease progression, as well as the impact on therapeutic efficiency (Gerlinger et al. 2012). In this research, we propose a computational method at the level of systems biology to reveal the heterogeneous composition of intratumoral subpopulations from multicellular tumor spheroids (MCTS) of the breast cancer cell line MCF7. To this end, we used the transcriptomic data generated with the high-throughput scRNA-seq technology by Muciño-Olmos and collaborators (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020). Starting from this data, the subpopulations, their phenotypes, and the possible trajectories that postulate their transformations were identified. Subsequently, a key transcriptional regulation network in the induction of intratumoral heterogeneity was reconstructed. From the proposed network, integrating 23 nodes and 40 interactions, the Boolean dynamic was done to identify the regulatory mechanisms between the sets of identified attractors that were established as the subpopulations. The results of the network simulation yielded the generation of three sets of attractors with different characteristics. One of them presented a proliferative phenotype, another an invasive phenotype, and another a combined phenotype, that is, both proliferative and invasive. Interestingly, these attractors or steady states converged with the results of the three subpopulations determined through cell trajectories. Finally, *in silico* perturbations of the *ESR1*, *TP53*, *CCND1*, *FOXA1*, *BRCA1*, and *STAT3* genes showed an inherent modulation of the subpopulations, and therefore, of intratumoral heterogeneity. In conclusion, the proposed regulatory network and its dynamic simulation allowed us to infer regulatory mechanisms between subpopulations through bioinformatic analysis of scRNA-seq data from MCTS.

Introducción

1. Cáncer de mama

De acuerdo con las cifras mundiales del 2020, el cáncer de mama tuvo la mayor incidencia con 2.3 millones de casos (Figura 1 panel A). Además, se posiciona en el tercer lugar de mortalidad causando 684,996 muertes, esto después del cáncer de pulmón y colorrectal en ambos sexos, y en primer lugar sólo en mujeres (Ferlay et al., 2021). Contemplando todos los tipos de cáncer en mujeres a nivel mundial, en ese año se registró un incremento del 24.5% de nuevos casos en este tipo de cáncer (Figura 1 panel B). Asimismo, se ha predicho que en un futuro las cifras no sean más alentadoras, llegando a aumentar hasta 3.2 millones de casos para el año 2030 (Ginsburg et al., 2017).

A)



B)

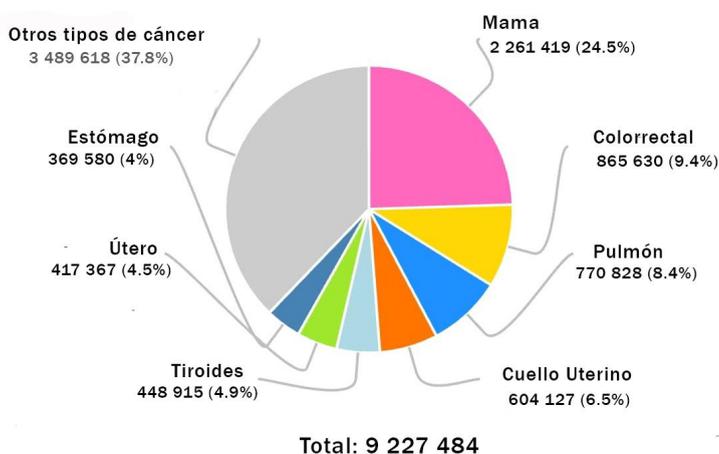


Figura 1. Panorama global del cáncer. A) Número estimado de casos incidentes y muertes en todo el mundo para ambos sexos. B) Número estimado de nuevos casos en 2020 para las mujeres. Tomada de The International Agency for Research on Cancer (IARC), 2020.

Aunque las razones por las cuales se está dando un aumento en la incidencia del cáncer de mama, y también global, son complejas, una de ellas es un aumento en la prevalencia y distribución de algunos factores de riesgo (Bray et al., 2018). Entre estos factores de riesgo en cáncer de mama, también conocidos como modificables, se encuentran: edad tardía de primer embarazo (mayor de 30 años), embarazos nulos, no amamantar, consumo de alcohol o tabaco, dietas altas en grasa y bajas en fibra, baja actividad física, obesidad, entre otros. Es decir, tanto factores en el estilo de vida como sociales están favoreciendo cambios importantes en la incidencia del cáncer. Aunado a lo anterior, solamente el 10% de todos los casos de cáncer de mama son causados por predisposición genética o antecedentes familiares de cáncer (Loibl et al., 2021; Łukasiewicz et al., 2021). Lo que es un hecho es la necesidad de estrategias para el aumento en la detección temprana del mismo y un genuino interés en investigación tanto a nivel básica como clínica para entender sus mecanismos de activación.

Las células cancerosas a medida que dejan atrás características de las células normales van adquiriendo capacidades funcionales cruciales para formar tumores malignos. Entre los procesos celulares que pueden compartir las células cancerosas se encuentra la proliferación y el crecimiento celular descontrolados, resistencia a la muerte celular, angiogénesis, activación de invasión y metástasis, inestabilidad genómica, reprogramación metabólica, evasión del sistema inmune, desbloqueo de la plasticidad fenotípica, reprogramación epigenética, inducción de senescencia y microbiomas polimórficos (Hanahan, 2022; Hanahan & Weinberg, 2011). Estas características ayudan a clasificar de manera general los procesos celulares que sustentan las células tumorales; sin embargo, estos no explican del todo esta compleja enfermedad, por lo que habrá que realizar estudios capaces de integrar estos procesos celulares, así como mecanismos moleculares, que contribuyan a entender de mejor manera la etiología y progresión de la misma.

El cáncer de mama puede originarse desde cualquier célula que forme el tejido mamario. Dicho tejido mamario está conformado por una red de lóbulos, los cuales se forman de lobulillos (pequeños sacos que producen la leche) unidos por conductos (ductos) que transportan la leche hasta el pezón. Tanto los lobulillos como los ductos están recubiertos por células epiteliales, delimitados por una membrana basal. Así, cuando se desarrolla una neoplasia, existe un crecimiento anormal e incontrolado de las células que revisten los conductos y los lobulillos. Desde un punto de vista anatómico-patológico, existe una clasificación para determinar el tipo de tumor, el cual puede ser un carcinoma no invasivo o un carcinoma invasivo. Los carcinomas no invasivos o "*in situ*" se originan en el interior del conducto o lobulillo, pero no traspasan la membrana basal como lo hacen los carcinomas

invasivos. Los carcinomas ductales invasivos, también conocido como ductales infiltrantes, son los tumores malignos que se presentan con más frecuencia (70-80%). Estos tumores se forman en el revestimiento del conducto mamario e invaden el tejido aledaño (ver Figura 2). Por su parte, los carcinomas lobulares invasivos también invaden el tejido circundante, pero éstos comienzan en los lobulillos (Candanedo-González & Pérez-Salazar, 2012). Cabe mencionar que esta clasificación está basada en estudios histológicos, útiles también para determinar el subtipo del carcinoma (ver Tabla 1).

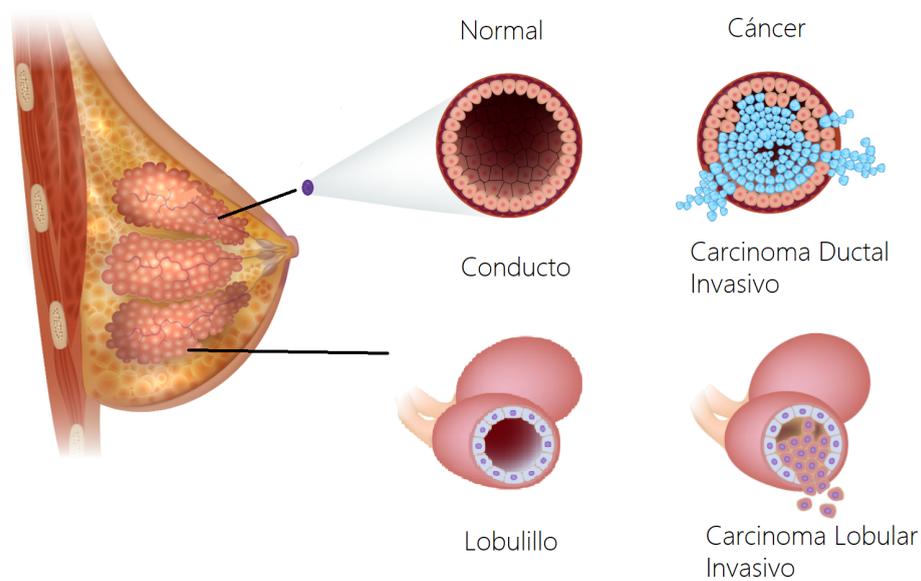


Figura 2. Carcinoma mamario invasivo. Tomada y editada de Understanding Breast Cancer Types, 2022.

Tabla 1. Tipos de tumor en cáncer de mama y subtipos histológicos.

¿Lesión invasiva?	Tipo de tumor	Subtipo
Sí	Carcinoma ductal invasivo	Ductal clásico, medular, papilar, tubular o mucinoso.
	Carcinoma lobular invasivo	
No	Carcinoma ductal <i>in situ</i>	Comedoniano, sólido, cribiforme, papilar o micropapilar.
	Carcinoma lobular <i>in situ</i>	

Tomada de Álvarez-Hernández et al., 2014.

La clasificación del cáncer de mama se ha basado en diferentes factores histopatológicos como el subtipo histológico, grado, tamaño tumoral, estado ganglionar, e invasión linfovascular, entre otros. Aunado a esta diversidad patológica, actualmente existe una clasificación basada en cambios moleculares que refleja de mejor manera la enfermedad. Las células en cáncer de mama presentan diferencias genéticas y se han clasificado en subtipos clínicos y moleculares que se basan en la positividad de marcadores proteicos y los perfiles de expresión del RNA mensajero (mRNA), representando cada uno diferentes características biológicas, clínicas y pronósticas. De esta manera, el cáncer de mama se ha clasificado clínicamente, principalmente, en 4 subtipos: luminal con receptores a estrógenos (RE) y progesterona (RP), con receptores a ERBB2 (HER2-positivo) y sin ninguno de estos receptores llamado triple negativo (TNBC). Además, dentro del grupo luminal se diferencian dos subtipos: luminal A puede tener RE y RP y es negativo a los receptores HER2, mientras que los luminales B son triples positivos (Alcaide-Lucena et al., 2021; Loibl et al., 2021).

Los modelos de líneas celulares han sido de gran utilidad en el estudio del cáncer de mama. Las líneas celulares son células que han podido crecer en un medio de cultivo de manera relativamente sencilla e ilimitada. La línea celular más estudiada y representativa para el tipo luminal A es MCF7, la cual se originó a partir de una mujer que tenía cáncer de mama metastásico (pleura y pared torácica). La mayoría de las investigaciones se han enfocado en esta línea por la semejanza que tiene con el de cáncer de mama luminal A (Lacroix & Leclercq, 2004). Por ejemplo, esta línea celular es una de las pocas que expresa niveles sustanciales del receptor a estrógenos alfa ($ER\alpha$) imitando la mayoría de los cánceres de mama humanos invasivos que expresan este receptor. Esto ha servido en el área clínica, donde se ha podido estudiar la respuesta a los tratamientos. Entre algunas de las características de MCF7 se encuentran las siguientes: es proveniente de un carcinoma ductal invasivo, sensible al estrógeno, con receptores a estrógenos y progesterona, se duplica en 30-40 horas y posee un fenotipo luminal A (Lee et al., 2015). A pesar de que estas células crecen en monocapa, también es posible formar una entidad en forma de esferas.

2. Esferoides como modelos de sistemas biológicos en cáncer de mama

A pesar de que los cultivos celulares en monocapa son herramientas de fácil manipulación, estos cultivos bidimensionales carecen de una arquitectura estructural, del mismo modo muestran deficiencias de la fisiología y del microambiente comparado con los tumores reales. En estudios donde se comparan los cultivos celulares 2D contra el tejido tumoral, se ha demostrado que los perfiles de expresión y la activación de vías de señalización difieren entre sí (Birgersdotter et al., 2005). Asimismo, no hay una semejanza de los tumores *in vivo* en cuanto los gradientes de pH, oxígeno y nutrientes, y varias de las interacciones celulares se pierden. Por otro lado, aunque los modelos en animales son importantes en la investigación del cáncer de mama, los estudios son costosos y suelen tomar más tiempo. Debido a lo anterior, surgieron nuevas aproximaciones, tales como los cultivos tridimensionales (3D) (Carranza-Rosales et al., 2018). Entre los diferentes cultivos en 3D se encuentran: las tumoresferas, esferoides multicelulares organotípicos (OMS), esferas tumorales derivadas de tejido (TDTs) y esferoides tumorales multicelulares (MCTS) (Figura 3). Algunas de las características de cada uno se describen a continuación. Las tumoresferas también llamadas mamoesferas (esfera que hace referencia a la mama), son modelos en donde hay una proliferación a partir de una célula troncal cancerosa en un medio sin adherencia y suplementado con diferentes factores de crecimiento. Estos sistemas han servido para estudiar la autorrenovación y supervivencia de células iniciadoras de cáncer de mama (Ponti et al., 2005). Por otra parte, los OMS se obtienen a partir del tumor, donde una parte del tejido tumoral se corta y se cultiva. Aunque estos modelos son los que más se asemejan a los tumores *in vivo*, los estudios en cáncer de mama son escasos. Al igual que los OMS, las TDTs son esferas que se forman a partir del tejido, empero su formación es a través de una disociación parcial mecánica o enzimáticamente del tejido, seguida de una compactación (Weiswald, Bellet & Dangles-Marie, 2015). Finalmente, los MCTS son agregados celulares esféricos que se han construido por diferentes técnicas, principalmente, a través de fuerzas físicas producidas por movimiento de los cultivos para que haya interacciones celulares o con el uso de superficies que no permitan la adherencia celular. El estudio de sistemas a nivel *in vitro* por medio de los MCTS ha permitido examinar de mejor manera terapias contra el cáncer. Por ejemplo, se han evaluado vehículos de fármacos oncológicos en cuanto a su distribución y otras propiedades fisicoquímicas, ya que los esferoides tienen una estructura tridimensional, lo que se traduce a una distribución celular más parecida a la observada en los tumores sólidos *in vivo*. Por otro lado, se ha descrito que los MCTS reflejan la formación de procesos

fisiológicos complejos como: diferentes tasas de proliferación, la expresión específica de genes, disposición de componentes del microambiente como matriz extracelular, interacciones célula-célula y célula-microambiente y resistencia a terapias (Laurent et al., 2013; Lazzari et al., 2017). A medida que el esferoide crece se ha visto que las células que se encuentran en la periferia son las más proliferantes donde hay una mayor concentración de nutrientes y oxígeno, mientras que las células que se localizan centralmente tienen un comportamiento apoptótico o necrótico. Además, se ha descrito una capa intermedia de células quiescentes (no proliferantes). Este mismo comportamiento se ha reportado en los tumores sólidos *in vivo*, como el cáncer de mama. Esto lo ha hecho un modelo prometedor para evaluar características específicas de un tumor, por ejemplo, la expresión génica, el metabolismo celular, gradientes bioquímicos, entre otros. Además, los MCTS trazan de mejor manera la dinámica tumoral, que se torna difícil de estudiar en la clínica por los métodos invasivos en la toma de muestras. De esta manera, los MCTS han brindado la oportunidad de estudiar el comportamiento de las células tumorales desde un enfoque de heterogeneidad de subpoblaciones celulares (Dufau et al., 2012).

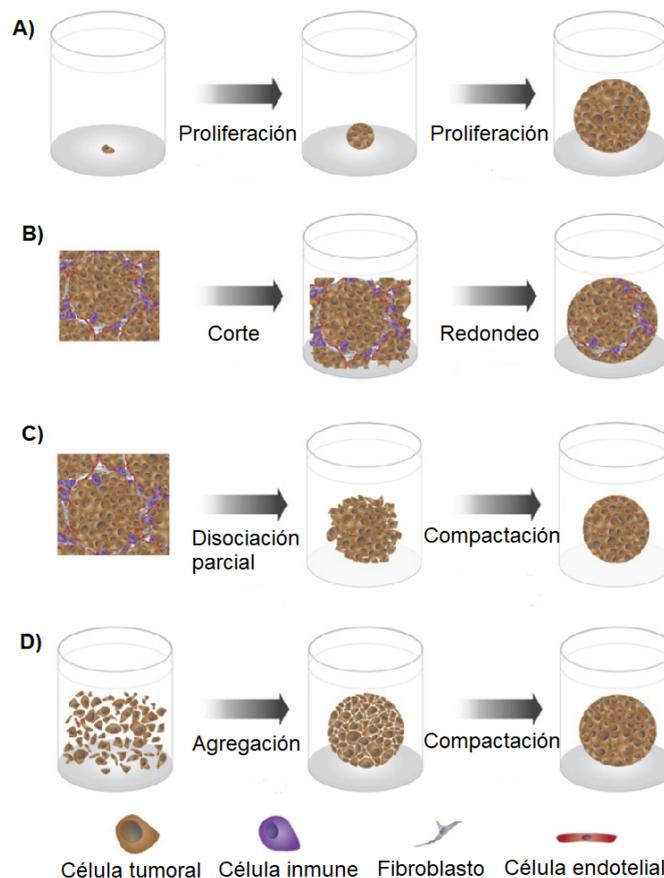


Figura 3. Formación de modelos esféricos. A) Tumoresferas. B) Esferoides multicelulares organotípicos (OMS) C) Esferas tumorales derivadas de tejido (TDTS). D) Esferoides tumorales multicelulares (MCTS). Tomada de Weiswald, Bellet & Dangles-Marie, 2015.

3. Heterogeneidad en cáncer de mama

La heterogeneidad poblacional observada en cáncer de mama se ha descrito de dos maneras: intertumoral e intratumoral (Figura 4). La primera se distingue por variaciones en los tumores entre los pacientes, por ejemplo; diferencias en los marcadores proteicos, con lo cual se han podido identificar diferentes subtipos clínicos (luminal A o luminal B) (Shi et al., 2017). La segunda es la heterogeneidad intratumoral, la cual hace referencia a que el tumor está conformado de distintos tipos celulares. Así, dentro del tumor se encuentran células tumorales, células del sistema inmune, fibroblastos y células endoteliales. Además, entre las células tumorales pueden existir diferencias fenotípicas que pueden diferir en la capacidad tumorigénica, resistencia a los tratamientos o capacidad de metástasis, llegando a formarse subpoblaciones celulares (Burrell et al., 2013). En esta tesis nos concentraremos a explorar las propiedades del segundo caso, es decir la heterogeneidad intratumoral.

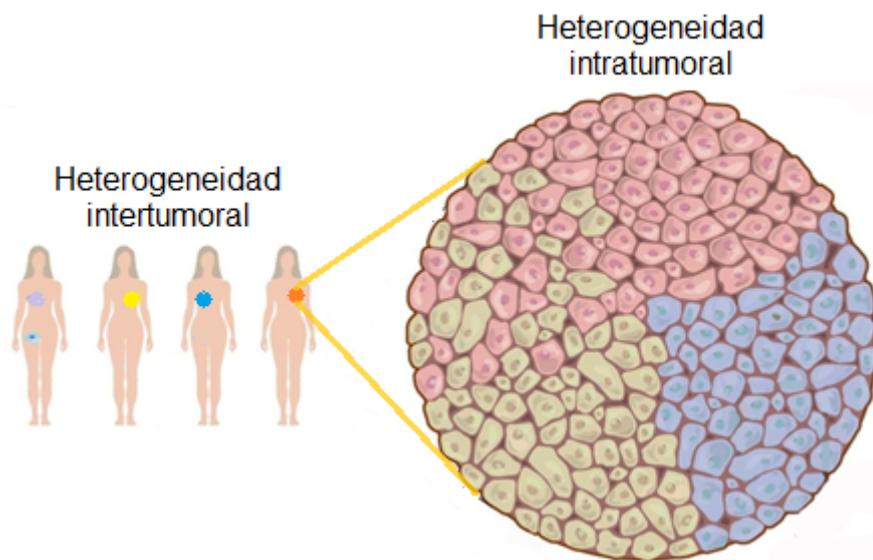


Figura 4. Heterogeneidad en cáncer de mama. Las personas con cáncer de mama presentan un subtipo de esta enfermedad (tumor representado por un color diferente: morado, amarillo, azul y naranja). Además, dentro de un tumor se observan diferentes tipos celulares. Tomada y modificada de Burrell et al., 2013.

