



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

**Mecanismos de memoria y herencia molecular
en redes de señalización y regulación génica
de células madre pluripotentes**

Tesis presentada para optar al título de Doctor de la
Universidad de Buenos Aires en el área de las Ciencias Físicas

Lic. Federico Sevlever

Directora de tesis: Dra. Alejandra C. Ventura
Co-Director de tesis: Dr. Santiago Miriuka
Consejero de Estudios: Dr. Hernán Grecco
Lugares de trabajo: Instituto de Fisiología, Biología Molecular y
Neurociencias, UBA - CONICET
Laboratorio de Investigaciones Aplicadas a las
Neurociencias, Fleni - CONICET
Fecha y lugar: Buenos Aires, 16 de Junio de 2022

Mecanismos de memoria y herencia molecular en redes de señalización y regulación génica de células madre pluripotentes

Resumen

Esta tesis consiste en el estudio del ciclo celular de células madre pluripotentes (CMP) en diferenciación y los mecanismos moleculares que pueden almacenar memoria de un estímulo externo. Pretende mostrar cómo existen factores hereditarios que determinan el ciclo, y estudiar los mecanismos de herencia y memoria molecular dentro de las redes de regulación génica. Se divide en 3 capítulos.

El primero aborda el estudio del ciclo celular y sus fases durante el proceso de diferenciación temprana en las CMP de embrión de ratón. Estas células fueron genéticamente modificadas para fluorescer en dos colores marcando las fases del ciclo, y ser filmadas por microscopía de time-lapse. Los análisis incluyen medidas del ciclo celular a nivel de célula única, construcción de dendrogramas de linajes celulares y el diseño y aplicación de algoritmos para probar si las relaciones familiares determinan la duración del ciclo celular. El resultado principal es que, durante la diferenciación, el ciclo se acorta y las células proliferan a mayor velocidad. Además, la duración del ciclo de una célula está determinada por las células progenitoras, probando que existen factores parentales que controlan la tasa de proliferación. Sin embargo, una de las dificultades a la hora de encontrar cuáles son estos factores, encontrar candidatos y compararlos es, por un lado, la falta de cuantificadores de memoria almacenada en las redes de regulación génica y, por otro, la falta de un catálogo de mecanismos moleculares capaces de almacenar memoria.

El segundo capítulo, por su parte, consiste en el desarrollo de un cuantificador matemático de la memoria molecular en redes complejas, aplicado principalmente a vías de señalización celular y redes de regulación génica, que guardan información de estímulos

transitorios. Una vez definido el cuantificador, se utilizó para buscar mecanismos o *motifs* de 1, 2 y 3 nodos que fueran capaces de almacenar información de estos estímulos. Los resultados principales consisten en que los ciclos de retroalimentación positiva son esenciales para dar multiestabilidad y memoria a las redes. Además, las redes capaces de oscilar de manera autónoma también pueden guardar información en la fase de la oscilación.

El tercer y último capítulo consiste en una serie de experimentos en las mismas células que se utilizaron en el primer capítulo. El objetivo es detectar la existencia de mecanismos capaces de almacenar memoria. Para esto, tratamos a las células con un estímulo de diferenciación transitorio y luego aplicamos el cuantificador desarrollado en el segundo capítulo. El resultado es que, luego de solo un día expuestas al estímulo de diferenciación, las células lo recuerdan y continúan con el proceso de diferenciación. Asimismo, es probable que esta memoria se deba a la existencia de alguna retroalimentación positiva entre los factores de transcripción marcadores de la pluripotencia.

Palabras Clave

Células madre, pluripotencia, ciclo celular, diferenciación, memoria molecular, red de regulación génica, biestabilidad.

Mechanisms of memory and molecular inheritance in signaling networks and gene regulation of pluripotent stem cells

Abstract

This thesis is focused on the study of cell cycle in pluripotent stem cells (PSC) during differentiation and on the molecular mechanisms which can store the memory of an external stimulus. It aims to show if there are hereditary factors that determine it, and to study the mechanisms of inheritance and molecular memory within gene regulatory networks. It is organized into three chapters.

The first chapter deals with the study of the cell cycle and its phases during the early differentiation process in mouse embryo PSCs. These cells were genetically modified to fluoresce in two colors marking the phases of the cycle, and to be filmed by time-lapse microscopy. Analyzes include cell cycle measurements at the single cell level, construction of cell lineage dendrograms, and the design and application of algorithms to test whether familial bonds determine cell cycle length. The main result is that, during differentiation, the cycle is shortened, and cells proliferate at a faster rate. Furthermore, the length of a cell cycle is determined by the progenitor cells, proving that there are hereditary factors that control the rate of proliferation. However, one of the needs in finding out what these factors are, obtain candidates and compare them is, on one hand, the lack of memory quantifiers stored in gene regulation networks and, on the other hand, some catalog of the molecular mechanisms capable of storing memory. This issue is addressed in the next chapter.

The second chapter deals with the development of mathematical quantifier to account for molecular memory in complex networks, applied mainly to signaling and gene regulatory networks which store information of transient stimuli. Once defined, we use the quantifier to look for 1,2 & 3 nodes mechanisms or *motifs* capable of stimuli information storage. The main results consist of positive feedbacks being essential to provide multistability and memory to networks. Furthermore, oscillating networks can also store information in their phase.

The third and last chapter is a description of experiments performed on the same cells used in the first chapter. The aim of these experiments is to detect mechanisms capable of storing memory. To accomplish this, we give a differentiation stimulus and apply the memory quantifier developed in the second chapter. The main result is that, after only one day exposed to a differentiation stimulus, cells remember it and continue its progression through the differentiation process. In addition, this memory is probably due to the existence of a positive feedback between pluripotency transcription factors.

Keywords

Stem cells, pluripotency, cell cycle, differentiation, molecular memory, gene regulatory networks, bistability.

Agradecimientos

Un doctorado es un trabajo que está pensado oficialmente para ser individual (el título de Dr. no es compartido). Sin embargo, esta individualidad se reduce exclusivamente a los papeles. No creo que sea posible hacer un doctorado de manera realmente individual, o al menos conservar la salud física y mental durante el proceso¹. Quiero agradecer a toda la gente que fue indispensable para completar esta etapa de mi vida, que considero tan difícil como formativa. No me van a alcanzar ni todas las combinaciones de palabras de este rico idioma, pero lo intentaré.

A Ale. Por dirigir mi investigación científica desde que empecé a hacerla allá por 2016, cuando era apenas un físico que lo único que creía saber de biología era que no tenía mucha matemática... Por formarme desde entonces, por mostrarme el fascinante camino de una línea científica que vive en la frontera de la física, de la biología, de la computación, de la matemática, de la química y de seguro más disciplinas que se me escapan. Por la infinita paciencia en el interminable proceso de aprender, de leer papers de otros campos y de acompañarme y guiarme en el pasaje de estudiar qué preguntas se hicieron otros y cómo se las respondieron a qué preguntas me puedo y quiero hacer yo y darme las herramientas para respondérmelas. Será un placer continuar haciéndonos preguntas.

A Ari. Por tus críticas tan fundamentales y tu manera de darlas sin salir nunca de la alegría y el placer de hacer mejor ciencia. No me explico cómo se pueden combinar un excelente docente, una paciencia infinita, un conocimiento envidiable, un amor por la ciencia que contagia y motiva y una humanidad tan considerada en una sola persona. No es fácil para un físico aprender biología, y mucho menos experimental. Sé que lo mucho que aprendí en este campo fue gracias a vos. Dudo que hubiera podido seguir este camino sin tu aporte y estoy seguro de que no podría haberlo hecho tan felizmente. Por más que lo intente, sé que no voy a poder devolverte nunca lo que me diste, así que

trataré de aplicarlo lo mejor que pueda con las generaciones venideras. Estoy feliz y ansioso de continuar el camino científico con tu mentoría.

A Santiago. Por alentarme siempre a aprender, no importa cuán lejos está el objetivo de mi comodidad. Por esa mirada siempre tan pragmática, estratégica y experimentada que aporta tanto a las decisiones determinantes que me tocan tomar en mi vida.

Al Conicet, al Estado Nacional y al Pueblo Argentino. Por seguir bancando (y bancándome) la ciencia.

Al grupo de la oficina 346. A la vieja guardia Juan DB, Jule y Juan M y la no tan vieja Sheshu, Cande y Pau. Por comentar, compartir, escuchar y hacer del laburo cotidiano algo mucho más ameno. Brindo por los viajes, las escuelas y los congresos compartidos. La ciencia no está bien si no es en equipo.

A Ali, Gus, Mati, Dai, Pau, Andy, Ger, Nico, Vicky, Estef, Maru, Pepe y Ale CL. Por todos los aportes en forma de lab meetings, por los mates que les he robado, por la curiosidad permanente en tratar de entender lo que hacía y por estar siempre que esperaba que termine de correr algún programa (e incluso cuando ya no estaba esperando). Por los Beernes.

A la banda del Lian. Por hacerme sentir siempre bienvenido, por siempre estar dispuestos a ayudar, a enseñar, a colaborar con un favor para un experimento, a festejar los logros y a aprender de los fracasos. Por formar un maravilloso equipo de trabajo e introducirme a maneras muy diferentes de organización grupal.

A Game Time, ese primer grupo con el que cuento en mi vida que sabe mantener la tradición de jugar al básquet una vez por semana. No saben lo necesarios que han sido para esta tesis.

A Ro, Emi, Jula y Bruno. Ese grupo cuyos integrantes son una guía para mí que siempre está para escuchar, aconsejar y, sobre todo, festejar como se debe cuando corresponde.

A Celes, Sole y Cami. Por acompañarme en mi inmersión a la biología molecular y bancarnos la catarsis mutua del arduo proceso que significa escribir una tesis doctoral.

A Agus, Jula, Chechu, Lau, Cami, Anto, Maru, Sol, Tina, Anita, Ro, Soph, Fiore y Cate. Por introducirme a los feminismos y cambiar mi ser de manera constante e irreversible. Estoy seguro de que sin ustedes haría ciencia de peor calidad, más sesgada e incluso dañina. Por enseñarme a empatizar.

A Ger, Andy y Dai. Por las clases, el pistacho, los recomendados gastronómicos y porque nunca falte el buen morfi en un encuentro, no importa si es helado y después chori. El orden de los factores no altera el producto.

A Cate, Yeyu, Lu y Axel. Esa banda completamente heterogénea que sabe encontrar siempre el momento para jugar.

A Machete. Ese grupo que lleva unido ya más años de los que se pueden recordar, y al que parece que el tiempo solo le hace bien. No importa qué subconjunto pueda ir a cenar hoy, siempre es garantía de un momento de risa y felicidad.

A Gonza en particular, por acompañarme en el doctorado y mostrarme también el camino de la biología. Por interesarse siempre en lo que estoy haciendo. Por ser un ejemplo cotidiano de crecimiento y superación y por también enojarse conmigo cuando había que enojarse.

A Cami, Anto y Axel. Esas personas donde cualquier combinación ha sabido armar un viaje de esos que marcan fuerte. Por la amistad y el amor, irremplazables para mí.

A la tía Silvita. Por seguir pensando que la voy a hacer millonaria.

A Charly, Kari y los sábados de súper acción. Por regalarme una actividad completamente diferente a la que estaba acostumbrado a realizar. Por las charlas, los tés, la técnica y, sobre todo, por enseñarme a robarle algo a la belleza de este mundo y plasmarla en un objeto.

A Tina y Berni, por agrandar la familia y dotarla de virtudes que, sin ustedes, difícilmente hubiera alcanzado.

A Soph, por el café. Porque siempre salgo mejor persona después de verla.

A los Partys. Por hacerme parte y por hacerme Party. Son la segunda familia que me abrió las puertas de su hogar cuando ni siquiera las busqué.

A la Prima, el Pig y Luisito. Dicen que la familia no se elige. "Vengan de a uno", como diría Luis. No puedo (ni quiero) imaginarme una navidad sin invisibles, por más caótico que resulte. A la prima por estar siempre y por siempre estar organizando el próximo encuentro.

A Juan y Maru. Por ser ejemplos de dinámicas paralelas a la mía, con casi exactamente las mismas condiciones iniciales, pero con intereses casi totalmente disímiles, aun así, complementarios. A Maru en particular, por corregirme algunas redacciones de este escrito (no está, claramente) y por estar dispuesta a corregir toda la tesis. Estoy seguro de que los de afuera ya no nos podrán devorar.

A Gustavo. Por ser, sesgos aparte, el padre más presente y dedicado que he conocido. Por enseñarme a darlo todo, a tener ambición y a ser siempre el actor principal de mis decisiones. Por mostrarme la importancia de poner un punto aparte y cambiar de capítulo para seguir escribiendo.

A Lila. Por estar siempre detrás, precedida por amor y cariño. Por atender. Por enseñarme a viajar y a conocer la naturaleza. Creo que mi interés por la ciencia y el mundo natural tiene origen en los viajes a los que me has llevado, en ver . Por encargarse, por disponer, por preocuparse. Por no dejar pasar nada y por ocuparse de todo.

A Agus. Por ser tan pero tan querible. Por acompañar mi vida y dejar que acompañe la tuya. Por enseñarme a amar y aprender conmigo a hacerlo. Por la casita. Por los viajes. Por las discusiones. Por pandemiar juntos. Por proyectar. Por mejorarme mucho más de lo que debés imaginar, desde Gerardo hasta Rayita, y lo que sea que vendrá. Por enseñarme a estar siempre cómodo, en el aspecto más amplio de la palabra. Por la felicidad gratuita y cotidiana que simplemente me regalás.

Índice

Resumen.....	2
Palabras Clave.....	4
Abstract.....	5
Keywords.....	7
Agradecimientos.....	8
Índice.....	13
Capítulo 1.....	17
Ciclo Celular	17
en Células Madre	17
Introducción.....	18
Células Madre	18
Identidad celular	19
Ciclo Celular	21
Definición y Fases	21
Ciclinas y Mecanismo molecular	22
Células Madre pluripotentes de ratón	23
Modelo de estudio	23
Vías de señalización y control	25
Medios de cultivo	26
Identidad celular, epigenética y ciclo celular	27
Metilación del ADN	27
Histonas	29
Acoplamiento con el Ciclo Celular	29
Acoplamiento con la densidad celular	31
Motivación, Estado del Arte y Objetivo.....	33
Motivación General	33
Motivación Particular	33
Objetivo	34
Esquema Experimental.....	35
Introducción	35
Reporteros fluorescentes del ciclo celular: Sistema Fucci ...	35
Marcador nuclear y segmentación	35
Microscopía de Time-Lapse	36
Extracción de Datos de los Videos	37

Segmentación, intensidad de fluorescencia y relaciones familiares	37
Base de datos	39
Análisis de Datos.....	43
Proliferación de las Colonias Celulares	43
Distribuciones del Ciclo Celular y sus fases	44
Estimación precisa del Ciclo Celular y sus fases: Método de Kaplan-Meier	46
Bases del método	47
Resultados	48
Herencia	51
Correlaciones del Ciclo y sus Fases	51
El ciclo celular está determinado por la familia: Algoritmo de Grassberger Procaccia	57
Estructura de las colonias.....	64
Tamaño	64
Densidad	66
Forma	66
Homogeneidad	68
Conclusiones y perspectivas.....	72
Autoría Correspondiente.....	74
Capítulo 2.....	75
Mecanismos de memoria molecular en redes biológicas.....	75
Introducción.....	76
Memoria de sistemas en general	76
Memoria Celular	76
Memoria molecular en redes biológicas de interacción química	77
Objetivo.....	80
Modelado matemático de redes de reacciones químicas.....	81
Ley de acción de masas	81
Dinámica de Michaelis-Menten	82
Cuantificador de memoria molecular.....	85
Ejemplo de aplicación	85
Fundamentos teórico-matemáticos	88
Mecanismos capaces de almacenar memoria.....	93
Métodos	93
Motifs	93
Determinación de motif con memoria	94
Elección de umbrales	95

Latin Hypercube Sampling	97
Integración de todas las combinaciones de redes de 1, 2 y 3 nodos	100
Combinaciones de redes	100
Modelo de ecuaciones diferenciales	101
Exploración de Parámetros	103
Base de datos	103
Algoritmo de filtrado de motivos <i>minimales</i>	104
Algoritmo	104
Resultados	107
Conclusiones	110
Memoria en osciladores.....	112
Resultados	115
Aporte de feedbacks positivos y negativos a la memoria.....	116
Algoritmo de comparación de vecinos	116
Motivación para aplicar en feedbacks	119
Resultados	120
Redes de 2 nodos	120
Redes de 3 nodos	122
Bandas de Control	122
Conclusiones y perspectivas.....	125
Capítulo 3.....	129
Biestabilidad y memoria en la diferenciación celular.....	129
Introducción.....	130
Redes de regulación génica	130
Construcción experimental	132
Módulos de Regulación en Cis	132
Mecanismos de mantenimiento de la pluripotencia	134
Feedbacks positivos y curvas dosis respuesta	136
Objetivo.....	140
Desarrollo Experimental.....	141
Diluciones 2^i	141
Hipótesis	141
Diseño experimental	141
Resultados	142
Pulso cuadrado de Dif	144
Hipótesis	144
Diseño experimental	144

Resultados	145
Aplicando el cuantificador de memoria	146
Reversión de Oct6	148
Línea celular con KO de Oct6 endógeno y agregado de Oct6 exógeno	148
Controles	150
Hipótesis	151
Diseño Experimental y Resultados	151
Métodos.....	153
RT-qPCR	153
Inmunofluorescencia	155
Protocolo experimental	155
Análisis	156
Sistema de Control de Oct6 exógeno	156
Discusiones y Perspectivas.....	158
Conclusiones, discusiones y perspectivas generales.....	160
Referencias.....	163

Capítulo 1

Ciclo Celular en Células Madre

Introducción

El objetivo de este capítulo estará centrado en el acoplamiento entre la diferenciación y el ciclo celular. Antes de enunciarlo, introducimos algunas nociones básicas de ambos campos.

Células Madre

Un organismo pluricelular está formado por células de diferentes identidades. Esto es, distinta morfología, expresión génica y, sobre todo, funcionalidad. Una neurona es diferente a un glóbulo rojo. Esta diversidad evolucionó, probablemente, porque permite la especialización de las funciones necesarias para la vida de un organismo. Un mamífero adulto consta de más de 200 tipos celulares que son consecuencia de muchos mecanismos de diferenciación celular. La diferenciación es el proceso de transformación que da lugar al cambio de identidad de una célula. Estos cambios ocurren de manera permanente en la formación de tejidos y, sobre todo, en el desarrollo embrionario. Las células se van multiplicando y diferenciando para construir los tejidos que corresponden a cada momento y lugar del cuerpo en construcción. El conjunto de identidades celulares puede organizarse en forma de árbol cuya raíz es el cigoto, la fusión del oocito y el espermatozoide, el origen de cualquier individuo. Allí se originan todas las diferenciaciones que ramifican el árbol hacia todas las identidades (Figura 1).

Así es como se van sucediendo diferenciaciones y las células van perdiendo potencial de diferenciación a medida que se especializan más y más. El potencial es la capacidad de una célula de diferenciarse. *Célula madre* no es ninguna de estas identidades en particular, sino que es la célula que tiene este potencial y la identidad dependerá del contexto, de la diferenciación que se esté estudiando. Mientras más identidades distintas puedan desarrollarse a partir de una célula madre, mayor es el potencial de ésta. De mayor a menor potencial, las células madre se clasifican en toti-, pluri-, multi- y nuli-potentes, siendo las totipotentes capaces de formar un organismo entero y nulipotentes las que no pueden cambiar de identidad (Figura 1).

Identidad celular

Las distintas identidades celulares se determinan en general midiendo la expresión génica³. Esto es, viendo qué mRNAs (por ARN mensajeros) y qué proteínas específicas de cada identidad están presentes en una dada condición o tratamiento. Algunas de estas proteínas específicas son factores de transcripción, proteínas que cumplen la función de unirse a sitios específicos en el ADN y regular la producción de otras proteínas o de sí mismas. Si un factor de transcripción está presente, significa en general que las proteínas cuya síntesis activa también lo estarán. O análogamente que no lo estarán aquellas cuya producción inhibe. Esto permite definir una identidad celular midiendo la concentración de solamente algunos factores de transcripción. En la sección Módulos de Regulación en Cis, del Capítulo 3 describimos este proceso de regulación con mayor detalle y en la de Redes de regulación génica se estiman algunas ventajas evolutivas del mismo.

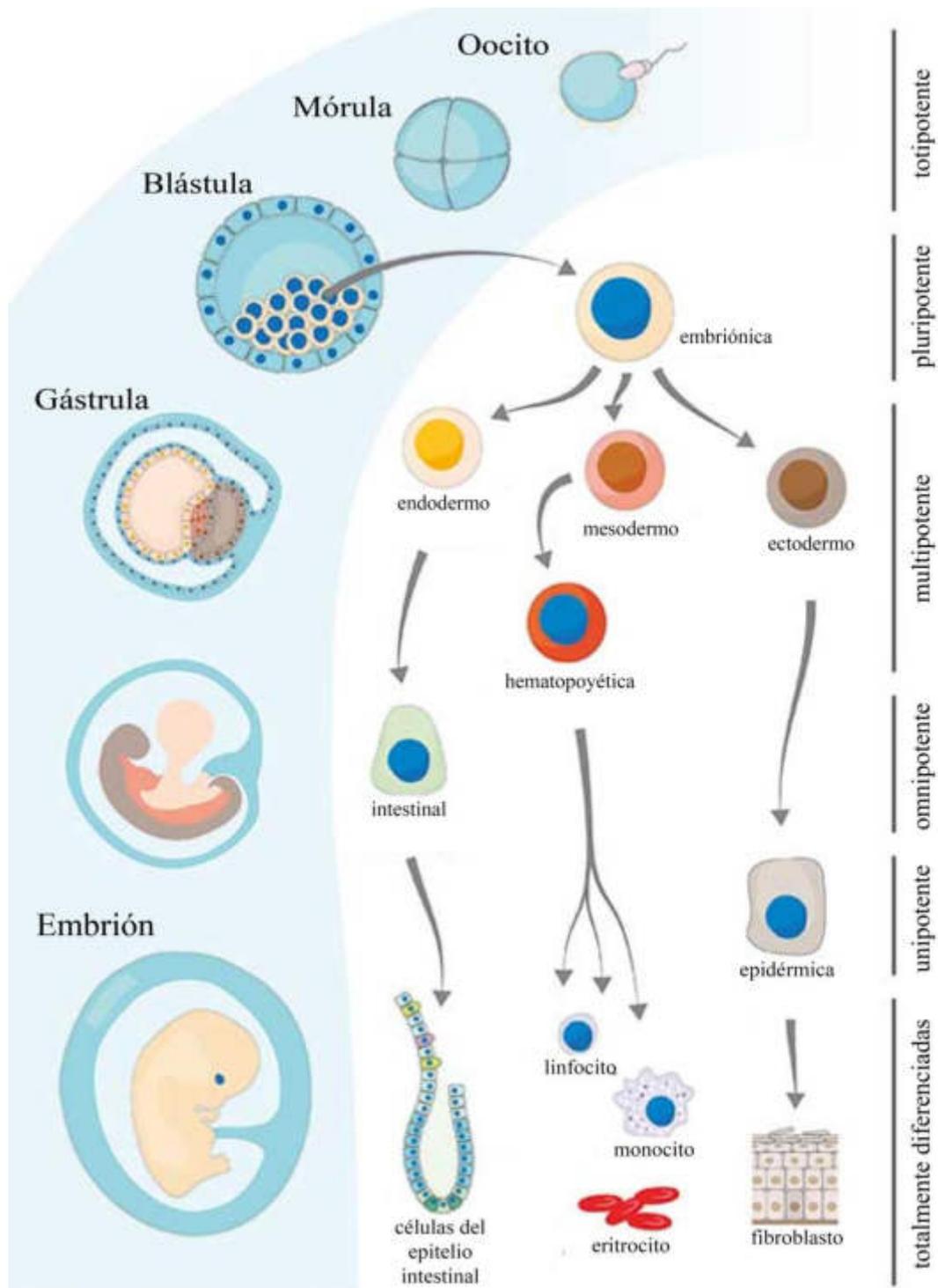


Figura 1: Imagen adaptada ¹. Esquema de algunos caminos posibles en la diferenciación celular. Ilustramos algunos estadios del desarrollo embrionario. Entre éstos, marcamos la masa celular interna del blastocito (las de núcleo azul y citoplasma amarillento), de donde se obtienen las células madre embrionarias, que pueden diferenciarse en todas las células del organismo adulto. A la derecha, los potenciales de diferenciación asociados.

¹ <https://www.enzolifesciences.com/science-center/technotes/2019/august/stem-cells-from-embryonic-origin-to-induced-pluripotency-an-overview/>

Ciclo Celular

Definición y Fases

El ciclo celular no es otra cosa que el ciclo de vida de una célula. Al ser un ciclo, la elección de comienzo y final es arbitraria, pero usualmente se lo define como el tiempo entre una división celular, que da origen a dos células, y la siguiente, donde una de estas dos vuelve a dividirse. Probablemente, la función más básica de la vida de una célula es replicarse, y para esto es preciso replicar todo el genoma y después repartir cada copia idéntica entre las dos células hijas. El ciclo se divide en 4 fases. La fase S, donde ocurre la Síntesis de la nueva copia del ADN; la fase M, donde se separan las copias y se divide la célula; y las fases G1 y G2 como intervalos o Gaps entre M y S (G1) y entre S y M (G2) (Figura 2)⁴.

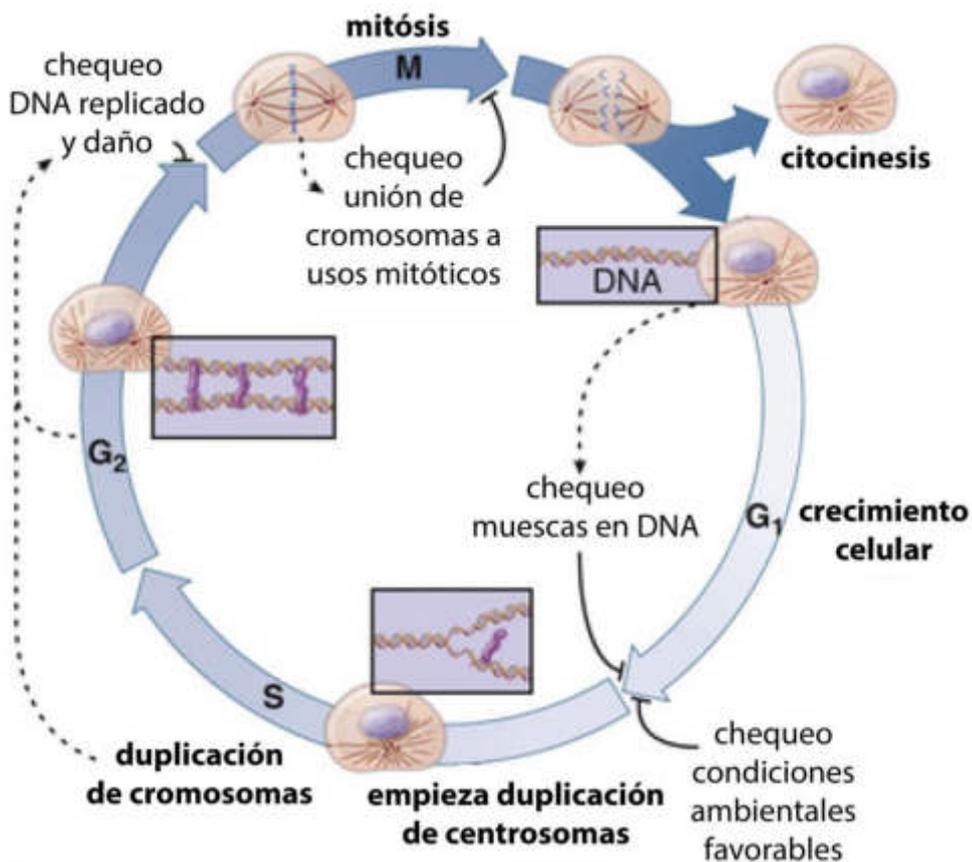


Figura 2: Imagen adaptada⁵. Esquema del ciclo celular y sus fases. Se marcaron algunos checkpoints que analiza la célula antes de continuar la progresión a través del ciclo.

Ciclinas y Mecanismo molecular

Las ciclinas⁴ son proteínas cuya síntesis y degradación están sincronizadas al ciclo celular. Así es como su concentración depende de la fase del ciclo, y funcionan como marcador de esta. Se unen de manera específica con diferentes proteínas quinasas (enzimas fosforiladoras) activando su función. Estas quinasas se denominan CDKs (*cyclin dependent kinases*). Cada una tiene un conjunto específico y diferente de proteínas diana cuya activación sincronizada en cada fase desencadena las distintas tareas necesarias para cada momento del ciclo.

El mecanismo que regula la progresión a través del ciclo está modelado en la Figura 3⁶. Consiste en un feedback negativo acoplado a dos feedbacks positivos, con el complejo Ciclina-CDK1 ocupando un rol central, formando parte de todos los feedbacks. Este sistema es capaz de oscilar de manera independiente. Los dos feedbacks positivos entre el complejo con las proteínas Wee1 (doble negativo) y Cdc25 (doble positivo) le dan biestabilidad. Así, existe un conjunto dado de concentraciones estables con Ciclina-CDK1 y Cdc25 activados y Wee1 inactivado y otro conjunto espejo con Ciclina-CDK1 y Cdc25 inactivados y Wee1 activado (Figura 3). Acoplando ahora el feedback negativo entre el complejo y APC-Cdc20, generamos la siguiente dinámica: se activa Ciclina-CDK1 y los feedbacks positivos, alcanzando la estabilidad con alta concentración del complejo. Esto activa luego a APC-Cdc20, que a su vez comienza a inhibir al complejo hasta que un tiempo después lo lleva a la estabilidad espejo, con el complejo inactivo. Finalmente, como APC-Cdc20 inactiva a su activador, termina por degradarse dando fin a un ciclo y comienzo al siguiente.

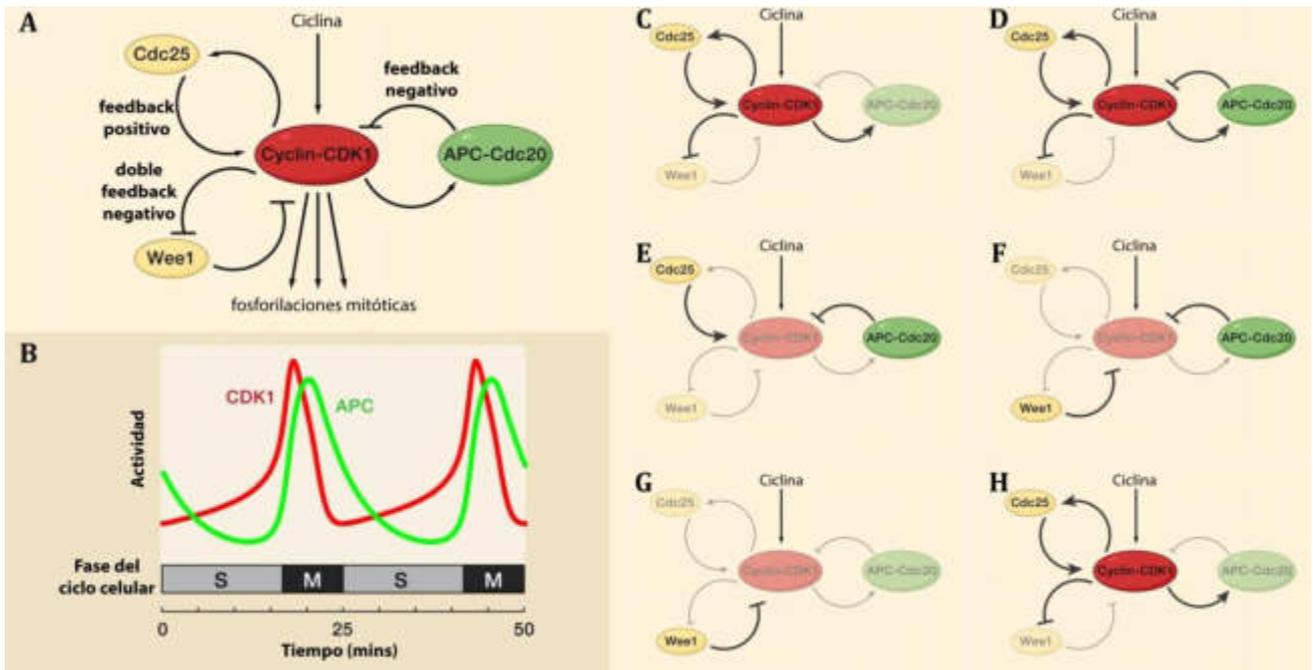


Figura 3: Imagen adaptada⁷.

A: Esquema simplificado del mecanismo de feedbacks negativo y positivos acoplados que oscila autónomamente y controla la progresión a través del ciclo.

B: Dinámica oscilatoria de los dos actores principales (el complejo Ciclina-CDK1 en rojo y el APC-Cdc20 en verde).

C-H: Dinámica y mecanismos representados de a pasos. En **C**, el complejo Ciclina-CDK1 está activo ambos feedbacks positivos lo estabilizan en actividad. En **D**, éste activa al complejo APC-Cdc20 y así su propia inhibición. En **E** y **F**, la ciclina se inactiva y con ésta ambos feedbacks positivos. En **G**, el complejo APC-Cdc20 se inactiva dado que perdió su fuente de activación. En **H**, la inhibición de APC-Cdc20 hace que el complejo Ciclina-CDK1 vuelva a activarse y comience nuevamente el ciclo.

Células Madre pluripotentes de ratón

Modelo de estudio

Todos los trabajos experimentales de esta tesis se realizaron en células madre embrionarias de ratón (mESC por mouse embryonic stem cells), cultivadas in vitro. Éstas son células pluripotentes, que significa que pueden diferenciarse en cualquier tipo de célula de un organismo adulto. Se obtienen de la masa celular interna (ICM por inner cell mass) del blastocito, un estadio del desarrollo que aparece al tercer día a partir de la fecundación⁸ (Figura 4).

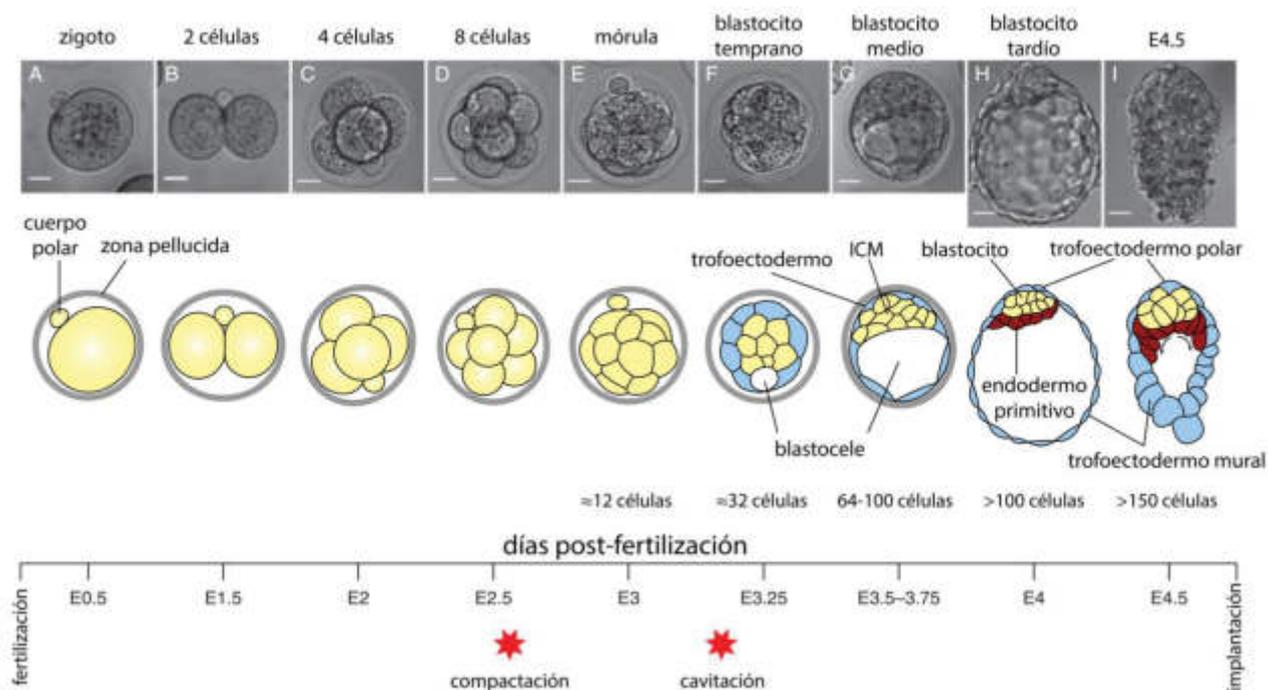


Figura 4: Imagen adaptada⁹. Fila superior: Fotografías de los estadios tempranos de desarrollo embrionario y sus nombres, desde la fertilización hasta la implantación en el útero. Fila intermedia: Esquema de los estadios tempranos de desarrollo embrionario. Los colores de las células indican diferenciaciones a distintos linajes. Fila inferior: Línea de tiempo correspondiente a los estadios, "E" es por embrión y el número que sigue marca la cantidad de días pasados desde la fertilización. Las mESC que utilizamos para esta tesis se obtienen de la masa celular interna (ICM) que aparece en la formación del blastocito⁸.

La identidad celular que corresponde a este estadio se denomina naïve (ingenuo) o ground state (estado basal). El primer nombre se debe a que estas células no cuentan con ninguna predisposición o diferenciación hacia algún linaje en particular. El segundo, a que, si se reinyectan en otro blastocito, logran acoplarse al desarrollo y formar organismos completos que cuentan con células propias e inyectadas en todos sus linajes (organismos quimera).

El proceso de diferenciación que estudia esta tesis es la transformación del estado naïve al primed. Ocurre naturalmente en el desarrollo entre los 3 y 6 días desde la fecundación. Las células primed tienen menor potencial de diferenciación y no logran formar organismos quimera si se las reinyecta en otro blastocito.

Vías de señalización y control

Introducción

Las vías de señalización son redes de interacciones químicas entre muchos compuestos en la membrana celular, el citoplasma y el núcleo. Funcionan como sensores químicos del entorno, procesadores de información y además gatillan respuestas específicamente programadas para cada entorno o estímulo externo. Usualmente, la cadena de reacciones comienza con una molécula de ligando en el medio extracelular que se une específicamente a un receptor afín, una proteína convenientemente engarzada en la membrana celular. Esta unión produce un cambio conformacional del receptor que desencadena a su vez otras reacciones hacia el interior celular, activando la vía⁴. Estas reacciones, a su vez, gatillan la respuesta, que puede ser muy variada. Por ejemplo, desde una diferenciación, cambio morfológico o transcripcional hasta la apoptosis (muerte celular programada). Así es como la simple presencia de un ligando genera una respuesta celular. Cada vía es un canal de comunicación entre la célula y el medio.

Control de la diferenciación naïve a primed

Los procesos de diferenciación en general son gatillados o están acoplados a la activación de las vías. La activación de la vía de ERK e inhibición simultánea de la vía de WNT gatillan la salida del estado naïve hacia el primed¹⁰. Este proceso ocurre naturalmente en los embriones, donde las células secretan el ligando de la vía de ERK (FGF4) y dejan de secretar el de la vía de WNT, de igual nombre (WNT).

Controlar la activación e inactivación de las vías es la manera que utilizamos en esta tesis para mantener o diferenciar las células. Las mESC en estado naïve pueden mantenerse artificial e indefinidamente in vitro utilizando un medio definido gracias a dos moléculas sintéticas inhibitoras, CHIRON (CHIR) y PD03 (PD), además de la proteína LIF (por *Leukemia Inhibitory Factor*)¹¹. Las dos primeras entran al citoplasma e inhiben cada una un nodo intermedio de las vías, las moléculas GSK3 y MEK, de las vías de WNT y ERK respectivamente (Figura 5). El LIF es el ligando de la vía de STAT3, que promueve la síntesis de factores de transcripción característicos de la pluripotencia.

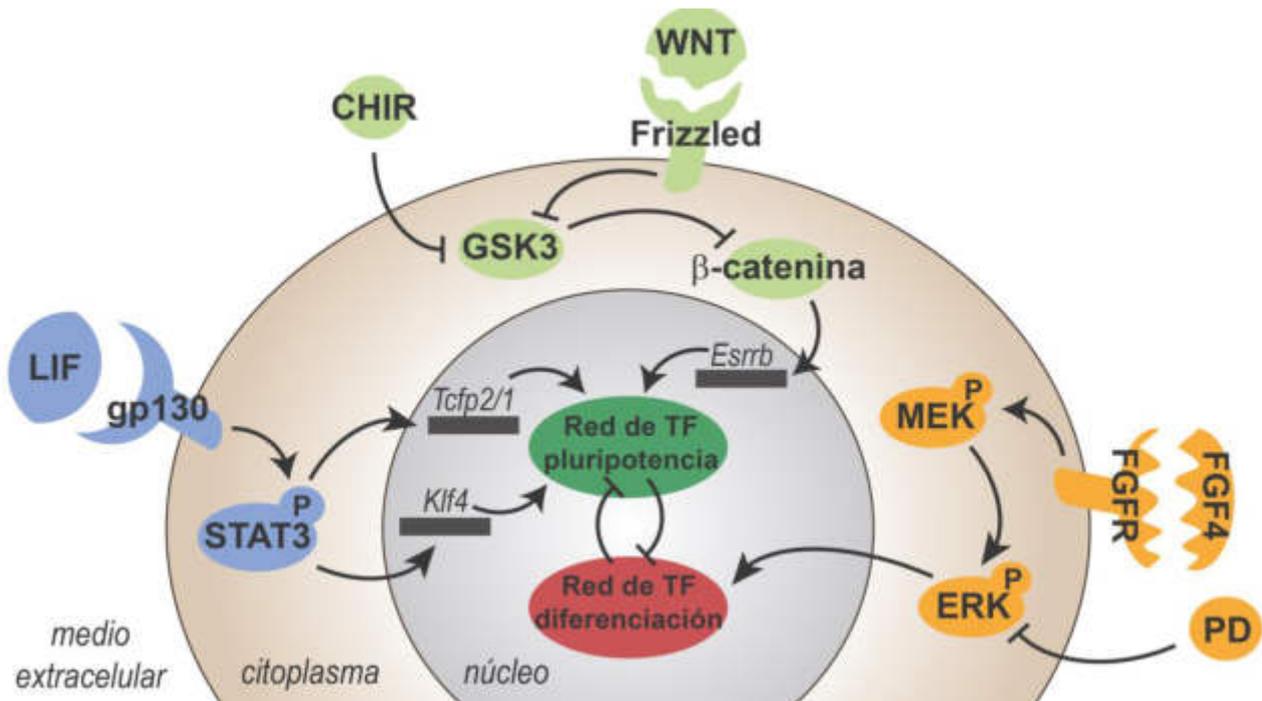


Figura 5: Imagen adaptada¹⁰. Esquema simplificado de las vías de señalización y el rol del LIF y las moléculas inhibitoras CHIR y PD en el mantenimiento de la pluripotencia.

Medios de cultivo

A lo largo de todo este capítulo utilizamos dos medios de cultivo celular, que llamaremos **Indif** y **Dif**. El primero es el medio que contiene las dos moléculas inhibitoras (CHIR y PD) de la diferenciación y el LIF (en la literatura, se denomina medio "2i+LIF"). Permite la proliferación de las células en el estado naïve. El segundo, es un medio muy similar, pero sin estos inhibidores, de manera que las células proliferan mientras se diferencian al estado primed.

En la Figura 6, mostramos la expresión de mRNAs medida por qPCR de 5 factores de transcripción marcadores del estado naïve y de otros 5 del estado primed. Estos datos fueron obtenidos en el laboratorio y aseguran que el protocolo de diferenciación funciona correctamente.

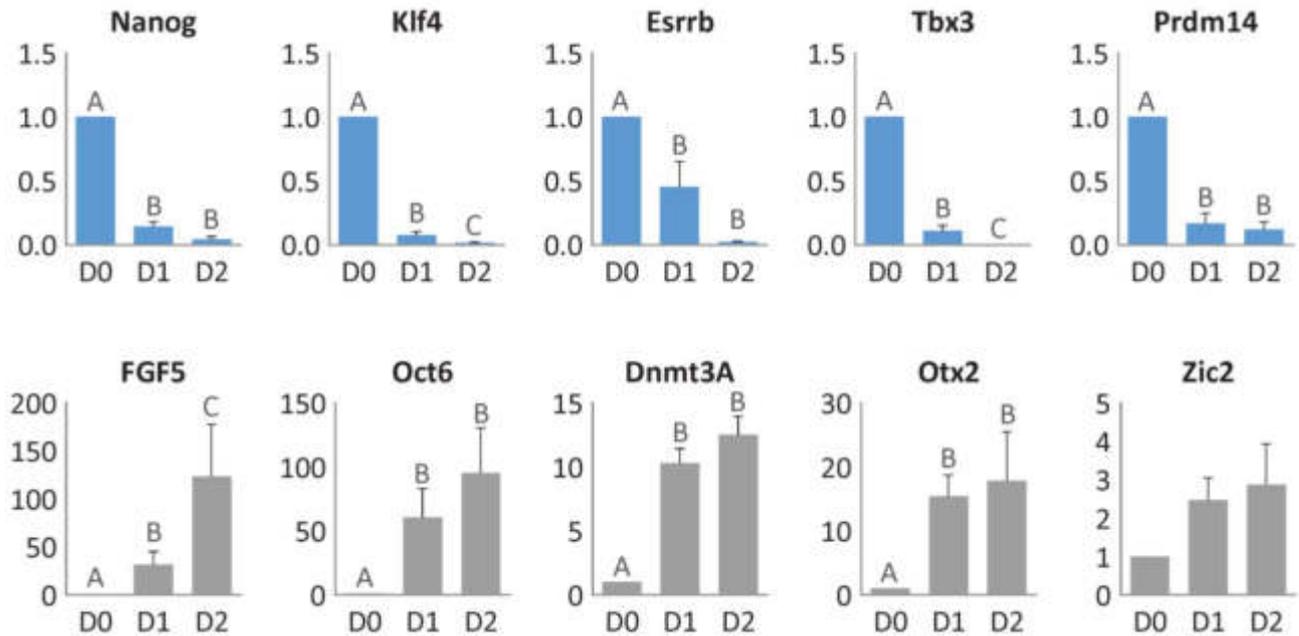


Figura 6: Expresión génica de 10 factores de transcripción. 5 marcadores del estado naïve (fila superior) y 5 del estado primed (fila inferior). Todas las expresiones se midieron por qPCR a días (D) 0, 1 y 2 desde el momento en que se colocó el medio *Dif*. Todas se encuentran normalizadas a D0. Las letras A, B, C indican expresiones estadísticamente diferentes con p -valor < 0.05 . Como era esperable, durante la diferenciación baja la expresión de los marcadores del estado naïve al tiempo que sube la de los marcadores del estado primed.

Identidad celular, epigenética y ciclo celular

La expresión génica no está regulada únicamente por los factores de transcripción. Existen más mecanismos que dan mayores posibilidades de control e incluso redundancia cuando activar o inhibir algún gen es esencial. Entre éstos se encuentran los que modifican la estructura de la cromatina para abrirla o cerrarla en regiones específicas del genoma, de manera de controlar la accesibilidad a genes específicos y así activar o inhibir su expresión. El estudio de estos mecanismos es la epigenética, y se ha demostrado que participan activamente en la diferenciación naïve a primed¹².

Metilación del ADN

Consiste en la unión de grupos metilo ($-CH_3$) a las bases nitrogenadas citosina y adenina del ADN. Estas uniones compactan la cromatina y usualmente se dan en bases dentro de promotores de genes. Los promotores son sitios en el ADN donde se unen específicamente las proteínas asociadas a la síntesis de ARN, funcionan como reguladores de la expresión génica.

Así, la metilación inactiva o silencia la expresión de los genes asociados a cada promotor.

La unión del grupo metilo es llevada adelante por distintas familias de enzimas DNMT (por DNA Metil Transferasas). Algunas cumplen funciones de mantenimiento y otras de novo. Las primeras se encargan de conservar los grupos metilo en las bases correspondientes durante la replicación del ADN, de manera que las dos células hijas hereden copias idénticas. Las segundas se encargan de cambiar los grupos metilo y con estos la expresión génica.

Histonas

Las histonas son proteínas cilíndricas donde se enrolla el ADN. Funcionan como pergaminos que se enrollan y desenrollan, regulando así la accesibilidad a genes específicos. Esta actividad depende de diversos cambios conformacionales en las histonas, entre estos hay metilaciones, acetilaciones y fosforilaciones.

Acoplamiento con el Ciclo Celular

La mitosis es, probablemente, el evento más disruptivo de la estructura de la cromatina (Figura 7). Primero, porque el ADN se replica completamente, y esto requiere del deslizamiento de la maquinaria de replicación por toda la doble hélice, que la divide en dos hélices simples como un cierre de cremallera (Figura 8). Segundo, porque una vez replicado en la fase S, los nucleosomas (histonas con ADN enrollado) se condensan para formar los cromosomas. Éstos son la estructura más comprimida del ADN, su forma empaquetada permite la separación de cada réplica en cada célula hija.

Estos dos procesos afectan probablemente a los mecanismos epigenéticos, acoplando el ciclo celular y la mitosis con la expresión génica. Una de las hipótesis de este capítulo será que estos cambios pueden afectar la identidad celular y aportar al proceso de diferenciación. Una observación que acompaña la hipótesis consiste en que, si un grupo de células naïve recibe el estímulo de diferenciación, la transición ocurre de manera asincrónica¹³ (Figura 9). La explicación consiste en que la asincronía está relacionada a que cada célula transitaba una fase del ciclo distinta al momento de recibir la señal de diferenciación.

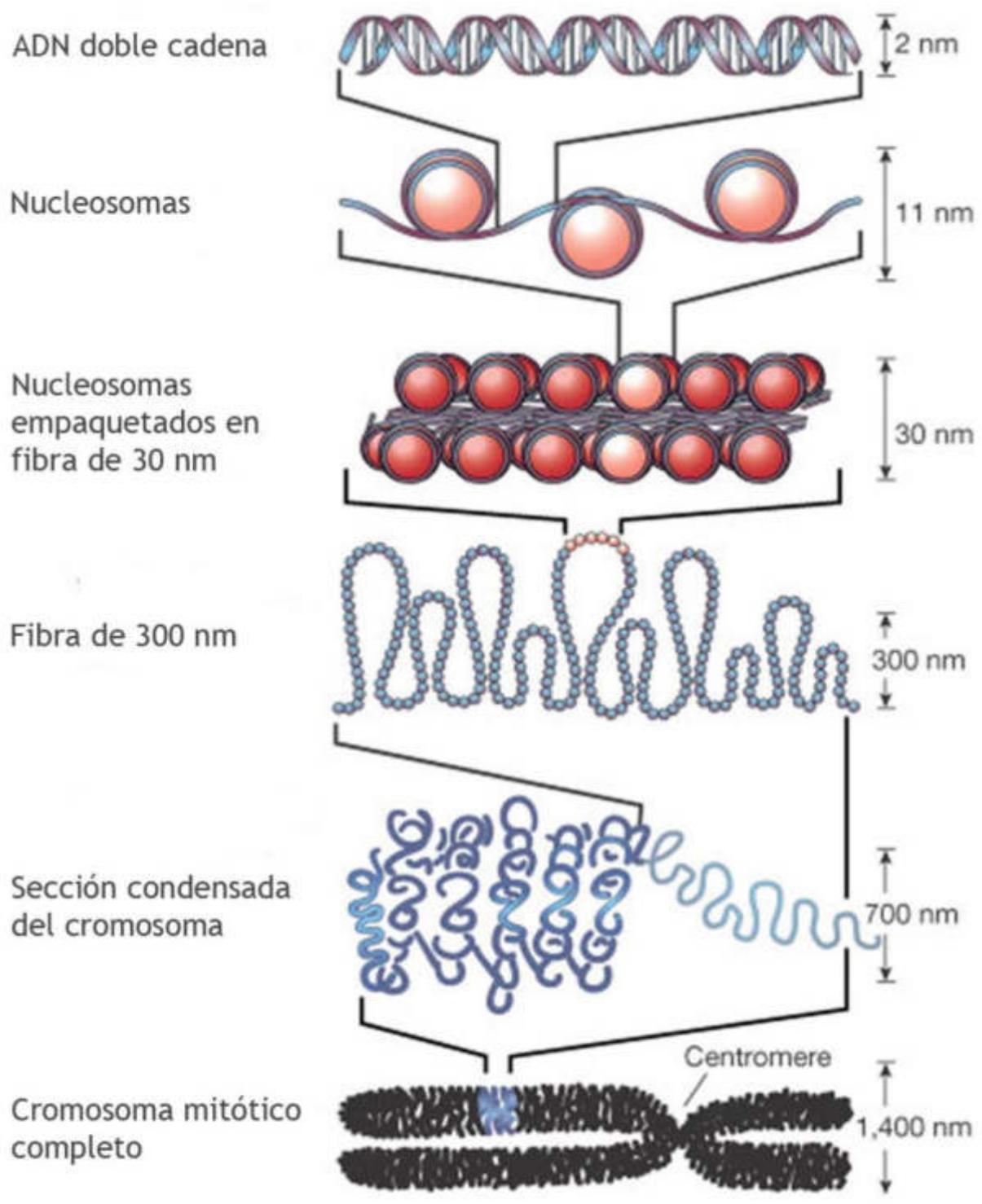


Figura 7: Imagen adaptada¹⁴. Estructuras del ADN desde la más expuesta a ser transcripta (superior) hasta la más condensada y cerrada a la transcripción (inferior).

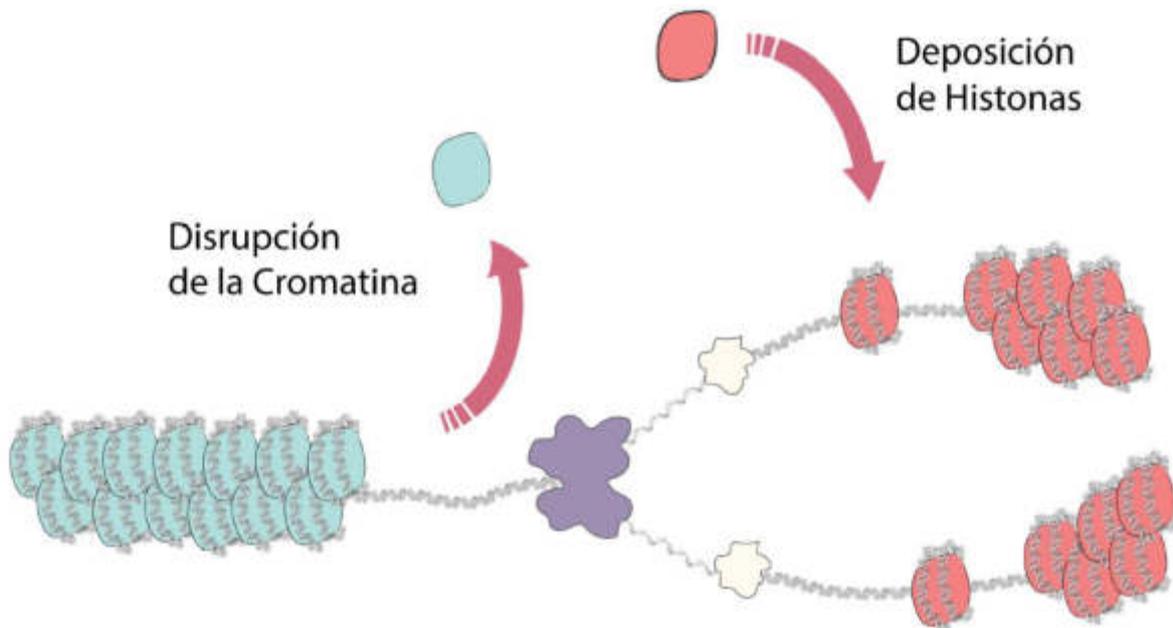


Figura 8: Imagen adaptada¹⁴. Esquema simplificado del mecanismo de replicación del ADN. La proteína central violeta (helicasa) divide la doble hélice en dos hebras simples y luego las proteínas blancas (ADN polimerasas) las completan para formar dos dobles hélices. En este proceso, las histonas que empaquetaban el ADN original (celestes) deben desligarse y luego adherirse nuevamente al ADN nuevo (rojas).

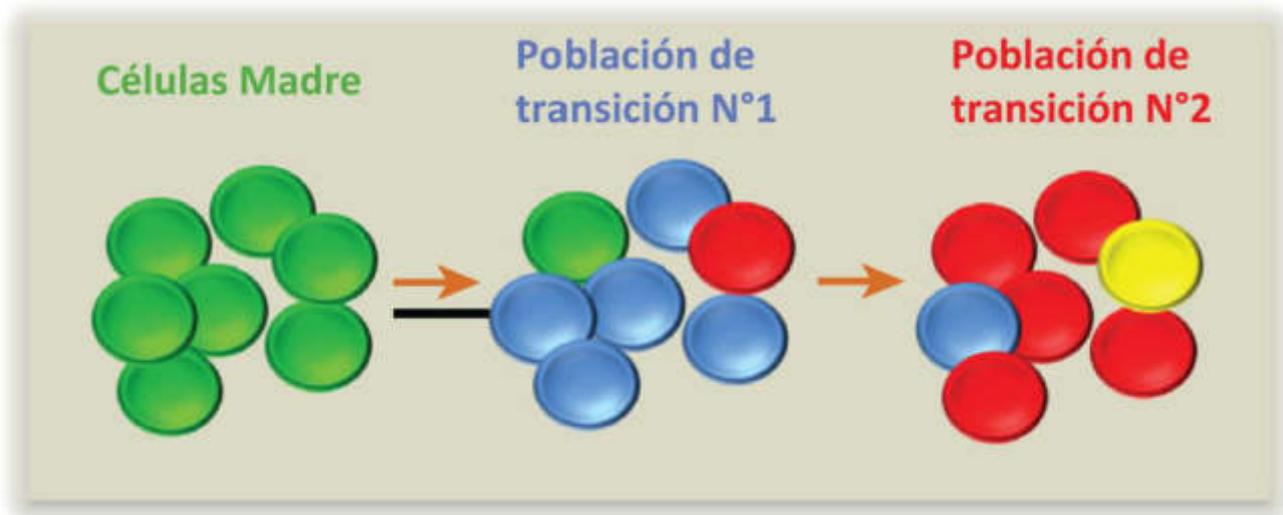


Figura 9: Imagen adaptada¹³. Esquema de la diferenciación asincrónica. La población de células verdes (izquierda) recibe el estímulo de diferenciación y comienzan a cambiar su identidad hacia las células azules, rojas y amarillas, en orden cronológico. Sin embargo, esta transición no ocurre homogéneamente en toda la población, sino que existe variabilidad en la velocidad de diferenciación de cada célula.

Acoplamiento con la densidad celular

La diferenciación es un proceso que rompe simetrías. El cuerpo del ratón comienza como un conjunto de células idénticas y éstas se van

diferenciando para formar tejidos u órganos. Esto implica que en el desarrollo se rompe la homogeneidad celular para dar paso a patrones espaciales de identidad celular^{15,16,17,18}. Si bien no se ha observado que las células en diferenciación de naïve a primed formen patrones espaciales, sí podríamos esperar que las velocidades de diferenciación o la proporción de células diferenciadas dependan de la posición en la colonia. Por otro lado, en trabajos similares^{18,19,20} se observó que las células indiferenciadas fomentan el estado pluripotente, creando un feedback positivo que mantiene mejor el estado naïve en regiones de alta densidad celular.

Esto ofrece otra explicación para la asincronía observada en la diferenciación, que consiste en lo siguiente. Teniendo en cuenta que cada célula naïve sintetiza factores de transcripción que mantienen la pluripotencia, existe una competencia de señales opuestas entre las de estos factores y las del medio de diferenciación¹³. Si agregamos que hay diferencias locales de densidad celular dentro de una colonia, podemos hipotetizar que, en los sectores más densos, las señales de pluripotencia continúan con ventaja por más tiempo, y que solo comienza la diferenciación desde la periferia de estos sectores hacia el centro. Esta será otra de las hipótesis de este capítulo.

Motivación, Estado del Arte y Objetivo

Motivación General

Uno de los objetivos a largo plazo del estudio en células madre es lograr controlar el mapa completo de todas las diferenciaciones, desde el óvulo fecundado hacia cualquier célula adulta especializada. Este paso abriría las puertas para la síntesis artificial de tejidos y órganos enteros, tecnologías que hoy son ciencia ficción, aunque ya existen casos recientes de impresión 3D de órganos^{21,22}. Estas impresoras utilizan tintas con células madre de diferentes identidades y medios de cultivo. La idea consiste en ubicar precisamente cada tipo celular con su medio en su lugar, de manera de reproducir el desarrollo natural del órgano deseado.

La gran promesa de este campo consiste en la medicina de reemplazos. De lograrse la impresión de órganos, bastaría con trasplantar un órgano dañado por uno nuevo, incluso desarrollado a partir de células del propio paciente, lo que minimizaría el rechazo inmunológico.

Motivación Particular

En la diferenciación naïve a primed, aún se ignora cómo se acopla la salida del estado naïve con el ciclo celular. Algunas preguntas sin responder son: ¿es necesaria la mitosis para esta diferenciación? ¿o puede una misma célula cambiar su identidad? ¿las células son igualmente sensibles a las señales de diferenciación a lo largo del ciclo? ¿o en alguna fase en particular pueden detectarlas y en el resto las ignoran?

Una hipótesis de este capítulo es que el proceso de diferenciación requiere de la mitosis. La idea es que, en la división celular, las células hijas heredan de la madre el programa de diferenciación al estado primed, estando ya predeterminadas al nacer. En esta situación, la célula progenitora debería tomar la decisión en algún momento de su ciclo: si sus hijas mantendrán su misma identidad naïve o comenzarán a diferenciarse.

Por otra parte, también se ignora cómo se acopla la densidad celular con la salida del estado naïve. Algunas preguntas sin responder son: ¿tardan más o es más costoso diferenciar sectores más densos de células?, dado que

la diferenciación es asincrónica, ¿existen patrones espaciales de diferenciación?

La segunda hipótesis de este capítulo consiste en que la síntesis de factores de transcripción que mantienen la pluripotencia compite con las señales de diferenciación del medio de cultivo, ralentizando el proceso de diferenciación en sectores de mayor densidad.

Objetivo

El objetivo general de este capítulo es descubrir si existe algún acoplamiento entre la diferenciación naïve a primed y el ciclo celular. En particular, detectar si hay algún factor hereditario en la mitosis que lo controle. Para esto, un objetivo anterior es caracterizar el largo del ciclo y sus fases durante la diferenciación.

Por otra parte, también buscamos determinar si existen patrones espaciales de diferenciación dentro de las colonias o si éstas son homogéneas.

Esquema Experimental

Introducción

La herramienta principal que utilizamos en este capítulo es la filmación microscópica de *time-lapse* de células en cultivo en dos condiciones: con medios **Indif** y **Dif**, y con marcadores fluorescentes de las fases del ciclo celular. El objetivo es poder medir y comparar la progresión a lo largo de ciclo durante el proceso de diferenciación contra la proliferación sin cambio de identidad.

Reporteros fluorescentes del ciclo celular: Sistema Fucci

El sistema Fucci²³ (*Fluorescent Ubiquitination-Based Cell Cycle Indicator*) consiste en una línea celular genéticamente modificada para expresar proteínas fluorescentes fusionadas (unidas covalentemente) a proteínas que se expresan específicamente en algunas fases del ciclo celular. De esta forma, las células fluorescerán en colores diferentes indicando fases específicas.

En la versión de Fucci utilizada, las células sintetizan las proteínas Cdt1 y Geminina humanas que se expresan específicamente en las fases G1 la primera y en S, G2 y M la segunda. Estas proteínas están fusionadas, respectivamente, a las proteínas mCherry y mVenus, que fluorescen en rojo y verde. De esta manera, las células rojas se encuentran en fase G1 mientras que las verdes, en S, G2 o M. (Figura 10)

Marcador nuclear y segmentación

Además de los marcadores del ciclo celular, a la línea celular se le introdujo otra modificación genética para que la histona H2B estuviera fusionada a la proteína mCelurean, que fluoresce en azul. De esta forma, las histonas unidas al ADN fluorescen en azul, y dado que el ADN se encuentra confinado al núcleo celular, solamente el núcleo brillará en azul. Esto es lo que permite marcar los núcleos de cada célula, etiquetarlas y seguirlas en los videos.

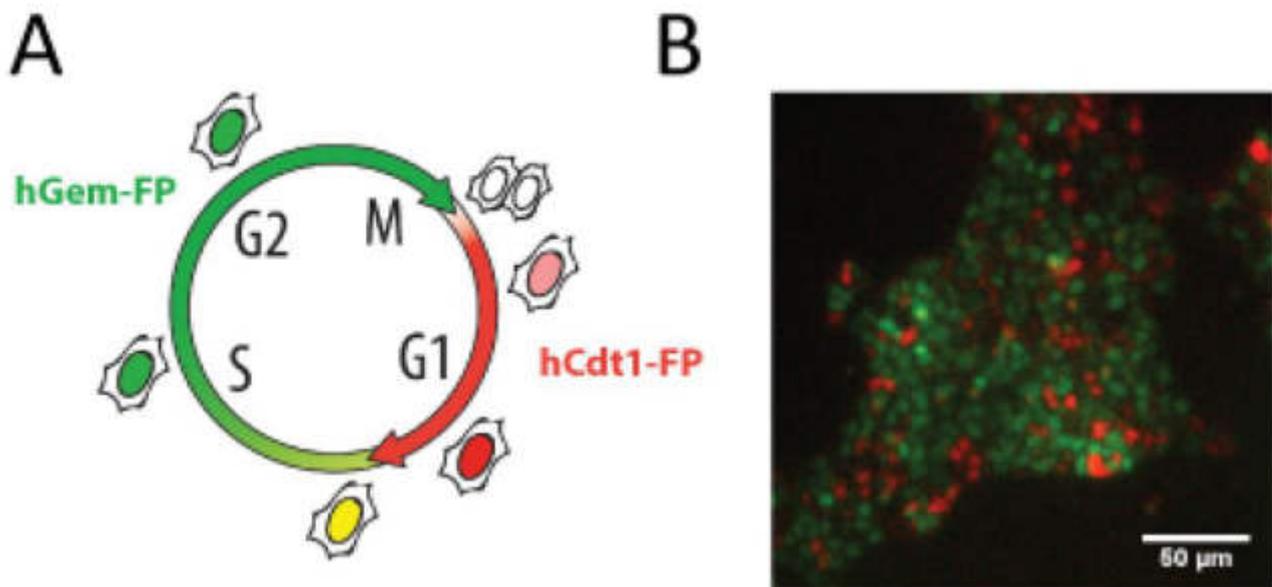


Figura 10: Imágenes adaptadas^{14,23}. **A:** Esquema del ciclo celular y sus fases. El color de las flechas del círculo indica el color del brillo fluorescente del sistema Fucci. **B:** Ejemplo de una foto de mESC transgénicas para el sistema Fucci en estado naïve, pueden reconocerse las células en la fase G1 (rojas) o en alguna de las fases S, G2 o M (verdes).

Microscopía de Time-Lapse

Las filmaciones tipo *Time-Lapse* se utilizan para filmar procesos cuyas dinámicas son lentas, y a simple vista parecen estáticos. Un ejemplo es el crecimiento de una planta². Una cámara común toma alrededor de 24 fotos o *frames* por segundo, que luego se reproducen a la misma frecuencia. Un video no es otra cosa que la sucesión ordenada de fotos, y mientras mayor es la cantidad de frames por segundo, más continuos se verán los movimientos en el video. Sin embargo, para dinámicas lentas, como son el movimiento de las nubes, de los astros y también el de las mESC, no es necesario tomar fotos a alta frecuencia, dado que no hay movimientos apreciables en el orden de los segundos.

En las filmaciones realizadas, se tomó una foto cada 15 minutos, frecuencia suficientemente alta para tener buena resolución del movimiento celular y suficientemente baja como para no sobreexponer a la luz las muestras y *bleechear* (romper con luz) los fluoróscoros.

El microscopio utilizado para las filmaciones se encuentra en la Universidad de Rockefeller de Nueva York. Es un confocal spinning-disc que

² [Video del crecimiento de una planta de porotos](#)

permite mantener las condiciones atmosféricas óptimas para el crecimiento celular (37°C y 5% CO₂) durante todo el tiempo de filmación. Además, tiene lugar para 3 placas de cultivo y la posición del campo visual de la cámara automatizada. Esto último permite hacer un barrido en muchos campos visuales previamente definidos y poder volver siempre al mismo exacto lugar. De esta forma, se señalan muchos campos visuales con una colonia celular cada uno, y se aprovechan los 15 minutos entre frames para fotografiar muchas colonias diferentes. Al final, en una sola sesión de filmación se registra el crecimiento de entre 20 y 30 colonias. Esto fue crucial para contar con un número alto y estadísticamente significativo de colonias y células.

Se hicieron dos sesiones de filmación de 45hrs cada una. Grabando videos de proliferación de 32 colonias en estado naïve, con medio **Indif** y otras 20 en diferenciación al estado primed, con medio **Dif** colocado al comienzo de la filmación.

Extracción de Datos de los Videos

Segmentación, intensidad de fluorescencia y relaciones familiares

Lo primero que se hizo con los videos crudos fue utilizar el canal azul del marcador nuclear fluorescente H2B-mCelurean para segmentar las células. En [este enlace](#), encontrarán un video ejemplo donde se muestra, primero, la filmación en crudo, segundo, la segmentación celular y, finalmente, el seguimiento.

Esto se logró automatizar en parte con el plugin Lineage Tracker²⁴ del software de análisis de imágenes ImageJ, que reconoce objetos en imágenes usando algoritmos de machine learning. Este programa recibe los videos completos y devuelve las fotos de cada frame segmentadas (Figura 11). Esto es, crea un archivo para cada video donde indica qué conjunto de píxeles de cada foto pertenece a cada célula, asignándole a todas las células del video una identidad única. Luego, para obtener la intensidad de fluorescencia de cada célula en función del tiempo, se mide el valor de intensidad del conjunto de píxeles correspondiente a cada célula en cada frame de cada canal (Figura 12).