Existen diferentes modelos que intentan describir el origen de las subpoblaciones dentro del tumor. Entre estos modelos resalta la teoría de células troncales cancerosas (CSC por sus siglas en inglés) y la teoría clonal. La teoría de CSC trata de explicar la heterogeneidad proponiendo que la formación del tumor se da a partir de un conjunto de células que adquieren la capacidad de iniciar y propagar el crecimiento del tumor, es decir, estas células con capacidad de autorrenovación se comportan como células madre. El modelo de desarrollo propuesto para estas CSC es de tipo jerárquico, de manera que los distintos tipos

celulares que se van formando dentro del tumor son a causa de las diferentes diferenciaciones, no obstante, se ha visto que este proceso no es suficiente para explicar lo que está pasando en cáncer. Así la jerarquía unidireccional asume que los tumores pueden generarse a través de una división simétrica, en la cual las células con capacidad de autorrenovación forman otras CSC o por una división asimétrica, donde en la replicación se produce tanto una CSC como una célula diferenciada (Prasetyanti & Medema, 2017). Además, estudios previos demostraron que en algunos tumores no fue suficiente únicamente erradicar a las CSC. Lo anterior, abrió el panorama hacia un aspecto interesante: la plasticidad celular. Este proceso de plasticidad aborda mecanismos reversibles, en donde una célula diferenciada puede desdiferenciarse hacia un estado similar al de sus progenitores (De Blander et al., 2021; Hanahan, 2022). En cáncer de mama se ha descrito que la pérdida de *GATA3*, un regulador crítico de la diferenciación luminal, genera una diseminación tumoral durante su progreso al provocar una pérdida de la diferenciación tumoral (Kouros-Mehr et al., 2008). Este ejemplo ilustra cómo la supresión de la expresión del factor de diferenciación puede facilitar la tumorigénesis al permitir que las células diferenciadas se desdiferencien, pero también se ha descrito la habilidad de las células progenitoras de evadir su curso hacia el estado de diferenciación terminal, lo que resulta en células atrapadas en una fase proliferativa (De Blander et al., 2021; Hanahan, 2022). Chiang y colaboradores reportaron mutaciones patogénicas en *IDH2* (isocitrato deshidrogenasa 2) en tumores de cáncer de mama y sugirieron que un mecanismo para conducir a la tumorigénesis podría ser una alteración en la diferenciación celular mediada por el oncometabolito D2HG (D-2-hidroxiglutarato) (Chiang et al., 2016). A su vez, la teoría de división clonal refleja una evolución de clones a causa de células aberrantes que permitieron la progresión del tumor al ser las mejor adaptadas al ambiente (Mavrommati et al., 2021; Wang et al., 2014). Las subpoblaciones de células que se generan dentro de un tumor surgen a través de variaciones genéticas, seguido de un crecimiento selectivo de clones que tienen una ventaja fenotípica dentro de un microambiente tumoral particular. Por ejemplo, la estructura espacial del tumor alberga diferentes gradientes de nutrientes permitiendo que algunas clones se encuentren en regiones más densas y tengan la capacidad de crecer más rápido. Asimismo, hay áreas donde los niveles de oxígeno cambian, como áreas hipóxicas, y ciertas clones pueden adaptarse a ello. Con lo cual, los tumores crecen de forma ramificada, y se conforman de subclonas genéticamente diferentes que exhiben comportamientos múltiples (Vendramin et al., 2021).

Dado lo anterior, se da la heterogeneidad intratumoral, donde existe una coexistencia de subpoblaciones molecular y fenotípicamente distintas (Vendramin et al., 2021). En términos de diversidad genética, el cáncer ha sido descrito como una enfermedad causada por

alteraciones del genoma. Dichas alteraciones reducen de forma significativa la capacidad de respuesta y los mecanismos reguladores de las células normales, hasta que las células proliferan descontroladamente e invaden tejidos normales. Entre los genes que participan en la carcinogénesis están los supresores de tumores que se encargan de restringir el crecimiento celular y evitan que las células se transformen a un estado maligno. En cáncer, estos genes pierden su función protectora cuando hay alteraciones en ambos alelos. Por otro lado, los oncogenes promueven una proliferación celular descontrolada con la mutación en un alelo, llevando a la célula al estado tumoral. Además de estas mutaciones existen rearrreglos cromosómicos, alteraciones en el número de copias, así como modificaciones epigenéticas que regulan la función de estos reguladores (Arruabarrena-Aristorena & Toska, 2022; Karp, 2011; Stratton et al., 2009). Sin embargo, los estudios en cáncer se han extendido a otros rubros para lograr entender de mejor manera el complejo panorama del mismo. Una de las herramientas ómicas utilizadas es la transcriptómica, la cual consiste en evaluar los cambios a nivel de expresión de RNA ofreciendo una perspectiva global de la actividad celular, siendo la secuenciación de RNA (RNA-seq) una de las tecnologías utilizada para los estudios en transcriptoma (Glaves & Tugwood, 2011).

4. Datos de célula única

Los datos de secuenciación de célula única (scRNA-seq) surgieron a partir de la técnica de RNA-seq. Esta técnica a diferencia de RNA-seq en masa proporciona un análisis célula por células, por lo que los cambios no expresan el promedio de la expresión de genes de todo el conjunto de células, sino el perfil de expresión genético de cada célula que integra un tejido o línea celular en cultivo (Figura 5).

4.1 Datos de célula única revelan heterogeneidad en cáncer

Recientemente, la secuenciación de RNA de célula única (scRNA-seq) ha permitido identificar subpoblaciones de células cancerosas en tumores de tejido humano y develar un mecanismo fundamental adicional: la heterogeneidad intratumoral (Figura 5) (Puram et al., 2017). Así, se ha descubierto que la heterogeneidad poblacional dentro del tumor tiene implicaciones importantes relacionadas con una progresión de la enfermedad y la resistencia de células cancerosas al tratamiento. Aún más importante, dicha heterogeneidad promueve una disminución de la efectividad de los tratamientos mediante la alteración del número de copias, mutaciones, cambios en los perfiles de expresión, entre otras (Basu et al., 2011; Chabon et al., 2016; Suda et al., 2016). Por ejemplo, Patel y colaboradores

examinaron los perfiles transcripcionales de glioblastomas y encontraron una heterogeneidad intratumoral donde se expresaron diferentes módulos transcripcionales con funciones específicas: ciclo celular, hipoxia, respuesta inmune/complemento y funciones de oligodendrocitos. Además, al clasificar a las células identificaron diferentes proporciones de subtipos celulares (proneural, clásico y mesenquimal) y determinaron que la heterogeneidad celular se asociaba con un peor pronóstico de supervivencia, es decir, existen subpoblaciones con diferencias transcripcionales que provocan mayor agresividad de la enfermedad (Patel et al., 2014). Por otro lado, Jin y colaboradores encontraron células malignas que tienen características funcionales tanto del sistema inmune como epiteliales en cáncer nasofaríngeo. Esta misma dualidad la asociaron con un pronóstico desfavorable para los pacientes, siendo los genes estimulados por interferón (*ISG*) los caracterizados principalmente (Jin et al., 2020). Interesantemente, este mismo estudio proponen la progresión del tumor por un mecanismo donde las células tumorales con una expresión alta en genes que conforman el MHCII (complejo mayor de histocompatibilidad tipo II) suprimen a las células T CD8+. Otro estudio que aborda la heterogeneidad de manera intratumoral se reportó en biopsias de melanomas (Tirosh et al., 2016). En este trabajo encontraron células tumorales con dos estados transcripcionales, uno de ellos se caracteriza con la expresión de *MITF*, un regulador transcripcional maestro de melanocitos.

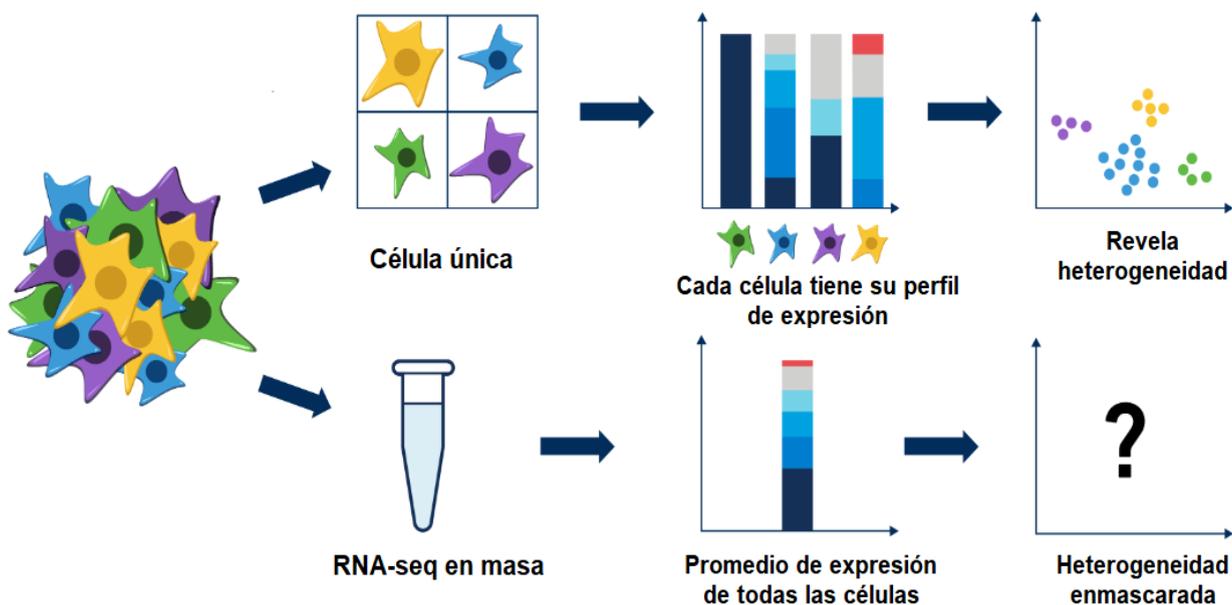


Figura 5. Diferencia entre la secuenciación de RNA de célula única y la secuenciación de RNA en masa. Tomada de *Single Cell Set-Up: Sample Preparation Tips*, 2017.

4.2 Reconstrucción de trayectorias celulares con datos de célula única

Los estudios con scRNA-seq han contribuido al entendimiento de dos campos de estudio interesantes en biología: 1) heterogeneidad celular en múltiples muestras, condiciones y enfermedades; y 2) la trayectoria que rige la diferenciación celular. La importancia biológica de los trabajos de diferenciación radica en la contemplación de las progenies celulares a partir de la célula inicial o madre, contribuyendo al entendimiento de los procesos por los cuales tienen que pasar las células para poder tener un fenotipo concreto. Estudios ya reportados en esta dirección han contribuido a realizar reconstrucciones de linajes celulares, como ejemplo estándar están los estudios de la diferenciación de células hematopoyéticas. Estos trabajos han asentado las bases para identificar no sólo poblaciones progenitoras, sino estadios intermedios, nuevas subpoblaciones y mecanismos moleculares que contribuyen a los destinos celulares (Nestorowa et al., 2016; Ruan et al., 2019; Zheng et al., 2017).

Desde el punto de vista bioinformático, la reconstrucción de trayectorias celulares es un reto interesante y un tema de vanguardia. La reconstrucción de trayectorias consiste en el análisis del progreso de un estado biológico A al estado B a través de la expresión de genes de célula única. La mayoría de los algoritmos desarrollados para esta tarea, parten de una colección heterogénea de células y exploran los patrones de expresión genéticos, los cuales están íntimamente relacionados con la transformación del proceso biológico en estudio. Estas inferencias de trayectorias celulares fueron propuestas bajo la idea de ordenar a las células en el transcurso de una variable denominada pseudotiempo, la cual simplemente cuantifica la lejanía del perfil de expresión entre dos estados determinados. Así, el pseudotiempo cuantifica una posición sobre una trayectoria, en donde los cambios de los perfiles de expresión se realizan como una función del progreso del proceso celular (Griffiths et al., 2018).

Recientemente, entre los métodos de reconstrucción de trayectorias a partir de datos de scRNA-seq se ha propuesto el algoritmo “Single-cell trajectories reconstruction, exploration and mapping of omics data”, conocido como STREAM (Chen et al., 2019). A diferencia de las propuestas anteriores, este algoritmo sobresale por sus algoritmos y sus métodos de visualización diáfanos. Además, una comparación de sus predicciones respecto a las obtenidas con otros algoritmos indica que STREAM permite realizar inferencias más precisas. También, brinda información adicional del proceso al identificar los genes de divergencia, es decir, los genes que promueven la formación de las subpoblaciones. La utilidad de este algoritmo se ha probado en sistemas biológicos de gran interés. Por

ejemplo, Chen y colaboradores analizaron células hematopoyéticas de ratón y encontraron el transcurso de diferenciación de estas células a través del pseudotiempo logrando establecer las trayectorias de las subpoblaciones que ya se habían identificado experimentalmente. Asimismo, identificaron los genes que guiaban la formación hacia células mieloides o hacia células eritroides, los llamados genes de divergencia (Chen et al., 2019). Una limitante importante es que a pesar de que STREAM infiere los genes de divergencia, el estudio de mecanismos de regulación transcripcional que induzcan la formación de diferentes fenotipos en cáncer aún está en fase de desarrollo.

5. Redes de regulación

5.1 Redes Booleanas

El complejo cableado de regulación transcripcional se ha visto alterado en diferentes enfermedades y en cáncer esta no es la excepción, por lo que existe una necesidad emergente de abordar esta enfermedad desde una perspectiva integral que puede ser estudiada bajo un análisis de biología de sistemas con un enfoque en redes. Típicamente, el conocimiento de las interacciones moleculares se resume en diagramas de complejidad variable, comúnmente conocidos como redes de interacción, las cuales representan un plano para entender la cooperatividad funcional entre genes. Las redes genéticas se pueden modelar para simular varios escenarios e incluso predecir comportamientos del sistema. Entre los modelos de regulación genética se encuentran los modelos Booleanos, los cuales permiten explorar el espacio de los posibles destinos o fenotipos celulares de una red de regulación biológica (Aldana et al., 2007; Banzhaf et al., 2020) por lo que este formalismo es una buena aproximación para explorar los fenotipos biológicos. Además, el desarrollo de este formalismo favorece dos aspectos importantes: su dinámica se asemeja al comportamiento de los sistemas biológicos y también son lo suficientemente simples como para entender y analizar. De acuerdo con lo anterior, en este trabajo, combinando datos de célula única en cáncer y la modelación de redes Booleanas, planteamos un estudio computacional para elucidar el desarrollo de subpoblaciones intratumorales, e identificar los circuitos de regulación genética que potencialmente explican las trayectorias celulares y de esa manera, caracterizar la heterogeneidad en cáncer de mama.

Antecedentes

Debido a que las líneas celulares en monocapa no revelan con fidelidad aspectos espaciales, las investigaciones en cáncer de tumores sólidos han aprovechado a los esferoides tumorales multicelulares (MCTS) como una forma más cercana a lo que sucede en un tejido humano. Recientemente, en nuestro grupo de investigación, se demostró que la línea tumoral MCF7 revela diferentes características cuando está en esferoide comparada con las células en monocapa (Muciño-Olmos, Vázquez-Jiménez, López-Esparza, et al., 2020). El estudio, indicó que los esferoides mostraban un metabolismo que se caracterizaba por la biosíntesis y degradación de aminoácidos, así como por el metabolismo de inositol y glicerofosfolípidos. En primera instancia, el estudio se enfocó en miRNAs vinculados con el metabolismo; sin embargo, los resultados dieron pauta para abordar otro aspecto fundamental de los tumores, la heterogeneidad intratumoral. De esta manera, en otro estudio de nuestro grupo de trabajo, también se analizaron los MCTS a través de la premisa en la que las células que conformaban los esferoides se comportaban igual o bien tenían actividades celulares diversas entre ellas a pesar de provenir de la misma línea celular (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020). Uno de los resultados demostró que, aun cuando las células secuenciadas provenían de dos días diferentes de formación de los MCTS: 6 y 19 días, donde a los 6 días los cultivos de MCTS estaban formados y agregados, a los 19 días los esferoides tenían un diámetro máximo, las células analizadas transcripcionalmente se agrupaban en tres grupos o clusters diferentes (Figura 6 panel A). Esto excluía una separación en los perfiles transcriptómicos originados por la temporalidad, es decir, no se obtuvieron dos agrupaciones: agrupación con células del día 6 y agrupación con células del día 19. Sino que las células con características similares se agruparon a pesar de las diferencias en las muestras. Dichas agrupaciones se conformaron de distintas proporciones de células del día 6 y 19. La agrupación A tenía alrededor del 85% de células del día 19, la agrupación C aproximadamente el 80% de células del día 6 y la agrupación B tenía una mezcla casi igual de células del día 6 y 19, ver Figura 6 panel B. Una vez que se identificaron estos clusters celulares se prosiguió a obtener información acerca de los papeles funcionales de cada grupo celular perteneciente a cada cluster. Fue así como, con un análisis de enriquecimiento de vías se encontró que cada una de las tres agrupaciones celulares se comportaba funcionalmente de distinta manera. Por un lado, el cluster A se vinculó con una respuesta de tipo inmune entre las que destacaron la respuesta de interferón, la vía de señalización de JAK, metabolismo de esfingolípidos, e incluso se asoció con aspectos de invasión de tejidos y metástasis asociándose a vías de modificación de matriz extracelular y la vía de señalización de KRAS. La agrupación C tenía un fenotipo

asociado con proliferación y control del ciclo celular, donde procesos como la replicación del DNA, transcripción y traducción fueron de gran relevancia. Por ejemplo, hubo vías enriquecidas en el procesamiento del DNA y RNA, la elongación de transcritos, el metabolismo de trifosfatos, pirimidinas y nucleótidos, además de una activación de la vía de señalización de P53 y MYC. Finalmente, la agrupación de células B, que los autores denominaron reservorio, compartía características funcionales tanto de la agrupación A como de la C. Esta última población es material de más preguntas e hipótesis, entre éstas, quizás podría ser una subpoblación que está transitando entre las subpoblaciones funcionalmente caracterizadas. Dicho trabajo concluye que no hay procesos biológicos que son definidos por los tiempos (6 y 19 días), y que las tres agrupaciones cambian en proporciones respecto al tiempo (Figura 6 panel C). Estas subpoblaciones quizás se complementan y en conjunto, de alguna manera, dotan de funciones al fenotipo maligno para asegurar la progresión y el desarrollo tumoral en MCTS. A pesar de estos avances de investigación enfocados en MCTS de la línea celular MCF7, los datos de scRNA-seq obtenidos en nuestro grupo de trabajo, brindan la oportunidad de revelar más planteamientos respecto a la heterogeneidad tumoral. Es por eso, que con el uso de estos datos transcriptómicos, este proyecto aquí planteado aborda la modulación de las subpoblaciones intratumorales de MCTS bajo una perspectiva de redes de regulación que brindará la identificación de blancos terapéuticos.

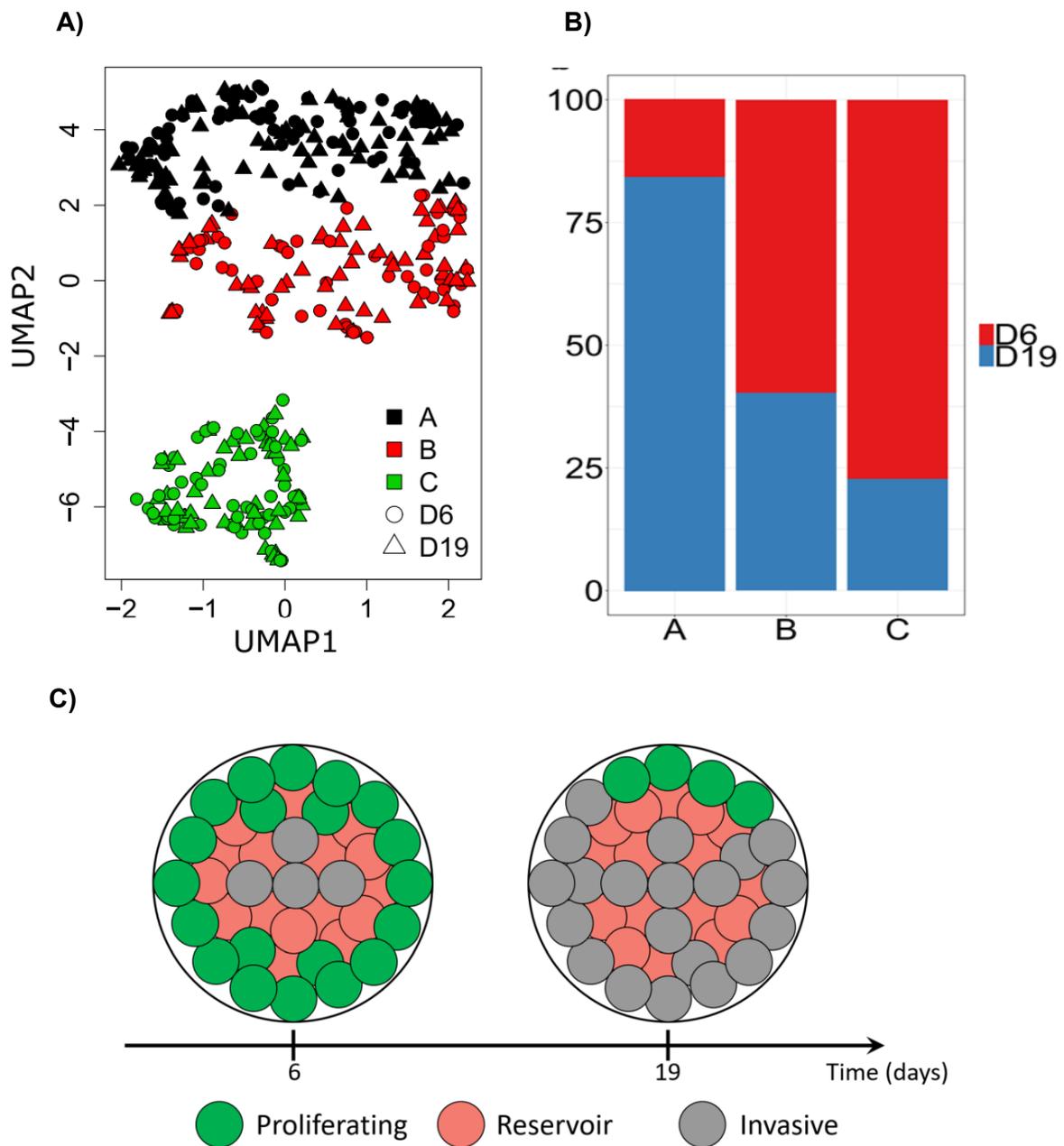


Figura 6. Subpoblaciones en esferoides tumorales multicelulares (MCTS) de la línea celular MCF7. En el panel A se observan las tres agrupaciones (negra, roja y verde) encontradas de los esferoides. Las células del día 6 están ejemplificadas con un círculo y las células del día 19 en triángulos. El panel B se muestran las proporciones normalizadas de células de los tiempos 6 y 19 días en cada cluster. En el panel C se ilustra una representación de los cambios de las subpoblaciones a través del tiempo. Tomada de Muciño-Olmos, Vázquez-Jiménez, Ávila-Ponce de León, et al., 2020.