Por otro lado, el mismo software detecta las divisiones celulares, y devuelve las relaciones de parentesco y los linajes completos. Con estas relaciones confeccionamos los dendrogramas de la Figura 13. Cada línea horizontal del dendrograma se corresponde con una célula, y su longitud es

proporcional a la duración del ciclo celular ($CC-L$). Esta duración la definimos como el intervalo temporal entre los frames inicial y final donde vive. La parte en rojo de las líneas se corresponde con la fase G1. Para esta tomamos el intervalo temporal entre el frame inicial donde vive y el tiempo del pico de máxima intensidad de fluorescencia del canal de mCherry asociado ($G1-L$). Por último, la duración del resto de las fases S, G2 y M ($SG2M-L$), la definimos como la resta entre $CC-L$ y $G1-L$.

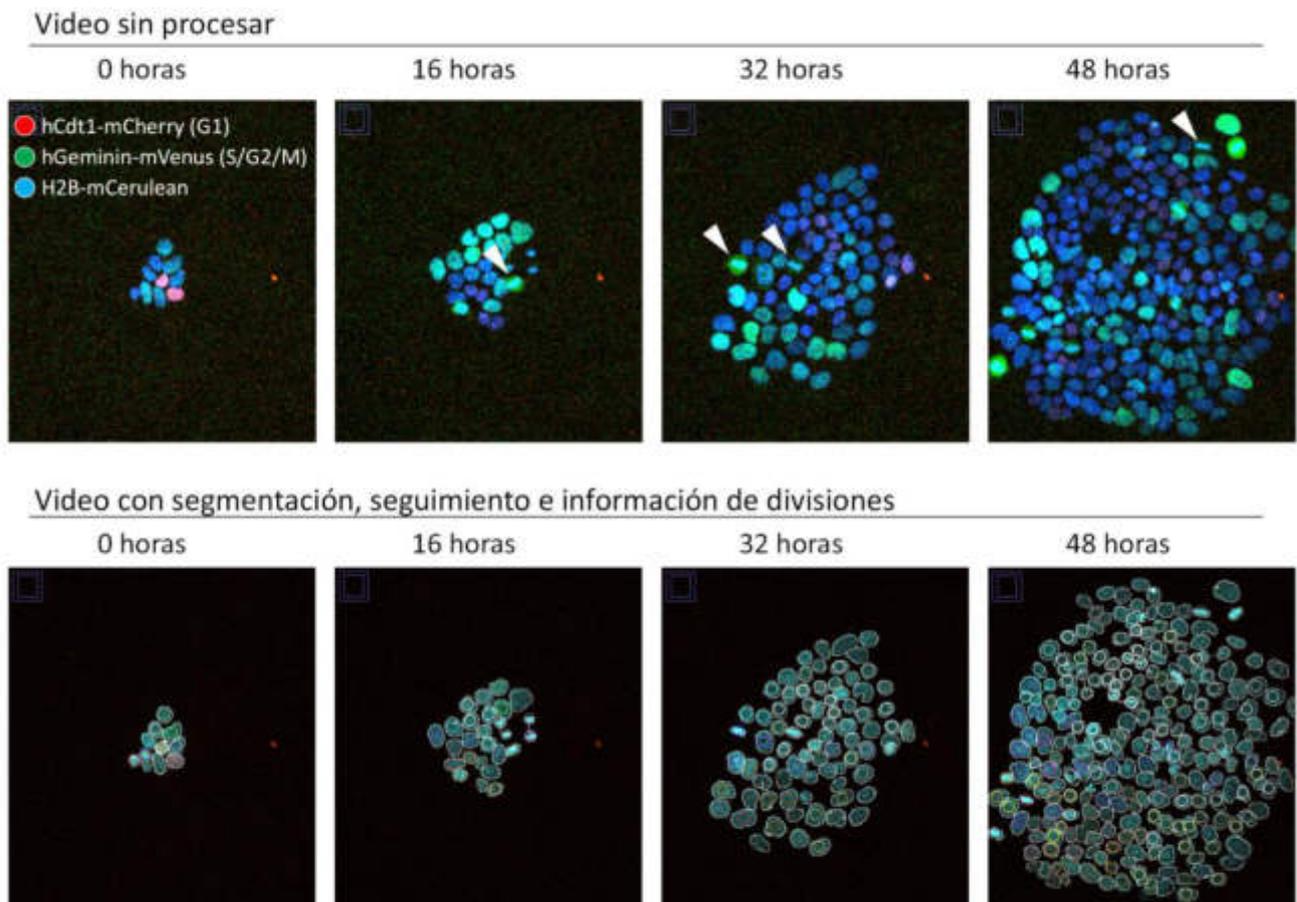


Figura 11: Imagen adaptada¹⁴. Fila superior: Fotos o frames de un video con el sistema Fucci + el marcador nuclear fluorescente H2B-mCerulean. Podemos reconocer las células en G1 (rojas), en S, G2 o M (verdes) y sus núcleos (azul). Fila inferior: Células segmentadas utilizando el marcador nuclear y el plugin Lineage Tracker²⁴ de ImageJ.

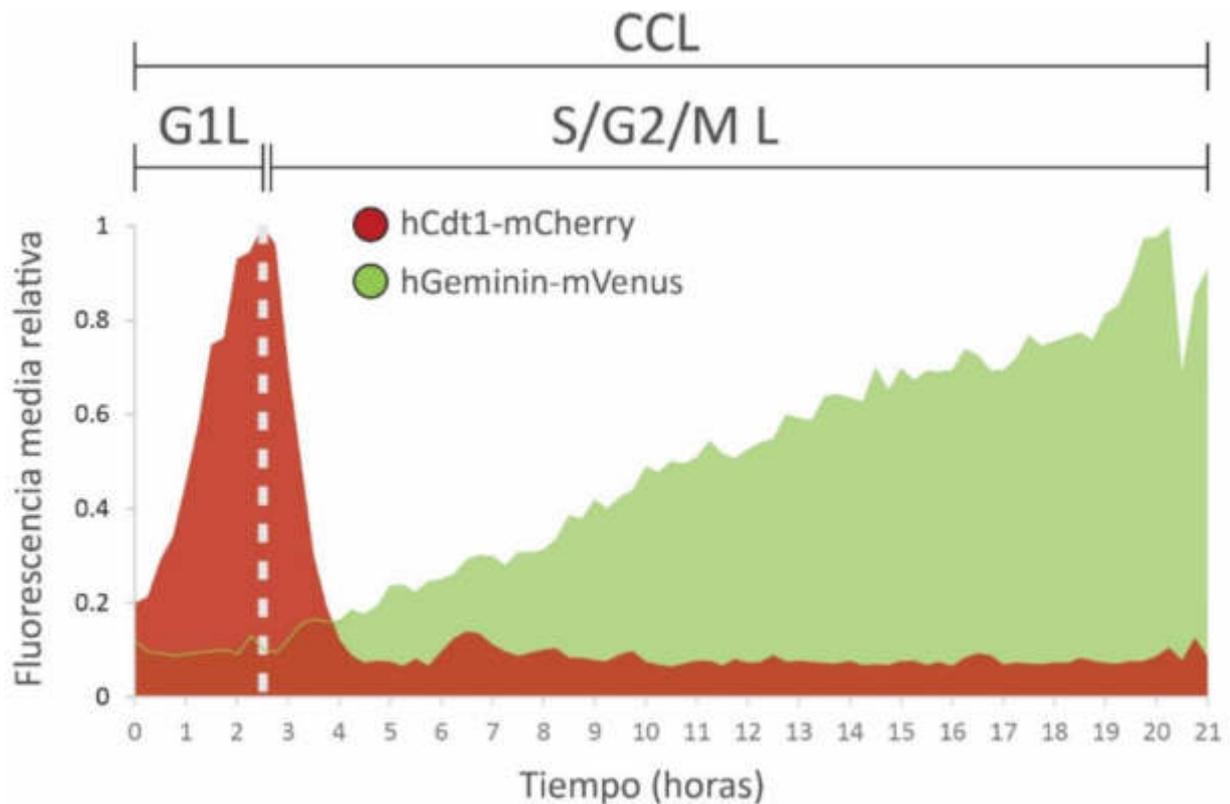


Figura 12: Imagen adaptada¹⁴. Ejemplo de la intensidad de fluorescencia de los canales verde y rojo para una célula de un video. La duración del ciclo total la definimos como el tiempo entre los frames inicial y final. El inicial es el primero en que se pueden separar los núcleos de la célula objetivo y su hermana y el final donde se pueden separar los núcleos de las hijas. La duración de G1 la definimos desde el inicio de la vida de la célula hasta el pico de intensidad del canal rojo. La duración de las fases S, G2 y M se define como el tiempo complementario entre el pico de G1 y el del frame final.

Base de datos

Gracias al plug in Lineage Tracker del programa ImageJ y los procesamientos explicados en la sección anterior, construimos la base de datos que muestra la Tabla 1. La misma incluye datos de 516 células en estado naïve y 623 en diferenciación al estado primed, repartidas en 2 réplicas con 72 linajes en 11 colonias y 50 linajes en 12 colonias respectivamente, como indican la Tabla 2 y la Tabla 3.

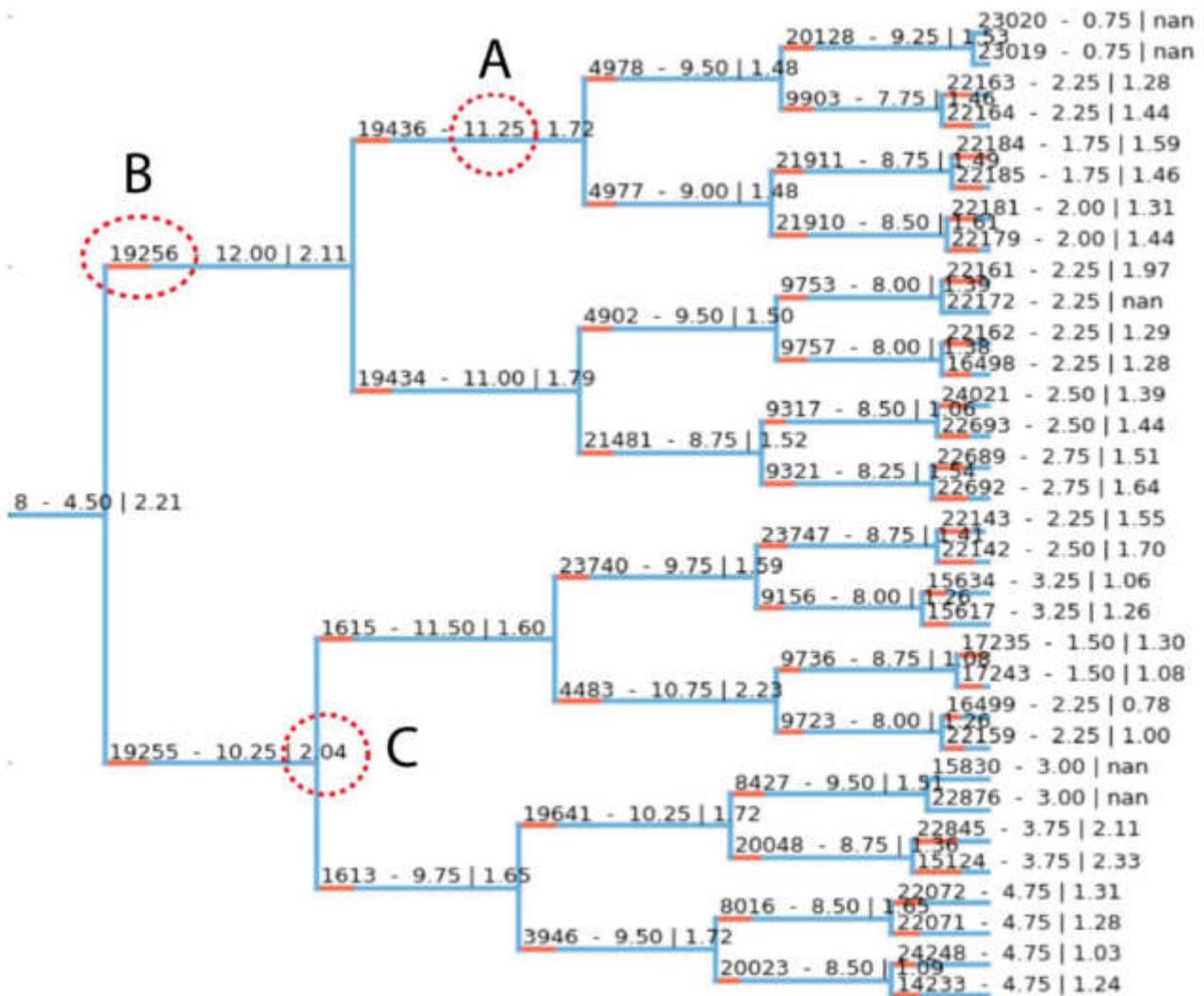


Figura 13: Imagen adaptada¹⁴. Gráfico que ilustra las relaciones familiares de todas las células en diferenciación de un linaje en un video (dendrograma). Cada línea horizontal representa una célula. Cada una tiene un código de identidad único (B). La primera de la izquierda se corresponde a una de las células que vivían al comenzar la filmación (1ra generación). Las últimas líneas a la derecha representan a las 32 células de la 6ta generación. El largo de las líneas es proporcional al largo del ciclo celular (A) y la parte roja a la duración de la fase G1 (C).

Es importante destacar que el software utilizado al momento del análisis no segmentaba de manera perfecta las células y fue necesario validar y corregir de manera manual el procedimiento. Este paso fue el limitante y es lo que impidió sumar la totalidad de las células filmadas al análisis.

Variable	Significado	Ejemplo
ID Celular	N° ID de cada célula	19436
Compleitud	Booleano (Verdadero o Falso). V si todo el ciclo está en el video y F si no.	Verdadero
Generación	Número que indica la generación a la que pertenece, siendo 1 si la célula vive al comienzo del video	3
Hermana	N° ID de célula hermana	19434
Hijas	N° ID de células hijas	{4978;4977}
Madre	N° ID de célula madre	19256
Primas	N° ID de células primas	{1615;1613}
CC-L	Duración del ciclo celular en horas	11.25
G1-L	Duración de la fase G1 en horas	1.72
SG2M-L	Duración de las fases S, G2 y M en horas	9.53
Posición	Lista de vectores (x,y) en función de los frames que indican la posición del centro de los núcleos para cada frame.	{(4386;1687), (4380;1699),..., (4216;1802)}
Frames	Lista de números de frames en los que vive la célula, siendo 1 el primer frame del video	{67;68;69;...;112}
Área Nuclear	Lista de Áreas (en píxeles) del núcleo en función de los frames.	{420;432;454;...;846}

Tabla 1: Base de datos extraída de los videos. Todas las variables indicadas fueron medidas para cada una de las más de mil células en las condiciones que muestra la Tabla 2.

Medio	Indif	Dif
Grupo		
Colonias	11	12
Linajes	72	50
N° de Células	1387	1601

Tabla 2: Cantidades de colonias, linajes y células totales filmadas para cada condición.

Réplicas	Wells	Colonias	Linajes	Células completas
1	1	3	21	115
	2	3	17	101
	3	6	25	293
2	1	3	23	228
	2	2	11	95
	3	6	25	411

Tabla 3: Distribuciones de células en linajes, colonias, wells y réplicas experimentales.

Análisis de Datos

Proliferación de las Colonias Celulares

La primera observación que surgió de observar los videos consiste en que las colonias de células en medio **Dif** tendían a crecer más rápido en tamaño y en cantidad de células que las colonias en medio **Indif**. Para cuantificar esta diferencia, contamos la cantidad de células promedio por colonia para cada frame a partir de la base de datos. Efectivamente, podemos ver en La Figura 14 que la proliferación aumenta con la diferenciación.

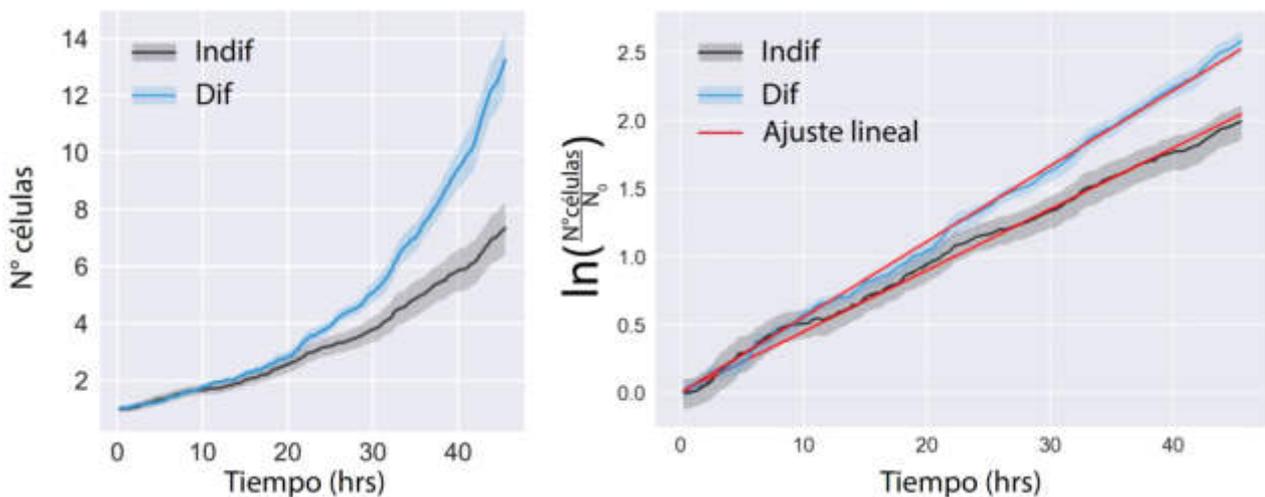


Figura 14: Promedio de la cantidad de células por colonia en función del tiempo, separadas en las cultivadas con medios **Indif** y **Dif**. Gráfico lineal (izquierda) y logarítmico (derecha) donde se agrega el ajuste lineal (ec. 2) para medir $CC-L$.

Teniendo en cuenta que estas células se reproducen por mitosis, uno de los modelos más simples que podemos plantear para el crecimiento celular está dado por la ec. 1, donde N_0 sería el número inicial de células, t el tiempo y $CC-L$ la duración del ciclo. Esta ecuación no es más que la descripción matemática de un crecimiento exponencial donde las células se duplican cada un tiempo equivalente a $CC-L$.

$$N^\circ \text{ células}(t) = N_0 2^{\frac{t}{CC-L}}$$

1

Podemos linealizar esta ecuación con el fin de ajustarlos usando la ec. 2 y medir $CC-L$.

$$\ln\left(\frac{N^{\circ}\text{células}(t)}{N_0}\right) = \frac{\ln(2)}{CC-L} t \quad 2$$

Los valores obtenidos fueron $(15.4 \pm 1.2)hrs$ y $(12.5 \pm 0.5)hrs$ para las condiciones **Indif** y **Dif** respectivamente. Esto es coherente con la mayor proliferación observada en diferenciación, dado que, si las células atraviesan su ciclo más rápido, también se dividen más rápido y proliferan más. Sin embargo, es una medida general que viene de un modelo muy sencillo y que no proporciona información de la distribución de $CC-L$ a nivel de células únicas. Estas distribuciones pueden obtenerse de los datos de los time-lapse y las presentamos en la siguiente sección.

Distribuciones del Ciclo Celular y sus fases

La Figura 15 muestra las distribuciones de $CC-L$, $G1-L$ y $SG2M-L$, segregados entre las células **Indif** y **Dif**. Puede observarse que el ciclo se acorta con la diferenciación y que además ambas fases G1 y SG2M se acortan.

Por otra parte, para comparar la evolución de este acortamiento del ciclo, graficamos su duración en cada célula en función del tiempo de nacimiento en el video (Figura 16). En estos gráficos, podemos apreciar varios resultados. El primero, que el $CC-L$ se mantiene en las células **Indif**, lo que era esperable dado que no están cambiando su identidad. El segundo, que el acortamiento en el $CC-L$ de las células **Dif** parece ser progresivo generación a generación.

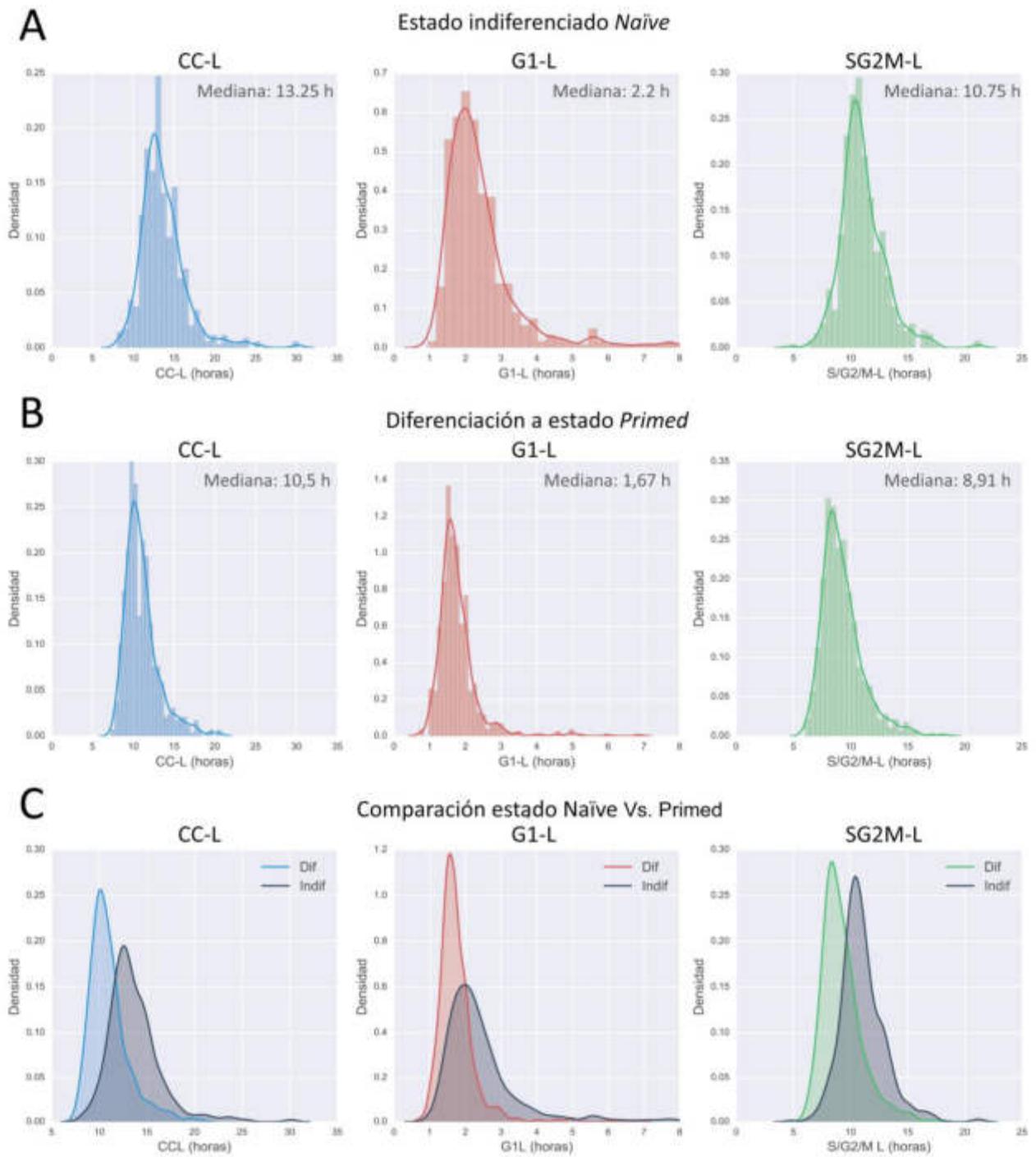


Figura 15: Imagen adaptada¹⁴. Distribuciones de la duración del ciclo celular (CC-L) y sus fases (G1-L y SG2M-L), para células indiferenciadas en medio **Indif** (A), en diferenciación con medio **Dif** (B) y comparando ambas condiciones (C).

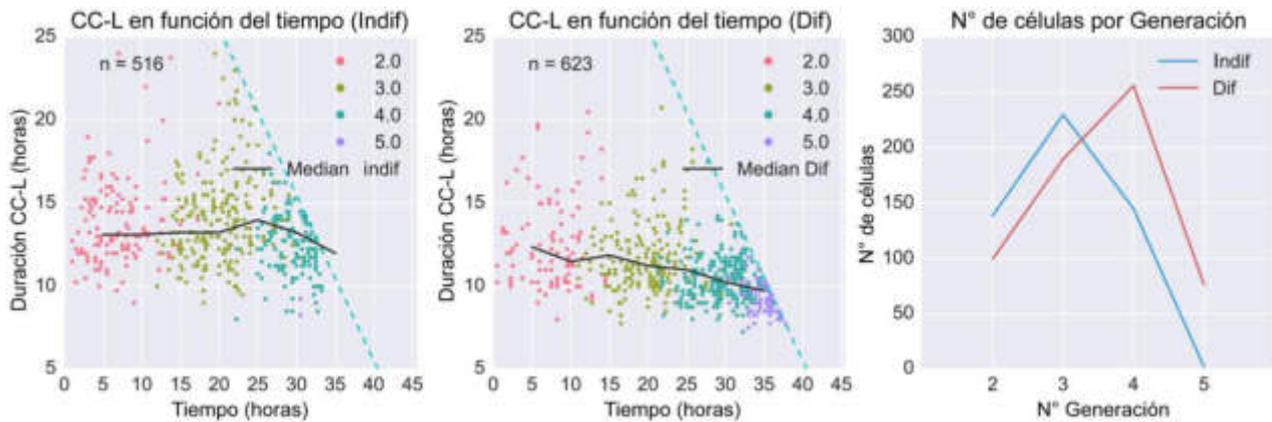


Figura 16: Imagen adaptada¹⁴. *CC-L en función del tiempo de nacimiento de cada célula, en condición **Indif** (izquierda) y **Dif** (centro). A la derecha, N° de células completas de cada generación. Podemos apreciar que, si bien este número sube debido a la proliferación, a partir de la 3ra (**Indif**) y 4ta (**Dif**) generación disminuye porque no llegan a completar el ciclo antes de que termine la filmación.*

Estos resultados preliminares indican que el proceso de diferenciación hace que las células atraviesen su ciclo y proliferen cada vez más rápido, tanto que incluso llegan a haber células de la 5ta generación en el tiempo de filmación en la condición **Dif**. Sin embargo, para cuantificar este acortamiento es indispensable tener en cuenta que, al término de las 45 horas del video, muchas células permanecen vivas y no quedan datos de cuándo se dividieron, impidiendo medir su *CC-L*. Esta cantidad de células no es despreciable dado que crecen de manera exponencial, y siempre hacia el final de los videos se espera que haya tantas como la sumatoria de todas las que vivieron a lo largo del mismo. Esto introduce un sesgo en la medición del *CC-L*: si las células viven menos tiempo, entonces es más probable ver su vida completa en el video que si vivieran más. Volviendo a las distribuciones de la Figura 15, implica que estamos dejando afuera más células de mayor longevidad que de menor, y entonces no sería correcto medir el *CC-L* (ni tampoco *G1-L* o *SG2M-L*) a partir de estas distribuciones. Para resolver el problema del sesgo aplicamos el método de Kaplan-Meier.

Estimación precisa del Ciclo Celular y sus fases: Método de Kaplan-Meier

Bases del método

El método de Kaplan-Meier es un análisis estadístico originado en la medicina²⁵. Tiene como objetivo calcular la esperanza de vida de una población donde el seguimiento de cada individuo se interrumpe antes de su muerte. Esta condición se cumple en muchos ensayos clínicos donde, o bien las enfermedades son largas, o bien los tratamientos exitosos y se concluye el estudio antes de la muerte.

El método consiste en calcular empíricamente la probabilidad de supervivencia de un paciente de un momento a otro para finalmente calcular la productoria y obtener la probabilidad de supervivencia en función del tiempo de vida. En nuestro caso, cada momento será un frame de video y la probabilidad de supervivencia entre un frame y el siguiente será análoga a la de no entrar en mitosis ($p(\text{no mitosis}, t)$), cuando una célula deja de vivir. Ésta se calcula a partir de la ec. 3, donde $n^\circ \text{celulas}(t)$ es el número de células aún sin hacer mitosis luego de un tiempo t , y $n^\circ \text{celulas mitóticas}(t)$ es el número de células que se dividen en este tiempo.

$$p(\text{no mitosis}, t) = \frac{n^\circ \text{celulas}(t) - n^\circ \text{celulas mitóticas}(t)}{n^\circ \text{celulas}(t)} \quad 3$$

Para calcular la probabilidad de supervivencia desde el inicio del video en adelante ($S(t)$) se hace la productoria de la ec. 4. Este paso asume que las mitosis de dos células distintas son independientes entre sí, hipótesis que no parece arriesgada si aplicamos el estudio sólo a células de la misma generación, donde no puede haber herencias entre sí.

$$S(t) = \prod_{t_i < t} \frac{n^\circ \text{celulas}(t) - n^\circ \text{celulas mitóticas}(t)}{n^\circ \text{celulas}(t)} \quad 4$$

Finalmente, $S(t)$ es la probabilidad acumulada de que las células no hagan mitosis hasta el tiempo t . Todas estas curvas comienzan en 1, dado que a tiempos cortos ninguna célula hace mitosis. Luego, son monótonamente decrecientes, dado que las células comienzan a dividirse y la probabilidad solo puede bajar. La ventaja de este método consiste en que el cálculo de estas probabilidades tiene en cuenta las células vivas cuya mitosis no se llega a observar al término de los videos. Estos datos aparecen en la variable $n^\circ \text{celulas}(t)$, donde se agrega información sobre cuántas células sí sabemos que al término del video aún no habían hecho mitosis y, por ende, vivieron más de cierto tiempo. Para comparar, si solamente tomáramos un promedio del $CC-L$ de todas las células, sí o sí deberíamos dejar afuera

del cálculo aquellas que al término de los videos seguían viviendo, al no tener el dato de su $CC-L$. Dado que células con menor $CC-L$ tienen más chances de hacer mitosis dentro del tiempo de filmación, tendríamos el sesgo.

Resultados

Aplicamos el método de Kaplan-Meier a todas las células de la base de datos, separándolas por generaciones y por condición **Indif** y **Dif**. Las curvas de $S(t)$ pueden observarse en la Figura 17, donde se aprecia, en la condición **Indif**, que no hay diferencias apreciables entre las curvas de distintas generaciones. Por otro lado, para la condición **Dif**, sí aparece un corrimiento de las curvas hacia tiempos más cortos a medida que aumentan las generaciones. Esto es coherente con las medidas preliminares del $CC-L$, con la ventaja de que ahora podemos estimar la mediana del $CC-L$ para cada generación. La misma es simplemente el tiempo donde $S(t) = 0.5$, que es cuando la probabilidad de no entrar en mitosis es del 50%, esto es, cuando la mitad de la población se dividirá. Estas medianas están graficadas en la Figura 18, tanto para el ciclo como sus fases.

Finalmente, encontramos que $G1-L$ y $SG2M-L$ mantienen una relación aproximadamente proporcional en el acortamiento del ciclo. A través de un ajuste lineal de las medianas, obtuvimos que $G1-L/SG2M-L = (0.28 \pm 0.05)$ (Figura 18).

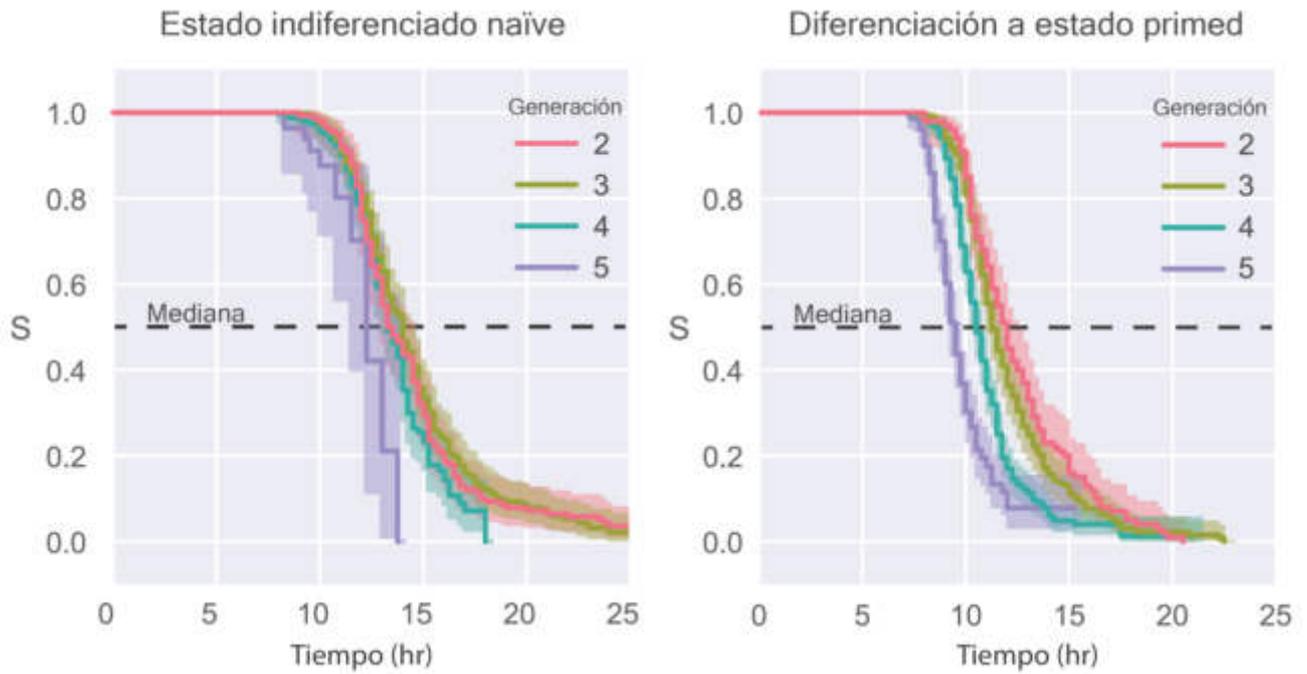


Figura 17: Probabilidad acumulada de que las células no se dividan (S) en función del tiempo de vida para las condiciones indiferenciada (izquierda) y en diferenciación (derecha). Se puede apreciar cómo las probabilidades son iguales y las curvas se solapan en el caso indiferenciado mientras que las probabilidades se acortan con las generaciones en el caso de diferenciación. El tiempo donde cada curva cruza la línea punteada es la mediana que usamos para los análisis de la Figura 18.