Justificación

La heterogeneidad poblacional es una característica distintiva en cáncer que tiene una enorme influencia en la progresión, agresividad y la efectividad de los tratamientos. Notablemente, el desarrollo combinado entre RNAseq de célula única y sus algoritmos bioinformáticos han establecido una base para identificar genes relevantes que se asocian en la formación de subpoblaciones. A pesar de que estos genes representan un área de oportunidad para modular la heterogeneidad, aún es necesario estudiar los circuitos genéticos para identificar la asociación entre estos y el fenotipo de heterogeneidad. Para contribuir a la resolución de este problema, en este proyecto se busca elucidar la formación de las subpoblaciones intratumorales a través de la reconstrucción y exploración de las trayectorias celulares y así determinar una red de regulación transcripcional específica de la línea celular de cáncer MCF7. La identificación de los genes que intervienen en la generación de subpoblaciones contribuirá a la caracterización *in silico* de principios que guían la heterogeneidad poblacional en un modelo de esferoides de una línea celular de cáncer de mama. Esta propuesta de investigación permitirá, a partir del análisis computacional, construir hipótesis verificables experimentalmente enfocadas a modular la heterogeneidad y desarrollar estrategias terapéuticas factibles en el control tumoral de la línea MCF7.

Hipótesis

La composición heterogénea de subpoblaciones celulares en esferoides tumorales multicelulares (MCTS) de la línea celular MCF7 puede modularse a través de una red de regulación.

Objetivos

Objetivo general

Determinar los genes y su interacción que intervienen en la formación de subpoblaciones en un modelo de MCTS de línea celular tumoral MCF7 a través de datos de scRNA-seq y modelado computacional de redes genéticas.

Objetivos específicos

1. Identificar los genes que podrían promover la divergencia de subpoblaciones mediante el algoritmo bioinformático de reconstrucción de trayectorias STREAM a partir de datos de scRNA-seq de la línea celular de cáncer de mama MCF7.
2. Reconstruir la red de regulación transcripcional a través de una búsqueda en la literatura y con el uso de software especializado de inferencia de regulación, IPA (Ingenuity pathway analysis).
3. Inferir la regulación de la formación de las subpoblaciones a partir de la red de regulación por medio de un modelado Booleano.

Métodos

Datos

Los datos de secuenciación de RNA de célula única, de esferoides multicelulares tumorales (MCTS) derivados de la línea celular de cáncer de mama MCF7, fueron obtenidos de Muciño y colaboradores con número de acceso GSE145633 (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020). En resumen, la formación tridimensional de los MCTS se llevó a cabo en agitación orbital constante hasta que las células fueron adhiriéndose entre ellas formando agregados y finalmente esferoides. Posteriormente, los MCTS fueron disgregados hasta que las células estuvieron en suspensión. Los tiempos de desarrollo de los MCTS se tomaron a los 6 y 19 días. Mientras que a los 6 días los cultivos de MCTS estaban formados y agregados, a los 19 días los esferoides tenían un diámetro máximo y contaban con pocos núcleos necróticos a diferencia de lo observado a los 20 días. Una vez disgregados los MCTS, se filtraron para excluir células muertas, se hizo la encapsulación en microgotas y se etiquetaron las secuencias de célula única de acuerdo con la tecnología ddSEQ. Finalmente, se realizó la secuenciación con la tecnología genómica de Illumina (Illumina Bio-Rad SureCell WTA 3' Library Prep Reference Guide) con una profundidad 500 000 lecturas por célula.

Procesamiento de los datos de secuenciación de RNA de célula única

Posterior a un filtro de la matriz de conteo, donde se descartaron las células con menos de 200 genes y/o más del 5% de genes mitocondriales, se obtuvo el perfil de expresión de 364 células para ambos tiempos, con 134 y 230 células para el día 6 y 19, respectivamente, con 23922 genes para cada célula. Usando STREAM (Single-cell trajectories reconstruction, exploration and mapping) v0.3.8, a través de una interfase en línea de comando con docker, se realizó la normalización por tamaño de librería y la transformación logarítmica de los datos por medio de la función “norm” y “log2” (Chen et al., 2019). Los algoritmos utilizados en el procesamiento de los datos, así como para los siguientes métodos, se pueden revisar en: https://github.com/resendislab/Heterogeneity_Population_singleCellRNAseq

Reconstrucción de trayectorias

La inferencia de trayectorias se realizó mediante el software STREAM ejecutándolo mediante el docker propuesto por los desarrolladores y considerando los parámetros por default (lle_neighbours: 0.1, EPG_lambda: 0.02, EPG_mu: 0.1, EPG_alpha: 0.02). De forma breve, se procedió de la siguiente manera. Primero se seleccionaron los genes que tuvieron más variabilidad entre los datos usando el método LOESS (Locally estimated scatterplot smoothing). En seguida, se obtuvo la reducción de dimensiones con el método MLLÉ (Modified locally linear embedding), donde el conjunto de datos de alta dimensionalidad (miles de genes) se convirtieron en una representación de baja dimensionalidad. Las trayectorias se infirieron en función del espacio de dimensión reducido, por medio de EIPiGraph. EIPiGraph se apoya de una función de elasticidad para establecer la mejor topología de árbol y así proyectar las células en la trayectoria, las cuales son analizadas en el curso del pseudotiempo (ver Figura 7 panel A).

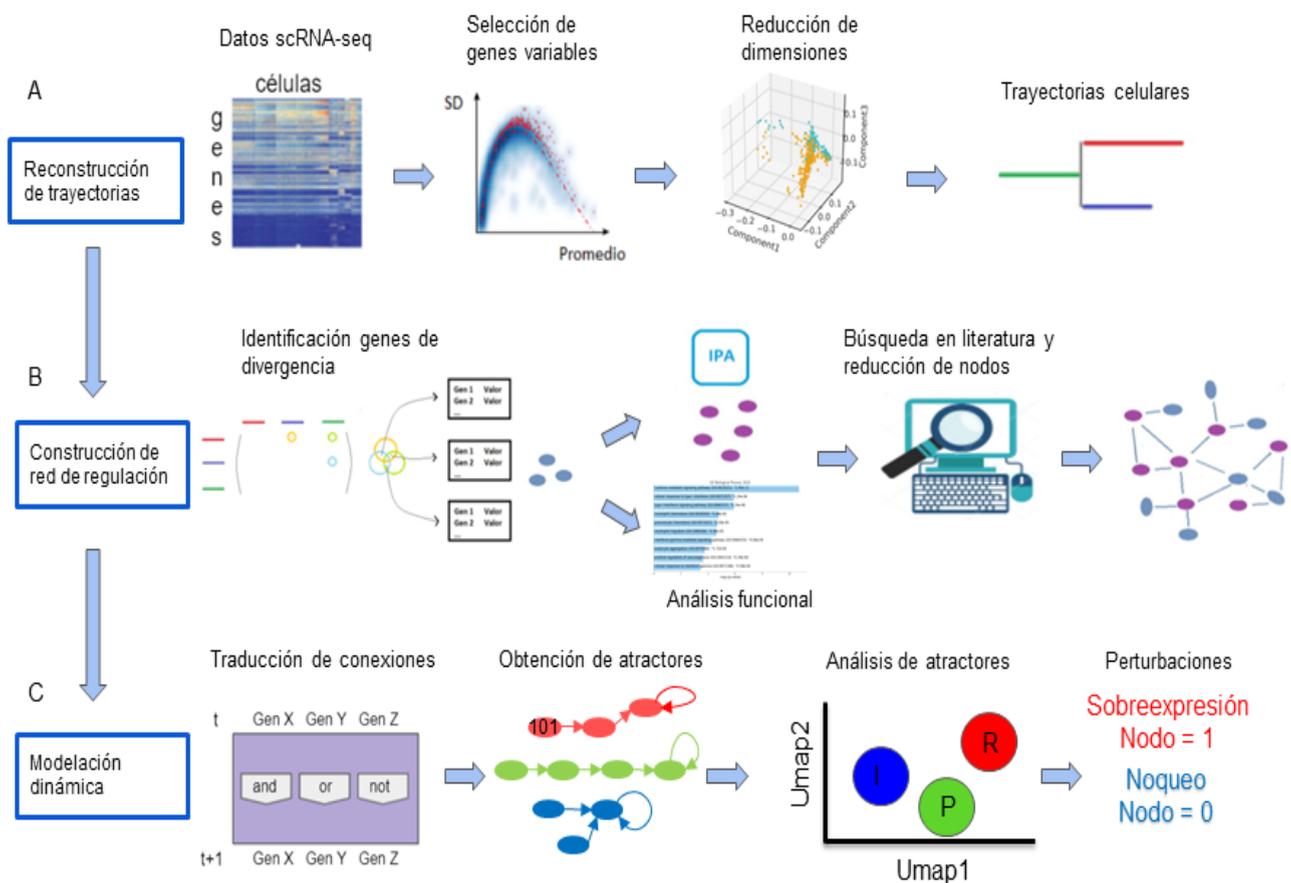


Figura 7. Descripción metodológica del análisis RNA-seq de célula única. Propuesta de análisis bioinformático, el cual constó de las etapas: inferencia de trayectorias, construcción de circuitos genéticos y análisis dinámico de la red. En el panel A se muestra la reconstrucción de trayectorias. A partir de la matriz de entrada, donde las filas son genes y las columnas son células, se seleccionaron

los genes que tuvieron más variabilidad entre los datos. En seguida, se hizo la reducción de dimensiones y las trayectorias se infirieron a través del espacio de dimensión reducido. Con las trayectorias definidas se identificaron y seleccionaron los genes que guían la formación de trayectorias. La red de regulación (B) se construyó con los genes que se identificaron como marcadores de divergencia de las subpoblaciones. Por un lado, se hizo un análisis funcional de estos genes y, por otro lado, con el uso de Ingenuity pathway analysis (IPA) se identificaron reguladores cuesta arriba, así como las conexiones regulatorias. La red se afinó a partir de una búsqueda bibliográfica en literatura y reducción de nodos. Finalmente, las conexiones se tradujeron a un lenguaje lógico y se realizó la dinámica de la red (C) por medio de un modelo Booleano obteniendo los atractores como los distintos fenotipos asociados a esa red y realizando perturbaciones para establecer las transiciones entre las subpoblaciones.

Identificación de genes de divergencia

Una vez obtenidas las trayectorias celulares se identificaron los genes marcadores que conducen la formación de subpoblaciones por medio del mismo algoritmo a través de la función “DE”. Como resultado de este análisis, se obtuvo una lista de genes expresados diferencialmente para cada comparación. Estos genes son llamados de divergencia y constituyen el conjunto de genes con potencial para guiar y transformar la transición de un estado poblacional a otro. Posteriormente, a los genes se les hizo un filtro quedándonos con aquellos con un z-score y $\log_{2}fc \geq 2$ y un p-value ≤ 0.01 . Para obtener los genes expresados de manera global para cada trayectoria respecto a las otras se construyeron listas de intersección donde un gen global se tomaba en cuenta cuando aparecía en las dos comparaciones (ver Figura 7 panel B).

Análisis de enriquecimiento

Los procesos biológicos sobre-representados para cada conjunto de genes de cada una de las subpoblaciones de las trayectorias se hizo con el software EnrichR (Kuleshov et al., 2016) disponible en <https://maayanlab.cloud/Enrichr/>. Los gene sets ocupados fueron los de la base de datos de Gene Ontology-Biological Process. La significancia estadística considerada para identificar el enriquecimiento de un proceso biológico fue con un FDR < 0.05 .

Reconstrucción de red de regulación

Con el fin de poder establecer los mecanismos de regulación de los genes, es decir, las reglas que gobiernan la dinámica, para la construcción de la red de regulación genética se

empleó el software IPA (Ingenuity pathway analysis) (Krämer et al., 2014). Esta herramienta bioinformática permitió encontrar los reguladores cuesta arriba de los genes de divergencia encontrados para las diferentes subpoblaciones en la reconstrucción de trayectorias con base a la expresión de los datos, lo cual se realizó con la función “upstream regulator” y con el filtro para factores de transcripción con un p-value < 0.01. La búsqueda y conexión entre los genes se hizo con los siguientes parámetros: species: human, mouse; tissues and cell lines: mammary gland, MCF7; relationship types: activation, inhibition and expression. Además, con búsquedas bibliográficas en la literatura científica, en la base de datos de NCBI, se agregó información de reguladores importantes en el contexto de cáncer de mama de tipo luminal. La red se sometió a una reducción de nodos (genes) de manera manual tomando en cuenta aquellos nodos que no influían en la regulación.

Traducción de conexiones regulatorias a funciones Booleanas

Las conexiones regulatorias de la red fueron traducidas a funciones Booleanas para posteriormente proceder a su simulación dinámica binaria. De acuerdo con la regulación para cada uno de los nodos de la red, las conexiones se tradujeron a compuertas lógicas tales como OR (|), AND (&) y NOT (!) que representan suma, producto y complemento en álgebra Booleana. Así para el caso donde está evidenciado que dos factores de transcripción coordinan la regulación de un gen, se utilizó el operador “&”; mientras que cuando existe la posibilidad de que uno u otro factor de transcripción regulen al mismo gen, se utilizó el operador “|”. Las reglas Booleanas correspondientes a cada nodo (gen o proteína) de la red y su evidencia en la literatura se resumen en la Tabla 2.

Tabla 2. Establecimiento de las reglas lógicas para las conexiones de la red de regulación.

Nodo	Regla lógica o Booleana	Fuente
AURKA	(CREB1 & ESR1) (FOXM1)	(Madak-Erdogan et al., 2011; Yang et al., 2017)
BIRC5	(CREB1 & ESR1) (! BRCA1)	(Madak-Erdogan et al., 2011; Promkan et al., 2013)
BRCA1	(BTG2 & NFE2L2 & ! TP53)	(Arizti et al., 2000; Karve & Rosen, 2012)
CCNB1	(CREB1 & ESR1) (! BRCA1)	(Madak-Erdogan et al., 2011; Promkan et al., 2013)
CCND1	(GATA3 ! PTTG1)	(Hatcher et al., 2014; Shan et al., 2014)
ESR1	(ESR1 & GATA3) (TP53)	(Eeckhoute et al., 2007; Shirley et al., 2009)
FOXA1	any(GATA3, BRCA1, ! TWIST1)	(Gong et al., 2015; Kouros-Mehr et al., 2006; Xu et al., 2017)

FOXM1	(FOXM1 & AURKA) (! TP53)	(Barsotti & Prives, 2009; Yang et al., 2017)
GATA3	(GATA3 & ESR1)	(Eeckhoutte et al., 2007)
IL20	(GATA3 & ESR1 & ! FOXA1)	(Lee et al., 2017)
MKI67	(CREB1 & ESR1)	(Madak-Erdogan et al., 2011)
MYC	FOXM1	(Yang et al., 2017)
NFE2L2	BRCA1	(Gorrini et al., 2013)
STAT1	STAT3	(Han et al., 2013)
SNAIL	(PTTG1 MYC)	(Cho et al., 2015; Yoon et al., 2012)
SOD2	(BTG2 & NFE2L2)	(Karve & Rosen, 2012)
S1007	(! BRCA1)	(Salama et al., 2008)
TWIST1	STAT3	(Cheng et al., 2008)

Modelado Booleano

A partir de la red de regulación reconstruida, integrada por 23 nodos y 40 interacciones, se realizó el estudio de su dinámica mediante un formalismo Booleano. Para ello, los estados posibles asociados a la red de estudio se representan por un vector. Dicho vector (v_i) se integra por valores binarios 1 y 0 (prendido y apagado respectivamente), los cuales representan la información de actividad de cada gen incluido en la red reconstruida, mediante funciones Booleanas (F), el estado de cada nodo se actualiza por unidades discretas de tiempo (t) hasta que los estados de cada nodo alcanzan su estado estacionario. El formalismo matemático se condensa de la siguiente manera:

$$v_i(t + 1) = F_i(v_1(t), v_2(t), \dots, v_N(t)) \quad (\text{Saadatpour et al., 2010})$$

Lo anterior, describe la conducta dinámica del nodo i -ésimo (v_i). F_i es una función Booleana que expresa cómo está siendo regulado el nodo i por los componentes que propiamente lo regulan. Una parte crucial de la dinámica se establece en las funciones Booleanas F . Para ello, las interacciones regulatorias de tipo activación o inhibición se transformaron a un idioma de operadores lógicos (ver métodos “Traducción de conexiones regulatorias a funciones Booleanas”) que forman parte del argumento de F_i . El estado del nodo i en una unidad de tiempo posterior está dado por $v_i(t + 1)$, la cual está en función de los estados previos. La evolución temporal de la red a través de un modelo de implementación asincrónica asume que en cada unidad de tiempo el sistema actualizará una función

aleatoriamente. Este proceso es iterativo y finaliza hasta que eventualmente se alcance un estado estacionario, o también llamado atractor. Los atractores pueden ser simples o cíclicos, en esta tesis nos enfocaremos al análisis únicamente de los atractores simples. Es decir, aquellos que cumplen la condición siguiente en el estado estacionario al tiempo t^* :

$$v_i(t + 1) = v_i(t^*) \text{ atractor simple}$$

Simulación y análisis de atractores con el modelo Booleano

La red de regulación transcripcional se actualizó asincrónicamente, haciendo uso del archivo de entrada .txt con las reglas booleanas establecidas (Tabla S1), con la paquetería de BoolNet implementada en R (Müssel et al., 2010). Los atractores encontrados a través de la función “getAttractors()” se proyectaron en un espacio de dos dimensiones con UMAP (Uniform manifold approximation and projection) con la paquetería en R “umap”. Enseguida, para determinar el número óptimo de clusters se utilizó la paquetería “NbClust”, la cual se basa en un conjunto de 30 métricas para determinar la mejor forma de agrupación a través de k-medias (Charrad et al., 2014). Finalmente, los estados de encendido y apagado de los atractores sirvieron para definir los fenotipos, es decir, ya que los genes toman un valor binario de 1 ó 0, los genes que tenían un valor 1 ó 0 para todos los atractores dentro un cluster (Figura 8), dieron pauta para establecer el sentido biológico de cada agrupación. Lo mencionado anteriormente se repitió un total de 10 veces y se reportó el promedio.

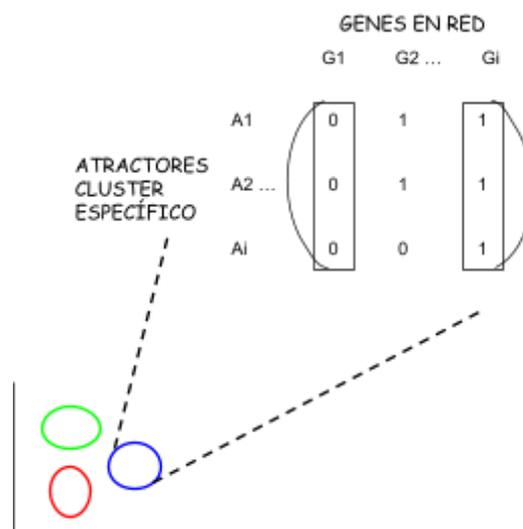


Figura 8. Estados de encendido y apagado de los genes para los atractores dentro de los clusters. Para cada cluster verde, rojo y azul (en esta figura se muestra como ejemplo el cluster azul), se identificaron los genes en encendido (1) y apagado (0) para definir los fenotipos.

Perturbaciones *in silico*

Las perturbaciones en la red de tipo knockout o sobreexpresión se hicieron con la función “fixGenes()” del paquete BoolNet. Para el caso del knockout, los genes se apagaron fijándolos en un valor de 0 durante toda la simulación; mientras que, para el caso de la sobreexpresión, se mantuvieron encendidos estableciendo el valor de 1. Los atractores encontrados, condicionando la expresión de los genes de la red, fueron nuevamente proyectados en un espacio de dos dimensiones con UMAP, se encontró el número óptimo de clusters y se analizaron los estados de los genes de cada cluster, como se describió en la sección “Simulación y análisis de atractores con el modelo Booleano”.

Resultados

Como parte del procesamiento de los datos de scRNA-seq, posterior a un filtro donde se descartaron las células que no eran informativas, con menos de 200 genes y/o más del 5% de genes mitocondriales, se realizó la normalización y la transformación logarítmica de los datos. Para ello se revisaron y probaron varios métodos de normalización propuestos en la literatura (Lytal et al., 2020). Después de comparar los métodos de normalización se seleccionó el método por tamaño de la librería. Este método fue el mejor y el más sencillo que se probó, ya que no utilizaba tantos parámetros como los otros algoritmos. Asimismo, era el método de normalización predeterminado en el algoritmo de inferencia de trayectorias que se iba a utilizar (STREAM). Cuando se compararon las proyecciones de los datos en una baja dimensionalidad no se observaron muchas diferencias en el espacio reducido de dos dimensiones respecto a los otros algoritmos (SCnorm, scran, Linnorm). Es importante mencionar que las diferencias se observaron cuando no se normalizaron los datos. En los datos sin normalizar, no se observaban los grupos celulares separados por días como con los datos normalizados, donde sí se observó una tendencia a formar agrupaciones (Figura 9).