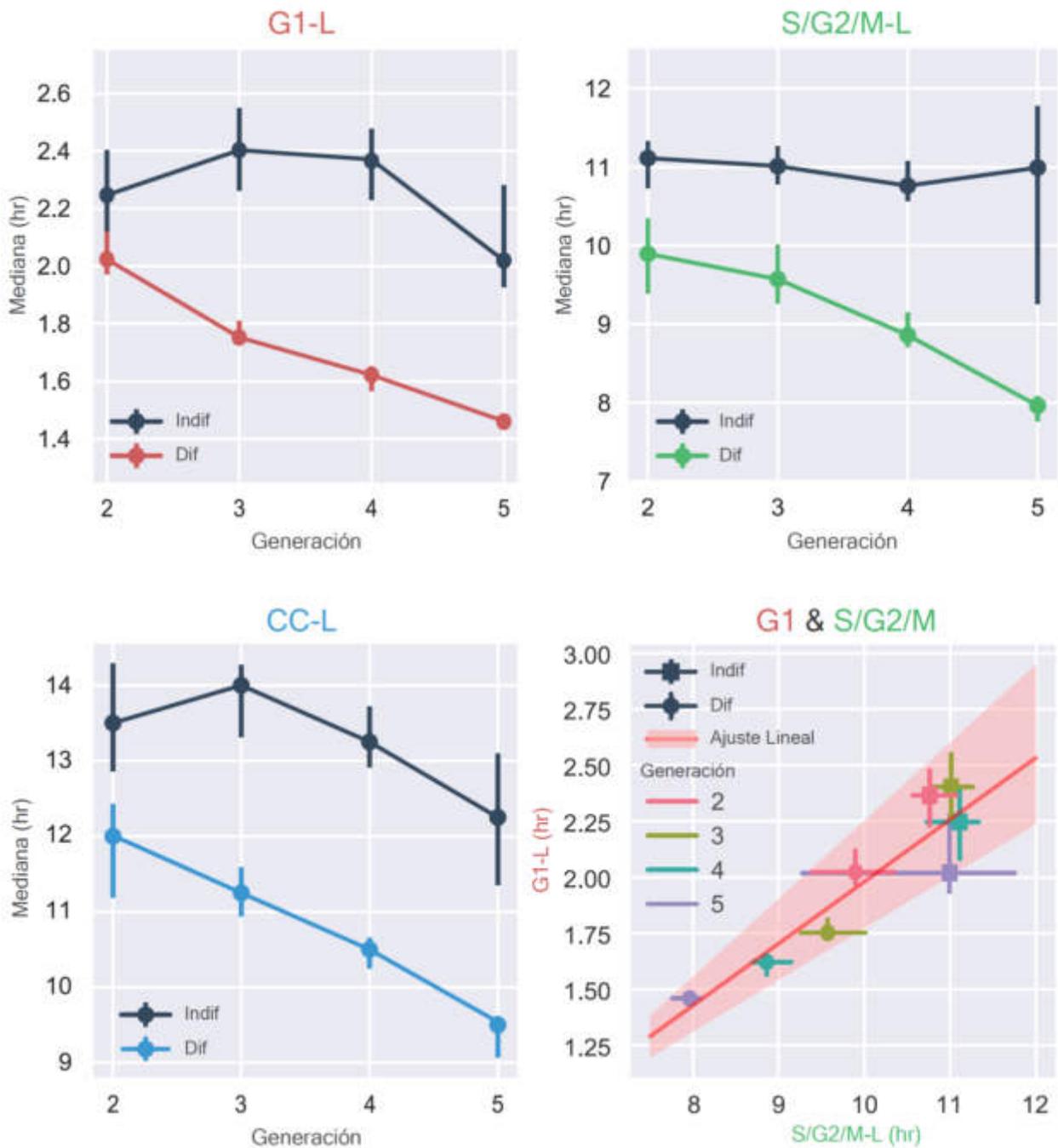


Figura 18: Duraciones del ciclo celular (abajo izquierda) y sus fases G1 (arriba izquierda) y S, G2 y M (arriba derecha), definidas a partir de la mediana por el método de Kaplan-Meier. Podemos apreciar que el ciclo se acelera y que tanto G1 como S, G2, M se acortan en la diferenciación, manteniendo una relación aproximadamente lineal (abajo derecha).

Herencia

Un aspecto interesante encontrado en las duraciones del ciclo y sus fases es la gran variabilidad células a célula. El $CC-L$, por ejemplo, puede durar entre 10 y 18 hrs, con células que llegan a vivir más de 20hrs. Una pregunta válida es si esta variabilidad se origina en la estocasticidad intrínseca de los procesos biológicos o si existe algún factor hereditario que determina si el ciclo y sus fases deben durar más o menos. Los primeros análisis dirigidos a responderla consisten en calcular las correlaciones entre las variables cruzadas $CC-L$, $G1-L$ y $SG2M-L$ de células con parentescos familiares, aprovechando las relaciones de linaje construidas.

Correlaciones del Ciclo y sus Fases

Los parentescos analizados fueron entre células hermanas, primas y madre-hija. El coeficiente calculado fue el de Spearman, que tiene en cuenta solo las relaciones de orden entre las variables. Un factor importante para este cálculo consiste en que el $CC-L$ va disminuyendo con las generaciones en la condición **Dif**. En la Figura 19, se aprecia cómo hay un corrimiento general de los datos hacia valores menores a medida que aumentan las generaciones. Esta correlación es debida al acortamiento del ciclo que acompaña la diferenciación, y no está dada por la relación de parentesco. Para descartarla, redefinimos cada variable restando la media generacional. Por otro lado, también existe cierta variabilidad intrínseca a cada colonia. Esto es, los $CC-L$ medios de cada colonia no son exactamente iguales, lo que también podría afectar el cálculo. Para evitarlo, usamos la estrategia de Bootstrap²⁶. Consiste en elegir aleatoriamente la misma cantidad de datos de cada conjunto, calcular el coeficiente de correlación, y repetir muchas veces esto. Al final, se obtiene una distribución de coeficientes (uno por cada repetición) de donde podemos estimar la media y el intervalo de confianza. La ventaja, en este caso, es que tomamos en cada repetición la misma cantidad de células de cada generación y de cada colonia, evitando cualquier sesgo hacia colonias de ciclos más largos o cortos.

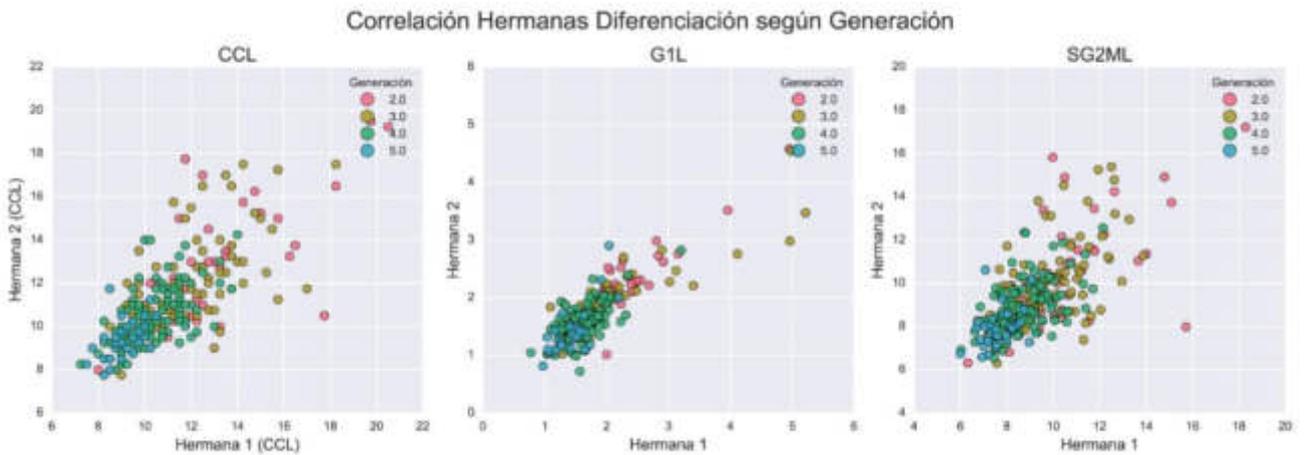


Figura 19: Imagen adaptada¹⁴. Correlación de las duraciones del ciclo y sus fases entre células hermanas, agrupadas por generación en la condición **Dif**. Podemos apreciar que, con el acortamiento del ciclo, los datos de las generaciones más grandes se corren hacia abajo a la derecha, y esto genera una correlación general independiente del linaje.

Herманas

La Figura 20 muestra los coeficientes de correlación de Spearman entre células hermanas, obtenidos con la estrategia de Bootstrap. Podemos ver que existe una alta correlación de $G1-L$ entre células hermanas, y algo más baja pero aún significativa de $SG2M-L$. Consecuentemente, la correlación de $CC-L$ es intermedia y también significativa. Por último, no parece haber dependencia con la condición de diferenciación.

Estos resultados indican que células hermanas tienen ciclos celulares parecidos, e implican que deben existir factores determinantes heredados de la célula progenitora. Teniendo esto en cuenta, calculamos la correlación entre células madre-hija.

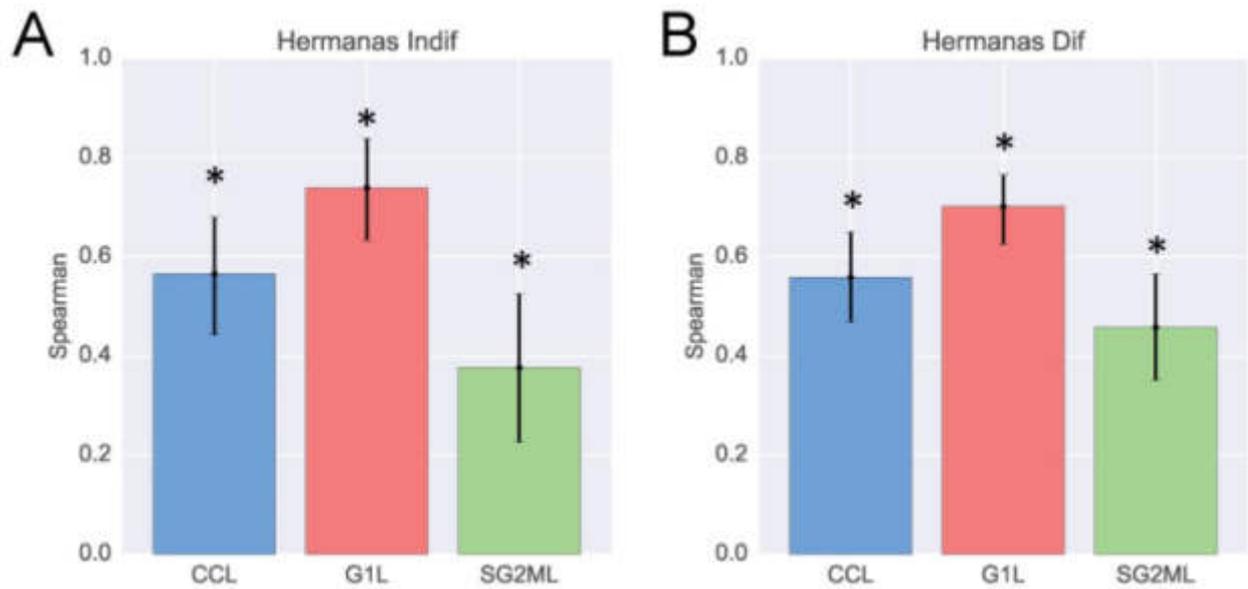


Figura 20: Imagen adaptada¹⁴. Coeficiente de correlación de Spearman entre células hermanas, restando las medianas generacionales y usando la estrategia de Bootstrap para elegir cantidades similares de células de cada colonia y linaje. En ambas condiciones (**A: Indif** y **B: Dif**), podemos apreciar que hay correlación positiva en el ciclo y sus fases, siendo más alta la correlación de G1.

Madre - Hija

La Figura 21 muestra los coeficientes de correlación de Spearman entre células madre-hija, obtenidos con la estrategia de Bootstrap. A diferencia de la correlación entre hermanas, en este caso no hubo correlaciones altas ni para el *CC-L* ni sus fases. Incluso ni fueron significativas en la condición **Indif**.

Estos resultados indican que no se puede determinar la duración del ciclo celular de una célula a partir de la duración del de su progenitora. Cabe mencionar que no invalida el resultado anterior. Si bien no será la duración del ciclo, debe existir otro factor hereditario para que haya tanta correlación entre células hermanas.

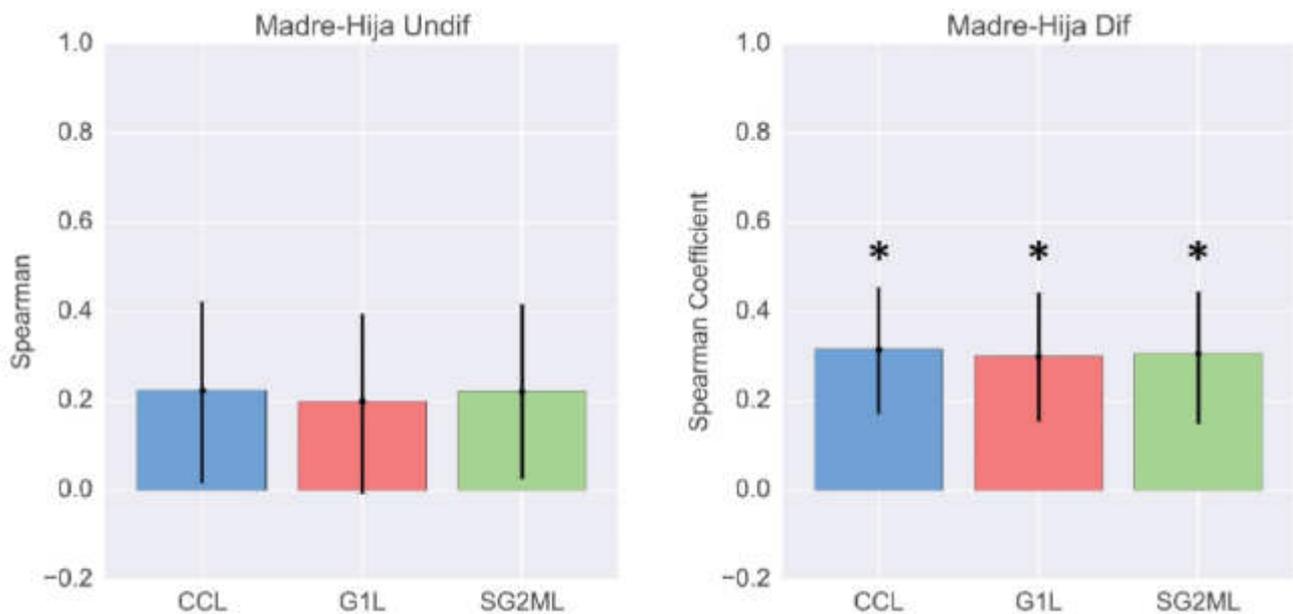


Figura 21: Imagen adaptada¹⁴. Coeficiente de correlación de Spearman entre células madres e hijas, restando las medianas generacionales y usando la estrategia de Bootstrap para elegir cantidades similares de células de cada colonia y linaje. En ambas condiciones (**Indif** izquierda y **Dif** derecha), podemos apreciar que no hay correlaciones altas ni para el ciclo ni sus fases. En la condición de indiferenciación, no fueron ni siquiera significativas.

Primas

Si bien el resultado negativo en la correlación madre-hija hace difícil encontrar correlación entre primas (células que no son hermanas pero comparten célula abuela, Figura 23), un trabajo similar y anterior a esta tesis en linfoblastos de ratón sí la encontró²⁷. Esto puede deberse al mismo factor hereditario que correlaciona células hermanas. Teniendo en cuenta que éstas son de la misma generación, al igual que las primas, podría ser que este factor hereditario correlacione los ciclos de manera intergeneracional.

La Figura 22 muestra los coeficientes de correlación de Spearman entre células primas, obtenidos con la estrategia de Bootstrap. Al igual que la correlación madre-hija, tampoco hubo correlaciones altas ni para el *CC-L* ni sus fases. Solamente fueron significativas para *G1-L*.

Estos resultados indican que, si bien son parecidos los ciclos de células hermanas, no lo son ni con los ciclos de su célula progenitora ni con los de sus células primas. Sin embargo, podemos hacer la comparación entre la correlación medida entre primas y la que esperaríamos obtener simplemente por el parentesco que las conecta como hijas de células hermanas

(Figura 23). Esto es, si medimos correlaciones significativas entre células hermanas y madre-hija, es esperable encontrar alguna correlación entre primas, debida únicamente a relaciones de parentesco directo²⁸. Esta correlación esperada puede calcularse como indica el miembro derecho de la ec. 5, que en este caso es menor a la correlación medida.

$$\rho_{\text{primas}} = \rho_{m-h}^2 \rho_{hnas} \quad 5$$

Esta desigualdad implica que existe algún factor hereditario que correlaciona los ciclos de células primas más de lo esperado. Es natural preguntarse si la célula abuela podría transmitir estos factores en parte. Para responder esta pregunta aplicamos el algoritmo de Grassberger Procaccia, un análisis general sobre la herencia en todo un linaje que detallamos en la siguiente sección.

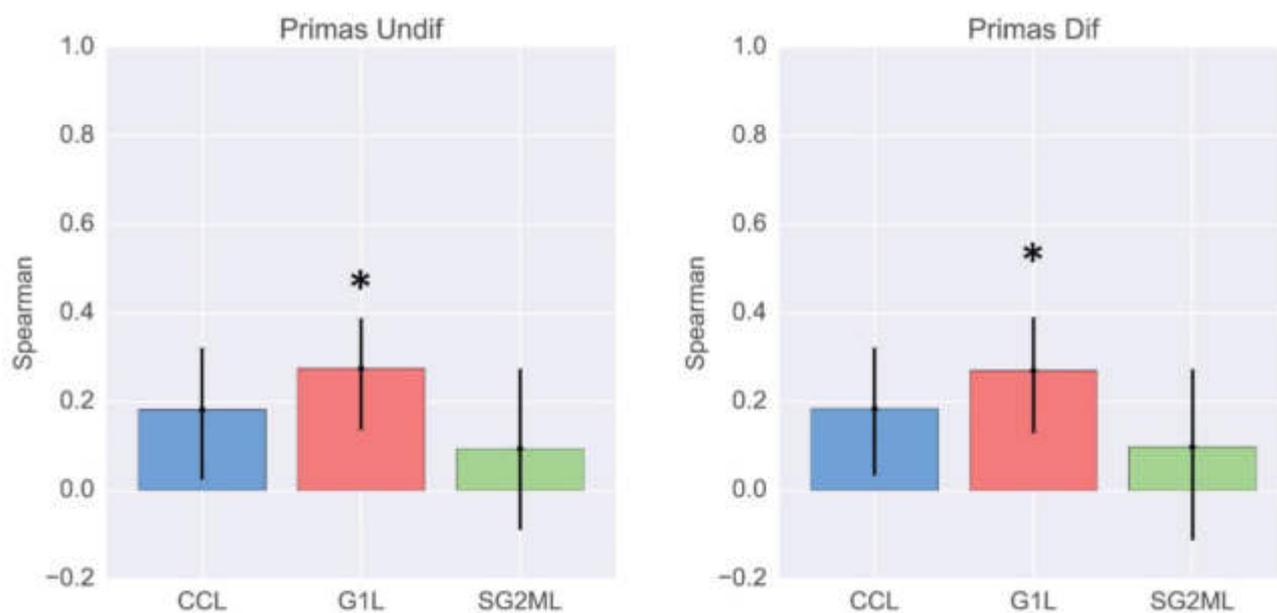


Figura 22: Imagen adaptada¹⁴. Coeficiente de correlación de Spearman entre células primas, restando las medianas generacionales y usando la estrategia de Bootstrap para elegir cantidades similares de células de cada colonia y linaje. En ambas condiciones (**Indif** izquierda y **Dif** derecha), podemos apreciar que no hay correlaciones altas ni para el ciclo ni sus fases. Solamente fueron significativas en el caso de G1.

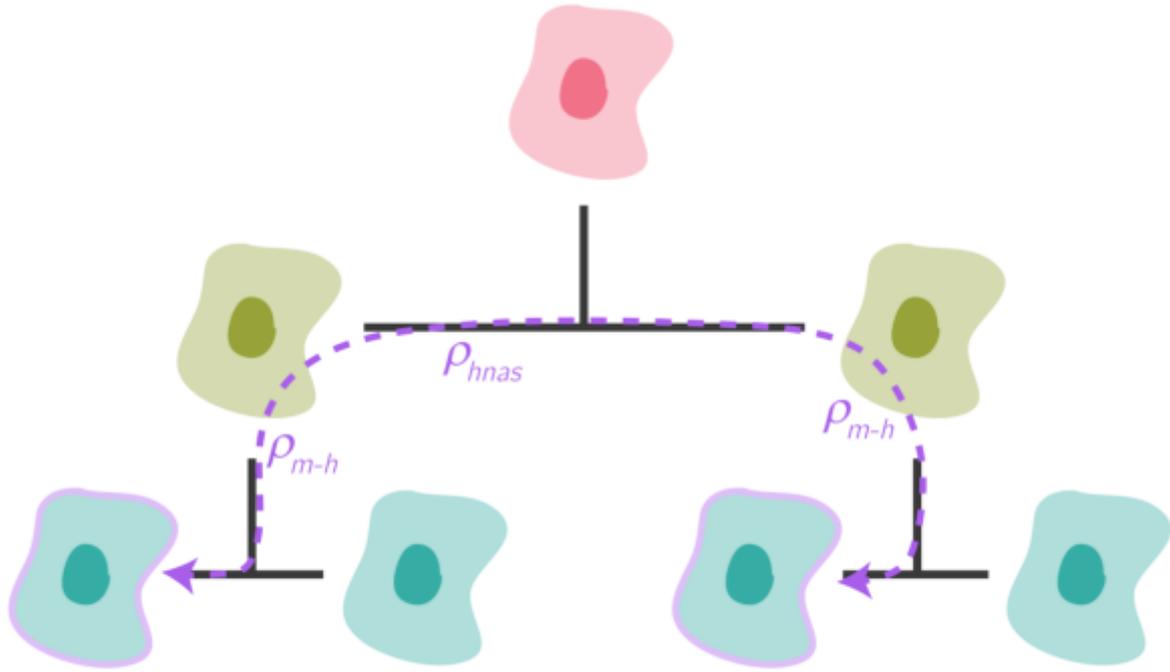


Figura 23: Camino que conecta células primas en algún linaje cualquiera, de donde proviene la ec. 5. Cada salto entre células debería mantener información dada por los coeficientes de correlación de Spearman ρ_{hnas} (entre hermanas) y ρ_{m-h} (entre madre-hija).

El ciclo celular está determinado por la familia: Algoritmo de Grassberger Procaccia

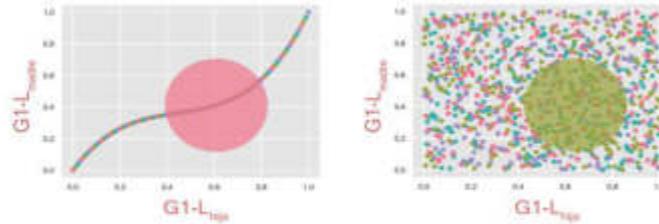
El algoritmo de Grassberger Procaccia²⁹ tiene como objetivo dilucidar si la variabilidad de un observable tiene origen estocástico o está determinada por algún mecanismo. Se aplica a variables que van tomando valores discretos en el tiempo y asume que, de existir un mecanismo, debería haber relaciones matemáticas que predigan el siguiente valor en el tiempo a partir de los anteriores.

En este caso, nuestras variables serán $CC-L$, $G1-L$ y $SG2M-L$, que van tomando valores discretos en las células de un mismo linaje en el tiempo. Así, la pregunta que abordamos es si existe un mecanismo que prediga la duración del ciclo o de las fases de una célula sabiendo cuánto para sus células parentales, incluyendo madre, abuela, bisabuela, etc.

Fundamentos del algoritmo

Imaginemos que graficamos $G1-L$ de muchas células hijas contra el de sus progenitoras madres. Si existiera alguna función matemática que determinara el $G1-L$ de la hija a partir del de la madre, los puntos en el gráfico deberían ordenarse en una curva de acuerdo con esta función (Figura 24, arriba izquierda). Por otro lado, si no hubiera nada determinante, los puntos se distribuirían en el espacio de manera aleatoria, llenando alguna región del gráfico más o menos uniformemente (Figura 24, arriba derecha). Si ahora ubicamos un círculo entre los puntos del gráfico y vamos agrandando su radio desde cero, cada vez más puntos quedarán encerrados dentro del círculo. La idea es contar la cantidad de puntos encerrados en función del radio y estudiar esta relación. Si los puntos se agruparan alrededor de una curva, entonces la cantidad encerrada debería aumentar linealmente con el radio, dado que éste es proporcional al segmento de curva dentro del círculo. Por otro lado, si los puntos llenaran el espacio, la cantidad debería aumentar con el área del círculo, proporcional al radio al cuadrado.

D = 2



D = 3

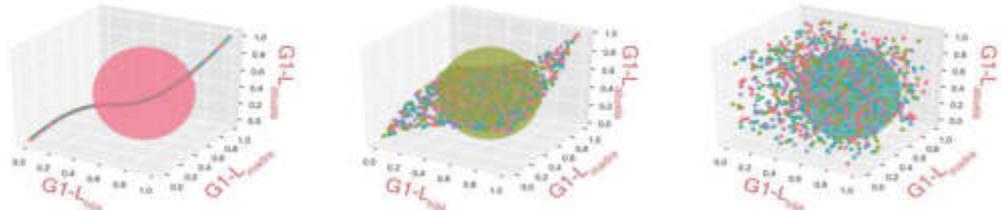


Figura 24: Imagen adaptada³⁰. Ejemplo simulado del agrupamiento de los datos de $G1-L$ de células madres con sus hijas ($D=2$) y de abuelas, madres e hijas ($D=3$).

$D=2$: las células de una sola generación determinan $G1-L$ de sus hijas (izquierda) o no lo hacen (derecha).

$D=3$: las células de una sola generación determinan $G1-L$ de sus hijas (izquierda); las de dos generaciones atrás lo determinan (centro) o no lo hacen (derecha).

La misma idea es válida si hacemos en un gráfico tridimensional de $G1-L$ de una célula hija contra sus células madre y abuela. La pregunta en este caso es si podríamos determinar la duración de la fase sabiendo las propias de la madre y la abuela. Si la madre determinara, entonces los puntos deberían agruparse alrededor de una curva (Figura 24 , abajo izquierda). Por otro lado, existe la posibilidad de que $G1$ esté determinado no solamente por la madre sino también por la abuela. Esto es, que $G1-L$ de la hija sea función del $G1-L$ de la madre y de la abuela. Esto agruparía los puntos alrededor de una superficie de dos dimensiones (Figura 24 , abajo centro). Aumentando el radio del círculo, mayor superficie quedaría encerrada por este, quedando una relación proporcional entre el radio al cuadrado y la cantidad de puntos dentro. Por último, y análogo al gráfico de dos dimensiones, si no hubiera relación, los puntos ocuparían todo el espacio tridimensional (Figura 24 , abajo derecha), de manera que la relación proporcional se daría entre el radio del círculo y el volumen encerrado, que va como el radio al cubo.

Estas diferencias en la relación entre cantidad de puntos encerrados y el radio del círculo son las que usaremos para definir si existe una relación en la duración de las fases de una célula con la de sus progenitoras. Para esto ordenaremos los $G1-L$ y $SG2M-L$ en vectores de 2,

3 y hasta 4 dimensiones, donde cada una se corresponde con una generación y cada vector tiene células del mismo linaje (emparentadas familiarmente), como muestran los vectores de la ec. 6

$$\begin{aligned}
 &(G1 - L_{bisabuela}, G1 - L_{abuela}, G1 - L_{madre}, G1 - L_{hija}) \\
 &\quad (G1 - L_{abuela}, G1 - L_{madre}, G1 - L_{hija}) \\
 &\quad\quad (G1 - L_{madre}, G1 - L_{hija})
 \end{aligned}
 \tag{6}$$

A partir de estos vectores hacemos los gráficos en 2, 3 y 4 dimensiones y contamos la cantidad de puntos dentro de un círculo en función de su radio (Figura 25). El objetivo es ajustar esta cantidad con Ar^d , donde A es un parámetro libre y el exponente d es el que nos indicará la dependencia con el radio. Si d fuera igual a la dimensión del gráfico, entonces el ciclo es estocástico o necesitamos más información de más generaciones hacia atrás para definirlo. Si d fuera menor a la dimensión del gráfico, entonces el número nos indica cuántas generaciones atrás determinan el ciclo celular.

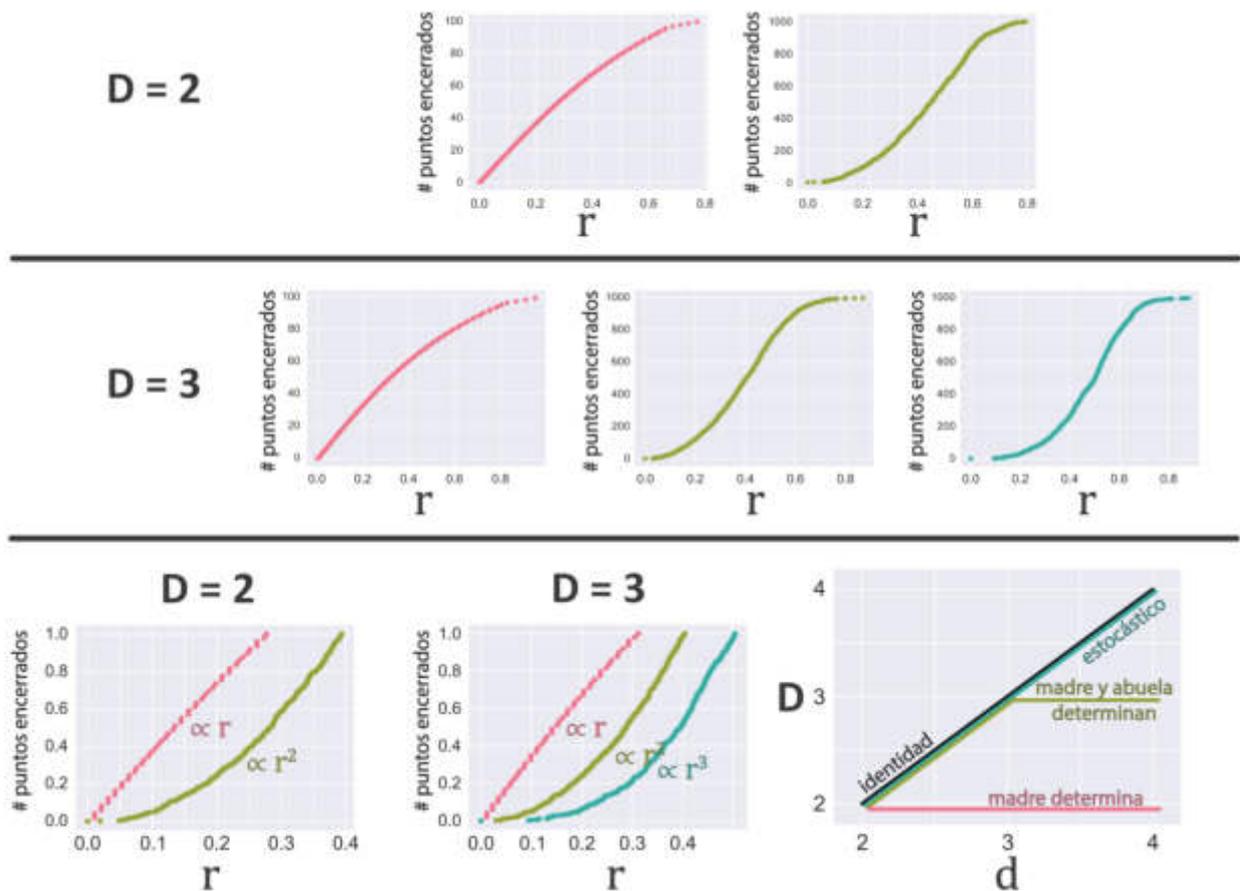


Figura 25: Ejemplo simulado de la cantidad (#) de puntos encerrados por las esferas mostradas en la Figura 24, en función del radio (r) ($D=2$ y $D=3$). La fila inferior muestra los mismos gráficos superpuestos ($D=2$ y $D=3$), normalizados y acotados al rango donde las dependencias con el radio van como r , r^2 o r^3 , donde el exponente depende del agrupamiento de los datos. El gráfico restante (derecha) muestra el resultado esperado para el exponente (d) en función de la dimensión (D) de los gráficos.

Resultados

La Figura 26 muestra los exponentes obtenidos para $G1-L$ y $SG2M-L$ de las condiciones **Indif** y **Dif** para 2, 3 y 4 dimensiones, es decir, para los 3 tipos de vectores de la ec. 6. En todos los casos, pareciera que el exponente no llega a 3, lo que indica que con la duración de las fases de células madre y abuela, pueden determinarse las de las hijas. Cabe destacar que, para células **Indif**, el estudio se redujo a 3 dimensiones porque no había suficientes linajes completos de células bisabuela-abuela-madre-hija. Esto sí ocurría en células **Dif**, dado que proliferan más rápido.

Es importante destacar que los resultados no son concluyentes en los casos de $G1-L$ **Indif** y $SG2M-L$ **Dif**. Esto se debe a que los exponentes obtenidos no se separan de las bandas de control. Estas bandas corresponden a los intervalos de confianza del 50% y 95%, e indican el porcentaje de las veces que el exponente podría caer en estos intervalos si repitiéramos el experimento debido a la variabilidad intrínseca en los datos (control *shuffle*). Sumado a estas bandas, se marcaron en los gráficos de la Figura 26 más controles. El control *random* graficado es el resultado de aplicar el protocolo sobre la misma cantidad de datos, pero elegidos aleatoriamente. En todos los casos se espera que la curva siga la función identidad y así fue. Estos dos controles están explicados en la siguiente sección.

El panel **C** de la Figura 26 muestra los resultados de aplicar el algoritmo sobre datos de posición y velocidad media de las células. La posición en el tiempo, obtenida del seguimiento automatizado, forma parte de la base de datos. La velocidad se calculó simplemente como el desplazamiento entre frame y frame dividido por los 15 minutos de tiempo entre éstos. La idea consiste en que la posición es un buen control positivo, dado que son células que mayormente permanecen en el lugar y entonces la posición de la célula progenitora resulta muy cercana a la de sus hijas. Por otra parte, el movimiento de las células es bastante errático y no existen direcciones privilegiadas de desplazamiento. Es por esto que tomamos a la velocidad como control negativo. El resultado tanto para el control negativo como para el positivo es el esperado, lo que indica que el protocolo está aplicado correctamente.

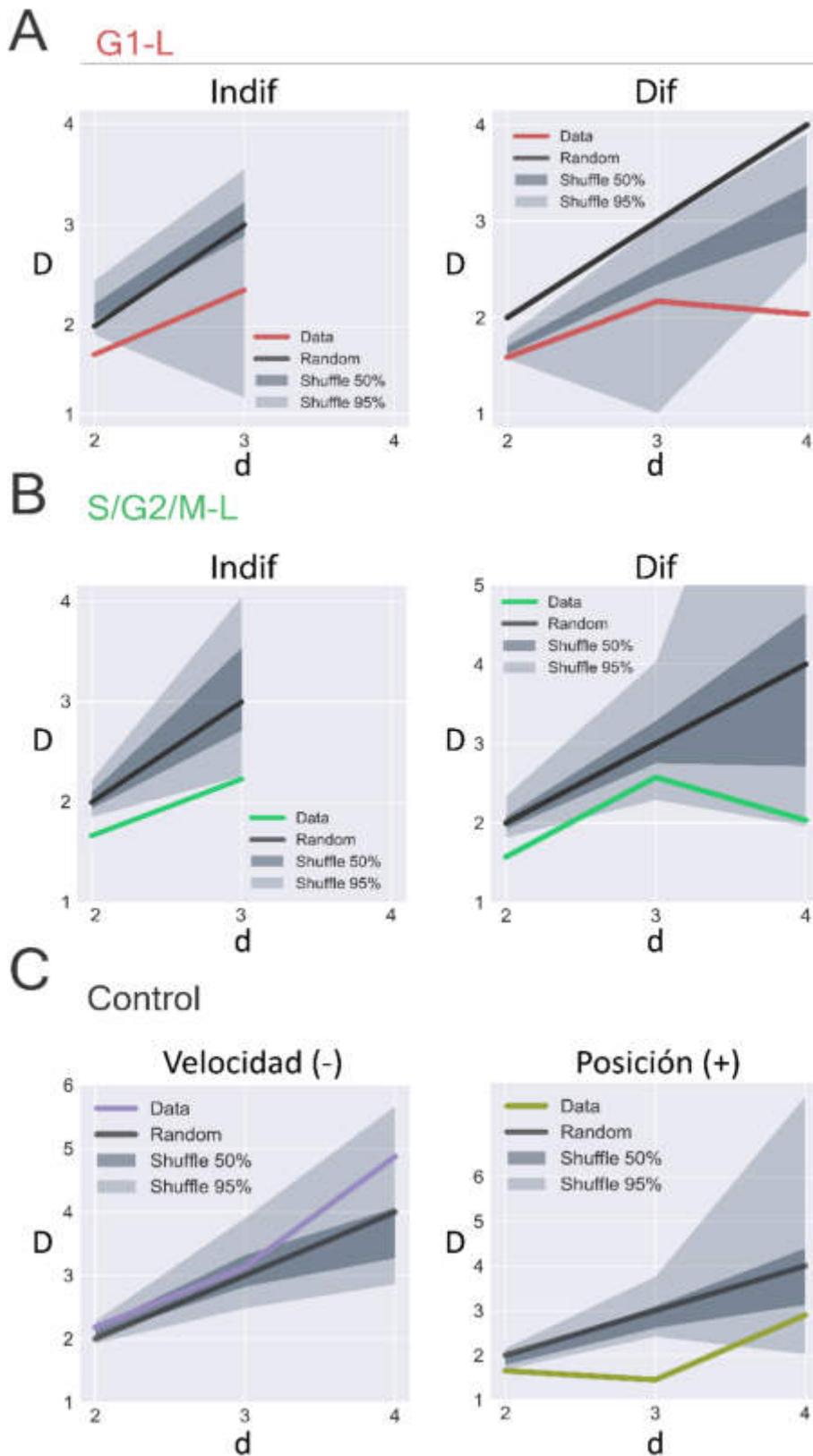


Figura 26: Resultados del algoritmo de Grassberger-Procaccia. **A** y **B**: G1-L y SG2M-L respectivamente, para las células en condiciones de pluripotencia (**Indif**) y en diferenciación (**Dif**). Dado que las células **Indif** viven más tiempo, el algoritmo solo se aplicó hasta 3 dimensiones debido a la falta de cantidad de células bisabuela-abuela-madre-hija. En todos los casos los resultados son compatibles con que las duraciones de las fases de una célula pueden determinarse a partir de las propias de la madre y abuela. Sin embargo, solo en los casos de SG2M-L **Indif** y G1-L **Dif** los datos se separan suficiente de las bandas de control como para afirmarlo. En este último caso, las bandas se separan del control random indicando que además existe algún factor externo determinante. **C**: Controles de aplicación negativo (Velocidad) y positivo (Posición). Ambos controles resultaron dentro de lo esperable.

Controles

Para cerciorarnos de no tener errores de aplicación en el algoritmo, lo primero que hicimos fue aplicarlo a datos estocásticos generados computacionalmente, lo que llamamos el control *random*. En este escenario, esperábamos que el exponente del radio coincidiera siempre con la dimensión de los gráficos, dado que los datos simulados llenaban toda el área, volumen o hipervolumen en los casos de 2, 3 y 4 dimensiones. Esto se verificó, como muestran estos controles en la Figura 26, donde los exponentes son siempre iguales a la dimensión.

Otro control, que llamamos *shuffle*, tiene dos objetivos. El primero, asegurar que no haya algún factor global que determine la duración de las fases a lo largo de las generaciones, independiente a las relaciones de parentesco. El segundo, ofrecer valores esperados para el exponente debido a fluctuaciones estocásticas. Si bien buscamos siempre que el exponente sea menor que la dimensión de los vectores no es claro cuánto menor debería ser para poder definir con seguridad que existe un factor determinante. Las bandas asociadas a este control, que muestra la Figura 26, fueron calculadas de la siguiente manera. Se tomaron los mismos vectores de datos de $G1-L$ y $SG2M-L$ y se mezclaron entre sí las coordenadas correspondientes a cada generación. El resultado son vectores similares a los de la ec. 6, con la diferencia que dentro de cada vector se perdieron las relaciones de linaje. Con estos vectores mezclados aplicamos el protocolo y obtuvimos exponentes asociados. Luego repetimos el proceso 10 mil veces, mezclado siempre los vectores de manera aleatoria, y obtuvimos una distribución de exponentes para cada dimensión. De estas distribuciones provienen los intervalos de la Figura 26. Cabe destacar que las mismas pueden incluir a los vectores de datos originales, dado que siempre que se mezcla algo existe alguna pequeña probabilidad de volver al orden original. Lo determinante es verificar que el exponente medido para los datos originales esté bien corrido de la distribución mezclada, en nuestro caso tomamos que esté por fuera del intervalo de 95% de confianza. Esto significa que los datos originales, ordenados por linaje, se agrupan en planos o curvas más que en el 95% de las veces que lo harían por azar si rompiéramos las relaciones de linaje.

Por otra parte, si existiera algún factor global, independiente del parentesco, que agrupara los datos en planos o curvas, podríamos medir

exponentes menores a la dimensión y sacar conclusiones erróneas. Sin embargo, al ser global, también debería afectar a los exponentes de los datos mezclados. La manera de corroborarlo consiste en ver si el control random se encuentra en el centro de las bandas shuffle o si éstas se corren hacia abajo. Lo último indicaría que sí existen factores globales que determinan las fases, y podemos ver que se cumple en el caso de *G1-L Dif* (Figura 26), probablemente debido al acortamiento general del ciclo con la diferenciación.

Estructura de las colonias

Los procesos de diferenciación celular son los que fabrican las diferentes identidades celulares de un organismo adulto. Estas identidades a su vez se autoorganizan espacialmente para formar tejidos y órganos. Por otra parte, se sabe que las células en estado naïve sintetizan factores de transcripción que mantienen la pluripotencia. De este modo, no sería raro encontrar que la diferenciación in vitro produce cambios en la morfología y estructura espacial de las colonias, y que podrían aparecer inhomogeneidades o patrones espaciales de diferenciación.

En esta sección analizamos los cambios que la diferenciación de naïve a primed produce en el tamaño y la densidad de las colonias, y buscamos si existe alguna dependencia de las identidades celulares con su posición dentro de la misma.

Tamaño

La primera observación consiste en que las células en diferenciación tienden a explorar más el espacio. Sumando la mayor proliferación, el resultado son colonias notoriamente más grandes (Figura 27). Sin embargo, no es trivial cómo definir el tamaño de una colonia que está formada por células separadas que no siguen un orden específico. Para hacerlo, tomamos el criterio del mínimo polígono convexo³¹. Este consiste en encerrar a la colonia con el polígono imaginario de mínima área y cantidad libre de lados que contenga a todas las células. La restricción es que sea convexo, es decir, con todos sus ángulos internos menores a 180° . Esto asegura que el polígono no tenga lados que penetren entre las células y permite definir un área y forma únicas para cada colonia en cada tiempo. Así, graficamos el área de las colonias en función del tiempo, y cuantificamos que cuánto más crecieron las colonias de células en diferenciación (Figura 27).

Como ya mencionamos, el mayor crecimiento de las colonias en diferenciación se deba en parte a que sus células exploran más espacio. Para cuantificarlo, utilizamos el mismo criterio del polígono convexo. El método consiste en proyectar toda la trayectoria en el tiempo de una célula en el plano espacial y encerrarla con el mínimo polígono. Definimos la exploración como el área de este polígono dividida por el tiempo de vida total (Figura 27). Calculando todas las exploraciones, construimos la

distribución para ambas condiciones y confirmamos que las células en diferenciación exploran más el espacio a su alrededor (Figura 27). Este resultado tendrá efectos también en la densidad y la forma de las colonias, estudiada en las siguientes secciones.

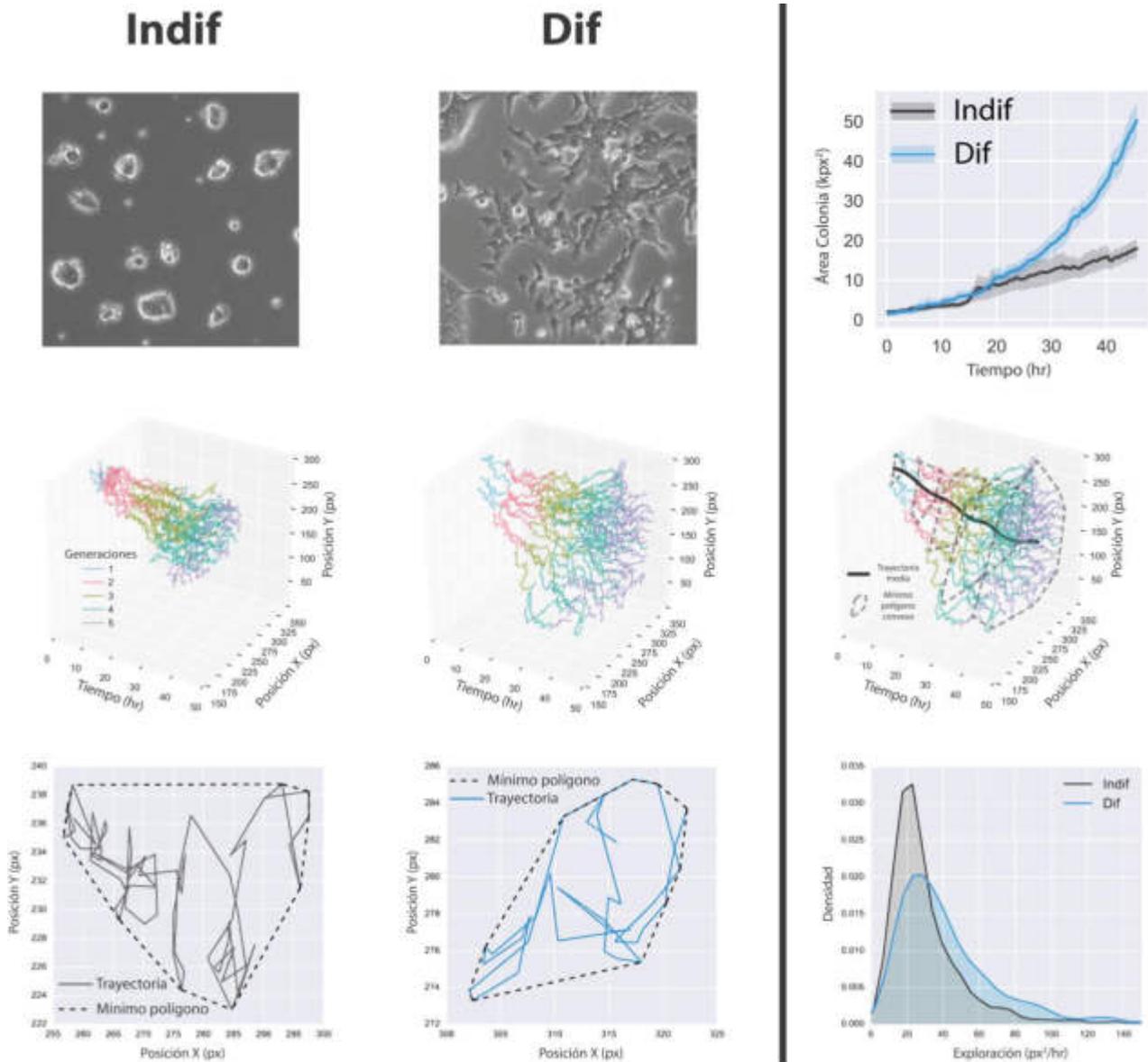


Figura 27: Fotos de células al microscopio de campo claro (fila superior, izquierda, imágenes adaptadas¹⁴). Podemos ver que las células en diferenciación son más grandes y tienden a explorar más. En la fila media, izquierda, graficamos las trayectorias de todas las células de un video por cada condición. Comprobamos que en los videos las colonias en diferenciación divergen más en el espacio. En la fila media, derecha, agregamos los mínimos polígonos convexos de 5 frames. A partir de éstos medimos el área de las colonias en función del tiempo (fila superior, derecha), donde cuantificamos que las colonias en diferenciación tienden a ocupar más espacio. En la fila inferior, izquierda, calculamos el mínimo polígono convexo a la trayectoria completa de una célula ejemplo por cada condición (imágenes adaptadas¹⁴). A partir de esta área, dividida por el tiempo de vida de la célula, definimos la exploración. Graficamos su distribución para ambas condiciones en la fila inferior, derecha, donde apreciamos que las células en diferenciación tienen tendencia a explorar más el espacio.

Densidad

Otra de las diferencias que encontramos en la estructura de las colonias es la densidad de las células. Es importante tener en cuenta que las células mismas aumentan su área con la diferenciación, probablemente porque se adhieren mejor a la placa contenedora (Figura 28, izquierda). Para tener esto en cuenta, definimos la densidad de dos maneras. Una es la cantidad de células de una colonia dividida por el área que ocupa ésta, medida con el método del mínimo polígono convexo. La otra, como el área sumada de todas las células de una colonia dividida por el área de la colonia. Podemos apreciar en la Figura 28 que las colonias en diferenciación son menos densas o más dispersas. Esto va en contra de lo esperado teniendo en cuenta que, en diferenciación, tanto el área nuclear de cada célula como el número total de células son mayores. La explicación parece ser entonces la mayor exploración de las células en diferenciación: si bien estas colonias tienen más células y cada una es más grande, el tamaño de la colonia es aún mayor debido a que cada célula explora más el espacio, y resultan colonias más dispersas.

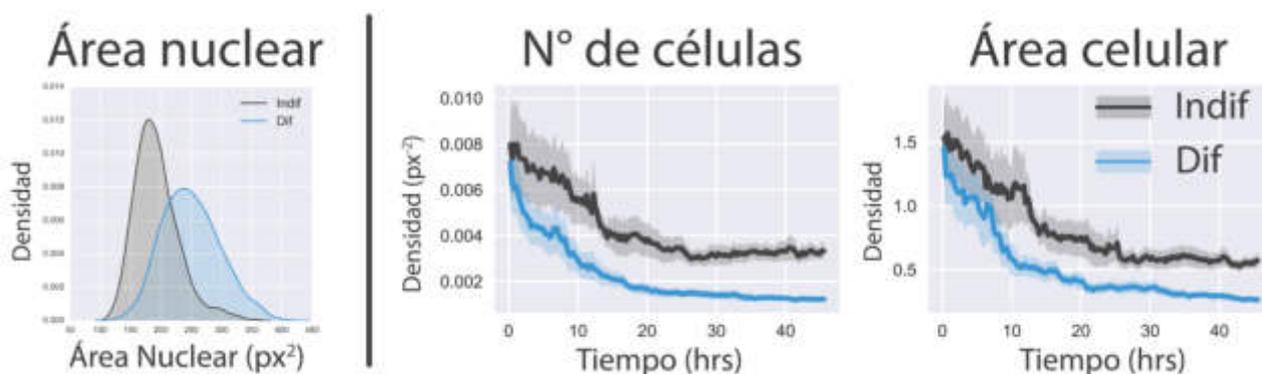


Figura 28: Distribución del área nuclear de las células en condiciones **Indif** y **Dif** (izquierda). Densidad en función del tiempo medida como cantidad de células por colonia (centro) y área sumada de todas las células por colonia (derecha). Podemos ver que la densidad parece estabilizarse en las dos condiciones, y que en diferenciación es menor.

Forma

Otra observación a simple vista en los videos consiste en que las colonias indiferenciadas mantienen una forma más circular. Esto se deba probablemente a que se mantienen más compactas que en diferenciación, donde presentamos que las colonias son más grandes y dispersas. Es esperable que las células que proliferan sin moverse activamente terminen formando

colonias aproximadamente circulares. Para cuantificar y verificar esta observación, calculamos la circularidad (ec. 7) de las colonias en ambas condiciones, un parámetro muy utilizado en la literatura para caracterizar la forma de colonias celulares^{32,33,19}.

$$C = 4\pi A/P^2 \quad 7$$

C es la circularidad, A y P son el área y perímetro del mínimo polígono convexo que envuelve a la colonia. La idea es que, si la colonia fuera un círculo, $P = 2\pi r$ y $A = \pi r^2$, con r el radio, de modo que $C = 4\pi \pi r^2 / 4\pi^2 r^2 = 1$. Por el contrario, si la colonia fuera muy irregular, por ejemplo, muy lineal, entonces $A/P^2 \rightarrow 0$. En síntesis, C es un parámetro que varía entre 0 y 1 siendo mayor cuanto más parecida a un círculo es la forma de la colonia.

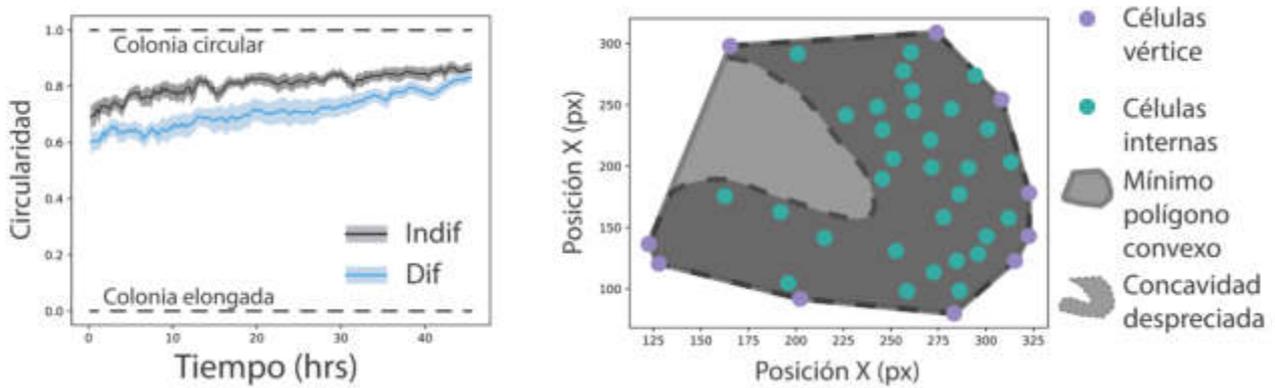


Figura 29: Circularidad promedio de todas las colonias en función del tiempo (izquierda). Podemos apreciar que aquellas en diferenciación son menos circulares. Debido a la definición de las áreas de las colonias a partir del mínimo polígono convexo, este efecto podría ser aún mayor dada la eliminación de las concavidades. A la derecha, un ejemplo de un frame de una colonia donde se marcaron las posiciones de las células, el polígono calculado y una posible concavidad no tomada en cuenta.

La Figura 29 muestra los resultados de la circularidad promedio de las colonias en ambas condiciones. En ambos casos aumenta con el tiempo, posiblemente debido al aumento en la cantidad de células que permite polígonos de mayor número de lados, más similares a un círculo. Podemos verificar que las colonias en diferenciación son menos circulares, lo que podía esperarse dada la mayor exploración de las células que las forman. Es importante remarcar que igualmente cuentan con circularidades grandes y que la diferencia, si bien es significativa, no es tanta con las colonias indiferenciadas. Esto puede deberse a que definimos la forma de las colonias

a partir del mínimo polígono convexo, y entonces descartamos cualquier concavidad que podrían tener (Figura 29). En conclusión, si bien los resultados son coherentes, la diferencia en circularidad podría ser incluso mayor.

Homogeneidad

El último aspecto que estudiamos en cuanto a la estructura de las colonias es la homogeneidad de las células que las forman. Nos apoyamos en la hipótesis presentada en la introducción (ver Acoplamiento con la densidad celular) que consiste en que las células indiferenciadas fomentan el estado pluripotente. Para probar esto, clasificamos a las células en dos grupos, centrales y periféricas, según el lugar que ocupan en la colonia y suponiendo que las periféricas estarían más expuestas al medio de cultivo y menos comunicadas con el resto de las células. Para cada grupo, construimos las distribuciones de $CC-L$, y el Área Nuclear, dos variables que sabemos que cambian significativamente con la diferenciación. El objetivo es ver si hay diferencias entre los grupos para cada condición de cultivo, con el fin de detectar si aparecen inhomogeneidades durante diferenciación.

La Figura 30 muestra las distribuciones de las variables analizadas, separadas para células centrales y periféricas y para las dos condiciones de cultivo. La separación entre centrales y periféricas la obtuvimos mediante tres métodos diferentes que detallamos en la siguiente sección. Estos los denominamos "vértices polígonos", "trayectoria media" y "n° de vecinos", donde en este último caso la separación en grupos está definida por la cantidad de células vecinas. Podemos observar que no se observan diferencias significativas en ninguna variable, para ninguno de los métodos. Esto indica que no parece haber diferencias en los cambios de identidad que dependan de la posición en la colonia. Por otra parte, podemos rescatar que en ambas condiciones el área nuclear de las células parece bajar con la cantidad de células vecinas. Esto se deba, probablemente, a la falta de superficie libre de contacto con el sustrato que tienen las células rodeadas de vecinas.

Métodos de separación por posición relativa a la colonia

Desarrollamos tres métodos con el objetivo de clasificar a las células de acuerdo con su posición relativa a la colonia. Por un lado, los métodos "vértices polígonos" y "trayectoria media" apuntan a encontrar diferencias en la homogeneidad dadas por diferente exposición a los medios de cultivo y sus señales de diferenciación o pluripotencia. El método restante de "n° de vecinos" apunta a clasificarlas de acuerdo con el nivel de comunicación intercelular.

El método de "vértices polígonos" clasifica las células en "centrales" y "periféricas", de manera binaria. Primero toma el mínimo polígono convexo que envuelve a la colonia, el mismo utilizado para calcular el área y la circularidad de las colonias (ver Tamaño y Forma). Luego, define como "periféricas" a aquellas células que se encuentran a menos de un umbral de distancia del perímetro de este polinomio (Figura 31). Las células restantes las define como "centrales".

El método de la "trayectoria media" también las clasifica en los mismos dos grupos. En lugar del polígono, en este caso calcula la posición media de la colonia como el centro geométrico de la misma, tomando a cada célula como un punto. Luego, define como "centrales" a aquellas que se encuentran a menos de un umbral de distancia de este punto (Figura 31). Y a las células restantes las define como "periféricas".

Cabe destacar que en estos dos métodos ("vértices polígonos" y "trayectoria media") tanto el polígono o el centro geométrico se calculan para cada frame de un video. Así, cada célula es clasificada en cada frame. En general, las células no cambian de grupo durante su vida, salvo algunas excepciones que fueron descartadas para estos análisis, para asegurarnos de no incluir células dudosas en los resultados.

El tercer método del "n° de vecinos" clasifica a las células en varios grupos de acuerdo con la cantidad de células vecinas que tiene. Para cada una, cuenta la cantidad de vecinas definidas como las que se encuentran más próximas que cierto umbral de distancia fijo. Esto lo repite para cada frame donde existe la célula que está clasificando (Figura 31). Finalmente, calcula el promedio de la cantidad de vecinas para todos los frames y redondea para clasificar. Cabe destacar que hubo células en condición **Indif** con hasta 5 vecinas mientras que, en la condición **Dif**, este número llegó a

4. Esto se deba probablemente a la diferencia en el área celular y en la adherencia al sustrato, ambas mayores en diferenciación. Las células que se adhieren mejor a la superficie se terminan amontonándose menos entre sí.

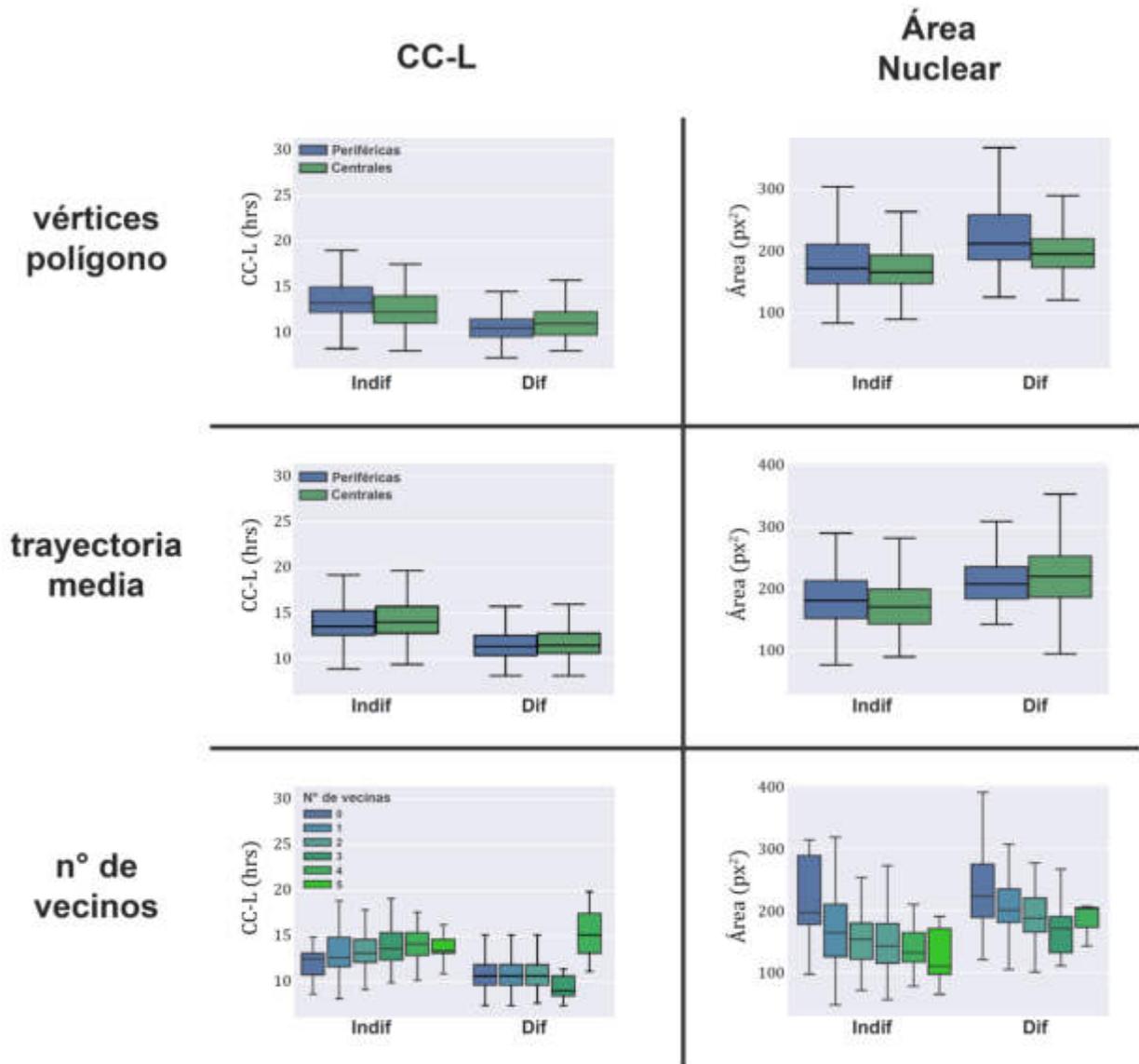


Figura 30: CC-L (columna izquierda) y Área Nuclear (columna derecha) de las células separadas por medio de cultivo y por posición relativa en la colonia utilizando tres métodos: vértices del mínimo polígono convexo (fila superior), distancia a la trayectoria media de la colonia (fila central) y n° de células vecinas (fila inferior). En ningún caso encontramos diferencias entre las células de acuerdo con su posición relativa a la colonia. Cabe destacar que el Área Nuclear sí parece depender del número de células vecinas.

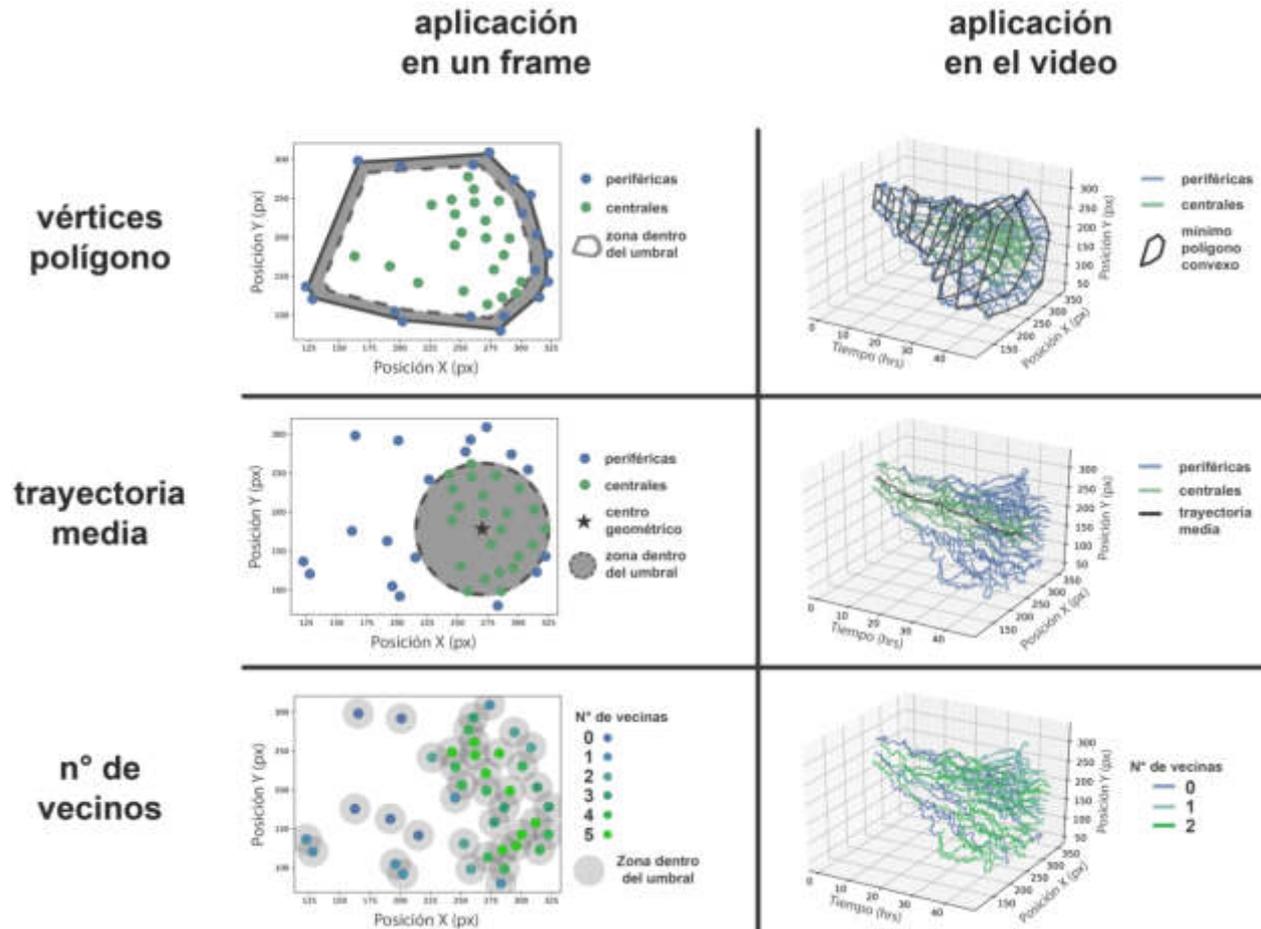


Figura 31: Ejemplos de aplicación de los tres métodos desarrollados para clasificar a las células de acuerdo con su posición relativa a la colonia. La columna izquierda muestra la aplicación para las posiciones de las células en un frame de video ejemplo. La columna derecha, muestra la clasificación para las posiciones en el tiempo de un video entero. Los 3 métodos son "vértices polígono" (fila superior), "trayectoria media" (fila central) y "n° de vecinos" (fila inferior). Para todos los métodos se utilizó el mismo frame (izquierda) y se marcaron las clasificaciones obtenidas y los umbrales elegidos. Éstos son, respectivamente, la distancia al borde del mínimo polígono convexo de la colonia, la distancia al centro geométrico de la misma y la distancia intercelular. En el último caso, correspondiente al "n° de vecinos" dibujamos círculos grises cuyo radio es la mitad del umbral elegido, de manera que las superposiciones de estos círculos indican que las células que representan son consideradas vecinas.

Conclusiones y perspectivas

En resumen, primero logramos medir la duración del ciclo celular y sus fases, quitando el sesgo del video de analizar las células con vida más corta. Esto, por un lado, es una de las primeras caracterizaciones robustas del ciclo celular en células naïve³⁴. Por otro lado, nos permitió ver cómo y cuánto se acertaba con cada generación durante el proceso de diferenciación, mediciones que habían sido hechas pero indirectamente^{35,36,37}. La principal ventaja de nuestro método consiste en que la medición por microscopía es directa y a nivel de célula única, permitiendo obtener la distribución poblacional de la duración del ciclo y sus fases, a diferencia de otros métodos a nivel poblacional, como la tinción con yoduro de propidio³⁸. El método de Kaplan-Meier resultó muy favorable para cuantificar el acortamiento del ciclo, lo que muestra que es un análisis provechoso y fácilmente adaptable para analizar este tipo de videos, y aportó una medición directa y comparable de las duraciones del ciclo celular. Concluimos que la salida del estado naïve viene asociada a un acortamiento del ciclo y aumento de la proliferación celular, evidencia que va en contra de literatura anterior³⁹, pero donde hicieron mediciones indirectas, y va en favor del aumento en proliferación observado en células del epiblasto post-implantatorio de ratón^{40,41}.

Segundo, pudimos aplicar el algoritmo de Grassberger-Procaccia y concluir que esta duración depende también del linaje familiar, y en particular de las células abuela y madre. Sin embargo, la cantidad de células necesarias para este análisis fue un limitante, dado que se necesitan linajes completos de varias generaciones en un mismo video. A futuro, sería interesante repetir el experimento con mayor cantidad de células y aprovechar que hoy en día existen técnicas de machine learning que permiten la segmentación celular automática⁴² y reducirían significativamente el tiempo de análisis de los videos. Las conclusiones, que indican una alta correlación de la fase G1 entre hermanas, apoyan la hipótesis de una división celular simétrica. Esta misma correlación ya había sido encontrada en estas mismas células³⁷ y también en linfocitos³⁰.

En tercer lugar, logramos caracterizar a las colonias en forma, tamaño y densidad. Si bien había reportes sobre la mayor dispersión de las células durante la diferenciación^{43,12}, no existía una caracterización completa de

la morfología de las colonias. Para esto fue indispensable el algoritmo del mínimo polígono convexo. Los resultados obtenidos muestran coherencia y robustez dado que se condicen el aumento de tamaño y densidad con el acortamiento del ciclo celular y la mayor movilidad celular observada. Esto último es coherente con reportes recientes del aumento de expresión de genes ligados a la adherencia⁴⁴. Aún más, van en línea con la implantación del epiblasto y los cambios morfogénicos que le siguen⁴⁰. Un estudio interesante que quedó pendiente es analizar el movimiento de cada célula para responder si eligen hacia dónde moverse o su trayectoria depende más del empuje general que realiza el crecimiento de la colonia.

Por otra parte, en cuanto al análisis de homogeneidad, es cierto que estos cultivos son bidimensionales (las células están pegadas a un sustrato, no flotan en medio) y esto hace que estén todas muy expuestas al medio de cultivo y las señales de diferenciación. Esto no representa fielmente la tridimensionalidad de un embrión, donde hay células internas que no tienen contacto directo con el medio. Sería interesante repetir este tipo de análisis en células flotando en forma tridimensional.