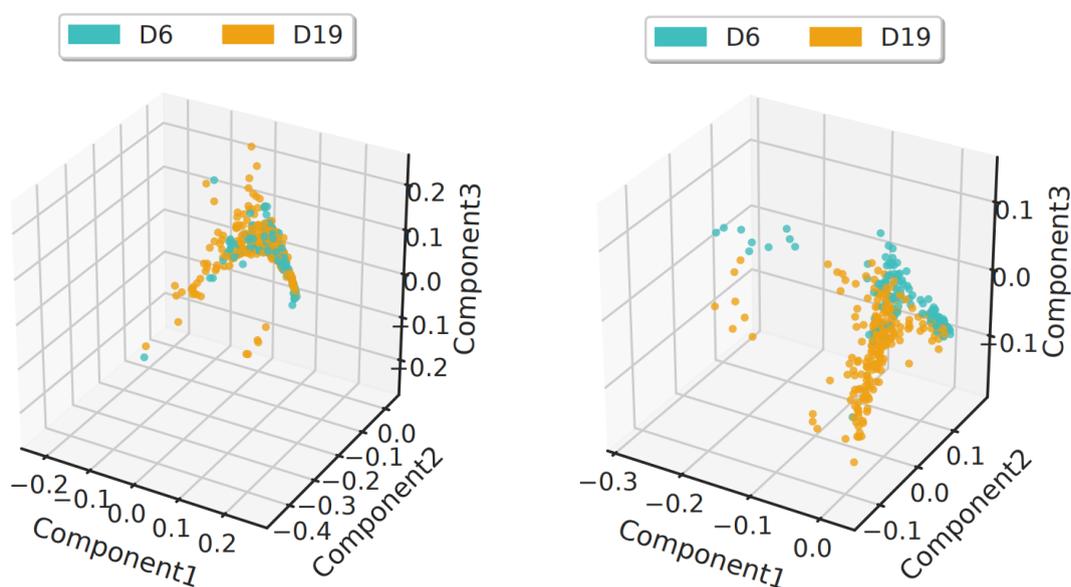


Figura 9. Comparación de datos antes y después de la normalización. A la izquierda se muestran los datos sin normalizar donde no se observa una separación de las muestras y a la derecha los datos después de efectuar la normalización por tamaño de la librería donde ya se muestra una separación. Cada componente indica el espacio de dimensión (dimensión1, dimensión 2 y dimensión 3). Técnica de reducción de la dimensionalidad hecha con el método MLLE (Modified locally linear embedding). D6 (células del día 6). D19 (células del día 19).

Los datos de célula única de los MCTS reflejan tres trayectorias celulares

Previamente en nuestro grupo de trabajo, al analizar las células del día 6 y 19 de formación de los MCTS no se encontraron dos agrupaciones o clusters que fueran indicativas de una separación debido a los dos diferentes tiempos, sino que se observó la formación de tres clusters con proporciones celulares diferentes (ver antecedentes). Con esta información en mente, surgió la pregunta de cómo serían las trayectorias celulares de estos esferoides y si al reconstruirlas obtendríamos tres trayectorias indicativas de los tres clusters. Fue así como con los datos de scRNA-seq normalizados y transformados logarítmicamente, se realizó la inferencia de trayectorias celulares con el algoritmo STREAM (Chen et al., 2019). En la Figura 10, se muestra el resultado de la reconstrucción de trayectorias. En ella se pueden observar la formación de tres subpoblaciones tumorales ejemplificadas en las trayectorias S3_S1, S1_S2 y S1_S0 que están en color verde, roja y azul, respectivamente. En este caso, observamos el inicio de la progresión a partir de la trayectoria verde (trayectoria S3_S1); sin embargo, es posible cambiar el punto de inicio (raíz) y obtener otras

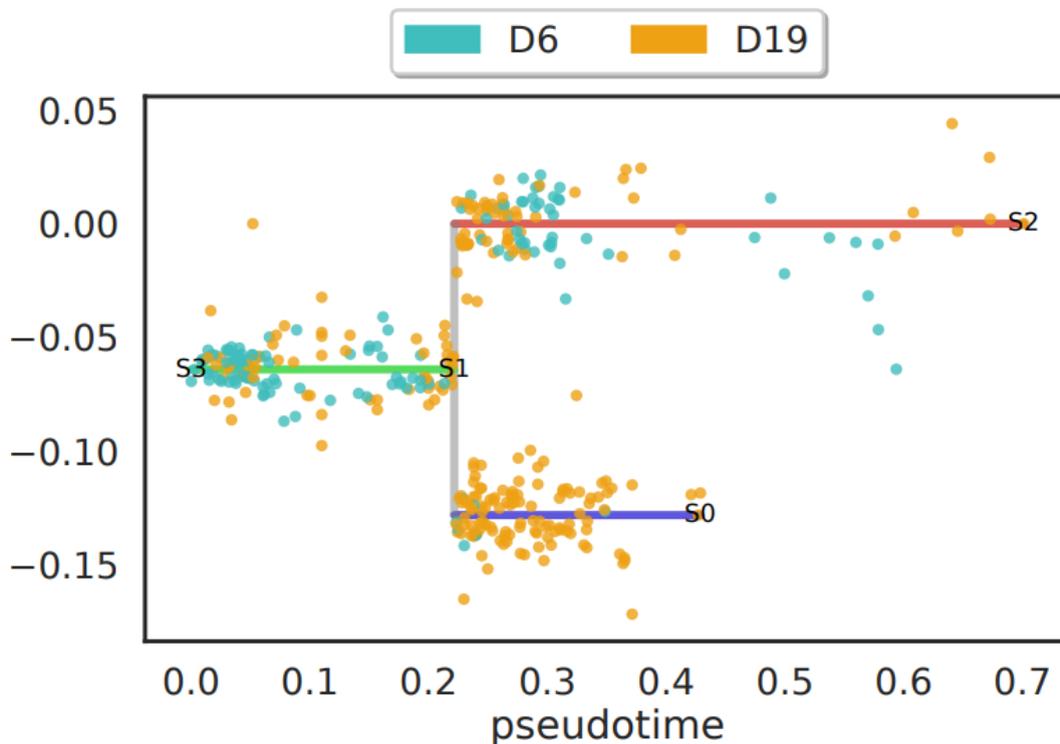


Figura 10. Reconstrucción de trayectorias de los esferoides de la línea celular MCF7 con datos de scRNA-seq. Cada punto representa una célula, en azul las células del día 6 (D6) y en naranja las células del día 19 (D19). En el eje de las Y se observa la distancia entre las trayectorias respecto a una célula tomada como punto de referencia. En el eje de las X representa la distancia (pseudotiempo) desde el nodo de inicio a lo largo de las diferentes trayectorias.

reconstrucciones (Figura S1). No obstante, al analizar las reconstrucciones nos dimos cuenta de que se tenía la misma información en cada una de ellas, es decir, observamos la formación de tres subpoblaciones celulares y las trayectorias eran invariantes en su estructura (verde, roja y azul). Dada la equivalencia de las trayectorias en el pseudotiempo, se decidió trabajar con la inferencia que se muestra en la Figura 10, en la cual el punto de inicio comienza con células del día 6 (D6) que era la primera muestra respecto a la otra muestra tomada a los 19 días (D19).

Por otro lado, Muciño-Olmos y colaboradores identificaron tres clusters o agrupaciones con diferentes proporciones celulares que fueron caracterizadas funcionalmente como: 1) cluster A invasiva/metastásica/inmune, 2) cluster C de proliferación, y 3) cluster B un reservorio que compartía funciones de los otros dos clusters (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020). Considerando esta información, se analizaron las proporciones celulares de la reconstrucción de trayectorias de la Figura 10 (ver Tabla 3), donde la trayectoria azul tuvo un 90% de células del día 19, la trayectoria roja una proporción casi equivalente de células del día 6 y 19 (59% y 41%, respectivamente), y la trayectoria verde una mayor proporción de células del día 6 (70%). Con esta información se prosiguió a realizar comparaciones específicas de las proporciones celulares entre las trayectorias y los clusters, encontrados previamente en nuestro grupo de trabajo, de la siguiente manera: trayectoria azul vs cluster A, trayectoria roja vs cluster B y trayectoria verde vs cluster C. El análisis estadístico demostró que las proporciones celulares de las trayectorias eran semejantes comparándolas con los cluster: cluster A tenía un 85% de células del día 19, el cluster B una mezcla de células idéntica al encontrado por medio de las trayectorias y el cluster C un 77% de células del día 6 (p valor mayor a 0.05 para todas las comparaciones, es decir, no fue significativo estadísticamente). Además, la proporción de células del día 6 y día 19 se invirtió entre la trayectoria celular azul y verde, con una mayor proporción de células del día 19 y menor del día 6; y una menor del día 19 y mayor del día 6, respectivamente. De esta manera, este análisis muestra cómo las proporciones celulares del día 6 y 19 para las tres subpoblaciones determinadas a través de la reconstrucción de trayectorias son parecidas a las proporciones encontradas en los tres clusters por Muciño-Olmos y colaboradores.

Tabla 3. Proporciones celulares normalizadas para las células tanto del día 6 como del día 19 para cada una de las trayectorias celulares o clusters.

	Células D6		Células D19		
	Proporciones celulares (%)		Proporciones celulares (%)		
Trayectoria/Cluster	Trayectorias	Clusters	Trayectorias	Clusters	Valor de p*
Azul/A	10	15	90	85	0.3
Roja/B	59	59	41	41	1.0
Verde/C	70	77	30	23	0.3

*El valor de p fue obtenido a través de la prueba Chi-cuadrada.

Una vez hechas las trayectorias celulares se obtuvieron los genes de divergencia, que son los genes que probablemente están promoviendo la formación de estas subpoblaciones. Para ello obtuvimos los genes diferencialmente expresados (DE) haciendo comparaciones entre las tres trayectorias (Figura S2). Como ejemplo se muestran el top 15 de los genes DE entre la trayectoria azul (S1_S0) comparada contra la trayectoria verde (S3_S1) (Figura 11). Posteriormente, se analizaron algunos genes DE dentro de cada subpoblación con la finalidad de vislumbrar algunas características funcionales. Para el caso de la trayectoria verde (S3_S1) se encontró a *MKI67*, que es un biomarcador de proliferación (Sun et al., 2017), altamente expresado en esta rama (Figura 12). En esta misma trayectoria se observó una expresión alta de *BIRC5*. Este gen codifica para la survivina, la cual se ha reportado que participa en diferentes funciones biológicas, entre ellas una inhibición de la apoptosis y control de la proliferación celular (Jha et al., 2012; Zhang et al., 2004). Adicionalmente, la expresión de *AURKA* también se vio elevada en esta rama. Este gen codifica para Aurora Cinasa A, la cual es capaz de formar un complejo con survivina con actividad en la segregación de cromosomas y la citocinesis en la división celular (Yamamoto et al., 2008). Para la trayectoria azul (S1_S0) se encontraron varios genes DE asociados con una respuesta inmune, como es el caso de los genes de la familia de los *IFIT* que son genes

inducidos por interferón pero que se asocian a procesos tumorales de invasión y resistencia de terapias (Post et al., 2018; Provance & Lewis-Wambi, 2019). Además, estos genes son regulados por *STAT1*, el cual se encuentra altamente expresado en la misma trayectoria (Figura 12). En conjunto, estos resultados nos sugirieron que posiblemente estábamos observando las trayectorias celulares de los tres clusters que Muciño-Olmos y colaboradores identificaron previamente en nuestro grupo de trabajo.

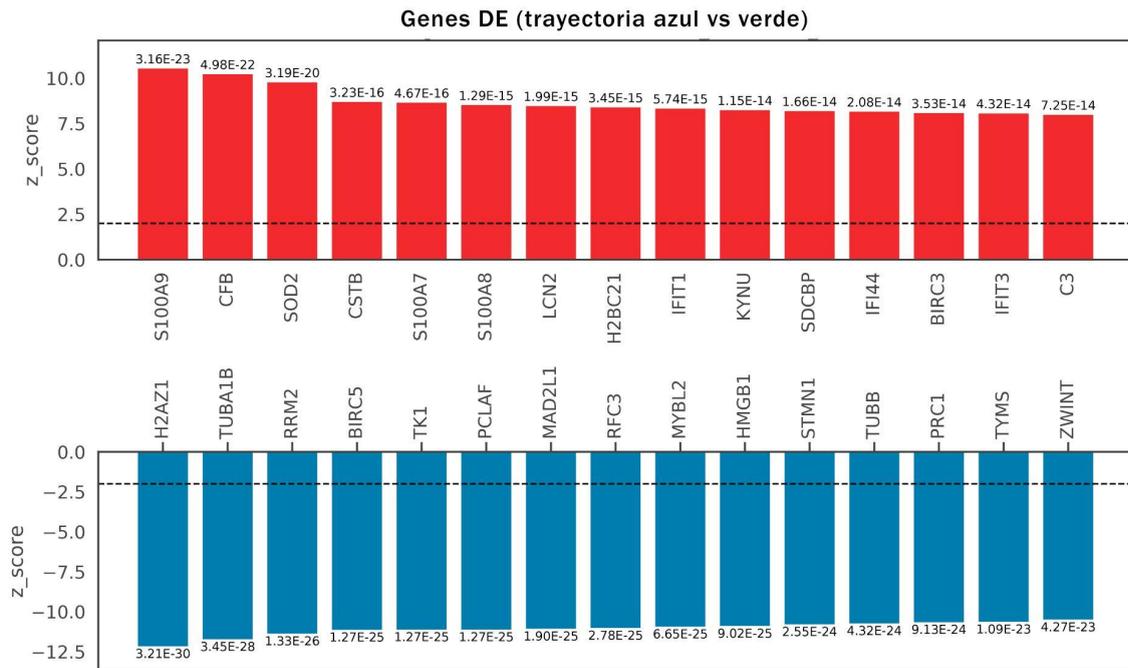


Figura 11. Genes de divergencia entre la trayectoria azul vs la trayectoria verde. En la parte superior se observan los genes expresados al alta en la trayectoria azul y en la parte inferior los genes expresados a la baja en comparación con la trayectoria roja. En este caso sólo se muestra el top 15. Los genes con un z-score > 2 y un p-value < 0.01 se consideraron genes diferencialmente expresados.

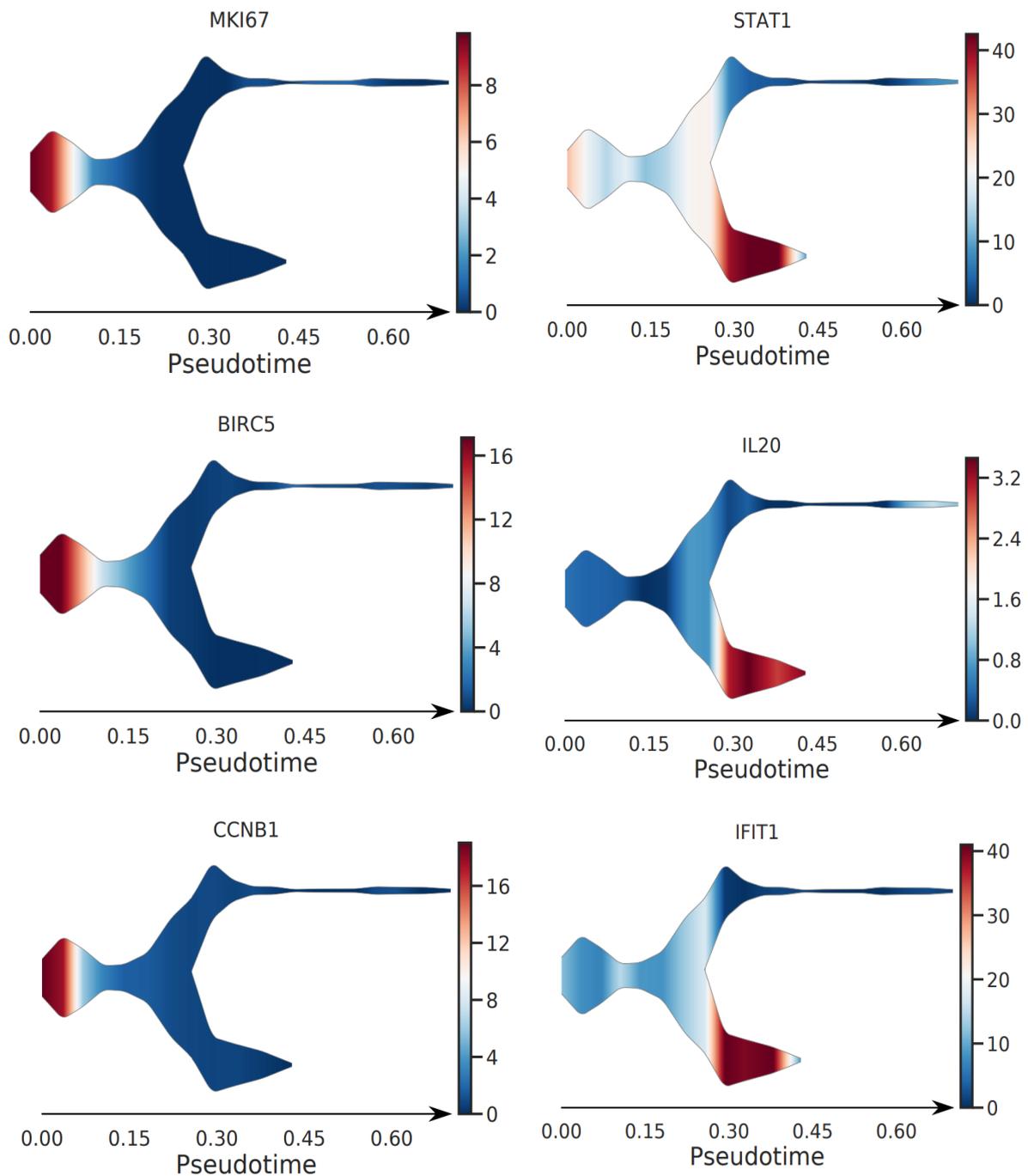
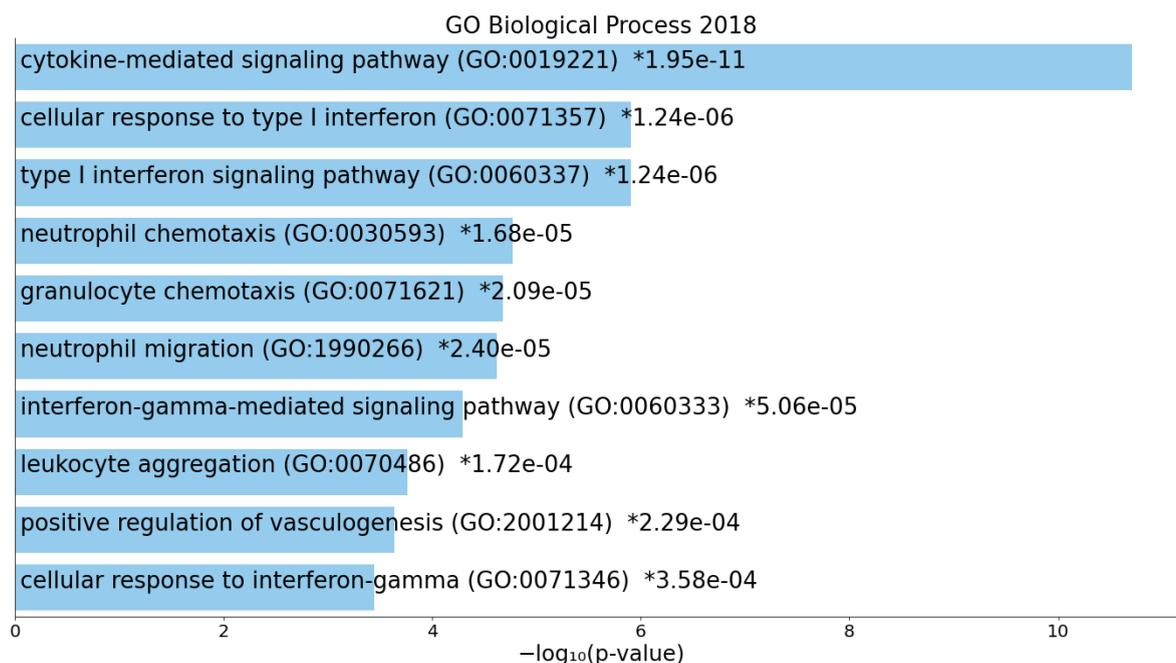


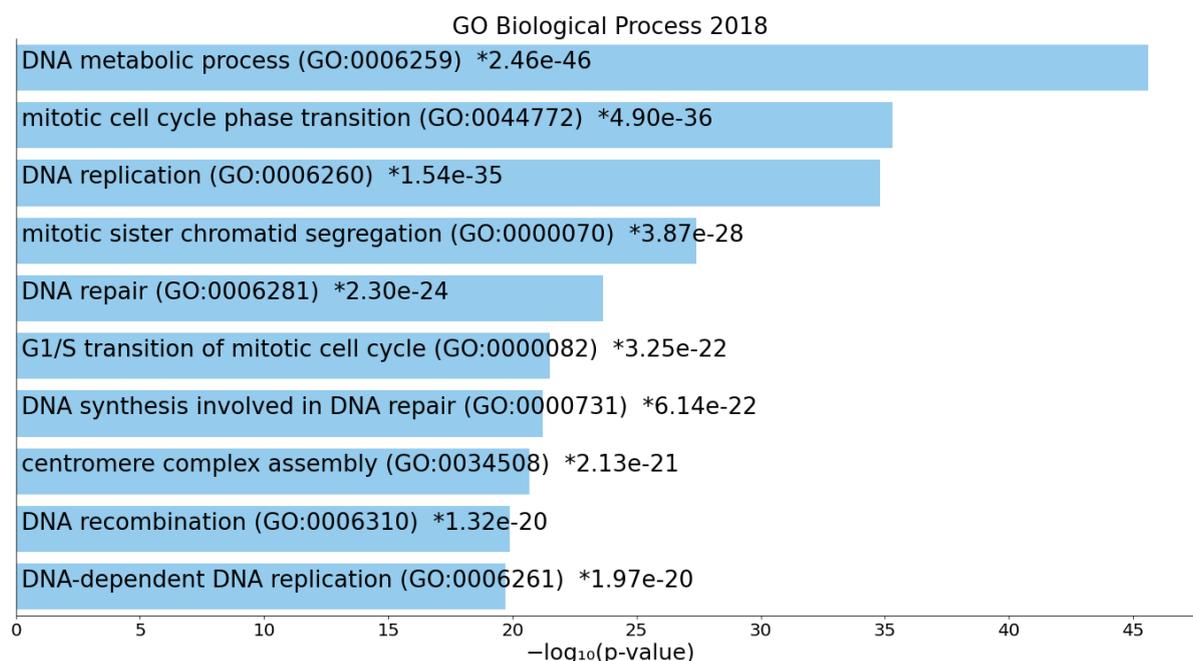
Figura 12. Genes expresados diferencialmente en la trayectoria verde y azul. A la izquierda se muestra una expresión alta de los genes *MKI67*, *BIRC5* y *CCNB1* en la trayectoria verde (S3_S1) vs azul (S1_S0). A la derecha una expresión alta de los genes *STAT1*, *IL20* e *IFIT1* en la trayectoria azul (S1_S0) vs roja (S1_S2). En el eje de las Y se muestra el \log_2 del coeficiente de la expresión del gen en las condiciones analizadas. El color rojo y azul hacen referencia a una alta y baja expresión, respectivamente.

Fue así como, para verificar que las actividades funcionales con las que estábamos proponiendo las subpoblaciones de las trayectorias celulares seguían siendo: Azul) invasiva/inmune, Verde) proliferación y Roja) funciones compartidas de azul y verde, se realizó un análisis funcional de vías por medio de Over-representation analysis (ORA), los resultados se muestran en la Figura 13. Entre los procesos biológicos sobre-representados para la trayectoria azul se observó la vía de señalización mediada por citocinas, la quimiotaxis de granulocitos, así como una respuesta de interferón (Figura 13 panel A). Para la trayectoria verde (Figura 13 panel B), el enriquecimiento que se obtuvo fue en actividades del ciclo celular y procesos del DNA, tal como replicación. Finalmente, para la trayectoria roja (Figura 13 panel C) se obtuvieron enriquecimientos funcionales en procesos tanto del sistema inmune como en la replicación del DNA.

A) Análisis para la trayectoria azul



B) Análisis para la trayectoria verde



C) Análisis para la trayectoria roja

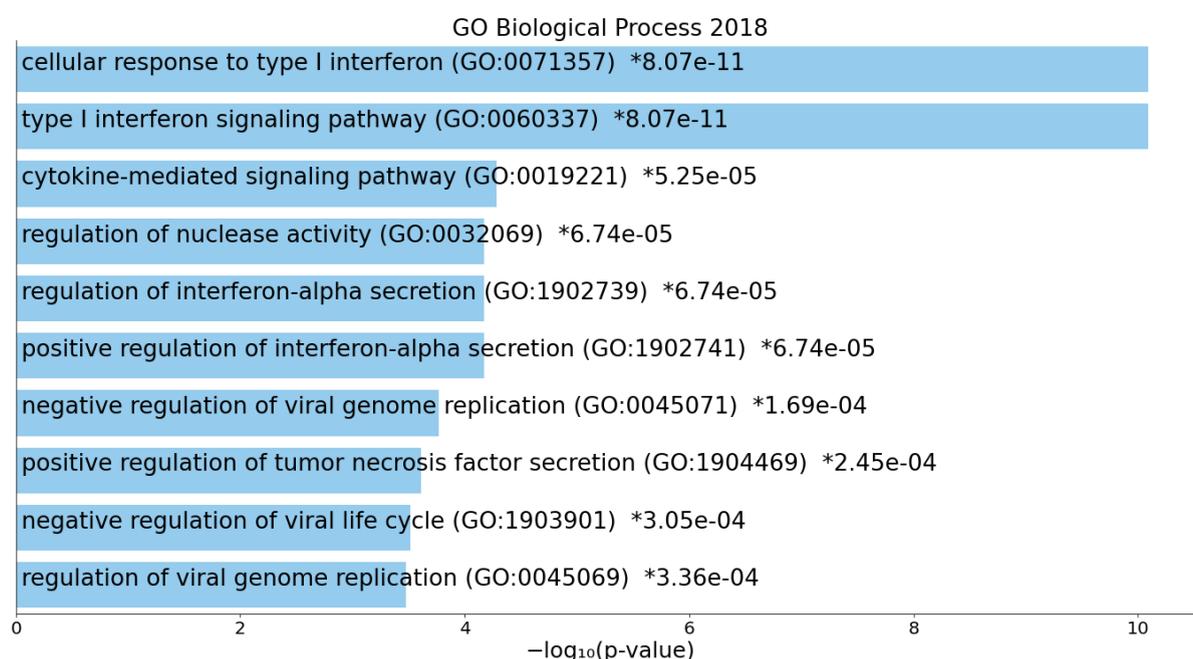


Figura 13. Análisis de enriquecimiento para el conjunto de genes de cada subpoblación. A) ORA para la subpoblación de la trayectoria azul. B) ORA para la subpoblación de la trayectoria verde. C) ORA para la subpoblación de la trayectoria roja. En barras se observa el proceso biológico sobre-representado. En el eje de las X se muestra el grado de significancia sobre el gene-set especificado. ORA (Over-representation analysis).

Reconstrucción de red de regulación transcripcional

Una vez obtenidos los genes DE, una premisa central en nuestro trabajo era que estos genes potencialmente tendrían un papel central en la transformación de las subpoblaciones. Con esta idea en mente, se prosiguió con la construcción de una red de regulación transcripcional curada en la literatura e involucrando a estos genes como elementos semilla. Ya que no existe un pipeline específico para construir la red de regulación, ésta se construyó seleccionando características basadas en la literatura y en los datos observados. Tomando en cuenta los datos, se utilizaron los genes de divergencia previamente identificados, seleccionando aquellos con un z-score y $\log_{2}fc \geq 2$ y un p-value ≤ 0.01 . Una vez hecho este filtro, se utilizó IPA para direccionar la reconstrucción (Krämer et al., 2014). IPA es una herramienta que contribuye a identificar reguladores cuesta arriba de acuerdo con la expresión del conjunto de datos y que se basa fundamentalmente en trabajos experimentales que están en la literatura. Cabe recalcar que la base de datos con la que cuentan es altamente curada por lo que las interacciones en la red son justificables experimentalmente, además de limitar la búsqueda al tipo celular en estudio, en este caso cáncer de mama luminal. Siguiendo estos pasos, se construyó la red de regulación que se muestra en la Figura 14. Esta reconstrucción incluye genes de divergencia entre las subpoblaciones, los reguladores y las conexiones entre ellos.

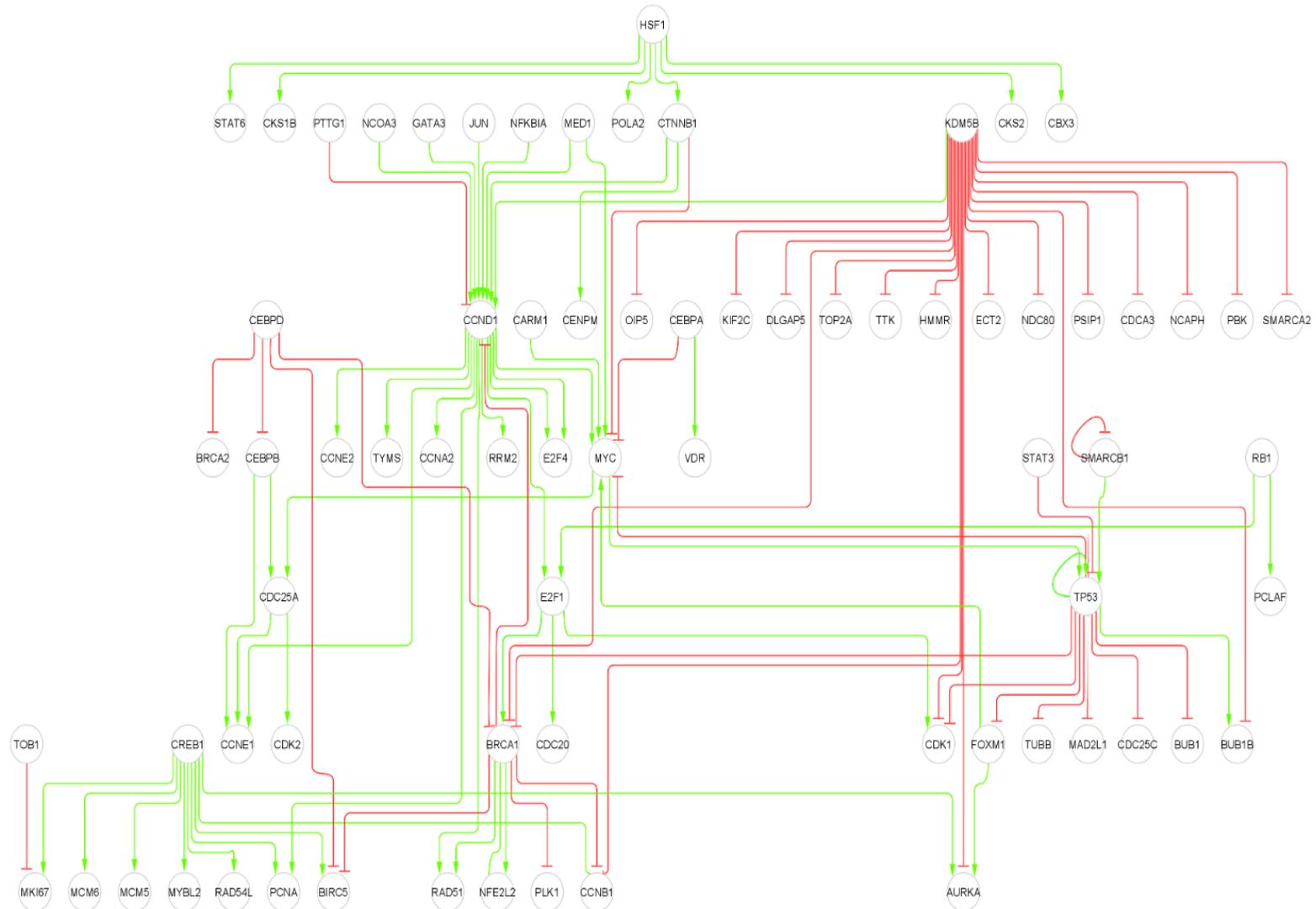


Figura 14. Red de regulación transcripcional. En círculos se observan los genes que representan los nodos de la red, mientras que las conexiones de tipo activación se representan con flechas (→) verdes y las de inhibición con flechas truncas (-) rojas.

Una vez obtenida la red de regulación de la Figura 14, se prosiguió con la reducción de nodos, tratando de identificar aquellos con relevancia en la definición del fenotipo. Este procedimiento se llevó a cabo descartando los nodos cuya regulación no repercuten en la modelación de la red, como los son los nodos que se encuentran en la periferia y que sólo están siendo regulados, como es el caso de *CBX3* (Figura 14). La reducción de nodos de una red se justifica desde el aspecto matemático. Ya que el objetivo de este proyecto fue modelarla a través de un modelo Booleano, una red con menos variables, pero que aun así sea predictiva, será más fácil de modelar y entender (Naldi et al., 2011). Asimismo, se redujo el número de nodos después de quedarnos con los genes que codifican proteínas que tuvieran actividad de factores de transcripción y que, por lo tanto, regulan la transcripción de genes dentro del contexto de cáncer de mama luminal. Eventualmente, se volvió a aplicar una reducción de nodos descartando aquellos que quedaron en la periferia como se hizo primero. De esta manera, la red de regulación transcripcional pasó de tener 73 a 15 nodos como se muestra en la Figura 15.

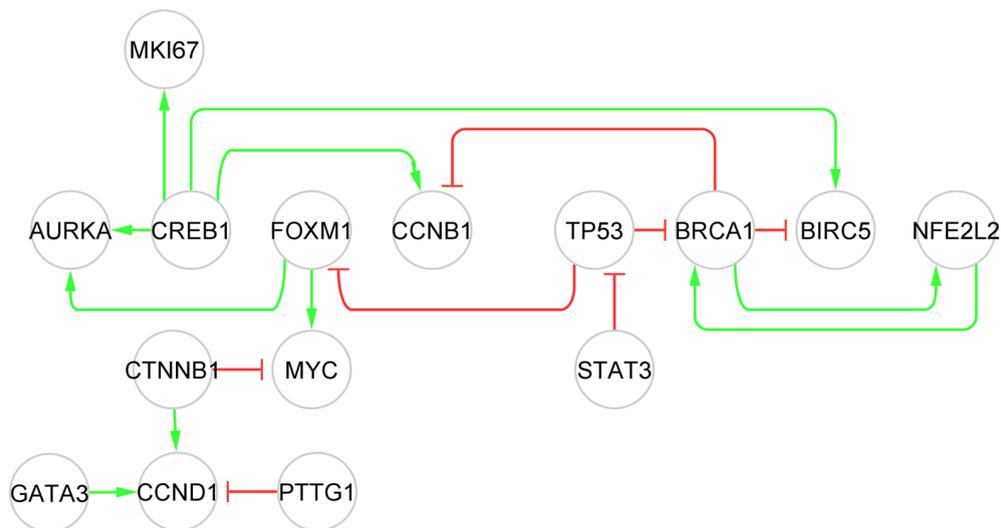


Figura 15. Red de regulación transcripcional reducida. En círculos se observan los genes que representan los nodos de la red, mientras que las conexiones de tipo activación se representan con flechas verdes (→) y las de inhibición con flechas truncadas (|) rojas.

Sin embargo, al analizar la red de regulación transcripcional reducida, apreciamos que la mayoría de los genes que conforman la red están asociados con procesos de proliferación y regulación del ciclo celular. Recordando, identificamos tres trayectorias o subpoblaciones, dentro de los datos de los esferoides de MCF7, que fueron caracterizadas funcionalmente como: azul) invasiva/inmune, verde) de proliferación y roja) un reservorio que compartía funciones de las otras subpoblaciones. Es decir, dentro de la red reducida había genes que predominantemente podían estar regulando solamente a la subpoblación de proliferación.

Una vez observado lo anterior, se realizaron ajustes a la red reducida para poder simular una red de regulación en la que se contemplaran reguladores para las tres subpoblaciones encontradas en la reconstrucción de trayectorias. Dicho lo anterior, por medio de búsqueda en la literatura, en bases de datos como NCBI, se añadieron algunos genes (ver Figura 16), bajo los siguientes criterios: relacionados con procesos invasivos, de respuesta del sistema inmune o proliferación en cáncer de mama luminal y que pudieran conectarse a la red, dándole prioridad a genes que resultaron DE en alguna trayectoria. Por ejemplo, *IL20*, un gen que salió diferencialmente expresado en la trayectoria azul (Figura 12), y cuya expresión se asocia principalmente con enfermedades inflamatorias como la artritis reumatoide y la aterosclerosis. Recientemente, varios estudios han demostrado que la IL-20 desempeña un papel crítico en la progresión tumoral, promoviendo la migración e invasión de células cancerosas de vejiga, mientras que en cáncer de mama se ha determinado que su expresión es activada por *GATA3* y *ESR1*, mientras que *FOXA1* la inhibe (Lee et al., 2017).

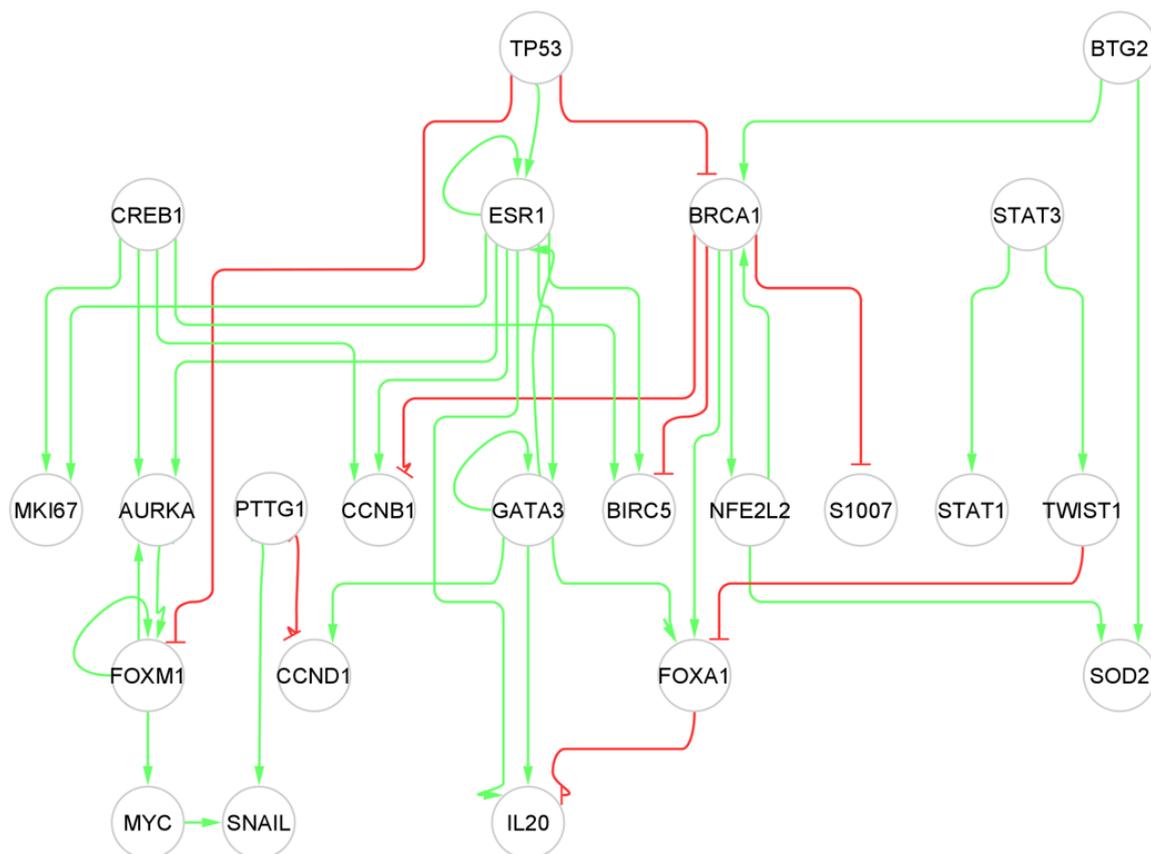


Figura 16. Red de regulación transcripcional propuesta. En círculos se observan los genes que representan los nodos de la red, mientras que las conexiones de tipo activación se representan con flechas verdes (→) y las de inhibición con flechas trancas (|) rojas.

Una vez propuesta la red de regulación, a través de los filtros y consideraciones anteriormente descritas, se prosiguió a traducir las conexiones entre los nodos en reglas lógicas también conocidas como compuertas Booleanas (ver Tabla 1) para poder llevar a cabo el modelado dinámico de la red y la identificación de sus atractores (ver métodos), los cuales tienen un valor biológico importante, al ser considerados fenotipos celulares.

La simulación Booleana de la red predice tres conjuntos de atractores

A partir de la simulación Booleana asincrónica de la red de regulación se obtuvieron 112 atractores, los cuales para poder visualizarlos y ver si se agrupaban de alguna manera se realizó UMAP (Uniform manifold approximation and projection) y clustering por medio de k-medias. Como se muestra en la Figura 17, se encontraron tres agrupaciones identificadas con un color diferente. Una vez obtenido estos grupos se procedió a caracterizar el fenotipo de cada uno de éstos. Con este objetivo, se analizaron los atractores por medio de los estados de encendido y apagado de los genes de la red y se identificaron patrones de expresión. Habiendo identificado estos patrones de expresión, se procedió a caracterizar el fenotipo biológico más cercano a cada uno de estos. Como ejemplo, para el cluster verde, los genes *CCNB1*, *BIRC5* y *FOXA1* están mayoritariamente encendidos para todos los atractores que conforman este cluster, los cuales sugieren estar asociados a un fenotipo de proliferación y control del ciclo celular, y cuya expresión resultó estar elevada para el caso de *CCNB1* y *BIRC5* en la subpoblación de la trayectoria proliferante (ver Figura 12), mientras que *STAT1*, *STAT3* y *TWIST1* se encuentran apagados. Asimismo, notamos que estos últimos genes permanecen mayoritariamente encendidos en los atractores del cluster azul y que se asocian con un estado de invasión y procesos inmunes. Además, los atractores de la agrupación en rojo mantienen un encendido de genes como: *AURKA*, *BRCA1*, *FOXA1*, *FOXM1*, *MYC*, *NFE2L2*, *SNAIL*, los cuales están involucrados en diferentes mecanismos relacionados con la proliferación, control del ciclo celular y la invasividad. Nuestro análisis sugiere que cada conjunto de atractores corresponde a una subpoblación identificadas experimentalmente, ya sea: invasiva, proliferante o reservorio.