Autoría Correspondiente

Para facilitar la lectura, todo el capítulo fue escrito en primera persona del plural, de manera que corresponde aclarar qué partes realicé como autor de este escrito. Toda la parte experimental de trabajo de mesa que llevó al armado de los videos fue un trabajo que realizó el Dr. Ariel Waisman, principalmente durante su doctorado, en conjunto con los colaboradores, el Dr. Martín Elías Costa y la Dra. María Soledad Cosentino y sus directores la Dra. Alejandra S. Guberman y el Dr. Santiago G. Miriuka, también codirector de esta tesis. Una vez filmados los videos y hecha la base de datos a partir de la segmentación y la confección de los linajes y dendrogramas, nos sumamos al proyecto con la Dra. Alejandra Ventura, directora de esta tesis, para aportar en los análisis posteriores de la base de datos. Estos son Kaplan Meier, Grassberger-Procaccia y los estudios de morfología, homogeneidad y crecimiento de las colonias celulares.

Capítulo 2

Mecanismos de memoria molecular en redes biológicas

Introducción

Memoria de sistemas en general

Generalmente, la memoria tiene definiciones muy dependientes del contexto. En computación, memoria es la cantidad de información que puede almacenar algún dispositivo y se mide en bits. Cada bit es una variable booleana (verdadera o falsa, 0 o 1) que funciona como elemento fundamental de memoria. En sistemas complejos, la memoria se relaciona con la bi- o multi-estabilidad de un sistema, con su histéresis, o con que la historia influya de alguna manera en el estado final.

Yendo al marco de la biología molecular, encontramos dos definiciones principales y parecidas de memoria. Una es la capacidad de almacenar información de un estímulo pasado^{45,46}. La otra, es que exista una respuesta estacionaria frente a un estímulo transitorio⁴⁷.

A priori, estas definiciones pueden parecer diferentes, pero tienen los mismos fundamentos en común. Un disco duro de computadora puede almacenar información en celdas magnéticas. Estas celdas son capaces de permanecer magnetizadas, incluso por años⁴⁸, luego de someterse a un estímulo magnético transitorio. Esta propiedad emerge de los dominios magnéticos de los materiales que, al ser alineados por un campo externo, generan su propio campo y éste mantiene la alineación de cualquier dominio, cerrando un ciclo de retroalimentación positiva o feedback positivo. El ciclo otorga a las celdas dos estados de equilibrio estables: magnetizadas o no. Son estos dos estados los que dan el soporte físico al bit, otorgan histéresis⁴⁸ al sistema y hacen que el estado final dependa de la historia (si recibió o no estímulo magnético externo).

Memoria Celular

Las células cuentan con muchísimos y diversos mecanismos y formatos físicos para almacenar información de estímulos pasados. Entre los más complejos, se encuentran la memoria neuronal y la inmunológica. La primera refiere a la que almacena, entre muchas otras cosas, nuestros propios recuerdos. Esta es capaz de almacenar información proveniente de todos los sentidos de manera integral, y se cree que el formato físico es la misma disposición física de las neuronas en el espacio, la cual determina qué neuronas serán activadas y bajo qué estímulos. La segunda refiere a la

capacidad del sistema inmune de almacenar información de casi cualquier patógeno que se pueda encontrar. El soporte físico es el ADN mismo, donde cada célula guarda la secuencia que sintetiza el anticuerpo correspondiente a cada patógeno.

En este capítulo, estudiaremos la memoria molecular de redes de interacción química en sistemas biológicos. Uno de los ejemplos de este tipo de memoria ya lo mencionamos en el Capítulo 1, y es la misma diferenciación celular. La diferenciación no es más que un proceso duradero, y muchas veces irreversible, gatillado por un estímulo transitorio. En el caso de las células madre embrionarias de ratón, la diferenciación del estado naïve al primed involucra dos mecanismos de memoria molecular bien diferentes. El primero es un mecanismo epigenético, consiste en el cambio conformacional de las histonas en la cromatina que cambian la expresión de los factores de transcripción asociados a la pluripotencia o a la diferenciación. Dos ejemplos son el aumento de H3K27me3 y H3K9me2 (conformaciones silenciadoras) en la zona promotora del gen *Essrb* y la disminución de H3K27me3 en la zona del gen *Otx2*³. El segundo mecanismo ocurre a nivel proteico, consiste en la autorregulación del complejo de pluripotencias Nanog-Oct4-Sox2⁴⁹. Este complejo activa su propia síntesis, manteniendo su concentración y con ella el estado de pluripotencia. Este tipo de memoria molecular es la que estudiaremos en este capítulo, la que tiene como soporte físico la existencia o falta activa de nodos o actores en una red de interacciones químicas biológicas.

Memoria molecular en redes biológicas de interacción química

Yendo en particular a la señalización celular y las redes de regulación génica, existe cada vez más evidencia de que la memoria es un factor determinante de muchos mecanismos moleculares^{50,51,18,52,53,54,55,56}. En estos trabajos, los sistemas celulares son tratados con uno y dos estímulos, separados por alguna ventana temporal. Las observaciones consisten en que las respuestas celulares difieren de acuerdo con si se da o no el primer estímulo, demostrando que cada sistema es capaz de recordarlo, al menos durante el tiempo que dura la ventana temporal de estudio.

Pensando ahora en los mecanismos subyacentes a estos tipos de memoria, es bien conocido que los feedbacks positivos moleculares pueden proveer

biestabilidad, histéresis y memoria^{18,53,54,46,45,57,58,59,60,61,62}. Imaginemos alguna proteína dentro de una célula que, por encima de cierta concentración, activa su propia síntesis (Figura 32). Si no hubiera concentración suficiente, esta proteína se degradaría hasta desaparecer, llegando a un punto estable de concentración nula. Por otro lado, si pudiéramos dar algún estímulo transitorio de proteína que haga que la concentración supere cierta cantidad necesaria, entonces activaríamos el feedback positivo y la concentración quedaría en algún estado estacionario estable no nulo, donde la auto-síntesis se equipararía a la degradación.

El sistema descrito es el feedback positivo más simple que existe, consiste en una sola variable (la concentración de proteína) y aun así tiene memoria. Sin embargo, incluso en este ejemplo sencillo, existen varias maneras de medir la memoria. Este problema se origina en las diversas definiciones expuestas. Por ejemplo, en los estudios de biología molecular, se usan medidas binarias que apuntan solamente a definir si un sistema tiene o no tiene memoria. Estas medidas tienen la desventaja que no informan cuánto tiempo dura la memoria ni permiten cuantificarla de alguna manera. ¿Cómo podríamos comparar si un sistema recuerda más o menos un estímulo u otro? ¿cómo podríamos comparar si un sistema tiene más o menos memoria que otro? ¿Es posible definir alguna medida que permita comparar cualquier tipo de sistemas moleculares?

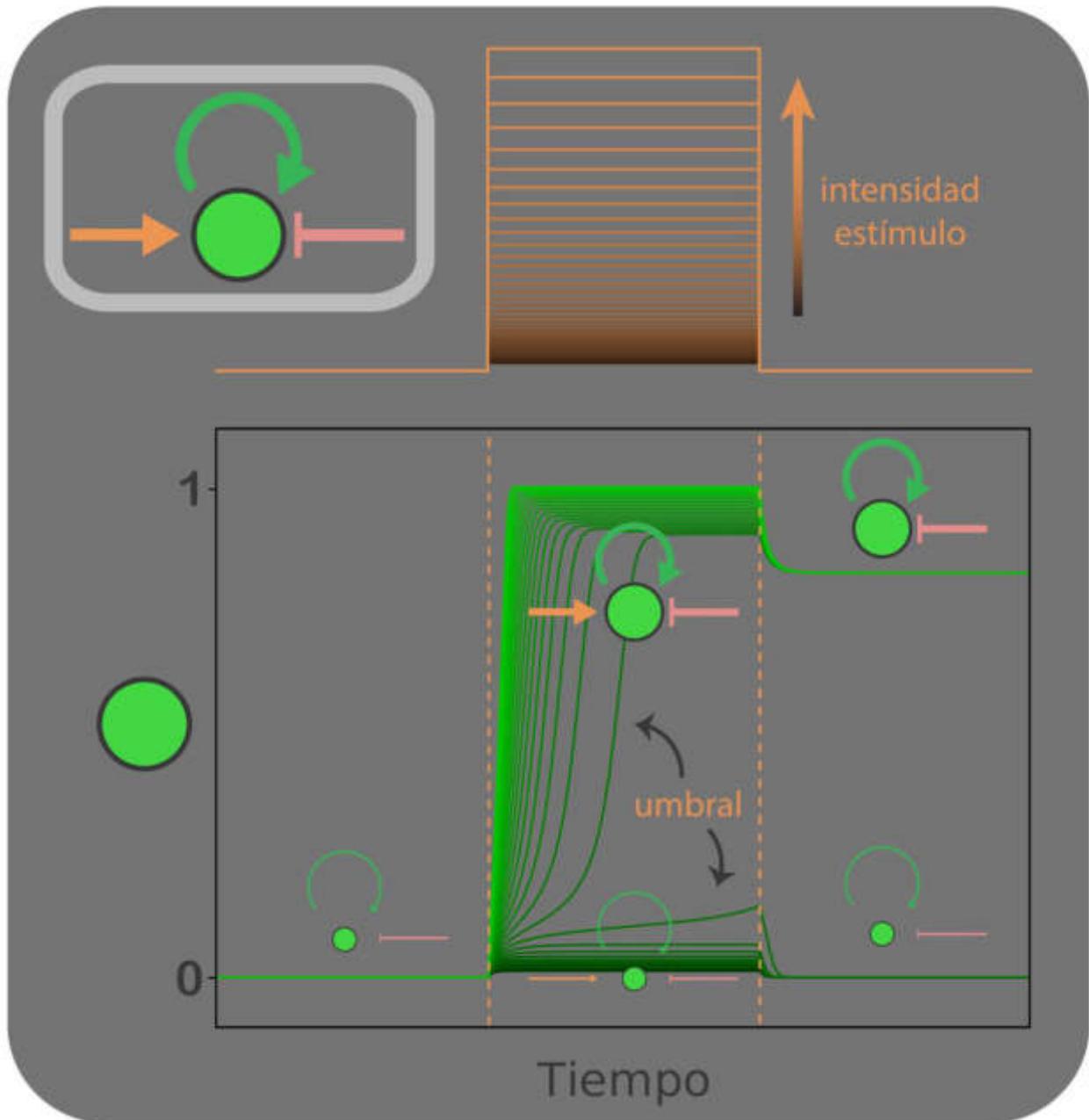


Figura 32: Modelo de una proteína que activa su propia síntesis (arriba izquierda). Al recibir un estímulo tipo pulso escalón transitorio (arriba), este puede ser suficientemente intenso como para encender la autoactivación de la proteína o no. Si el estímulo supera este umbral de intensidad, la proteína podrá mantener su propia síntesis aun cuando el estímulo haya pasado (abajo).

Objetivo

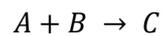
El objetivo de este capítulo es encontrar una medida de memoria molecular para redes de interacción química biomolecular, que permita comparar cuantitativamente la memoria de redes con diferente topología. Luego, usarla para encontrar computacionalmente todos los mecanismos o *motifs* en redes de 1, 2 y 3 nodos capaces de guardar información de un estímulo.

Modelado matemático de redes de reacciones químicas

Ley de acción de masas

El método principal que usaremos en este capítulo es el modelado matemático de las redes de reacciones químicas como sistemas de ecuaciones diferenciales. Este tipo de modelos se fundan en la "Ley de acción de masas", que dice lo siguiente⁶³:

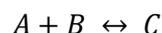
Sean dos sustancias químicas, A y B, que reaccionan al colisionar formando el producto C, entonces la tasa de acumulación de C será el producto entre el número de colisiones entre los reactivos por unidad de tiempo y la probabilidad de que una colisión supere la energía libre de activación de la reacción.



La ec. 8 traduce el enunciado en términos matemáticos, donde $[X]$ representa la concentración del producto o reactivo X y el número de colisiones por unidad de tiempo será proporcional a las concentraciones de los reactivos, con un factor de proporcionalidad k que dependerá de la geometría de las partículas y la temperatura.

$$\frac{d[C]}{dt} = k[A][B] \quad 8$$

Además, si se trata de una reacción reversible, es posible que C se disocie en A y B, con lo cual debe agregarse un término negativo y proporcional a la concentración de C, como expresa la ec. 9, donde los factores de proporcionalidad k_+ y k_- se denominarán tasas de pegado y despegado, respectivamente, cuando se trate de reacciones de unión, y tendrán unidades de $\text{Concentración}^{-1} \cdot \text{tiempo}^{-1}$ y de tiempo^{-1} .



$$\frac{d[C]}{dt} = k_+[A][B] - k_-[C] \quad 9$$

A partir de la ec. 9, podemos igualar la tasa a 0 para obtener la concentración de C en el estado estacionario (ec. 10), donde definimos la

constante de equilibrio o de disociación K (ec. 11) a partir de las tasas de pegado y despegado.

$$0 = k_+[A][B] - k_-[C]$$

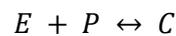
$$[C] = \frac{k_+[A][B]}{k_-} = [A][B]/K \quad 10$$

$$K \equiv k_-/k_+ \quad 11$$

Dinámica de Michaelis-Menten

La aproximación de Michaelis-Menten, muy utilizada en la literatura^{64, 65, 66}, es un modelo de la dinámica de las reacciones enzimáticas que tiene como objetivo expresar las tasas de fosforilación y desfosforilación de una proteína.

El modelo parte de las dos reacciones siguientes, donde una proteína P y una enzima E se unen reversiblemente para dar lugar a un complejo C . Luego, la enzima puede fosforilar a la proteína (P_f) y se separa irreversiblemente.



La primera hipótesis de la aproximación consiste en que la primera reacción ocurre mucho más rápido que la segunda, y que entonces el complejo se encuentra en un estado cuasiestacionario. Esto permite expresar la concentración del complejo a partir de la ec. 10, donde E y P (ec. 12) son análogos a A y B .

$$[C] = \frac{[E][P]}{K_E} \quad 12$$

La ec. 12 está expresada en términos de la enzima libre, la parte que no está unida al complejo C . Como en general la variable que se puede controlar o medir experimentalmente es la enzima total, conviene expresarlo en términos de ésta (ec. 13).

$$[C] = \frac{([E_T] - [C])[P]}{K_E} \quad 13$$

Dejando el complejo, llegamos a la ec. 14.

$$[C] = \frac{[E_T][P]}{[P] + K_E} \quad 14$$

Usando ahora la ley de masas para calcular la tasa de cambio de P_f , postulamos la segunda hipótesis de la aproximación, que consiste en que la concentración de P_f debe ser baja, o menor que la del complejo (ec. 15).

$$\frac{d[P_f]}{dt} = k_+[C] - k_-[P_f] \cong k_+[C] \quad 15$$

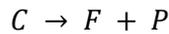
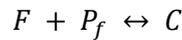
Reemplazando el valor del complejo llegamos a la tasa de fosforilación (ec. 16).

$$\frac{d[P_f]}{dt} \cong k_+ \frac{[E_T][P]}{[P]+K_E} \quad 16$$

Finalmente, podemos suponer que la cantidad total de proteína no cambia (o lo hace muy lento respecto a su fosforilación) y normalizar para que $[P_T] = 1$ (ec. 17).

$$\frac{d[P_f]}{dt} \cong k_+[E_T] \frac{[P_T]-[P_f]}{([P_T]-[P_f])+K} = k_+[E_T] \frac{1-[P_f]}{(1-[P_f])+K_E} \quad 17$$

Análogamente, podríamos haber calculado la tasa de desfosforilación (ec. 18), siguiendo los mismos pasos pero para la tasa de cambio de la proteína desfosforilada y partiendo de las dos reacciones siguientes, donde F es la enzima fosfatasa:



$$\frac{d[P]}{dt} \cong k_-[F_T] \frac{1-[P]}{(1-[P])+K_F} \quad 18$$

Sin embargo, esta expresión está en términos de la proteína desfosforilada. Por otro lado, las tasas de cambio de la proteína fosforilada y sin fosforilar son opuestas (ec. 19).

$$\frac{d[P]}{dt} = \frac{d}{dt}(1 - [P_f]) = -\frac{d[P_f]}{dt} \quad 19$$

Juntando las expresiones de las ecs. 18 y 19, llegamos a la tasa en términos de la proteína fosforilada (ec. 20).

$$\frac{d[P_f]}{dt} \cong -k_-[F_T] \frac{[P_f]}{[P_f]+K_F} \quad 20$$

Finalmente, juntando ambas tasas obtenemos la expresión típica para una proteína fosforilada por una quinasa E y desfosforilada por una

fosfatasa F (ec. 21), donde $k_{+,-}$ son las tasas de fosforilación y desfosforilación y $K_{E,F}$ las constantes de equilibrio asociadas.

$$\frac{d[P_f]}{dt} = k_+[E_T] \frac{1-[P_f]}{(1-[P_f])+K_E} - k_-[F_T] \frac{[P_f]}{[P_f]+K_F} \quad 21$$

Cuantificador de memoria molecular

El cuantificador que proponemos para la memoria molecular está basado en las diferentes definiciones de memoria expuestas en la introducción de este capítulo. El mismo está diseñado para medir cuán bien o por cuánto tiempo un sistema es capaz de recordar que recibió algún estímulo. Consiste en simplemente calcular la diferencia de estados de un sistema habiendo y no recibido el estímulo, como indica la ec. 22, donde $\bar{X}(t)$ representa la posición en el espacio de fases de un sistema cualquiera.

$$\bar{M}(t) = \bar{X}_{estimado}(t) - \bar{X}_{sin\ estímulo}(t) \quad 22$$

Ejemplo de aplicación

La Figura 33 muestra un ejemplo de aplicación en un sistema sencillo, un feedback positivo de dos nodos A y B , formado por una inhibición mutua. Si A inhibe a B , que a su vez inhibe a A , A está desinhibiéndose a sí mismo, cerrando el ciclo de retroalimentación positiva. Modelando este sistema usando la dinámica de Michaelis Menten (ec. 21), obtenemos las ecs. 23 para las tasas de cambio de A y B activos o fosforilados. Cada una cuenta con dos términos. El primero positivo para la activación, controlado por los parámetros k , K e I representando un estímulo sobre A . El segundo, negativo para la desactivación, controlado por K_{AB} y la concentración de B o A , quienes se inhiben mutuamente.

$$\begin{aligned} \frac{dA}{dt} &= Ik \frac{(1-A)}{K+(1-A)} - B \frac{A}{K_{AB}+A} \\ \frac{dB}{dt} &= k \frac{(1-B)}{K+(1-B)} - A \frac{B}{K_{AB}+B} \end{aligned} \quad 23$$

Cabe destacar que este sistema es simétrico con respecto al intercambio de $A \leftrightarrow B$, a excepción del estímulo sobre A (que no tiene B). Consiste en una competencia entre los nodos donde el que comience con mayor concentración terminará inhibiendo al otro. Así, tiene 3 puntos fijos o estacionarios. Uno estable donde "gana A ", otro donde "gana B ", también estable, y el tercero, inestable, en una situación intermedia, un "empate". La Figura 33A muestra estos puntos como círculos verde, magenta y blanco/negro respectivamente, además de muchas trayectorias en el espacio de fases. Todas terminan en algún punto fijo estable, verde si iniciaron con mayor A que B y magenta en el caso contrario.

La idea del cuantificador $\bar{M}(t)$ se basa en reconocer que es indistinto analizar la trayectoria de un sistema que recibió un estímulo transitorio a olvidarse del estímulo y tomar solamente la trayectoria del sistema sin el estímulo pero partiendo de la condición inicial donde el estímulo lo dejó en el instante que terminó.

La Figura 33B muestra cómo sería $\bar{M}(t)$ para la diferencia entre dos pares de trayectorias que parten de distintas condiciones iniciales, análogo a comparar trayectorias con y sin estímulo. En un caso, estas condiciones son opuestas en el espacio de fases (están muy lejos), pero terminan llegando a la misma posición, el punto fijo verde. En el otro, las condiciones iniciales son casi la misma (están muy cerca), pero a uno y otro lado del eje de simetría (en una hay un poquito más de **A** y en otra de **B**), y terminan en posiciones opuestas, una en el punto fijo verde y otra en el magenta. En términos de memoria, estos dos ejemplos son opuestos. En el primer caso, pese a venir de lugares bien diferentes, si uno midiera la posición final de cada trayectoria no podría distinguirlos. El sistema perdió toda información respecto del pasado de cada una y esto está expresado en que $\bar{M}(t) \xrightarrow{t \rightarrow \infty} 0$. En el segundo caso, si bien ambas trayectorias partieron de lugares parecidos, uno podría medir el estado final y deducir si la trayectoria fue una o la otra. El sistema conserva información del pasado en este caso y $\bar{M}(t) \xrightarrow{t \rightarrow \infty} 1.4 > 0$.

La Figura 33C muestra dos aplicaciones del cuantificador de memoria, para un estímulo chico y otro grande. En ambos casos, el sistema parte del punto fijo magenta y allí recibe el estímulo sobre **A**. En el primer caso, el mismo no es suficiente para que **A** "le gane" a **B** y entonces el sistema vuelve al mismo punto luego del estímulo. En el segundo caso, el estímulo es mayor y sí alcanza para llevar al sistema al punto fijo verde, donde hay una cantidad suficiente de **A** para continuar inhibiendo a **B** incluso una vez pasado el estímulo, y de manera permanente. De esta forma, estos casos también son opuestos en cuanto a memoria frente al estímulo. En el primer caso, no queda rastro en el sistema de haber sido estimulado luego de un tiempo, y podemos registrarlo con $\bar{M}(t) \xrightarrow{t \rightarrow \infty} 0$. En el segundo, por el contrario, $\bar{M}(t) \xrightarrow{t \rightarrow \infty} 1.4 > 0$.

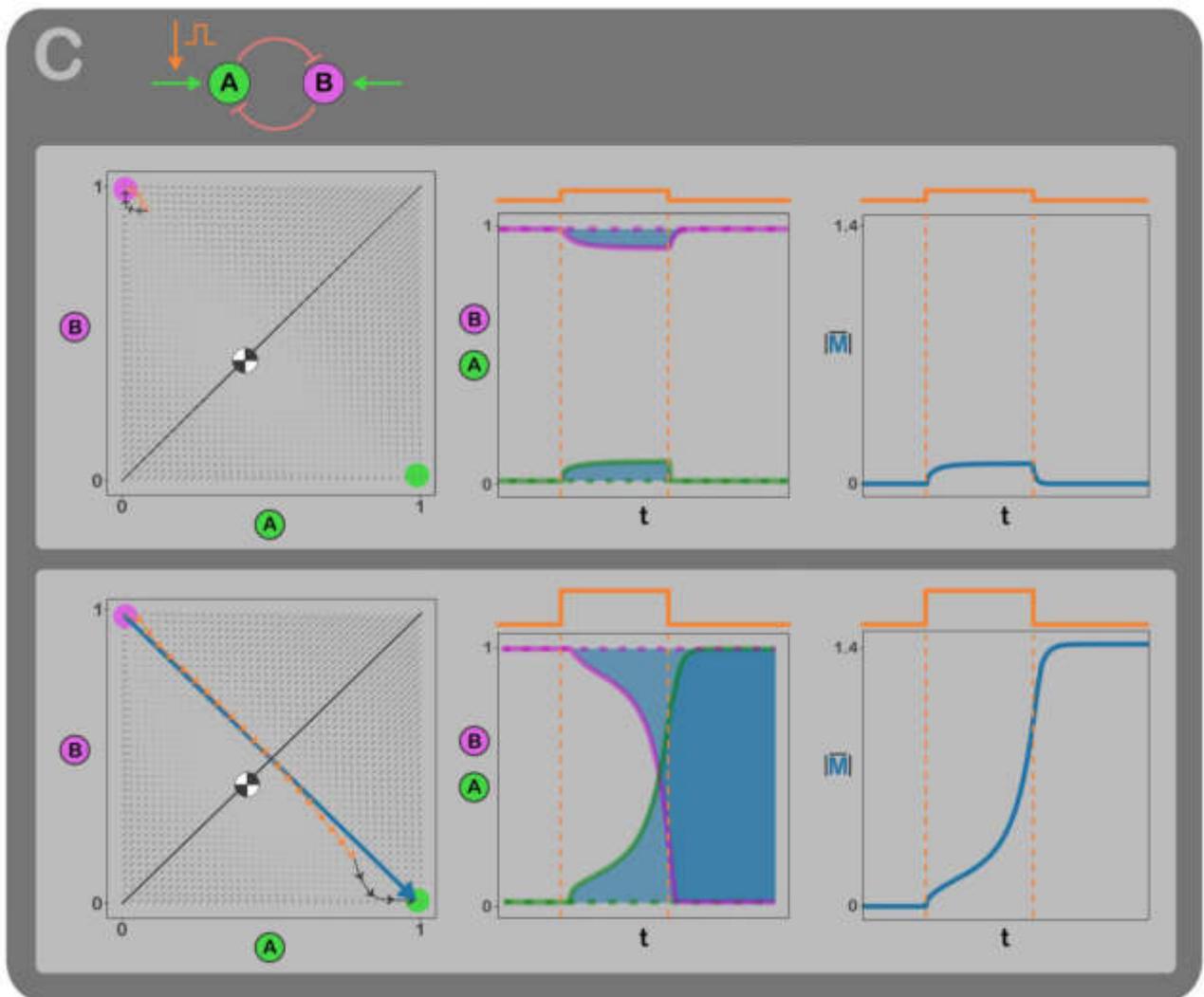
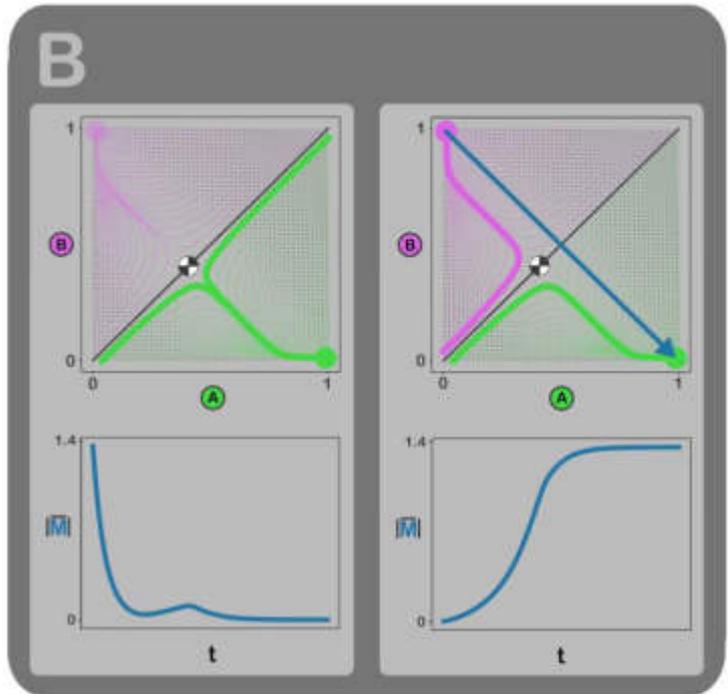
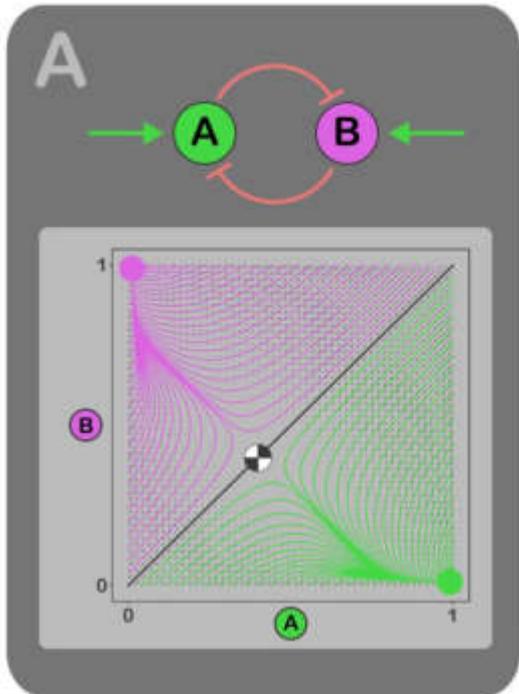


Figura 33.

A: Modelo ejemplo, feedback positivo formado por una doble inactivación (arriba). Trayectorias del sistema en su espacio de fases para un barrido de condiciones iniciales que recorren sus bordes, el cuadrado $(0,0),(1,0),(1,1),(0,1)$. Los círculos indican los puntos fijos, el verde el punto fijo estable correspondiente a donde "gana A", en magenta, donde "gana B", y en blanco y negro, el punto silla, donde negro indica la dirección de atractora y blanco la repulsora. La recta diagonal negra divide al espacio de fases. Todas las trayectorias que comiencen abajo a la derecha terminarán en el punto verde. Aquellas que lo hagan arriba a la izquierda, en el magenta.

B: Dos ejemplos de pérdida y ganancia de información de la condición inicial. A la izquierda, dos trayectorias que comienzan en puntos opuestos del espacio de fases (pero del mismo lado de la diagonal), terminan convirgiendo al mismo punto fijo. A la derecha, otras dos que comienzan en puntos muy cercanos (pero de lados opuestos), terminan en lugares opuestos del espacio de fases. En el primer caso, no se puede distinguir de dónde provino cada trayectoria midiendo el estado final. En el segundo caso, sí es posible. Abajo se muestra la memoria en el tiempo de cada caso, medida como la diferencia, en el tiempo, entre las dos trayectorias provenientes de dos condiciones iniciales distintas.

C: Dinámicas para un estímulo tipo pulso cuadrado chico (arriba) y grande (abajo). A la izquierda, las trayectorias en el espacio de fases. Ambas comienzan en el punto fijo magenta. El estímulo chico, sobre A, no alcanza para llevar al sistema a cruzar la diagonal, de manera que vuelve al mismo punto fijo y no quedará información sobre el estímulo. Por el contrario, el estímulo grande sí lo cruza y el sistema termina cambiando de punto fijo. Los gráficos centrales muestran las dinámicas de A y B, con (línea continua) y sin (línea punteada) estímulo. La diferencia entre estas dos dinámicas es el cuantificador de memoria propuesto, en azul. Éste está graficado, para estos casos, en los gráficos de la derecha. Podemos apreciar que la memoria para el estímulo chico se anula para tiempos largos, mientras que no lo hace para el estímulo grande.

Fundamentos teórico-matemáticos

Como dijimos anteriormente, el cuantificador de memoria está basado en las diversas definiciones de memoria expuestas en la introducción de este capítulo. Tomemos la que dice que memoria es la capacidad de almacenar información del pasado. En el caso de querer recordar un estímulo transitorio, esto será posible si el sistema evoluciona de manera diferente en caso de que lo haya recibido con respecto a si no recibió el estímulo. Si el sistema terminara en el mismo punto o estado, no habría nada que pudiéramos medirle que nos indique si hubo o no algún estímulo. Siguiendo esta línea, podemos empezar cuantificando la separación entre trayectorias que vienen de distintas condiciones iniciales, retomando el concepto de que es idéntico pensar en un estímulo a tomar la trayectoria desde la posición o estado donde el estímulo la dejó al terminar.

En la Figura 34A, mostramos muchas trayectorias del mismo modelo que el de la Figura 33, haciendo un barrido de condiciones iniciales, desde $(0,1)$ hasta $(1,0)$, esto es, desde A inactivo y B activo hasta A activo y B

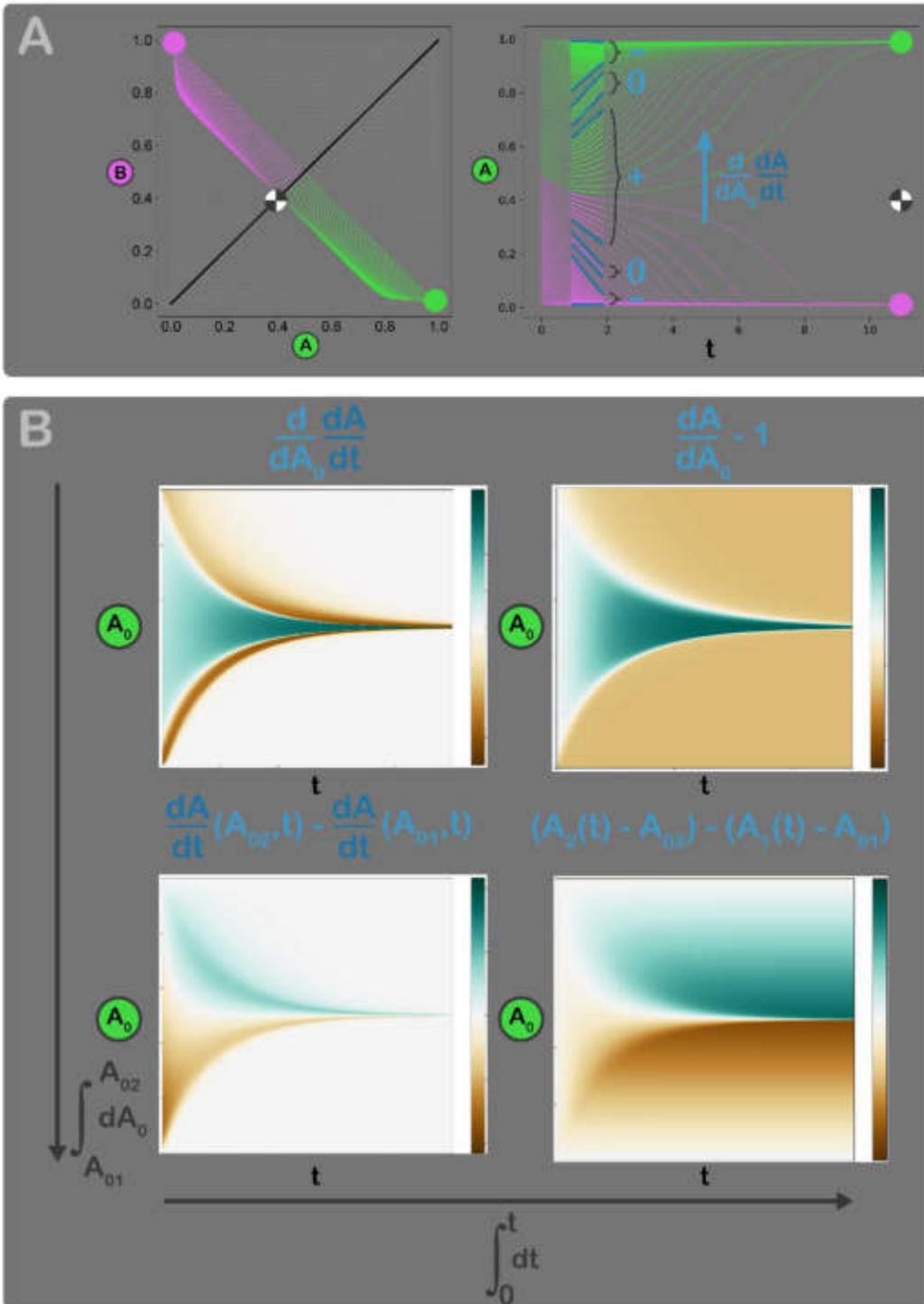
inactivo. En las curvas temporales (derecha), podemos apreciar que hay zonas donde las trayectorias se juntan, zonas donde son aproximadamente paralelas y zonas donde se separan. Esta convergencia o separación de las curvas puede cuantificarse midiendo cómo cambia la pendiente entre trayectorias cercanas. Matemáticamente, la primera derivada $\frac{dA}{dt}$ nos da la pendiente de las curvas y la segunda $\frac{d}{dA_0}$ nos da cómo esta pendiente cambia moviéndose entre curvas (ec. 24). Si hay convergencia de trayectorias, este cuantificador será negativo, si hay separación, será positivo y si son paralelas será nulo.

$$\frac{d}{dA_0} \frac{dA}{dt}$$

24

De esta forma, tenemos un cuantificador que nos indica, a cada tiempo y para cada condición inicial (o estímulo) cuánto divergen o convergen las trayectorias en el entorno y cómo el sistema está ganando o perdiendo capacidad de distinguir ese estado inicial o estímulo de otros parecidos. Esto último ligado a nuestra capacidad experimental de distinguir estados finales diferentes del sistema. Esto es, cuán precisamente podríamos medir la concentración de una molécula A y distinguir que está en uno u otro estado.