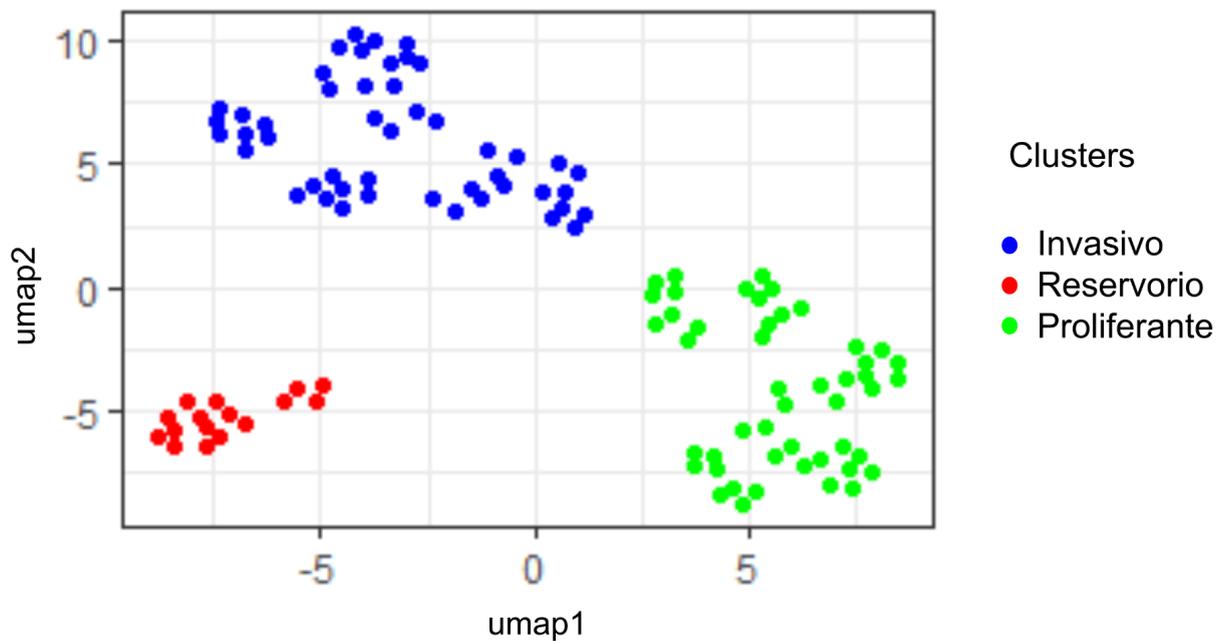


Figura 17. Clustering de los atractores encontrados a través del modelado Booleano de la red. Los clusters se identifican con cada color y se etiquetaron de acuerdo con los patrones de los genes. Cada punto representa un atractor que se obtuvo al realizar la dinámica de la red.

Modulación de la heterogeneidad a través de perturbaciones *in silico*

Con el fin de determinar la relevancia que la expresión de algunos genes tiene en la determinación de las subpoblaciones definidas y, por lo tanto, en la heterogeneidad celular que se observa en los MCTS, se prosiguió a realizar perturbaciones *in silico*. La pregunta central que nos motivó a este estudio fue explorar si el perturbar el nivel de expresión de algunos genes podría influir en la heterogeneidad poblacional que observamos experimentalmente. Al igual que en la biología molecular, las mutaciones *in silico* probadas en esta tesis se realizaron mediante noqueos (knockouts) o sobreexpresión de genes. Desde el punto de vista computacional, esto se logró imponiendo para cada estado un apagado o encendido de los genes con los valores binarios 0 ó 1, respectivamente. De las perturbaciones hechas para cada nodo presente en la red, se obtuvieron cambios en las subpoblaciones cuando se perturbaron los genes *ESR1*, *TP53*, *CCND1*, *FOXA1*, *BRCA1*, y *STAT3* (Figura 18), con respecto a la simulación sin perturbaciones (Figura 17), la cual se usó como control.

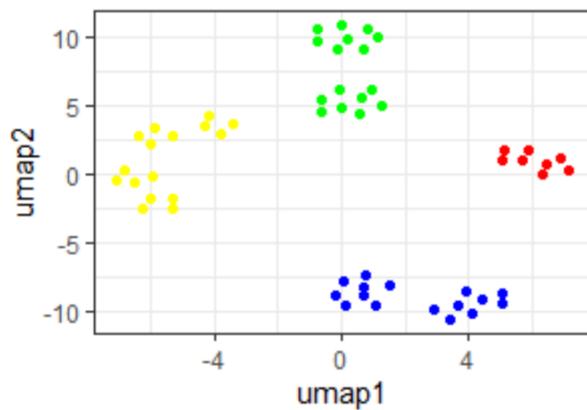
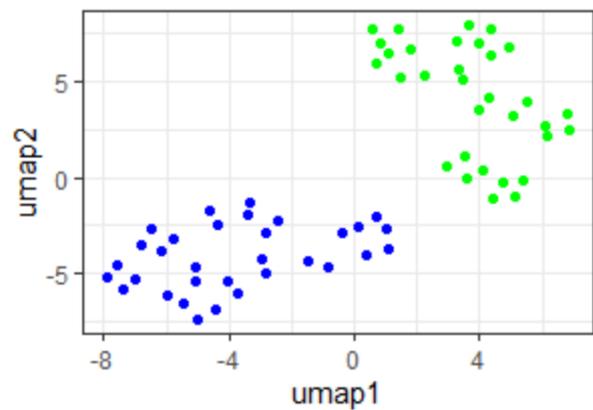
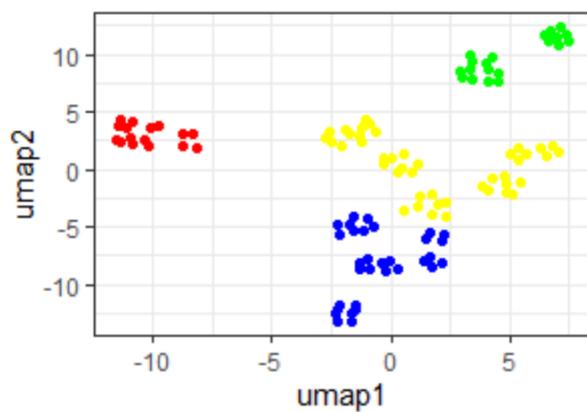
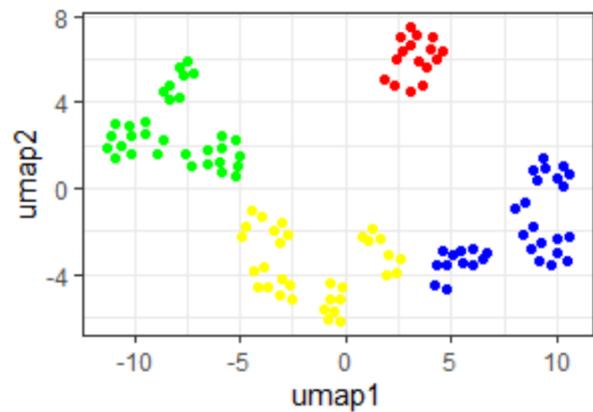
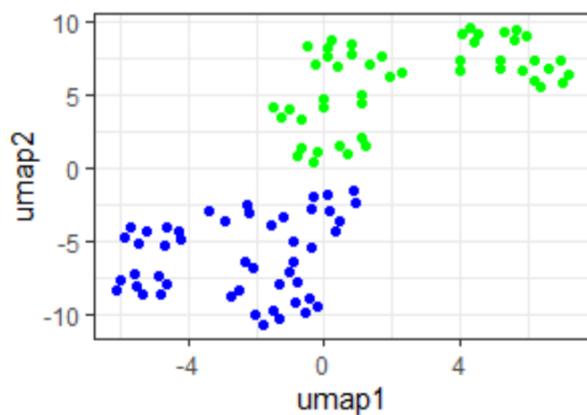
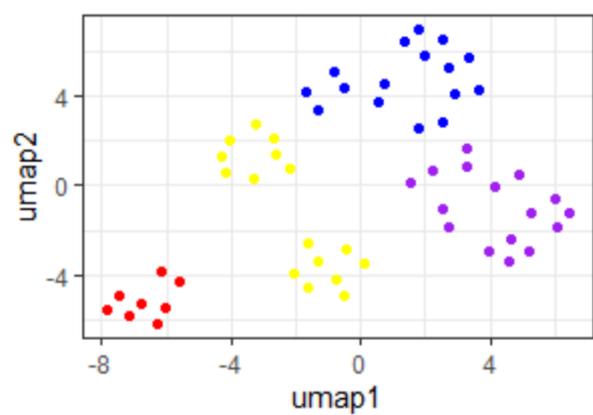
A) KO ESR1**B) OV TP53****C) OV CCND1****D) OV FOXA1****E) KO BRCA1****F) OV STAT3**

Figura 18. Perturbaciones *in silico* de la red de regulación. Las figuras muestran las proyecciones en el espacio UMAP (Uniform manifold approximation and projection) de las agrupaciones de los atractores después de perturbar la red, ya sea estableciendo un noqueo (KO) o sobreexpresión (OV) de los genes. Los colores representan las subpoblaciones: verde-proliferante, azul-invasiva, roja-reservorio, amarillas y moradas nuevas.

Los atractores encontrados, fueron nuevamente proyectados en un espacio de dos dimensiones con UMAP, se obtuvieron los clusters y se analizaron los estados de los genes de cada cluster (ver métodos). Como se muestra en la Figura 18, la sobreexpresión de *CCND1*, *FOXA1* y *STAT3*, y el noqueo de *ESR1*, generaron una subpoblación nueva representada por un color amarillo, mientras que al sobreexpresar *TP53* o noquear *BRCA1* se eliminó la subpoblación roja. Cabe destacar que las subpoblaciones amarillas para *CCND1* y *FOXA1* presentaron un patrón de encendido para genes asociados con la proliferación (*BIRC5*, *CCNB1*, *CCND1*, *ESR1* y *FOXA1*), mientras que el noqueo de *ESR1* hubo un encendido de *BIRC5*. Además, en la sobreexpresión de *STAT3*, se generaron dos subpoblaciones representadas en color morado y amarilla (Figura 18 panel F) con un patrón de encendido parecido al reservorio y otra al invasivo, respectivamente. De esta manera, se identificaron aquellos genes que al perturbarlos eliminaron alguna subpoblación o generaron alguna subpoblación novedosa. Las implicaciones biológicas de las perturbaciones se presentan en la sección siguiente.

Discusión

Computacionalmente se observa que por medio de dos metodologías diferentes se obtuvieron tres clusters (previamente identificadas en nuestro grupo de trabajo) que fueron equivalentes a las tres trayectorias reconstruidas, ya sea a través de las proporciones celulares (Tabla 3) o por el análisis funcional de los genes DE para cada trayectoria (Figura 13). Por medio de las proporciones celulares también fue evidente como hay trayectorias con mayores proporciones celulares de cada tiempo. Así la trayectoria verde tiene una mayor proporción de células del día 6 y una menor de células del día 19, cuyo comportamiento se invierte en la trayectoria celular azul, la cual tiene una mayor proporción de células del día 19 y una menor de células del día 6. Esto posiblemente muestra como las células pueden mostrar perfiles de expresión similares a pesar de provenir de tiempos diferentes. Cabe recalcar que las trayectorias sugieren la idea de que las subpoblaciones se transforman entre ellos y que posiblemente estén surgiendo de la rama verde, pero en realidad resulta complicado determinar la direccionalidad porque no tenemos grupos con muestras de un sólo día, sino una mezcla de distintas subpoblaciones que prevalecen en los dos tiempos. Es decir, no se observa una transición totalmente dependiente del tiempo. Quizás estos cambios en proporciones sean una respuesta a los cambios del microambiente en el interior del esferoide y a las necesidades funcionales. Interesantemente, se puede pensar que existe una memoria en términos de funcionalidad que fuertemente influye en el desarrollo tumoral, donde a pesar de que hubo una evolución temporal, existen células que presentan perfiles de condiciones pasadas, por ejemplo; acotándonos a las células de la trayectoria verde (Figura 10), las del día 19 se están expresando como las del día 6, ya sea quizás por cuestiones funcionales o de adaptación tumoral. Podría ser que este comportamiento celular se justifique desde un enfoque de plasticidad celular que se ha observado en cáncer, donde las células tumorales se desdiferencian hacia sus progenitoras o bien existe una diferenciación bloqueada para congelar las células en desarrollo en estados progenitores (Hanahan, 2022).

Con la premisa de que estas subpoblaciones eran diferentes entre sí, al analizar los genes que se estaban expresando en cada trayectoria celular o subpoblación, se pudo identificar cuáles eran sus papeles a nivel funcional, por medio de los procesos biológicos sobre-representados. Para la trayectoria azul se observa la vía de señalización mediada por citocinas, la quimiotaxis de granulocitos como neutrófilos, agregación de leucocitos, así como una respuesta de interferón tipo 1 y gamma (Figura 13 panel A). En cáncer estos procesos del sistema inmune se relacionan con transformaciones tumorales, la

supervivencia, la invasión y la metástasis (Lee et al., 2014). Como ejemplo, una sobreexpresión de IL-6 en células MCF7 produjo la transformación hacia un fenotipo mesenquimal (Sullivan et al., 2009). Así mismo, el IFN γ (interferón gamma) puede facilitar el inicio del tumor y, posteriormente, promover cambios en el fenotipo de células tumorales a través de la promoción de microambientes tumorales inmunosupresores (Mojic et al., 2017). Por otro lado, el gen *S100A7*, altamente expresado en la trayectoria azul (Figura 11) aumenta el crecimiento tumoral de mama y la metástasis mediante la activación de vías proinflamatorias y metastásicas (Nasser et al., 2012). Con toda información se evidencia que las células de la trayectoria azul se relacionan con procesos invasivos y evasión del sistema inmune. Para la trayectoria verde (Figura 13 panel B), el enriquecimiento que se muestra en actividades del ciclo celular como la transición G1/S y hacia mitosis, segregación de cromátidas en mitosis, así como procesos del DNA, tal como replicación, metabolismo y reparación del DNA. En las células tumorales se ha descrito que el punto de chequeo al final de la fase S está desregulado, y que las células alcanzan a pasarlo, aunque su material genético no esté totalmente duplicado o haya errores contribuyendo a la inestabilidad genética (Boyer et al., 2016). En cáncer está ampliamente descrito que la replicación del DNA se asocia a células proliferantes (Mazurek et al., 2012). Además, obtuvimos genes de divergencia en esta trayectoria (Figura 12) tales como *BIRC5* y *CCNB1* involucrados en el control de la proliferación celular. Para el caso de *BIRC5* se ha reportado que su expresión es dependiente del ciclo celular y un aumento de la misma se ha visto reflejada en células de diversos tipos de cáncer incluido el cáncer de mama (Zhang et al., 2004). Así, lo anterior nos está sugiriendo una subpoblación que está llevando a cabo constantes divisiones celulares y que es, por tanto, proliferante. Finalmente, para la trayectoria roja se obtuvieron enriquecimientos funcionales en procesos tanto del sistema inmune como en la replicación del DNA (Figura 13 panel C). Dicho análisis, manifiesta que en términos de funcionalidad las células de la trayectoria azul, verde y roja tienen un fenotipo invasivo/evasión sistema inmune, proliferante y compartido, respectivamente. Además, dichos procesos biológicos sobre-representados para cada conjunto de genes de cada una de las subpoblaciones de las trayectorias concuerda de manera general con los clusters previamente reportados por nuestro grupo de trabajo (Muciño-Olmos, Vázquez-Jiménez, Avila-Ponce de León, et al., 2020).

Ya que los datos transcriptómicos brindan información de lo que se está expresando, se decidió reconstruir una red de regulación transcripcional que pudiera describir el comportamiento del sistema que, en este caso, se conforma de tres subpoblaciones celulares. Las herramientas bioinformáticas han sido de gran ayuda para el científico, ya

que brindan la capacidad de trabajar con un gran tamaño de información, así como de emplear modelos que serían complicados de probar sin una computadora. Los modelos Booleanos han sido útiles para predecir comportamientos de redes. En nuestro caso, fue posible obtener la dinámica de la red de regulación para los esferoides tumorales de la línea celular de cáncer de mama MCF7. De esta manera, obtuvimos tres conjuntos de atractores, los cuales restablecieron los fenotipos de las tres trayectorias celulares, siendo uno de los hallazgos más sorprendente. Cabe recalcar que este modelo tiene un gran potencial en redes de regulación para integrar trabajos experimentales y comprender de mejor manera sistemas biológicos con mecanismos de regulación complejos, como en cáncer. Sin embargo, ya que la red es basada en conocimiento que constantemente se está desarrollando, así como la necesidad de una validación experimental, las redes están abiertas a ser mejoradas y de esta manera favorecer el entendimiento del comportamiento del proceso biológico en estudio.

Aunado a esto, fue posible modular las subpoblaciones tumorales de manera *in silico* por medio de perturbaciones de tipo noqueo o sobreexpresión. A continuación, discutimos los resultados obtenidos y sus implicaciones en biología del cáncer. El análisis vertido en este trabajo fue desarrollado alrededor de los genes *ESR1*, *TP53*, *CCND1*, *FOXA1*, *BRCA1*, y *STAT3*, todos estos con una evidencia experimental alrededor de su papel en el fenotipo canceroso.

- ***ESR1***

La terapia hormonal es una de las principales terapias en cáncer de mama de tipo luminal (receptores positivos a estrógenos y progesterona), siendo uno de los principales medicamentos el Tamoxifeno. Este medicamento actúa como un agonista parcial por los receptores de estrógenos, como consecuencia la activación transcripcional de genes que se relacionan con el crecimiento y proliferación tumoral disminuyen (Farrar & Jacobs, 2021). Dado lo anterior, en primera instancia se podría deducir que al inhibir totalmente al gen *ESR1*, el cual codifica al receptor de estrógenos ER α , ayudaría a disminuir el crecimiento y proliferación tumoral; sin embargo, lo que se observó de nuestro modelo es que la heterogeneidad aumenta, se generó una subpoblación más (Figura 18 panel A). No obstante, al analizar las agrupaciones se identificó un fenómeno antiproliferativo, donde por un lado el número de atractores disminuyó en cada agrupación comparada con la proyección sin perturbación y, además, la nueva agrupación parece tener características menos dañinas al estar expresando menor cantidad de genes asociados a proliferación. Estas observaciones posiblemente indican que esta inhibición genética está provocando

una muerte tumoral. En consistencia con este resultado, un estudio en cáncer de mama reportó que ER α reprimió la respuesta apoptótica que lleva a cabo P53 (Bailey et al., 2012), por lo que al eliminar la expresión del gen *ESR1* probablemente los mecanismos apoptóticos mediados por P53 estén contribuyendo con este fenómeno.

- ***TP53***

El supresor tumoral P53 está involucrado en distintas funciones celulares como la apoptosis y control del ciclo celular (Silwal-Pandit et al., 2017). En el caso de *TP53* se han descrito diferentes frecuencias mutacionales que dependen del tipo de cáncer. En el cáncer de mama, las mutaciones de *TP53* están presentes en el 25%-30% de todos los casos, y estas cifras a su vez dependen del subtipo, siendo el 17% del tipo luminal A; 41% del luminal B; un 50% en los tumores HER2 y un 88% en carcinomas tipo basal (Bertheau et al., 2013), es decir, para el caso luminal A, la frecuencia es la más baja. Asimismo, se ha reportado una expresión wild-type de la proteína P53 en las células de cáncer de mama tipo luminal, cuyo aumento de su expresión se relaciona con una actividad antiproliferativa en células MCF7 (Andreani et al., 2010). Consistente con este reporte, en nuestro estudio, al establecer una sobreexpresión del gen *TP53*, se observó una disminución de la heterogeneidad, en la que se eliminó la agrupación del reservorio (grupo rojo) y disminuyó el número de atractores para la población proliferante y la invasiva (ver Figura 18 panel B). Es notorio como se puede observar que esta perturbación repercute no sólo en el grupo proliferante sino también en el invasivo y reservorio, por lo que existe una importancia de este supresor en la regulación de la heterogeneidad intratumoral mostrada a través de este modelo.

- ***CCND1* y *FOXA1***

Las sobreexpresiones de *CCND1* y *FOXA1* mostraron resultados parecidos. El gen *CCND1*, el cual codifica para la proteína Ciclina D, es un oncogén implicado en la proliferación celular a través del control del ciclo celular durante la transición de la fase G1-S. La sobreexpresión de *CCND1* está relacionada con varios tipos de cáncer, incluyendo cáncer de mama (Elsheikh et al., 2008; Valla et al., 2022). La sobreexpresión en nuestro modelo mostró la generación de otra agrupación de atractores u otra subpoblación (grupo amarillo), indicando que este gen aumentó la heterogeneidad poblacional (ver Figura 18 panel C). Asimismo, esta nueva subpoblación tiene características más proliferantes al tener encendido más genes relacionados con este fenotipo. Interesantemente, este mismo aumento de encendido de genes involucrados en proliferación se observó en la subpoblación proliferante del inicio (grupo verde). Un comportamiento similar encontramos

para el caso de *FOXA1*, cuya sobreexpresión origina una subpoblación más (Figura 18 panel D) que también expresó genes relacionados con el fenotipo proliferante. Recientemente, se reportó que la sobreexpresión de este gen se relaciona con una resistencia terapéutica en cáncer de mama ER positivo, además, al reducir su expresión se observó una disminución de células tumorales en la formación de un modelo de mamoesferas (Kumar et al., 2021).

- ***BRCA1***

BRCA1 es un supresor tumoral que está involucrado en diferentes mecanismos como la reparación del DNA dañado. Para el caso de *BRCA1*, se observó que cuando se abatía la expresión de este gen, el conjunto de atractores descrito como reservorio se eliminaba, mientras que el conjunto de atractores invasivo y proliferante permanecían (Figura 18 panel E). Lo que planteamos es que los atractores del reservorio se transformaron ya sea a un fenotipo invasivo o a uno proliferante. Recientemente, se ha empezado a describir una plasticidad celular en cáncer, la cual se asocia con la adquisición de diferentes firmas transcripcionales y, consecuentemente, otros fenotipos. Por ejemplo, se ha demostrado que las células cancerosas senescentes, las cuales entran en un estado irreversible, son capaces de salir de ese estado y contribuir con la señalización proliferativa, así como de evitar la apoptosis, inducir la angiogénesis, estimular la invasión y la metástasis, y suprimir la inmunidad tumoral (De Blander et al., 2021; Hanahan, 2022).

- ***STAT3***

STAT3 juega un papel importante en la inflamación. En cáncer se ha visto implicado en la inflamación inicial del tumor, así como en regulación de la expresión de una serie de genes que están involucrados en la agresividad del cáncer a través de la invasividad y metástasis, incluido el cáncer mama (Lo et al., 2007). La sobreexpresión de *STAT3* generó la formación de dos nuevas subpoblaciones (morada y amarilla), como se muestra en la Figura 18 panel F. La subpoblación en morado mostró un patrón de encendido de genes parecida a la subpoblación invasiva inicial, mientras que la subpoblación en amarillo refleja patrones más parecidos a la subpoblación de reservorio. Por lo que estos resultados indican un aumento en la heterogeneidad que se vincula con un fenotipo con capacidades menos proliferativas, pero con mayor afinidad a realizar metástasis.

Si bien un aumento de la heterogeneidad intratumoral se ha asociado con un peor pronóstico de la enfermedad del cáncer, consideramos que es necesario contemplar el panorama completo del nicho tumoral. En nuestro análisis se muestra que el noqueo de

ESR1 genera una subpoblación más; sin embargo, la perturbación provoca una muerte tumoral de acuerdo con nuestras predicciones. Por otro lado, las perturbaciones de *TP53* y *BRCA1* provocan que dos subpoblaciones se conserven, mientras que para las perturbaciones de *ESR1*, *CCND1*, *FOXA1* y *STAT3* surgen subpoblaciones nuevas. Dicho comportamiento puede explicarse por medio de la reconstrucción de la red y las reglas Booleanas planteadas; no obstante, es importante mencionar que el sistema es complejo y los mecanismos que lo regulan no son intuitivos, por lo que se requerirán de estudios posteriores que profundicen en esto. Cabe destacar, que se obtuvieron perturbaciones en estos nodos con una inherencia en la dinámica de la red, provocando cambios en las subpoblaciones, por lo que la aplicación de modelos, como el Booleano utilizado en este proyecto, son de gran utilidad para generar hipótesis y enfocar el estudio hacia lo más significativo.

Finalmente, entre las limitaciones de este proyecto están las siguientes: 1) se podría pensar que en las trayectorias celulares se tiene una dinámica temporal fiel; sin embargo, lo que observamos es un orden celular, de ahí el término pseudotiempo, 2) el modelo Booleano contempla estados binarios 1 ó 0; empero, biológicamente lo que se realiza con más frecuencia son abatimientos en la expresión sin llegar a un noqueo. Además, no se contemplan mecanismos como el procesamiento del RNAm o el post-procesamiento de la proteína, se asume que los genes se expresan o inhiben totalmente, 3) el estudio del microambiente tumoral ha mostrado cada vez más implicaciones en el desarrollo del cáncer. Aunque en este trabajo no se tomaron en cuenta variables como la interacción con otras células, por ejemplo, células del sistema inmune, con este modelo es posible agregar elementos de interés que permitan darnos información del nicho tumoral a través de una simulación más parecida a lo que ocurre de manera *in vivo*.

Conclusión

La dinámica de la red de regulación propuesta para los esferoides caracteriza a tres conjuntos de atractores cuyos estados de encendido/apagado se vinculan con los fenotipos: proliferante, invasivo y reservorio.

En este trabajo hemos reconstruido por primera vez una red de regulación transcripcional que es capaz de reproducir los fenotipos observados en un modelo de MCF7 de cáncer de mama. Al hacer perturbaciones *in silico* sobre esta red de regulación fue posible identificar genes que potencialmente pueden alterar la heterogeneidad intratumoral de los MCTS de MCF7. De forma concreta se encontró que:

- Al noquear la expresión de *BRCA1* o sobreexpresar *TP53* se observó una disminución de la heterogeneidad.
- Al noquear la expresión de *ESR1* o sobreexpresar la expresión de *CCND1*, *FOXA1* o *STAT3* se pudo observar un aumento de la heterogeneidad.

Los resultados en este trabajo nos permiten proponer y sustentar que la formación de la red de regulación descrita en este estudio es una aproximación poderosa y simple para describir la formación de las subpoblaciones evidenciadas experimentalmente con datos de scRNA-seq. La estrategia de investigación utilizada aquí constituye un andamiaje computacional capaz de diseñar posibles estrategias experimentales para explorar la heterogeneidad en líneas celulares de cáncer y descubrir los elementos de control mediante la inducción o represión de algunos genes que la integran. Finalmente, consideramos que esta red es una aportación original que plantea, bajo una plataforma *in silico*, construir hipótesis y diseñar experimentos alrededor del problema de heterogeneidad del cáncer en la línea celular de MCF7.

Perspectivas

La heterogeneidad intratumoral es claramente una de las líneas de trabajo que debemos de seguir estudiando para poder entender el cáncer. En este trabajo nosotros sugerimos un modelo computacional para poder analizar la relación entre los perfiles de transcripción y los fenotipos asociados a la heterogeneidad en un sistema *in vitro* de la línea celular MCF7 cultivadas en MCTS. El pipeline que se construyó en esta tesis será de gran utilidad para esclarecer patrones de regulación de la heterogeneidad que permitan proponer un tratamiento más preciso, y guiar posibles estrategias en tratamientos personalizados. De manera relevante, estas perturbaciones servirán para predecir respuestas y establecer una terapia blanco o una combinación de éstas. Para esto último, se podrían simular mutaciones simultáneas para determinar el efecto de las perturbaciones en paralelo en el fenotipo de las subpoblaciones intratumorales. Además, la importancia de esta aproximación *in silico* tiene la ventaja de postular hipótesis que pueden ser verificadas experimentalmente. Así, de acuerdo con las inferencias de la modulación de la heterogeneidad por medio del modelo Booleano, la validación experimental *in vitro* de los mecanismos regulatorios que se plantean en este proyecto serán útiles. Para demostrar las hipótesis, las sobreexpresiones o abatimientos de la expresión de los genes se podrían llevar a cabo experimentalmente con métodos con CRISPR para la edición de genes, RNAs de interferencia o la construcción de vectores. Posteriormente, verificar los efectos que estos ensayos puedan tener en las células.

Aunado a las perturbaciones *in silico* realizadas en este trabajo, llama la atención el comportamiento de dos genes *BTG2* y *CREB1*. Por un lado, *BTG2* tiene un papel antiproliferante. Cuando se realizó el abatimiento de su expresión, se esperaba que las subpoblaciones aumentaran; sin embargo, lo que se observó fue que la subpoblación reservorio se eliminó, mientras que se redujeron los atractores en la invasiva y proliferante. Por otra parte, *CREB1* se ha visto implicado en la expresión positiva de distintos genes involucrados en el crecimiento de células de cáncer de mama. Cuando se abatió su expresión se generó otra subpoblación que se asemejó a la proliferante y que parece está tendiendo a tener también un fenotipo invasivo. Además, cuando se realizó la sobreexpresión de *CREB1*, las subpoblaciones de reservorio y proliferante se erradicaron, quedando en el sistema la invasiva y surgiendo una nueva. Estos resultados contraintuitivos apuntan a que el sistema no se comporta de manera lineal, sino que es un sistema complejo. Debido a esto se requiere seguir explorando las reglas booleanas planteadas en este proyecto, o bien de otros mecanismos de regulación que no se incluyeron en la

reconstrucción de la red de regulación. Ambos aspectos quedan como perspectivas de este trabajo.

En términos de otros mecanismos de regulación que pudieran considerarse para extender la red propuesta, la regulación transcripcional se ha extendido a otros estudios enfocados en la epigenética. Desde el punto de vista molecular, los cambios en la apertura o cierre de la cromatina, que se traducen con una expresión o represión de los genes, ya sea por modificaciones postraduccionales de las histonas, la metilación del DNA, RNAs no codificantes, remodeladores de cromatina, entre otros, han contribuido al entendimiento de la regulación transcripcional. Esta regulación además de asociarse con la variabilidad celular, recientemente se ha vinculado con la heterogeneidad celular intratumoral, donde los patrones de expresión dependen del paisaje epigenético celular, el cual tiende a ser dinámico. Lo anterior pone de manifiesto que los reguladores epigenéticos cada vez toman más fuerza y alcance en el entendimiento de la heterogeneidad en cáncer, por lo que resultaría interesante enfocarse y elucidar mecanismos que contribuyan a la generación de nuevos blancos terapéuticos epigenéticos usados para disminuir la heterogeneidad intratumoral.

Finalmente, ya que el modelo Booleano contempla estados discretos de activación o inhibición de la expresión, queda la perspectiva de extender este modelo a un sistema continuo. Con esta implementación se podrían tomar en cuenta distintos niveles de expresión de un gen, es decir se ahondaría en una dinámica cuantitativa. No obstante, aunque se contemplaría en detalle los mecanismos moleculares, tales como velocidades cinéticas precisas para cada reacción bioquímica, es importante mencionar que para muchos sistemas biológicos no siempre se cuenta con toda la información. La construcción de los modelos continuos, y sus implicaciones biológicas en el cáncer de mama queda como perspectiva de este trabajo de tesis.

Material suplementario

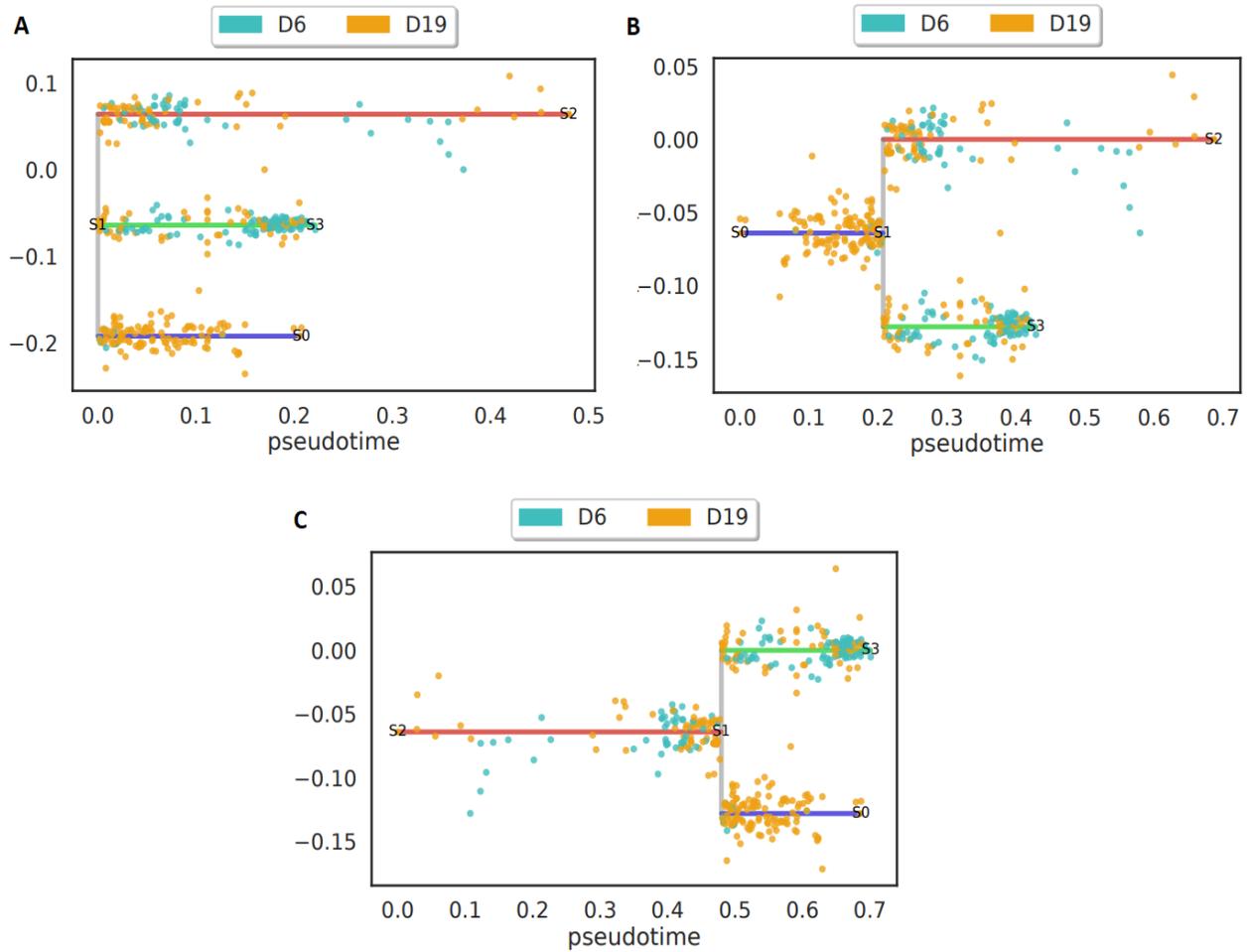
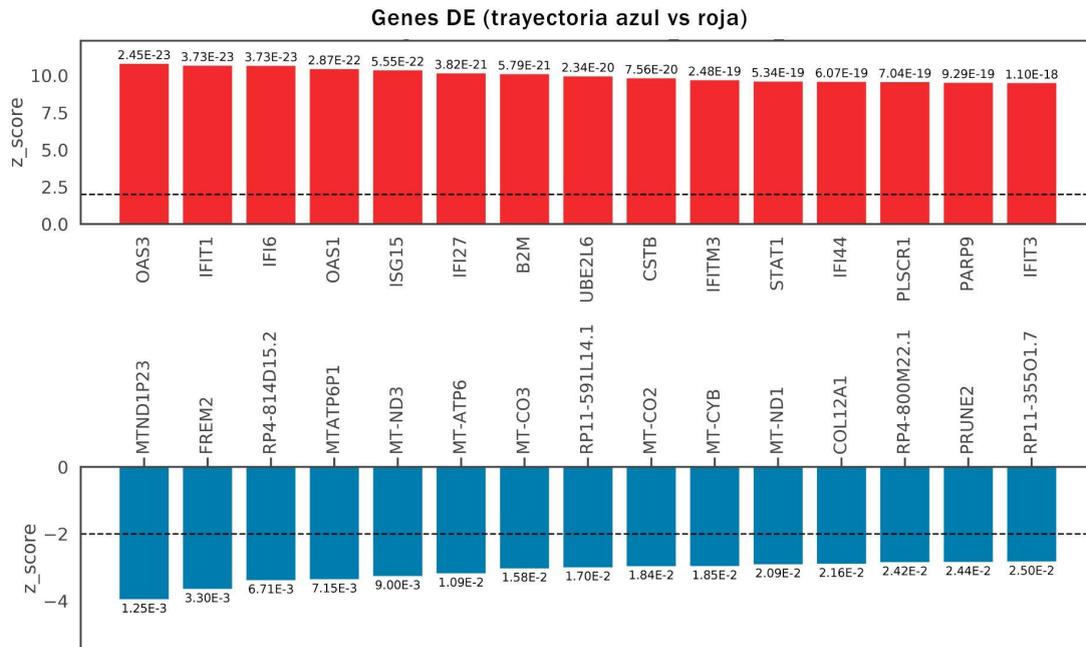


Figura S1. Reconstrucción de trayectorias variando el punto de inicio. Para el panel A las trayectorias inician desde la raíz S1, en el panel B las trayectorias inician desde la raíz S0 y en el panel C desde la raíz S2. Cada punto representa una célula, en azul las células del día 6 (D6) y en naranja las células del día 19 (D19).

A)



B)

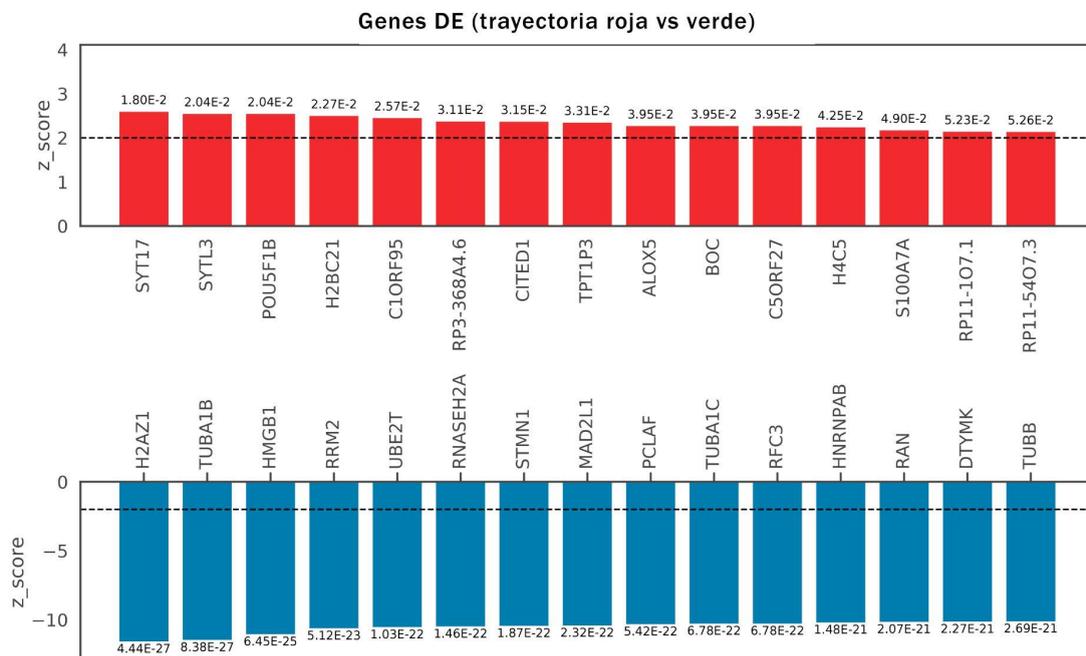


Figura S2. Top 15 de genes de divergencia obtenidos con STREAM. En el panel A se observan los genes diferencialmente expresados (DE) entre la trayectoria azul vs la trayectoria roja. En el panel B se muestran los genes DE entre la trayectoria roja vs la trayectoria verde. Los genes expresados al alta se denotan en color rojo y los genes expresados a la baja en color azul. En este caso sólo se muestra el top 15. Los genes con un z-score > 2 y un p-value < 0.01 se consideraron genes diferencialmente expresados.

Tabla S1. Archivo txt utilizado en BoolNet donde se muestran las reglas lógicas utilizadas para la modelación.

```
targets, factors
AURKA, (CREB1 & ESR1) | (FOXM1)
BIRC5, (CREB1 & ESR1) | (! BRCA1)
BRCA1, (BTG2 & NFE2L2 & ! TP53)
BTG2, BTG2
CCNB1, (CREB1 & ESR1) | (! BRCA1)
CCND1, (GATA3 | ! PTTG1)
CREB1, CREB1
ESR1, (ESR1 & GATA3) | (TP53)
FOXA1, any(GATA3, BRCA1, ! TWIST1)
FOXM1, (FOXM1 & AURKA) | (! TP53)
GATA3, (GATA3 & ESR1)
IL20, (GATA3 & ESR1 & ! FOXA1)
MKI67, (CREB1 & ESR1)
MYC, FOXM1
NFE2L2, BRCA1
PTTG1, PTTG1
STAT3, STAT3
STAT1, STAT3
SNAIL, (PTTG1 | MYC)
SOD2, (BTG2 & NFE2L2)
S1007, (! BRCA1)
TP53, TP53
TWIST1, STAT3
```

Referencias

- Alcaide-Lucena, M., Rodríguez-González, C. J., de Reyes-Lartategui, S., Gallart-Aragón, R., Sánchez-Barrón, M. T., García-Rubio, J., & Torné-Poyatos, P. (2021). Molecular classification of breast cancer. Treatment and prognosis implications. *Cirugía Andaluza*, 32(2), 155–159.
- Aldana, M., Balleza, E., Kauffman, S., & Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, 245(3), 433–448.
- Álvarez-Hernández, C., Vich Pérez, P., Brusint, B., Cuadrado Rouco, C., Díaz García, N., & Robles Díaz, L. (2014). [Update of breast cancer in Primary Care (III/IV)]. *Semergen / Sociedad Española de Medicina Rural y Generalista*, 40(8), 460–472.
- Andreani, A., Bellini, S., Burnelli, S., Granaiola, M., Leoni, A., Locatelli, A., Morigi, R., Rambaldi, M., Varoli, L., Calonghi, N., Cappadone, C., Zini, M., Stefanelli, C., Masotti, L., & Shoemaker, R. H. (2010). Substituted E-3-(3-indolylmethylene)-1,3-dihydroindol-2-ones with antitumor activity. In depth study of the effect on growth of breast cancer cells. *Journal of Medicinal Chemistry*, 53(15), 5567–5575.
- Arizti, P., Fang, L., Park, I., Yin, Y., Solomon, E., Ouchi, T., Aaronson, S. A., & Lee, S. W. (2000). Tumor suppressor p53 is required to modulate BRCA1 expression. *Molecular and Cellular Biology*, 20(20), 7450–7459.
- Arruabarrena-Aristorena, A., & Toska, E. (2022). Epigenetic Mechanisms Influencing Therapeutic Response in Breast Cancer. *Frontiers in Oncology*, 12, 924808.
- Bailey, S. T., Shin, H., Westerling, T., Liu, X. S., & Brown, M. (2012). Estrogen receptor prevents p53-dependent apoptosis in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), 18060–18065.
- Banzhaf, M., Resendis-Antonio, O., & Zepeda-Mendoza, M. L. (2020). Uncovering the dynamic mechanisms of the *Pseudomonas aeruginosa* quorum sensing and virulence networks using Boolean modelling. *IEEE Transactions on Nanobioscience*.
<https://doi.org/10.1109/TNB.2020.2977820>
- Barsotti, A. M., & Prives, C. (2009). Pro-proliferative FoxM1 is a target of p53-mediated repression. *Oncogene*, 28(48), 4295–4305.
- Basu, S., Kwee, T. C., Gatenby, R., Saboury, B., Torigian, D. A., & Alavi, A. (2011). Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(6), 987–991.