Figura 34.



A: Trayectorias en el espacio de fases para el modelo de la Figura 33A, para un barrido de condiciones iniciales desde **A** alto y **B** bajo hasta **A** bajo y **B** alto (izquierda). Dinámica del nodo **A** para las mismas trayectorias. Podemos apreciar que, en el entorno del punto silla, las trayectorias divergen, mientras que, en el entorno de los puntos estables, convergen. Es posible cuantificar esta convergencia o separación como el cambio en la pendiente de las trayectorias, a medida que nos movemos entre condiciones iniciales.

B: Pasaje entre la convergencia o separación y el

cuantificador de memoria. Arriba a la izquierda, el cambio de las pendientes (en escala de colores) en función del tiempo y de la condición inicial. Hacia la derecha, integrando respecto del tiempo, obtenemos la convergencia neta entre dos tiempos. Hacia abajo, integrando respecto de la condición inicial, obtenemos la convergencia neta entre dos condiciones iniciales. Abajo a la derecha, el cuantificador de memoria, resultado de la doble integral.

En la Figura 34B, mostramos cuánto vale este cuantificador a cada tiempo y para cada condición inicial, y cómo pasamos de mirarlo en un entorno a hacerlo de manera global. Esto es, cuánto se separan las curvas en total desde el tiempo inicial (cuando termina el estímulo transitorio), hasta algún tiempo genérico y entre alguna condición inicial y otra. Matemáticamente, esto consiste en integrar la expresión de la separación o convergencia (ec. 24) respecto del tiempo y de la condición inicial. Como éstas son variables independientes, el orden de integración puede ser cualquiera, y cada camino tiene sus valores intermedios que podemos interpretar. Mirando el gráfico de la Figura 34B, arriba a la izquierda, vemos que las condiciones iniciales más cercanas a $\frac{1}{2}$ son las que más tardan en juntarse. Esto se debe a que esta es la condición que comienza con cantidades similares de A y B , de manera que la competencia es pareja y les lleva más tiempo "ganar" a A o B . Este tiempo extra hace que, si midiéramos ahí el sistema, éste aún no habría colapsado a uno de los puntos fijos estables, y podríamos deducir aproximadamente su condición inicial.

Integrando primero respecto del tiempo (gráfico de arriba a la derecha en la Figura 34B), estamos calculando la convergencia o separación neta de las curvas en un entorno de condiciones iniciales, desde 0 hasta un tiempo genérico. El gráfico muestra que, salvo las curvas de condiciones iniciales cercanas a $\frac{1}{2}$, todos los entornos de las curvas se acercan entre sí. Esto es por la convergencia de todas las curvas a simplemente dos puntos fijos.

Integrando primero respecto de las condiciones iniciales (gráfico de abajo a la izquierda en la Figura 34B), estamos calculando la convergencia o separación neta de las curvas entre dos condiciones iniciales, $A_{0_1} = 0$ y A_{0_2} genérica. El gráfico muestra que las curvas que comenzaron con A_0 menor a $\frac{1}{2}$ se acercan mientras que las de A_0 mayor a $\frac{1}{2}$ se alejan, y que el tiempo en que lo hacen es más chico a medida que más lejos de $\frac{1}{2}$ estaban.

Finalmente, integrando respecto de ambas variables (gráfico de abajo a la derecha en la Figura 34B), llegamos a la convergencia o separación neta, desde 0 hasta un tiempo genérico y entre $A_{0_1} = 0$ y A_{0_2} genérica. El gráfico muestra que todas las curvas que empezaron con A_0 menor a $\frac{1}{2}$ convergen y que todas las que lo hicieron con A_0 mayor a $\frac{1}{2}$ divergen. Esto es, las curvas quedan divididas en dos grupos, las primeras serán indistinguibles de la condición con $A_{0_1} = 0$ y las segundas sí podrán diferenciarse.

La expresión que queda de integrar dos veces la ec. 24 está expresada en la ec. 25.

$$\int_0^t \int_{A_{01}}^{A_{02}} \frac{d}{dA_0} \frac{dA}{dt}(A_0, t) dA_0 dt = \int_0^t \left(\frac{dA}{dt}(A_{02}, t) - \frac{dA}{dt}(A_{01}, t) \right) dt =$$

$$(A(A_{02}, t) - A_{02}) - (A(A_{01}, t) - A_{01}) \quad 25$$

Y volviendo a la analogía entre estímulo transitorio y condición inicial, podemos pensar que ambos sistemas parten de la misma condición, pero $A(A_{02}, t)$ es la curva que recibe el estímulo transitorio. Así, la expresión se reduce a la ec. 26.

$$M_A(t) = A(t)_{estimulado} - A(t)_{sin estímulo} \quad 26$$

Incluyendo ahora la componente faltante del nodo B , que sería análoga, queda la expresión que da la ec. 27, que es el cuantificador de memoria propuesto.

$$\bar{M}(t) = (A(t)_{estimulado} - A(t)_{sin estímulo} ; B(t)_{estimulado} - B(t)_{sin estímulo}) \quad 27$$

Mecanismos capaces de almacenar memoria

Una vez que definimos nuestro cuantificador de memoria, decidimos utilizarlo para encontrar todos los mecanismos o *motifs* de 1, 2 y 3 nodos que sean capaces de almacenar memoria frente a un estímulo tipo pulso cuadrado. Antes de pasar al algoritmo que diseñamos para esto, introduciremos las definiciones de motif y de cómo definiremos que un motif dado tiene o no memoria.

Métodos

Motifs

El término de motif fue acuñado por primera vez en 2007⁶⁷ y se utiliza en el marco de la teoría de redes complejas, en nuestro caso, en redes de interacciones químicas, formadas por nodos y enlaces, representando agentes reactivos y las reacciones.

Un motif es un pequeño subgrafo o subconjunto de nodos y enlaces dentro de una red, que tiene alguna característica o propiedad, emergente de su dinámica interna, y que puede dotar a la red con ésta. Una analogía podría ser que serían las partes o los átomos que forman las propiedades dinámicas y las capacidades de procesamiento de una señal, como serían los transistores y microcircuitos integrados en una placa electrónica.

Así, existen muchos motifs descubiertos a partir de entonces con diversas propiedades dinámicas^{68,69}. Estos descubrimientos permitieron reducir algunos problemas de redes complejas en partes más simples, modelables, comprensibles. Por ejemplo, hoy sabemos que todos los sistemas adaptativos, aquellos que responden a estímulos estacionarios sólo de manera transitoria, seguro involucran algún tipo de *feedback loop* negativo o *incoherent feedforward loop*⁷⁰.

En este capítulo, tomaremos el concepto de motif y, utilizando el cuantificador de memoria propuesto, trataremos de encontrar todos aquellos de 1, 2 y 3 nodos que sean capaces de recordar un estímulo. Cabe destacar que, si nuestro cuantificador no fuera completamente general y aplicable a

cualquier sistema dinámico, no sería posible aplicarlo a motifs de distinta cantidad de nodos y diferentes parámetros de manera comparable.

Determinación de motif con memoria

Diremos que un dado motif tiene memoria frente a un estímulo tipo pulso cuadrado si existe al menos una región de su espacio de parámetros donde nuestro cuantificador detecte memoria. Esto significa que integraremos cada motif muchas veces para explorar exhaustivamente su espacio de parámetros. Dando el estímulo cuadrado y no dándolo en cada caso para computar la memoria. Al final, nos fijamos si el motif elegido recordó el estímulo y si así fue determinamos que tiene memoria.

En cada motif, tomamos 10 mil sets de parámetros para la exploración usando el algoritmo de *Latin Hypercube Sampling*⁷¹, que tiene el objetivo de elegir los parámetros al azar pero distribuidos homogéneamente en el espacio. Lo explicaremos en detalle en el siguiente subtítulo. Para cada uno de los 10 mil sets, integramos el motif con y sin el pulso cuadrado, hasta que la dinámica alcanza el estado estacionario ($t \rightarrow \infty$), y allí computamos la memoria ($\bar{M}(t \rightarrow \infty)$) usando nuestro cuantificador dado por la ec. 22. De esta manera, para cada motif tendremos una distribución de 10 mil valores de $\bar{M}(t \rightarrow \infty)$, y determinaremos que el motif tiene memoria en algún nodo A si $\bar{M}_A(t \rightarrow \infty) > \text{umbral}_M$ en al menos umbral_{sets} casos. Estos dos umbrales tienen el objetivo de filtrar casos donde la integración no alcanza un estado estacionario, y su elección será explicada en el siguiente párrafo.

La Figura 35 ilustra el algoritmo diseñado para determinar si un dado motif tiene memoria. Ésta compara dos motifs simples de un único nodo donde la única diferencia es una autorregulación positiva. Este feedback positivo genera casos biestables donde el sistema permanece "prendido" incluso cuando el estímulo cesó, de manera que sí tiene memoria. Podemos observar ambas distribuciones de $\bar{M}(t \rightarrow \infty)$ y vemos que, si bien en la gran mayoría de los sets $\bar{M}(t \rightarrow \infty) = 0$, existe un pico de memoria alta para el motif capaz de tener biestabilidad (distribución azul oscura). Podemos ver que este pico contiene casos que superan ambos umbrales y entonces será clasificado entre los que tienen memoria, a diferencia del motif sin la autorregulación, que no supera ninguno de los umbrales.

Cabe mencionar que, al quedarnos únicamente con $\bar{M}(t \rightarrow \infty)$, vamos a obtener solo un número por cada nodo (el correspondiente al último tiempo de integración), de manera que perdemos mucha información temporal que el cuantificador guarda. Esto lo diseñamos pensando en detectar los sistemas que tengan memoria por cambiar de punto fijo, de manera que la búsqueda estará sesgada en favor de encontrar sistemas multiestables. Cualquier mecanismo que sea capaz de recordar el estímulo de manera dinámica (sin llegar a un punto fijo) quedará descartado de esta búsqueda.

Elección de umbrales

Los umbrales fueron elegidos a partir de las distribuciones celestes de la Figura 35. Notemos que, aunque sean pocos, existen casos para el motif sin la autorregulación donde $\bar{M}(t \rightarrow \infty) > 0$. Estos casos no se deben a una memoria real del sistema, sino que son errores de integración. El origen está en que, dado el amplio barrido de parámetros, existen sets con escalas temporales muy diferentes. Entre estos sets algunos pocos son mucho más lentos que la media y la integración concluye antes de que puedan alcanzar el estado estacionario. Es por esto que usamos el integrador de Runge-Kutta de paso adaptativo, que varía el paso temporal para adaptarse a la escala propia del sistema. Aun así, el problema persiste y tiene origen en cómo definimos que el sistema alcanzó el estacionario. Decimos que esto ocurre cuando la diferencia de todas las variables entre una iteración y la anterior no es casi apreciable (no supera 10^{-5} , ec. 28)

$$|X(t_i) - X(t_{i-1})| < 10^{-5} \quad 28$$

Esto hace que existan casos de dinámica muy lenta que no alcanzan el estacionario. Si quisiéramos hacer más estricta la condición estacionaria, reemplazando por ejemplo el 10^{-5} por 10^{-6} , tenemos el problema de que todos los sets que concluían correctamente lo siguen haciendo, pero tardan mucho más en integrarse. Haciendo una prueba para estimar estos tiempos, pasamos de aproximadamente 3min para integrar mil sets a 45min. Esta diferencia es limitante si tenemos en cuenta que la exploración de todos los motifs de 2 nodos requiere integrar 10 mil sets de 16038 motifs diferentes.

Para solucionar el problema, integramos nuevamente los dos motifs de la Figura 35, pero esta vez con un umbral 100 veces más estricto para la definición de estacionario: 10^{-7} . Con las nuevas integraciones etiquetamos

cada set como efectivamente estacionario (azul oscuro) o erróneamente estacionario (celeste). Esto nos permitió reconocer, dentro de ambas distribuciones de $\bar{M}(t \rightarrow \infty)$, que todas aquellas donde $\bar{M}(t \rightarrow \infty) > 0$ para el motif sin la autorregulación estaban incorrectamente integradas, mientras que sí aparece un pico de memorias altas, correctamente integrado, para el motif con la autorregulación.

Así, optamos por mantener la definición de estacionario y aprovechar que los casos erróneos tienen una distribución de memoria diferente a los que efectivamente tienen memoria. Esta distribución no tiene un pico, sino que decae exponencialmente, y son muy raros o nulos los casos erróneos con alta memoria (Figura 35). El mayor de estos fue un caso con $\bar{M}(t \rightarrow \infty) \cong 0.78$, y es por esto que fijamos $umbral_M = 0.8$ y $umbral_{sets} = 2$. Esto implica que, para determinar que un motif tiene memoria, debe contar con más de 2 de los 10 mil sets que tengan más de 0.8 de memoria.

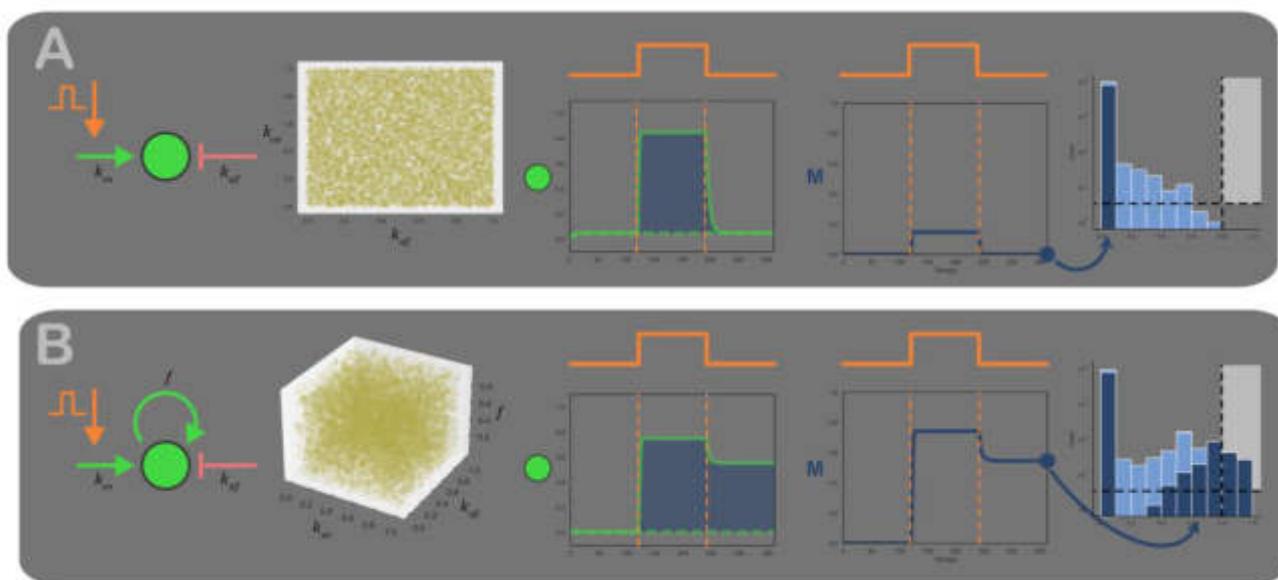


Figura 35: Algoritmo para determinar si un motif tiene o no memoria. Comparamos dos motifs control, el panel **A**, es un nodo simple, sin ninguna interacción. El del panel **B**, contiene una autoactivación. De izquierda a derecha, los modelos, el barrido de 10 mil parámetros para cada uno (notas que son 2 parámetros en el panel **A** y 3 en el **B**), las dinámicas con y sin estímulo, la memoria en función del tiempo y las distribuciones de memorias a largo plazo, para todos los sets de parámetros. El modelo del panel **A** no puede tener memoria a largo plazo, a diferencia del **B**. Los gráficos de las distribuciones, a la derecha, muestran en azul los sets integrados hasta el estado estacionario real y en celeste aquellos que no llegaron al estacionario y resultan en memorias no nulas. La zona clara indica la región a la derecha y por encima de los umbrales que filtran estos sets. Si el motif en cuestión contiene barras dentro de estas regiones, determinamos que tiene memoria.

Antes de continuar, es importante destacar que la elección de umbrales la hicimos a partir de una prueba en dos redes diferentes, y los elegimos de manera muy estricta, de manera de asegurarnos de descartar toda red que pueda no haberse integrado bien, a expensas de no encontrar redes que sí cuentan con memoria.

Latin Hypercube Sampling

El algoritmo de *Latin Hypercube Sampling*⁷¹ (LHS) es una forma de elegir conjuntos de parámetros de manera aleatoria en espacios de alta dimensión, como son frecuentemente los espacios de parámetros de mecanismos complejos. Su principal ventaja es que permite llenar el espacio a explorar de manera más uniforme que la elección completamente azarosa. Esto evita tener que tomar grandes cantidades de puntos de muestreo e integrar el sistema más veces de lo necesario, acortando los tiempos de cómputo.

Supongamos que tenemos que elegir n sets de los parámetros de la ec. 29, donde cada valor para cada uno debe estar dentro de cierto intervalo. La manera más simple quizás, sería elegir n números random para cada parámetro, en su intervalo, y terminar con n conjuntos. El problema de esta estrategia consiste en que, si tenemos muchos parámetros, la distribución de los puntos elegidos no sería uniforme, y quedaríamos con grandes sectores sin explorar, a menos que tomáramos n^p conjuntos, donde p es la cantidad de parámetros. Esta expresión crece muy rápido con p , que en nuestro caso será $p = 9$. Si quisiéramos por ejemplo dividir cada intervalo de cada parámetro en 10, y tener al menos un valor en cada subintervalo, deberíamos tomar 10^9 conjuntos. Integrar esta cantidad para cada uno de los 16038 motifs, resultaría en tiempos de cómputo imposibles.

(p_1, p_2, p_3, p_4)

29

El LHS consiste en dividir todos los intervalos de cada parámetro en la cantidad de conjuntos en total que se quieran tomar, creando una grilla cuadrada en el espacio de parámetros. Luego, ir eligiendo puntos en el espacio de manera random, pero eliminando cada fila y columna de la grilla para futuras elecciones, como ilustra la Figura 36A. Esto asegura que elijamos siempre un único valor de cada parámetro en cada una de las n divisiones, al tiempo que mantiene que el total de conjuntos sea n , independientemente de la cantidad de parámetros. Cabe destacar que el

gráfico es en dos dimensiones para dos parámetros, pero el algoritmo puede aplicarse para cualquier cantidad, donde habría que ir eliminando una línea de fila o columna por cada dimensión. En la Figura 36B, mostramos un ejemplo de las distribuciones de a pares para conjuntos de 4 parámetros, donde la diagonal contiene los histogramas de cada uno. Comparando con la elección random, podemos apreciar que el LHS genera distribuciones más uniformes en el espacio.

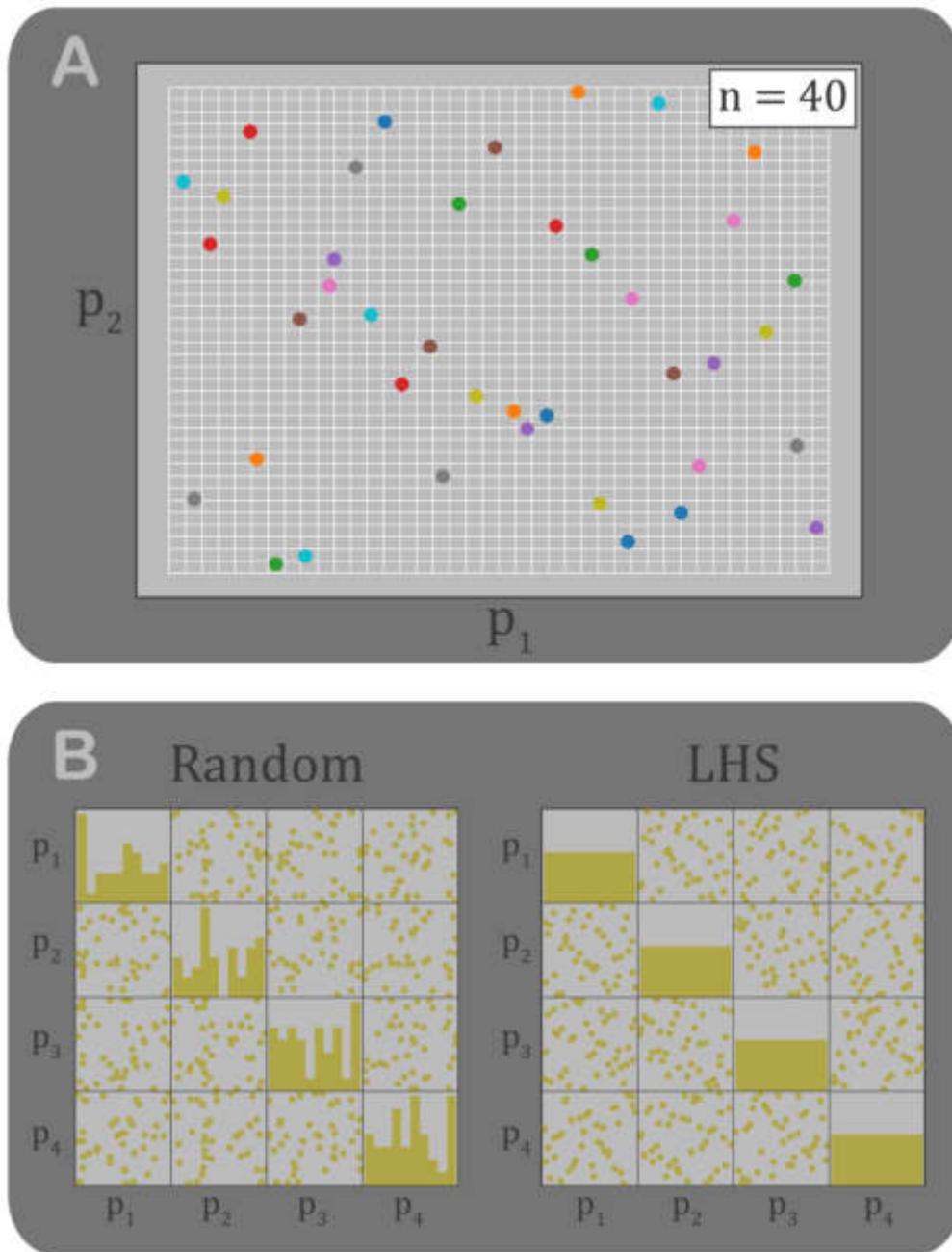


Figura 36: Esquema de elección de parámetros a través del algoritmo de Latin Hypercube Sampling.

A: Scatter-plot de 40 puntos seleccionados para un modelo de dos parámetros (p_1, p_2) . Cada rango de variación de cada parámetro se divide en tantos intervalos iguales como puntos (40) y se eligen los puntos aleatoriamente, pero uno solo por cada fila/columna que resulta de esta división. De esta forma, el espacio se llena de manera más uniforme que en la elección completamente aleatoria.

B: Scatter-plots para una elección de 40 sets de 4 parámetros (p_1, p_2, p_3, p_4) . En las diagonales, las distribuciones de cada parámetro. A la izquierda, elección puramente aleatoria. A la derecha, con el método de Latin Hypercube Sampling. Podemos apreciar que las distribuciones son más homogéneas y el espacio de parámetros también se llena más homogéneamente. Igualmente, este efecto se apreciaría mejor en espacios de dimensión más grandes, difíciles de graficar en el plano del papel.

Integración de todas las combinaciones de redes de 1, 2 y 3 nodos

Combinaciones de redes

Con el objetivo de encontrar todos aquellos motifs capaces de almacenar información de un estímulo tipo pulso cuadrado, integramos todas las redes de 1, 2 y 3 nodos dos veces cada una. Una dando el estímulo y otra sin darlo, para computar nuestro cuantificador de memoria de acuerdo a la ec. 22. La cantidad total de redes de 1, 2 y 3 nodos son 3, 54 y 16038 respectivamente. Estos números surgen de contar todas las posibles combinaciones de enlaces o interacciones entre nodos. Teniendo en cuenta que un dado nodo A , origen de un enlace, puede activar, inhibir, o no cambiar a otro nodo B , destinatario del enlace, existen 3 combinaciones posibles de interacción entre un nodo origen y otro destinatario. Incluso estos dos pueden ser el mismo, dado que queremos incluir motifs con autorregulaciones. Así, estas 3 combinaciones del nodo A se repiten por cada uno de todos los nodos de la red, incluyendo al propio A , que contará entonces con $3 \times 3 \times 3$ combinaciones en una red de 3 nodos A , B y C . Como queremos estudiar todos los posibles motifs, incluiremos interacciones recíprocas donde A pueda activar a B y B a A . De esta forma, la cuenta es análoga para los nodos B y C , que también contarán con $3 \times 3 \times 3$ combinaciones. Esto nos da, finalmente, 3^9 combinaciones, es decir, 19683 redes de interacción de 3 nodos diferentes.

Sin embargo, entre éstas existen redes donde el nodo A , que será el que reciba el estímulo, estará completamente desconectado del nodo C . En estas redes, el estímulo no puede cambiar en absoluto la dinámica de C , y entonces las descartaremos. Imaginando que los motifs están dentro de redes más grandes (como vías de señalización o redes de transcripción), supondremos que A recibe el estímulo, que atraviesa el motif y C será el que transmita la dinámica resultante al resto de la red. Si no hubiera conexión de interacciones desde A hasta C (ya sea directamente o a través de B), ese motif no sería capaz de procesar información del estímulo, o al menos transmitirla al resto de la red. Este tipo de redes que descartamos están ilustradas en la Figura 37. Descartando estas, quedan las 16038 redes de 3 nodos, 54 de 2 nodos y 3 de 1 nodo.

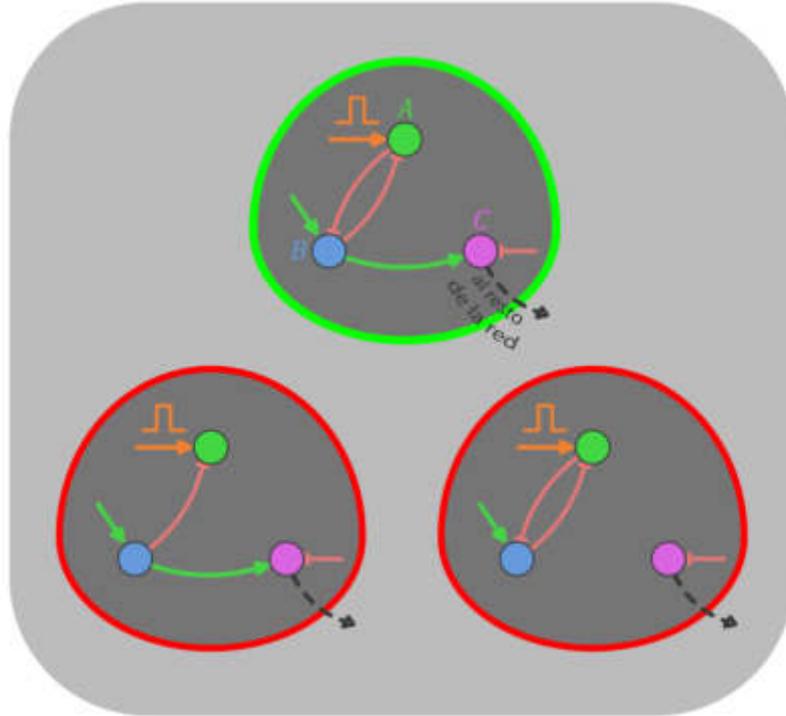


Figura 37: Ejemplos de combinaciones de redes de 3 nodos A , B y C . A siempre será el nodo que recibe el estímulo y C es el que imaginaremos conectado al resto de la red biológica. Así, descartaremos las combinaciones donde C no reciba información alguna del estímulo. En la red de arriba, el estímulo puede afectar a C indirectamente a través de B , y es una de las combinaciones que estudiaremos. En las de abajo, el estímulo no puede afectar a C . A la izquierda, porque no hay enlaces desde A hacia B , a la derecha, porque no hay enlace hacia C .

Modelo de ecuaciones diferenciales

Para cada una de estas redes, definimos el modelo como un sistema de ecuaciones diferenciales, usando la dinámica de Michaelis Menten para todas las interacciones. La ec. 30 expresa el sistema genérico, donde, para la tasa de cambio de cada nodo, sumamos sobre todos los nodos que lo activan ($\sum_{X=\{A,B,C\}}$) e inhiben ($\sum_{Y=\{A,B,C\}}$) de acuerdo con la combinación que corresponda. En la ecuación del nodo A , sumamos el término del estímulo (I).

$$\begin{aligned}
 \frac{dA}{dt} &= \sum_{X=\{A,B,C\}} Xk_{XA} \frac{1-A}{1-A+K_{XA}} - \sum_{Y=\{A,B,C\}} Yk_{YA} \frac{A}{A+K_{YA}} + Ik_{IA} \frac{1-A}{1-A+K_{IA}} \\
 \frac{dB}{dt} &= \sum_{X=\{A,B,C\}} Xk_{XB} \frac{1-B}{1-B+K_{XB}} - \sum_{Y=\{A,B,C\}} Yk_{YB} \frac{B}{B+K_{YB}} \\
 \frac{dC}{dt} &= \sum_{X=\{A,B,C\}} Xk_{XC} \frac{1-C}{1-C+K_{XC}} - \sum_{Y=\{A,B,C\}} Yk_{YC} \frac{C}{C+K_{YC}}
 \end{aligned} \tag{30}$$

Cabe destacar que existen combinaciones donde alguno de los nodos no es activado / desactivado por ningún otro. En estos casos, la ecuación diferencial correspondiente a este nodo tendrá solamente términos o

positivos o negativos, de manera que la dinámica será monótonamente creciente o decreciente, respectivamente. Por otro lado, elegimos las condiciones iniciales siempre en total inactivación. Esto es, todos los nodos comenzarán siempre completamente inactivos. De esta manera, los nodos monótonos no tendrían más opción que mantenerse inactivos (si ningún otro los activa) o activarse al máximo y permanecer allí (si ninguno los desactiva). Este factor también impide que los nodos transmitan la información del estímulo a través de la red. Para evitar este problema, y siguiendo la estrategia utilizada en la literatura⁷², agregamos un término negativo o positivo en la ecuación diferencial de los nodos correspondientes, como indica la ec. 31, donde X sería el nodo que no es activado por ningún otro e Y el que no es inactivado. Estas adiciones pueden pensarse como enzimas quinasas y fosfatasas externas, de concentración constante, que mantienen una activación o desactivación basal fija.

$$\begin{aligned}
 &+ E_{act}k_{act} \frac{1-X}{1-X+K_{act}} \\
 &- E_{inh}k_{inh} \frac{1-Y}{1-Y+K_{inh}}
 \end{aligned}
 \tag{31}$$

Para dar un ejemplo de un sistema de ecuaciones concreto, mostramos cómo queda un sistema de ecuaciones en la ec. 32, donde coloreamos en rojo y verde los términos de las enzimas basales, añadidos en las tasas de B (que no es activado por ningún nodo) y C (que no es inactivado). La combinación de enlaces está ilustrada en la Figura 38.

$$\begin{aligned}
 \frac{dA}{dt} &= -Bk_{BA} \frac{A}{A+K_{BA}} + Ik_{IA} \frac{1-A}{1-A+K_{IA}} \\
 \frac{dB}{dt} &= -Ak_{AB} \frac{B}{B+K_{AB}} + E_{act}k_{act} \frac{1-B}{1-B+K_{act}} \\
 \frac{dC}{dt} &= Bk_{BC} \frac{1-C}{1-C+K_{BC}} - E_{inact}k_{inact} \frac{1-C}{1-C+K_{inact}}
 \end{aligned}
 \tag{32}$$

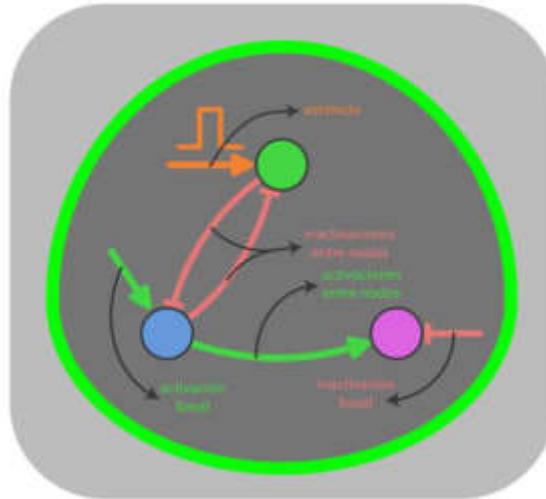


Figura 38: Ejemplo de combinación posible de una red de 3 nodos. El nodo A recibe el estímulo, en naranja. Los enlaces de inactivación los dibujamos en rojo y los de activación en verde. El nodo B no es activado por ningún otro, de manera que se agrega la activación basal. El nodo C no es inactivado por ningún otro y agregamos una inactivación basal.

Exploración de Parámetros

La cantidad de parámetros final de cada red depende de la cantidad de interacciones. Como podemos ver en la ec. 32, habrá dos parámetros por cada interacción, la intensidad k_{XY} y el umbral de activación K_{XY} . Esto es así tanto para las interacciones entre nodos como para las enzimas basales y el estímulo.

Independientemente a la cantidad variable de parámetros de cada red, elegimos 10 mil sets para cada red, usando el algoritmo de Latin Hypercube Sampling, detallado anteriormente. Integramos cada red para cada uno de estos sets, llegando a un total de más de 160 millones de sistemas de ecuaciones diferenciales. Para hacerlo, tuvimos que recurrir al servicio de computadoras virtuales de Google, que nos permitió correr, en poco menos de un mes, un algoritmo que nos hubiera llevado más de un año completarlo usando solamente las computadoras disponibles en el laboratorio.

Base de datos

Para cada set de cada red de 1, 2 y 3 nodos (Figura 39A), guardamos el vector de memoria, donde cada coordenada representa la memoria de un nodo, como ilustra la tabla de la Figura 39B. Este vector proviene de la ec. 22. Cada componente puede ser negativa o positiva, implicando que el

estímulo transitorio desactivó o activó el nodo correspondiente de manera irreversible.

A partir de esta tabla, construimos las distribuciones de memoria para cada nodo de cada red, como explicamos en la Figura 35, a diferencia que ahora las distribuciones para cada nodo pueden ir entre -1 y 1, indicando que el estímulo inactiva o activa completamente al nodo.