- Bertheau, P., Lehmann-Che, J., Varna, M., Dumay, A., Poirot, B., Porcher, R., Turpin, E., Plassa, L.-F., de Roquancourt, A., Bournstyn, E., de Cremoux, P., Janin, A., Giacchetti, S., Espié, M., & de Thé, H. (2013). p53 in breast cancer subtypes and new insights into response to chemotherapy. *Breast*, *22 Suppl 2*, S27–S29.
- Birgersdotter, A., Sandberg, R., & Ernberg, I. (2005). Gene expression perturbation in vitro—a growing case for three-dimensional (3D) culture systems. *Seminars in Cancer Biology*, *15*(5), 405–412.
- Boyer, A.-S., Walter, D., & Sørensen, C. S. (2016). DNA replication and cancer: From dysfunctional replication origin activities to therapeutic opportunities. *Seminars in Cancer Biology*, *37-38*, 16–25.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424.
- Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. In *Nature* (Vol. 501, Issue 7467, pp. 338–345). <https://doi.org/10.1038/nature12625>
- Candanedo-González, F., & Pérez-Salazar, E. (2012). El papel de la progesterona en cáncer de mama. *Gaceta Mexicana de Oncología* *11* (Mayo): 6.
- Carranza-Rosales, P., Guzmán-Delgado, N. E., Carranza-Torres, I. E., Viveros-Valdez, E., & Morán-Martínez, J. (2018). Breast Organotypic Cancer Models. *Current Topics in Microbiology and Immunology*. https://doi.org/10.1007/82_2018_86
- Chabon, J. J., Simmons, A. D., Lovejoy, A. F., Esfahani, M. S., Newman, A. M., Haringsma, H. J., Kurtz, D. M., Stehr, H., Scherer, F., Karlovich, C. A., Harding, T. C., Durkin, K. A., Otterson, G. A., Purcell, W. T., Camidge, D. R., Goldman, J. W., Sequist, L. V., Piotrowska, Z., Wakelee, H. A., ... Diehn, M. (2016). Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nature Communications*, *7*, 11815.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: AnRPackage for Determining the Relevant Number of Clusters in a Data Set. In *Journal of Statistical Software* (Vol. 61, Issue 6). <https://doi.org/10.18637/jss.v061.i06>
- Cheng, G. Z., Zhang, W. Z., Sun, M., Wang, Q., Coppola, D., Mansour, M., Xu, L. M., Costanzo, C., Cheng, J. Q., & Wang, L.-H. (2008). Twist is transcriptionally induced by activation of STAT3 and mediates STAT3 oncogenic function. *The Journal of Biological Chemistry*, *283*(21), 14665–14673.
- Chen, H., Albergante, L., Hsu, J. Y., Lareau, C. A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A. N., Bauer, D. E., Aryee, M. J., Langenau, D. M., Zinovyev, A., Buenrostro, J.

- D., Yuan, G.-C., & Pinello, L. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nature Communications*, *10*(1), 1903.
- Chiang, S., Weigelt, B., Wen, H.-C., Pareja, F., Raghavendra, A., Martelotto, L. G., Burke, K. A., Basili, T., Li, A., Geyer, F. C., Piscuoglio, S., Ng, C. K. Y., Jungbluth, A. A., Balss, J., Pusch, S., Baker, G. M., Cole, K. S., von Deimling, A., Batten, J. M., ... Schnitt, S. J. (2016). IDH2 Mutations Define a Unique Subtype of Breast Cancer with Altered Nuclear Polarity. *Cancer Research*, *76*(24), 7118–7129.
- Cho, M.-H., Park, J.-H., Choi, H.-J., Park, M.-K., Won, H.-Y., Park, Y.-J., Lee, C. H., Oh, S.-H., Song, Y.-S., Kim, H. S., Oh, Y.-H., Lee, J.-Y., & Kong, G. (2015). DOT1L cooperates with the c-Myc-p300 complex to epigenetically derepress CDH1 transcription factors in breast cancer progression. *Nature Communications*, *6*, 7821.
- De Blander, H., Morel, A.-P., Senaratne, A. P., Ouzounova, M., & Puisieux, A. (2021). Cellular Plasticity: A Route to Senescence Exit and Tumorigenesis. *Cancers*, *13*(18). <https://doi.org/10.3390/cancers13184561>
- Dufau, I., Frongia, C., Sicard, F., Dedieu, L., Cordelier, P., Ausseil, F., Ducommun, B., & Valette, A. (2012). Multicellular tumor spheroid model to evaluate spatio-temporal dynamics effect of chemotherapeutics: application to the gemcitabine/CHK1 inhibitor combination in pancreatic cancer. *BMC Cancer*, *12*, 15.
- Eeckhoute, J., Keeton, E. K., Lupien, M., Krum, S. A., Carroll, J. S., & Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Research*, *67*(13), 6477–6483.
- Elsheikh, S., Green, A. R., Aleskandarany, M. A., Grainge, M., Paish, C. E., Lambros, M. B. K., Reis-Filho, J. S., & Ellis, I. O. (2008). CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. *Breast Cancer Research and Treatment*, *109*(2), 325–335.
- Farrar, M. C., & Jacobs, T. F. (2021). Tamoxifen. In *StatPearls*. StatPearls Publishing.
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer. Journal International Du Cancer*. <https://doi.org/10.1002/ijc.33588>
- Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., ... Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine*, *366*(10), 883–892.
- Ginsburg, O., Bray, F., Coleman, M. P., Vanderpuye, V., Eniu, A., Rani Kotha, S., Sarker, M., Huong, T. T., Allemani, C., Dvaladze, A., Gralow, J., Yeates, K., Taylor, C., Oomman, N.,

- Krishnan, S., Sullivan, R., Kombe, D., Blas, M. M., Parham, G., ... Conteh, L. (2017). The global burden of women's cancers: a grand challenge in global health. In *The Lancet* (Vol. 389, Issue 10071, pp. 847–860).
[https://doi.org/10.1016/s0140-6736\(16\)31392-7](https://doi.org/10.1016/s0140-6736(16)31392-7)
- Glaves, P. D., & Tugwood, J. D. (2011). Generation and analysis of transcriptomics data. *Methods in Molecular Biology*, 691, 167–185.
- Gong, C., Fujino, K., Monteiro, L. J., Gomes, A. R., Drost, R., Davidson-Smith, H., Takeda, S., Khoo, U. S., Jonkers, J., Sproul, D., & Lam, E. W.-F. (2015). FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer. *Oncogene*, 34(39), 5012–5024.
- Gorrini, C., Baniyasadi, P. S., Harris, I. S., Silvester, J., Inoue, S., Snow, B., Joshi, P. A., Wakeham, A., Molyneux, S. D., Martin, B., Bouwman, P., Cescon, D. W., Elia, A. J., Winterton-Perks, Z., Cruickshank, J., Brenner, D., Tseng, A., Musgrave, M., Berman, H. K., ... Gauthier, M. L. (2013). BRCA1 interacts with Nrf2 to regulate antioxidant signaling and cell survival. *The Journal of Experimental Medicine*, 210(8), 1529–1544.
- Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Molecular Systems Biology*, 14(4), e8046.
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), 31–46.
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646–674.
- Han, W., Carpenter, R. L., Cao, X., & Lo, H.-W. (2013). STAT1 gene expression is enhanced by nuclear EGFR and HER2 via cooperation with STAT3. *Molecular Carcinogenesis*, 52(12), 959–969.
- Hatcher, R. J., Dong, J., Liu, S., Bian, G., Contreras, A., Wang, T., Hilsenbeck, S. G., Li, Y., & Zhang, P. (2014). Pttg1/securin is required for the branching morphogenesis of the mammary gland and suppresses mammary tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), 1008–1013.
- Jha, K., Shukla, M., & Pandey, M. (2012). Survivin expression and targeting in breast cancer. In *Surgical Oncology* (Vol. 21, Issue 2, pp. 125–131).
<https://doi.org/10.1016/j.suronc.2011.01.001>
- Jin, S., Li, R., Chen, M.-Y., Yu, C., Tang, L.-Q., Liu, Y.-M., Li, J.-P., Liu, Y.-N., Luo, Y.-L., Zhao, Y., Zhang, Y., Xia, T.-L., Liu, S.-X., Liu, Q., Wang, G.-N., You, R., Peng, J.-Y., Li, J., Han, F., ... Zeng, M.-S. (2020). Single-cell transcriptomic analysis defines the interplay between tumor cells, viral infection, and the microenvironment in

- nasopharyngeal carcinoma. *Cell Research*, 30(11), 950–965.
- Karp, G. (2011). *Biología celular y molecular: conceptos y experimentos* (6a. ed.). McGraw Hill Mexico.
- Karve, T. M., & Rosen, E. M. (2012). B-cell translocation gene 2 (BTG2) stimulates cellular antioxidant defenses through the antioxidant transcription factor NFE2L2 in human mammary epithelial cells. *The Journal of Biological Chemistry*, 287(37), 31503–31514.
- Kouros-Mehr, H., Bechis, S. K., Slorach, E. M., Littlepage, L. E., Egeblad, M., Ewald, A. J., Pai, S.-Y., Ho, I.-C., & Werb, Z. (2008). GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell*, 13(2), 141–152.
- Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D., & Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell*, 127(5), 1041–1055.
- Krämer, A., Green, J., Pollard, J., Jr, & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, 30(4), 523–530.
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97.
- Kumar, U., Ardasheva, A., Mahmud, Z., Coombes, R. C., & Yagüe, E. (2021). FOXA1 is a determinant of drug resistance in breast cancer cells. *Breast Cancer Research and Treatment*, 186(2), 317–326.
- Lacroix, M., & Leclercq, G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Research and Treatment*, 83(3), 249–289.
- Laurent, J., Frongia, C., Cazales, M., Mondesert, O., Ducommun, B., & Lobjois, V. (2013). Multicellular tumor spheroid models to explore cell cycle checkpoints in 3D. *BMC Cancer*, 13, 73.
- Lazzari, G., Couvreur, P., & Mura, S. (2017). Multicellular tumor spheroids: a relevant 3D model for the in vitro preclinical investigation of polymer nanomedicines. In *Polymer Chemistry* (Vol. 8, Issue 34, pp. 4947–4969). <https://doi.org/10.1039/c7py00559h>
- Lee, A. V., Oesterreich, S., & Davidson, N. E. (2015). MCF-7 Cells—Changing the Course of Breast Cancer Research and Care for 45 Years. *Journal of the National Cancer Institute*, 107(7). <https://doi.org/10.1093/jnci/djv073>
- Lee, H. J., Song, I.-C., Yun, H.-J., Jo, D.-Y., & Kim, S. (2014). CXC chemokines and chemokine receptors in gastric cancer: from basic findings towards therapeutic targeting. *World Journal of Gastroenterology: WJG*, 20(7), 1681–1693.
- Lee, J. Y., Park, Y. J., Oh, N., Kwack, K. B., & Park, K.-S. (2017). A transcriptional complex

- composed of ER(α), GATA3, FOXA1 and ELL3 regulates IL-20 expression in breast cancer cells. *Oncotarget*, 8(26), 42752–42760.
- Lo, H.-W., Hsu, S.-C., Xia, W., Cao, X., Shih, J.-Y., Wei, Y., Abbruzzese, J. L., Hortobagyi, G. N., & Hung, M.-C. (2007). Epidermal growth factor receptor cooperates with signal transducer and activator of transcription 3 to induce epithelial-mesenchymal transition in cancer cells via up-regulation of TWIST gene expression. *Cancer Research*, 67(19), 9066–9076.
- Loibl, S., Poortmans, P., Morrow, M., Denkert, C., & Curigliano, G. (2021). Breast cancer. *The Lancet*, 397(10286), 1750–1769.
- Łukasiewicz, S., Czezelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. *Cancers*, 13(17).
<https://doi.org/10.3390/cancers13174287>
- Lytal, N., Ran, D., & An, L. (2020). Normalization Methods on Single-Cell RNA-seq Data: An Empirical Survey. *Frontiers in Genetics*, 11, 41.
- Madak-Erdogan, Z., Lupien, M., Stossi, F., Brown, M., & Katzenellenbogen, B. S. (2011). Genomic Collaboration of Estrogen Receptor α and Extracellular Signal-Regulated Kinase 2 in Regulating Gene and Proliferation Programs. In *Molecular and Cellular Biology* (Vol. 31, Issue 1, pp. 226–236). <https://doi.org/10.1128/mcb.00821-10>
- Mavrommati, I., Johnson, F., Echeverria, G. V., & Natrajan, R. (2021). Subclonal heterogeneity and evolution in breast cancer. *NPJ Breast Cancer*, 7(1), 155.
- Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., & Stillman, B. (2012). DDX5 Regulates DNA Replication and Is Required for Cell Proliferation in a Subset of Breast Cancer Cells. *Cancer Discovery*, 2(9), 812–825.
- Mojic, M., Takeda, K., & Hayakawa, Y. (2017). The Dark Side of IFN- γ : Its Role in Promoting Cancer Immuno-evasion. *International Journal of Molecular Sciences*, 19(1).
<https://doi.org/10.3390/ijms19010089>
- Muciño-Olmos, E. A., Vázquez-Jiménez, A., Avila-Ponce de León, U., Matadamas-Guzman, M., Maldonado, V., López-Santaella, T., Hernández-Hernández, A., & Resendis-Antonio, O. (2020). Unveiling functional heterogeneity in breast cancer multicellular tumor spheroids through single-cell RNA-seq. *Scientific Reports*, 10(1), 12728.
- Muciño-Olmos, E. A., Vázquez-Jiménez, A., López-Esparza, D. E., Maldonado, V., Valverde, M., & Resendis-Antonio, O. (2020). MicroRNAs Regulate Metabolic Phenotypes During Multicellular Tumor Spheroids Progression. *Frontiers in Oncology*, 10, 582396.
- Müssel, C., Hopfensitz, M., & Kestler, H. A. (2010). BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. In *Bioinformatics* (Vol. 26, Issue 10,

- pp. 1378–1380). <https://doi.org/10.1093/bioinformatics/btq124>
- Naldi, A., Remy, E., Thieffry, D., & Chaouiya, C. (2011). Dynamically consistent reduction of logical regulatory graphs. In *Theoretical Computer Science* (Vol. 412, Issue 21, pp. 2207–2218). <https://doi.org/10.1016/j.tcs.2010.10.021>
- Nasser, M. W., Qamri, Z., Deol, Y. S., Ravi, J., Powell, C. A., Trikha, P., Schwendener, R. A., Bai, X.-F., Shilo, K., Zou, X., Leone, G., Wolf, R., Yuspa, S. H., & Ganju, R. K. (2012). S100A7 enhances mammary tumorigenesis through upregulation of inflammatory pathways. *Cancer Research*, *72*(3), 604–615.
- Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G., & Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*, *128*(8), e20–e31.
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Louis, D. N., Rozenblatt-Rosen, O., Suvà, M. L., Regev, A., & Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, *344*(6190), 1396–1401.
- Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M. A., & Daidone, M. G. (2005). Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Research*, *65*(13), 5506–5511.
- Post, A. E. M., Smid, M., Nagelkerke, A., Martens, J. W. M., Bussink, J., Sweep, F. C. G. J., & Span, P. N. (2018). Interferon-Stimulated Genes Are Involved in Cross-resistance to Radiotherapy in Tamoxifen-Resistant Breast Cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *24*(14), 3397–3408.
- Prasetyanti, P. R., & Medema, J. P. (2017). Intra-tumor heterogeneity from a cancer stem cell perspective. *Molecular Cancer*, *16*(1), 41.
- Promkan, M., Dakeng, S., Chakrabarty, S., Bögl, O., & Patmasiriwat, P. (2013). The effectiveness of cucurbitacin B in BRCA1 defective breast cancer cells. *PloS One*, *8*(2), e55732.
- Provance, O. K., & Lewis-Wambi, J. (2019). Deciphering the role of interferon alpha signaling and microenvironment crosstalk in inflammatory breast cancer. *Breast Cancer Research: BCR*, *21*(1), 59.
- Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., & Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, *171*(7), 1611–1624.e24.

- Ruan, H., Liao, Y., Ren, Z., Mao, L., Yao, F., Yu, P., Ye, Y., Zhang, Z., Li, S., Xu, H., Liu, J., Diao, L., Zhou, B., Han, L., & Wang, L. (2019). Single-cell reconstruction of differentiation trajectory reveals a critical role of ETS1 in human cardiac lineage commitment. *BMC Biology*, *17*(1), 89.
- Saadatpour, A., Albert, I., & Albert, R. (2010). Attractor analysis of asynchronous Boolean models of signal transduction networks. *Journal of Theoretical Biology*, *266*(4), 641–656.
- Salama, I., Malone, P. S., Mihaimeed, F., & Jones, J. L. (2008). A review of the S100 proteins in cancer. In *European Journal of Surgical Oncology (EJSO)* (Vol. 34, Issue 4, pp. 357–364). <https://doi.org/10.1016/j.ejso.2007.04.009>
- Shan, L., Li, X., Liu, L., Ding, X., Wang, Q., Zheng, Y., Duan, Y., Xuan, C., Wang, Y., Yang, F., Shang, Y., & Shi, L. (2014). GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation. *Oncogene*, *33*(24), 3205–3216.
- Shirley, S. H., Rundhaug, J. E., Tian, J., Cullinan-Ammann, N., Lambertz, I., Conti, C. J., & Fuchs-Young, R. (2009). Transcriptional regulation of estrogen receptor-alpha by p53 in human breast cancer cells. *Cancer Research*, *69*(8), 3405–3414.
- Shi, Y., Tsang, J. Y. S., Ni, Y., & Tse, G. M. (2017). Intratumoral Heterogeneity in Breast Cancer: A Comparison of Primary and Metastatic Breast Cancers. *The Oncologist*, *22*(4), 487–490.
- Silwal-Pandit, L., Langerød, A., & Børresen-Dale, A.-L. (2017). TP53 Mutations in Breast and Ovarian Cancer. *Cold Spring Harbor Perspectives in Medicine*, *7*(1). <https://doi.org/10.1101/cshperspect.a026252>
- Single Cell Set-Up: Sample Preparation Tips*. (2017, December 18). <http://www.biocompare.com/Bench-Tips/345311-Single-Cell-Set-Up-Sample-Preparation-Tips/>
- Stratton, M. R., Campbell, P. J., & Andrew Futreal, P. (2009). The cancer genome. In *Nature* (Vol. 458, Issue 7239, pp. 719–724). <https://doi.org/10.1038/nature07943>
- Suda, K., Murakami, I., Sakai, K., Tomizawa, K., Mizuuchi, H., Sato, K., Nishio, K., & Mitsudomi, T. (2016). Heterogeneity in resistance mechanisms causes shorter duration of epidermal growth factor receptor kinase inhibitor treatment in lung cancer. *Lung Cancer*, *91*, 36–40.
- Sullivan, N. J., Sasser, A. K., Axel, A. E., Vesuna, F., Raman, V., Ramirez, N., Oberyszyn, T. M., & Hall, B. M. (2009). Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. *Oncogene*, *28*(33), 2940–2947.
- Sun, X., Bizhanova, A., Matheson, T. D., Yu, J., Zhu, L. J., & Kaufman, P. D. (2017). Ki-67

Contributes to Normal Cell Cycle Progression and Inactive X Heterochromatin in p21 Checkpoint-Proficient Human Cells. *Molecular and Cellular Biology*, 37(17).

<https://doi.org/10.1128/MCB.00569-16>

The International Agency for Research on Cancer (IARC). (2020). *Global Cancer Observatory*. Retrieved June 8, 2022, from <https://gco.iarc.fr/>

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., 2nd, Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352(6282), 189–196.

Understanding Breast Cancer Types. (2022). Retrieved June 8, 2022, from

<https://www.rockymountaincancercenters.com/breast-cancer/types-hormone-receptors>

Valla, M., Klæstad, E., Ytterhus, B., & Bofin, A. M. (2022). CCND1 Amplification in Breast Cancer -associations With Proliferation, Histopathological Grade, Molecular Subtype and Prognosis. *Journal of Mammary Gland Biology and Neoplasia*, 27(1), 67–77.

Vendramin, R., Litchfield, K., & Swanton, C. (2021). Cancer evolution: Darwin and beyond. *The EMBO Journal*, 40(18), e108389.

Wang, Y., Waters, J., Leung, M. L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., Multani, A., Zhang, H., Zhao, R., Michor, F., Meric-Bernstam, F., & Navin, N. E. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, 512(7513), 155–160.

Weiswald, L. B., Bellet, D., & Dangles-Marie, V. (2015). Spherical cancer models in tumor biology. *Neoplasia*, 17(1), 1-15.

Xu, Y., Qin, L., Sun, T., Wu, H., He, T., Yang, Z., Mo, Q., Liao, L., & Xu, J. (2017). Twist1 promotes breast cancer invasion and metastasis by silencing Foxa1 expression. *Oncogene*, 36(8), 1157–1166.

Yamamoto, H., Ngan, C. Y., & Monden, M. (2008). Cancer cells survive with survivin. In *Cancer Science* (Vol. 99, Issue 9, pp. 1709–1714). <https://doi.org/10.1111/j.1349-7006.2008.00870.x>

Yang, N., Wang, C., Wang, Z., Zona, S., Lin, S.-X., Wang, X., Yan, M., Zheng, F.-M., Li, S.-S., Xu, B., Bella, L., Yong, J.-S., W-F Lam, E., & Liu, Q. (2017). FOXM1 recruits nuclear Aurora kinase A to participate in a positive feedback loop essential for the self-renewal of breast cancer stem cells. In *Oncogene* (Vol. 36, Issue 24, pp. 3428–3440). <https://doi.org/10.1038/onc.2016.490>

Yoon, C.-H., Kim, M.-J., Lee, H., Kim, R.-K., Lim, E.-J., Yoo, K.-C., Lee, G.-H., Cui, Y.-H., Oh, Y. S., Gye, M. C., Lee, Y. Y., Park, I.-C., An, S., Hwang, S.-G., Park, M.-J., Suh, Y., &

- Lee, S.-J. (2012). PTTG1 oncogene promotes tumor malignancy via epithelial to mesenchymal transition and expansion of cancer stem cell population. *The Journal of Biological Chemistry*, 287(23), 19516–19527.
- Zhang S.-Q., Qiang S.-Y., Yang W.-B., Jiang J.-T., & Ji Z.-Z. (2004). [Expression of survivin in different stages of carcinogenesis and progression of breast cancer]. *Ai zheng = Aizheng = Chinese journal of cancer*, 23(6), 697–700.
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8, 14049.