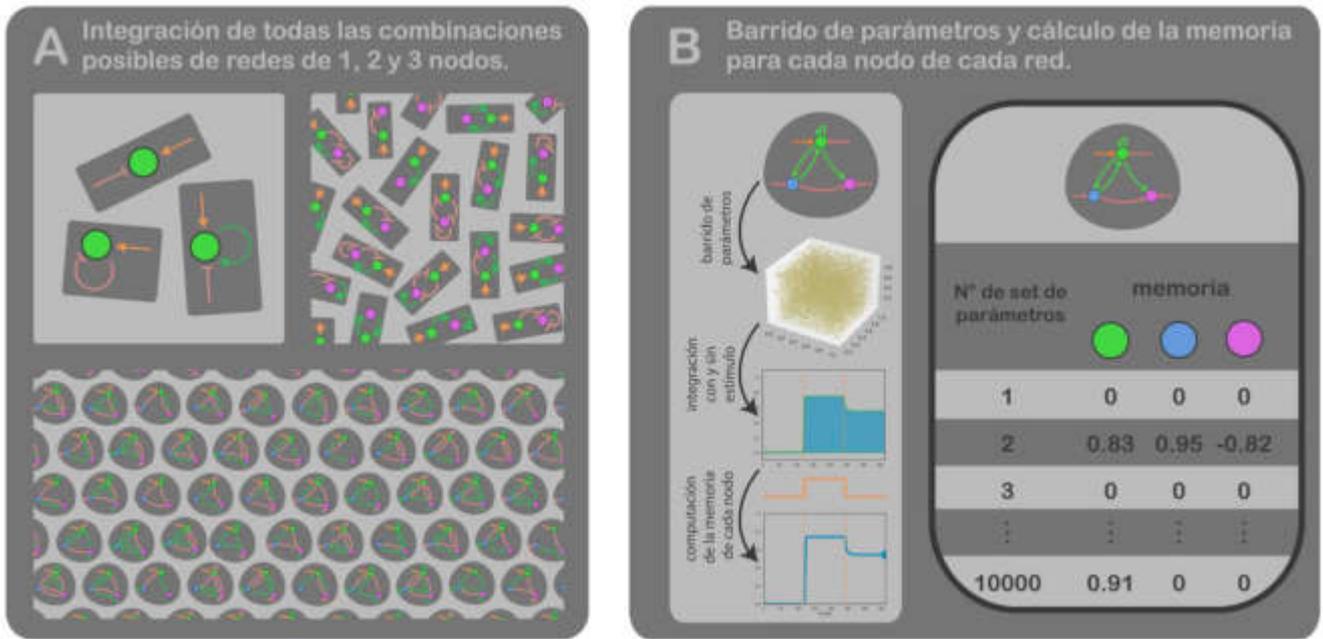


Figura 39.

A: Esquema de las combinaciones de redes de 1, 2 y 3 nodos estudiadas.

B: Generación de la base de datos para una red ejemplo. Izquierda: tomamos la red, elegimos 10 mil sets de parámetros con el algoritmo de Latin Hypercube Sampling, integramos con y sin el estímulo y computamos la memoria a largo plazo. Derecha: la base de datos obtenida es una tabla para cada red. Cada fila es un vector de 3 coordenadas, una para cada componente de cada nodo de la memoria a largo plazo.

Algoritmo de filtrado de motifs *minimales*

Algoritmo

Tomando la base de datos obtenida de la integración, y aplicando los umbrales (fijos) $umbral_M$ y $umbral_{Sets}$, detallados en Elección de umbrales, obtuvimos diferentes conjuntos de redes con memoria positiva o negativa en sus nodos. El algoritmo de filtrado de motifs minimales tiene por objetivo filtrar las combinaciones redundantes encontradas y dejar solamente el conjunto de motifs que tengan memoria, positiva o negativa, en los nodos pedidos. Un ejemplo del mecanismo está ilustrado en la Figura 40. Aquí,

pedimos los motifs *minimales* de 3 nodos que activan irreversiblemente todos al recibir el estímulo. Esto es, primero buscamos en la base de datos y filtramos todas las redes cuyos 3 nodos tengan una $M_{A,B,C} > \text{umbral}_M$ en un $N^\circ \text{ de sets} > \text{umbral}_{sets}$. Este primer tipo de filtro arroja en general alrededor de unos cientos de entre las 16038 redes de 3 nodos. En la Figura 40, mostramos un subconjunto de estas redes (panel azul) para ilustrar el segundo filtro. Podemos ver, en este subconjunto azul, que todas las redes cuentan con al menos un feedback positivo, ya sea la autoactivación de A o la doble activación de A y B , y en todas, C es activado por alguno de estos nodos. Estas combinaciones son las buscadas, ya que existirán combinaciones de parámetros donde un estímulo sobre A activará alguno de los feedbacks positivos y esto, a su vez, activará irreversiblemente a los 3 nodos. Sin embargo, queda en evidencia que en este conjunto existe una redundancia en los motifs fundamentales encontrados, donde muchas redes tienen memoria en los 3 nodos por contener un mismo motif repetido.

Para resolver este problema, definimos los motifs *minimales* como aquellos tal que, si les elimináramos cualquier enlace, dejarían de tener memoria en alguno de los nodos pedidos. Esto se computa de la siguiente manera: tomemos por ejemplo la red de borde magenta (panel azul, arriba al centro) y observemos qué redes podemos formar si eliminamos un solo enlace. Este conjunto nuevo es el panel magenta de la Figura 40, que contiene 4 nuevas redes, una por cada enlace que contenía la original y que fue eliminado. Podemos comparar el conjunto magenta con el azul y notar que dos de las 4 redes nuevas ya se encontraban en el conjunto azul, es decir, que también cuentan con memoria en sus 3 nodos. Por ende, es posible quitarle al menos un enlace a la red magenta original y no perder la memoria en los 3 nodos. Esto implica que esta red no es minimal y la descartaremos. El segundo filtro que aplicamos consiste en descartar estas redes no minimales con este método, armando las redes que se generan eliminando un enlace y comparando con el conjunto original.

Siguiendo este método, si lo aplicamos a las dos redes de borde verde del panel azul en la Figura 40, podemos ver que no obtenemos ninguna red que conserve la memoria en los 3 nodos, ya sea porque destruimos el feedback positivo o porque algún nodo queda desconectado de éste.

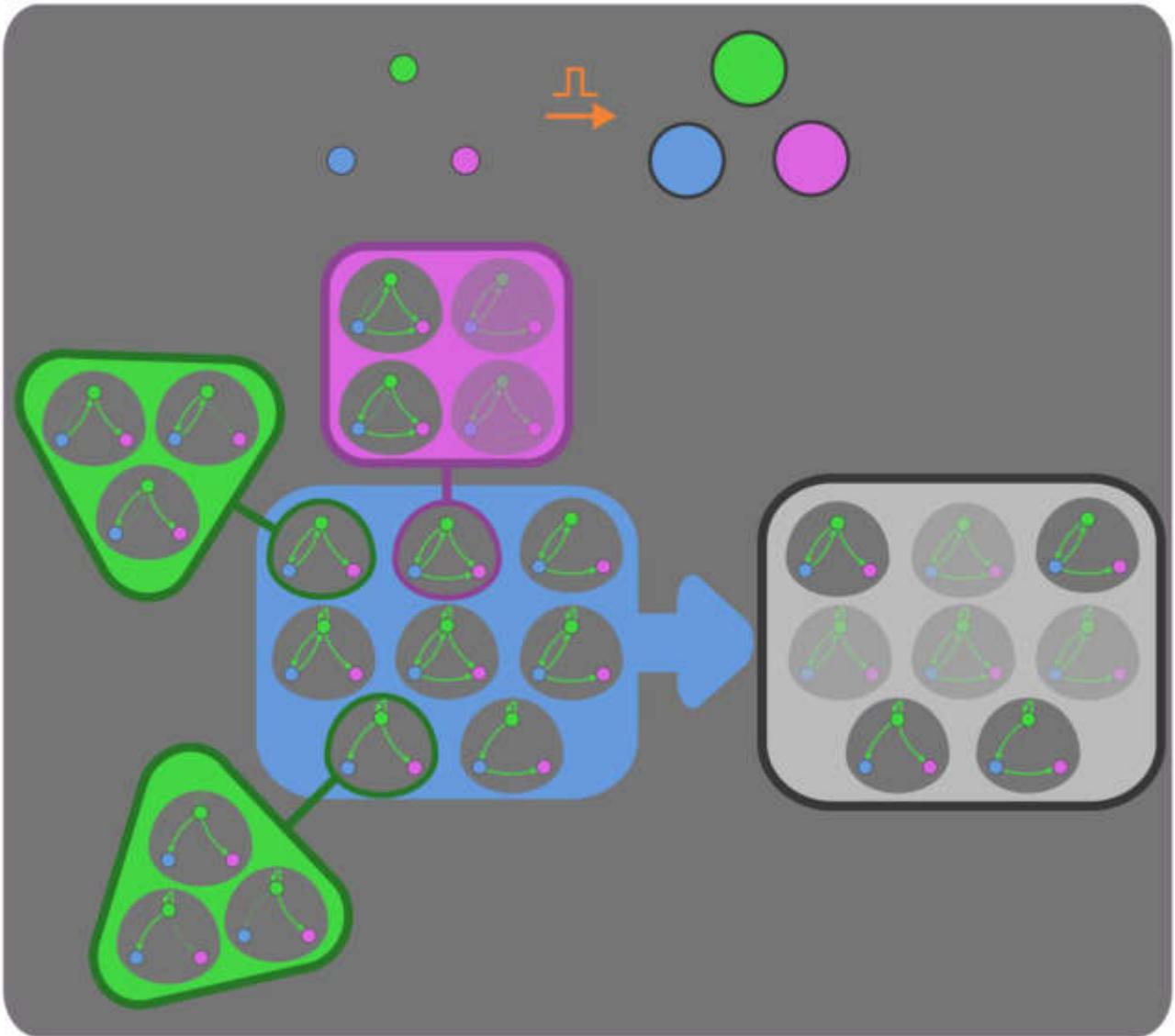


Figura 40: Esquema del filtrado de motifs con memoria. A partir de la elección del conjunto de redes con memoria (panel azul), generamos, para cada red del conjunto, todas las combinaciones posibles quitando un solo enlace (paneles verdes y magenta). El panel magenta muestra dos de las 4 redes generadas descartadas por pertenecer al conjunto original del panel azul. Así, la red magenta será descartada por considerarse un motif redundante. Las redes en verde generan conjuntos que no contienen ninguna de las redes originales, ninguna de éstas tiene memoria, por lo que consideramos que es un motif con memoria, no redundante.

Finalmente, el panel gris de la Figura 40 muestra los motifs minimales, las redes que sí atravesaron este segundo filtro. Estas forman el conjunto de todas las redes tales que el estímulo activa irreversiblemente a los 3 nodos y no hay redundancia, o no podemos obtener una de estas redes desde otra simplemente eliminando enlaces.

Resultados

En la Figura 41, paneles **A** y **B**, podemos ver todos los motifs minimales encontrados para redes de 1 y 2 nodos respectivamente. Las redes de 1 solo nodo son las tres mostradas en la Figura 39**A**, de éstas, solamente encontramos un motif minimal con memoria, el que tiene un feedback positivo debido a la autoactivación.

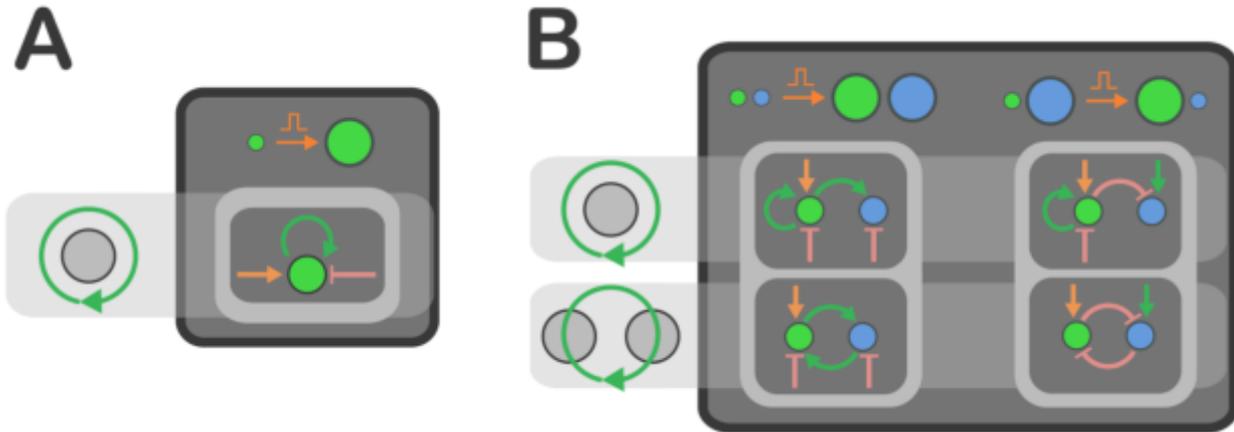


Figura 41: Motifs con memoria encontrados para redes de 1 (**A**) y 2 (**B**) nodos. Las imágenes superiores de los paneles representan el tipo de memoria pedido en cada caso. **A**: que el estímulo active al nodo **A** irreversiblemente. **B**: que active a **A** y **B** (izquierda) y que active a **A** e inactiva a **B** (derecha). No encontramos motifs donde se inactive el nodo **A**, debido probablemente a que siempre recibe un estímulo de activación. Todos los casos son feedbacks positivos y aparecen todas las combinaciones de éstos donde el nodo **A**, que recibe el estímulo, conforma el feedback.

Para las redes de 2 nodos, encontramos 4 motifs minimales, 2 de éstos donde se activan ambos nodos y otros 2 donde se activa **A** e inactiva **B**. Para los casos complementarios, donde se inactiven ambos o se inactive **A** y active **B**, no encontramos ningún motif. Esto se debe a que el estímulo siempre es de activación sobre el nodo **A**, imposibilitando que este nodo pueda inactivarse por el estímulo, independientemente de la red.

Podemos observar que los motifs minimales encontrados son nuevamente feedbacks positivos. Si miramos los casos donde se activan **A** y **B** (Figura 39**B**, columna izquierda), vuelve a aparecer la autoactivación de **A**, sumando ahora que **A** activa a **B**. Además, aparece la doble activación entre **A** y **B**, que forma un feedback positivo entre ambos nodos. La columna de la derecha muestra los motifs minimales donde se activa **A** e inactiva **B**. Estos son

también una autoactivación de *A* y un feedback positivo entre *A* y *B*. La diferencia en ambos es que, en el primer caso, *A* inactiva a *B*, de manera que éste pasa de estar basalmente activado a inactivado irreversiblemente. En el segundo caso, aparece la doble inactivación.

Los resultados para las redes de 3 nodos están ilustrados en la Figura 42, clasificados en cuatro paneles de acuerdo con si se activan todos los nodos (**A**), se activa solamente *A* e inactivan *B* y *C* (**B**), se activan *A* y *C* e inactiva *B* (**C**) o se activan *A* y *B* e inactiva *C* (**D**). Nuevamente, encontramos feedbacks positivos en todos los casos, sumando ahora la posibilidad de feedbacks entre 3 nodos. En los feedbacks encontrados de un solo nodo (primera fila) vemos que la autoactivación es siempre sobre *A*, debido a que es el que recibe siempre el estímulo. Cabe destacar que siempre aparece esta autoregulación en 3 motivos minimales diferentes, donde el cambio está en cómo *A* activa o inactiva a *B* y *C*. Observando el panel **A**, esta activación puede ser tanto directa (*A* activa a *B* y a *C* por enlaces directos) o indirecta, tanto a través de *B* (*A* activa a *B* y este, a su vez, a *C*) como de *C*.

Para el resto de los paneles, encontramos en la primera fila otros 3 motivos simétricos por cada panel. Esta simetría consiste en que un nodo puede almacenar memoria tanto activándose como inactivándose irreversiblemente con el estímulo. Esto es, si intercambiamos las proporciones activas (X) por inactivas ($1-X$) de los nodos y también los signos de todas las interacciones que los involucran, cambiando activaciones por inactivaciones y viceversa, entonces la red intercambiada también será un motivo minimal. Esta simetría en los resultados se cumplió en todos los casos hallados, sean de 1, 2 o 3 nodos, y es una prueba robusta e independiente del buen funcionamiento del algoritmo.

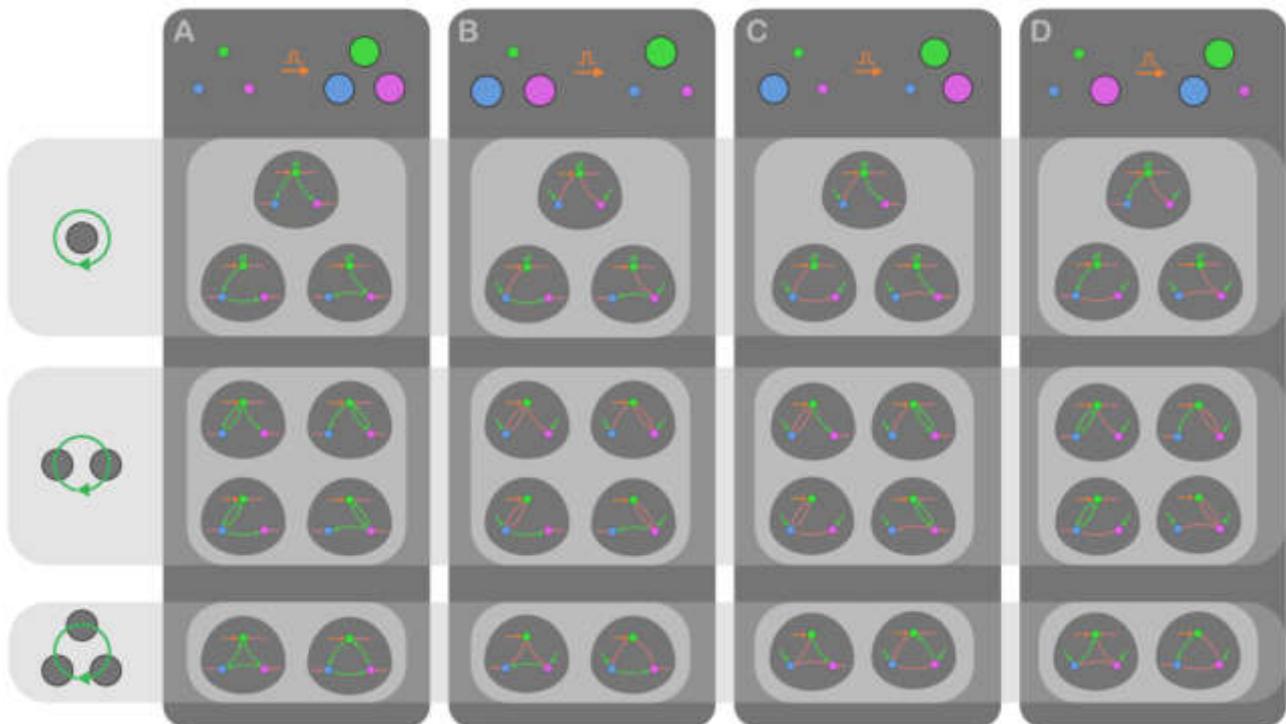


Figura 42: Motifs con memoria encontrados para redes de 3 nodos. Nuevamente, todos los casos son feedbacks positivos y aparecen todas las combinaciones de éstos donde el nodo *A*, que recibe el estímulo, conforma el feedback. **A**: podemos apreciar la simetría de intercambio de nodos *B* y *C*, todos los motifs graficados del lado izquierdo del panel son simétricos con los derechos. Lo mismo ocurre en el panel **B**. En los paneles **C** y **D**, la simetría es cruzada (los izquierdos de **C** con los derechos de **D**, y viceversa), dado que aquí se pidieron activaciones e inactivaciones opuestas para *B* y *C*. Puede apreciarse la simetría de intercambio de un nodo activo por inactivo, que relaciona por misma posición relativa a las redes de todos los paneles. Consiste en intercambiar los signos de todos los enlaces que lleguen o salgan del nodo que cambia de activarse a inactivarse entre paneles.

La segunda fila de la Figura 42 muestra los motifs minimales con feedbacks positivos en dos nodos. Podemos notar otra vez que en todos los casos el nodo *A* que recibe el estímulo forma parte del feedback. En el panel **A**, vemos que existen 4 posibilidades: que el feedback doble se forme entre *A* y *B*, o entre *A* y *C*, y en cada caso hay dos versiones donde es *A* o el otro nodo el que activa al tercero fuera del feedback. El resto de los paneles contienen los mismos motifs simétricos.

La tercera fila de la Figura 42 muestra los motifs minimales con feedbacks positivos de los tres nodos. En el panel **A**, podemos observar que hay dos versiones, en sentido horario y antihorario de las interacciones. Esta posibilidad aparece en tres nodos recién ya que es el mínimo número necesario para definir un sentido de giro, y muestra una nueva simetría: si encontramos un motif cíclico horario o antihorario entre alguna cantidad

de nodos, es esperable encontrar también su versión opuesta. El resto de los paneles contienen, nuevamente, los mismos motifs simétricos.

Por último, cabe mencionar otras simetrías que aparecen en los resultados. En los paneles **A** y **B**, donde pedimos la activación e inactivación de **B** y **C** respectivamente y en simultáneo, existe una simetría de intercambio entre **B** y **C**. Dado que **A** es quien recibe el estímulo, **B** y **C** juegan un papel equivalente. Es esperable entonces que si encontramos un motif minimal donde, supongamos, **B** activa a **C**, entonces aparezca otro motif donde **C** active a **B** de manera simétrica. Los paneles **A** y **B** nos muestran esta simetría entre todos sus motifs minimales: cada motif de la columna izquierda es el mismo de la derecha, intercambiando **B** con **C**. Esta misma simetría aparece también entre los paneles **C** y **D**. La única diferencia está en que, en este caso, en el panel **C**, pedimos que **B** se inactive y **C** se active y viceversa para el panel **D**. La equivalencia entonces es entre los motifs minimales de la columna izquierda del panel **C** con los de la derecha del panel **D** y viceversa. Están cruzados porque **B** y **C** juegan un rol equivalente cuando ambos se activan o inactivan.

Conclusiones

En síntesis, independientemente de la cantidad de nodos, encontramos el conjunto de todos los feedbacks positivos posibles que incluyen al nodo que recibe el estímulo. Sumando las simetrías que aparecen en los resultados, podemos concluir que el algoritmo funciona correctamente y que el método y el cuantificador de memoria son robustos.

Sin embargo, el hecho de que solo aparezcan feedbacks positivos indica que el método no encontró ningún motif que guarde memoria haciendo uso de alguna dinámica alternativa, diferente a la multiestabilidad. Esto podría deberse, por un lado, a que esta dinámica no tiene por qué existir. Implicando que cualquier memoria, fundamentalmente, se origina en algún feedback positivo. Por otro lado, podría también ser consecuencia de algún sesgo en el método, que haga que solamente encontremos feedbacks positivos. Teniendo en cuenta que el cuantificador utiliza solamente información del estado estacionario, es coherente que aparezcan motifs multiestables y que no hayamos encontrado dinámicas que recuerden el estímulo sin necesariamente llegar a un punto fijo. Resumiendo, la pregunta ahora es si podemos hallar motifs que den memoria y no sean multiestables.

En las dos secciones que siguen, estudiaremos los osciladores como ejemplo de dinámicas que pueden guardar memoria sin tener multiestabilidad. Después, y para cerrar el capítulo, buscaremos cuantificar primero cuánta memoria pueden darle a la red todos los feedbacks positivos (y cuáles son los mejores y peores) y comparar con los feedbacks negativos.

Memoria en osciladores

El objetivo de este análisis es mostrar que sistemas que oscilan pueden también almacenar memoria, sin necesidad de tener multiestabilidad ni feedbacks positivos. Es un mecanismo diferente a los encontrados hasta ahora, que se basan en multiestabilidades, donde el estímulo lleva a los sistemas irreversiblemente de un punto fijo a otro. En este caso, la dinámica oscilatoria no convergerá a ningún punto fijo. Esta búsqueda está motivada en que existe una relación fundamental entre sistemas oscilantes y feedbacks positivos⁷³, que sabemos que pueden dar memoria y que también pueden generar oscilaciones.

Analizamos entonces los sistemas oscilantes. Comencemos por el más simple, el único motif de 2 nodos capaz de oscilar (Figura 43A). Un ejemplo de la dinámica con y sin el estímulo para este sistema está graficado en la (Figura 43B). Podemos observar que el estímulo, una vez que cesó, no cambia cualitativamente las oscilaciones, y solamente genera cambios importantes en la dinámica por el período en que está presente. Sin embargo, comparando las dinámicas con y sin el estímulo, sí aparece un cambio en la fase de las oscilaciones. Este cambio de fase es intrínseco a la duración del estímulo, y contiene información que bien podría medir y utilizar un sistema biológico como recuerdo del estímulo. La Figura 43C muestra las trayectorias cíclicas de este sistema con y sin el estímulo, que quedan desfasadas. El vector de memoria \bar{M} , nuestro cuantificador original de la ec. 22, quedaría girando en el espacio de fases, indicando que el sistema guarda memoria y que esta información oscila entre ambos nodos (vector azul en Figura 43C).

Con este mecanismo en mente, descartamos que sea esencial la multiestabilidad para tener memoria. Sin embargo, dado que este sistema necesita del feedback positivo para oscilar, no podemos responder si los feedbacks positivos son esenciales. Más aún, no existen redes de 2 nodos capaces de oscilar sin tener al menos un feedback positivo⁶⁹. Las oscilaciones necesitan de un feedback negativo y un delay. Con solo 2 nodos, es imposible obtener el delay en el tiempo de respuesta de un tercer nodo, pero el feedback positivo aporta una histéresis que da el delay necesario. En esta línea entonces, nos preguntamos si esta histéresis, si el feedback positivo, siempre es esencial.

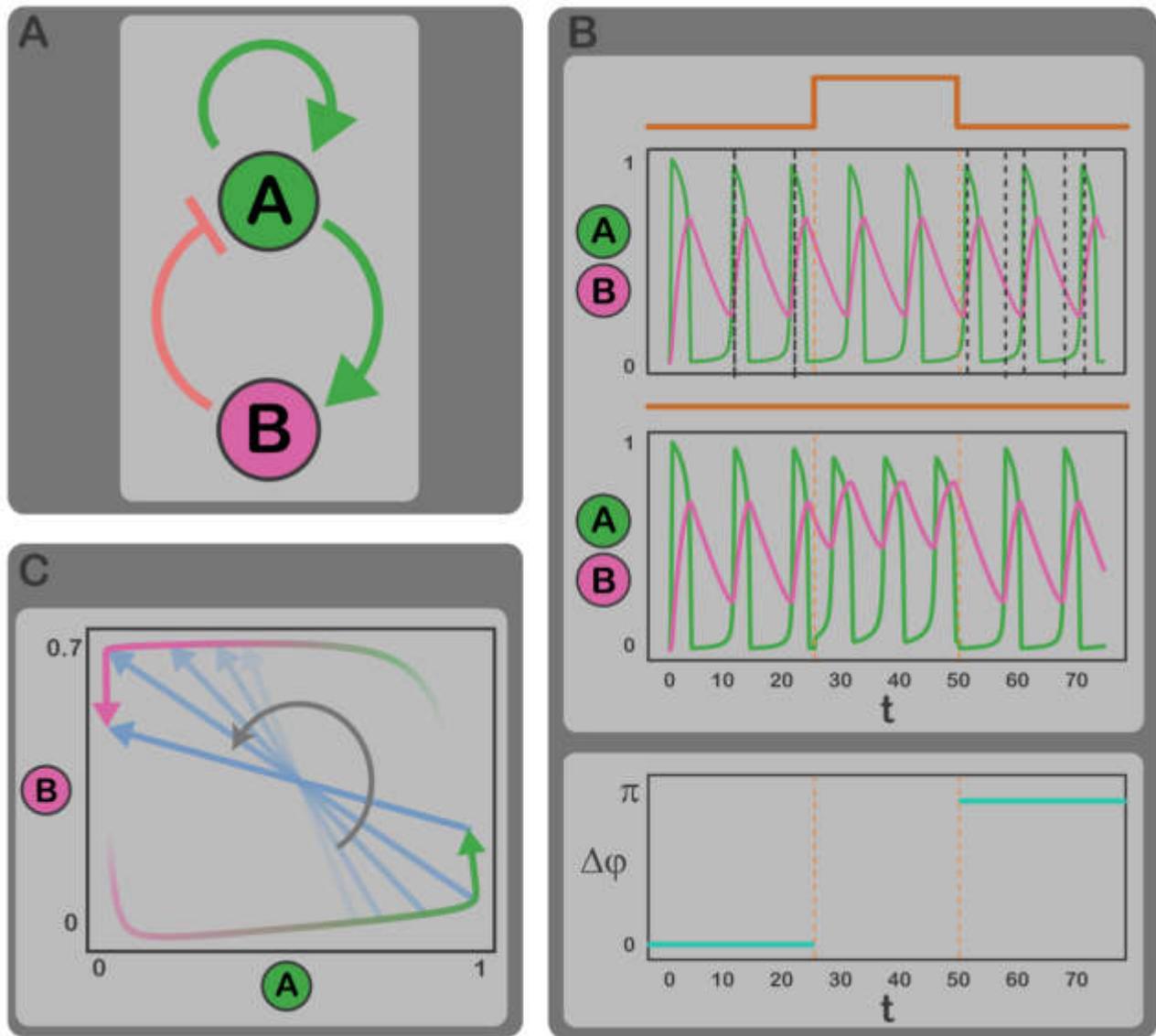


Figura 43. **A:** Motif modelo de dos nodos, el sistema más simple capaz de oscilar⁶⁹. **B:** Arriba, dinámica con (superior) y sin (inferior) el estímulo cuadrado. Podemos apreciar que el estímulo afecta la amplitud de las oscilaciones de manera transitoria, pero, en el caso de la fase ($\Delta\phi$, abajo), lo hace irreversiblemente. **C:** Trayectorias en el espacio de fases de las dos dinámicas graficadas en **B**. Los colores indican el nodo más activo de acuerdo con la fase de oscilación y las flecha la dirección en que cada una recorre el ciclo. Las flechas azules muestran, para algunos tiempos sucesivos, cómo va es el vector de memoria y la flecha circular gris, la dirección en que éste va rotando. Cabe destacar que este sistema es capaz de almacenar información del estímulo en la diferencia de fase irreversible que éste genera.

Para abordar esta pregunta, buscamos todas las redes oscilantes de 3 nodos que no tienen feedbacks positivos. Son las 8 ilustradas en la Figura 44 (paneles superiores). Las clasificamos en dos grupos por simetría: cada grupo contiene redes análogas, donde se puede llegar de una a la otra intercambiando nodos. Como son equivalentes, analizamos solamente dos, una de cada grupo. A cada una la integramos para 10 mil sets de parámetros elegidos por Latin Hypercube Sampling. Integramos todos los sets y los

clasificamos de acuerdo con si eran o no osciladores autónomos (que oscilan sin recibir estímulo periódico). Con esta clasificación, volvimos a integrar con y sin el estímulo transitorio tipo pulso cuadrado, para computar la memoria de cada set de cuerdo a nuestro cuantificador de la ec. 22. A diferencia de mirar $\bar{M}(t \rightarrow \infty)$, como hicimos anteriormente y no encontramos sistemas oscilantes, esta vez tomamos la integral en un tiempo fijo T del vector de memoria, una vez que el sistema entró en régimen estacionario. Esto apunta a tener una medida no nula de \bar{M} cuando oscila.

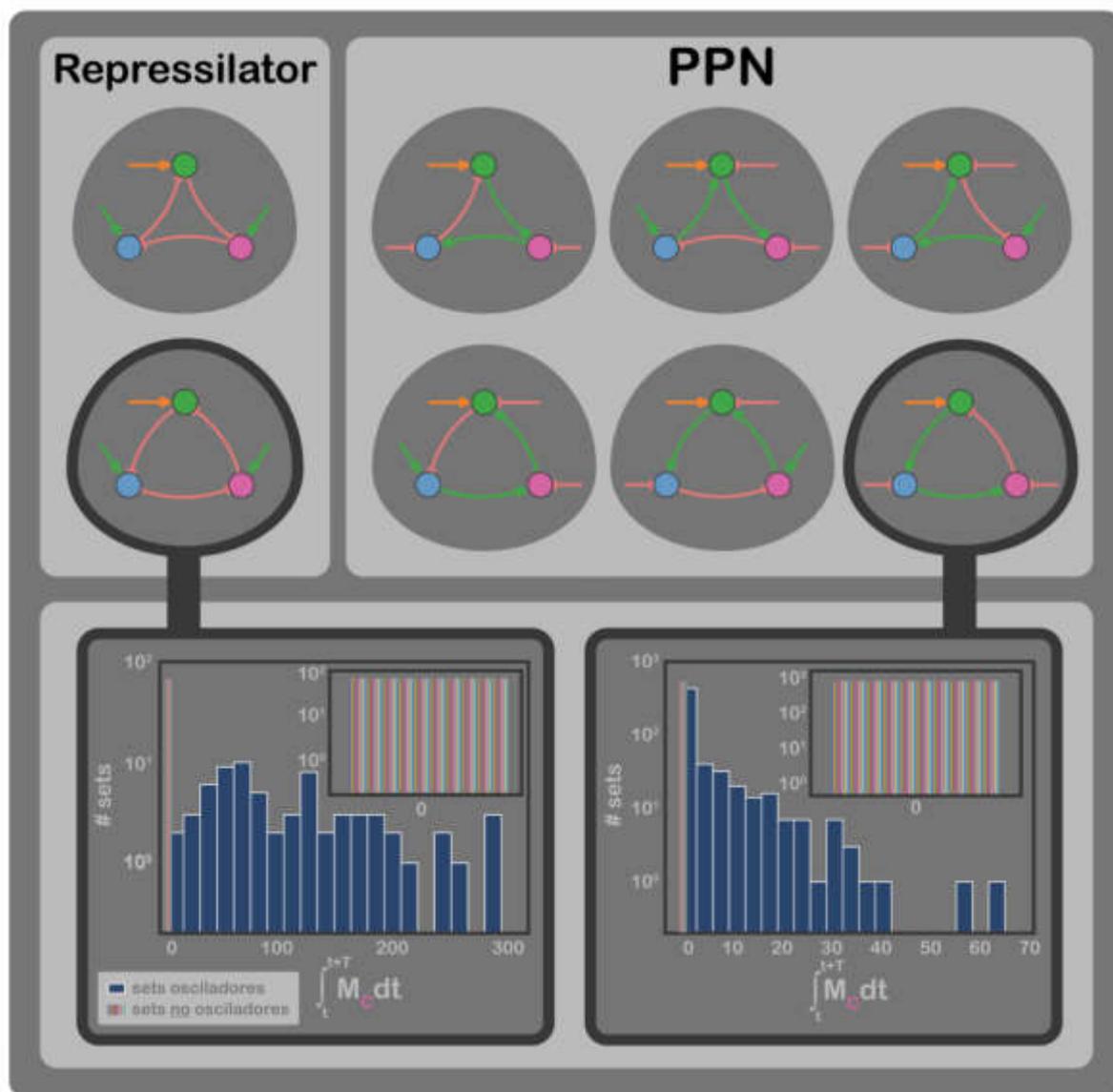


Figura 44. Paneles superiores: motifs de redes de 3 nodos capaces de oscilar sin contar con feedbacks positivos. A la izquierda, el repressilator, en sus simetrías horario y antihorario. Es un motif que ha sido hallado en la naturaleza e incluso se han logrado oscilaciones insertándolo en células sintéticamente⁷⁴. A la derecha, el feedback negativo más clásico formado por dos activaciones y una inactivación (PPN por Positivo Positivo Negativo), en sus 6 simetrías de horario y antihorario y de rotación de los enlaces. Panel inferior: histogramas de memoria de los sets osciladores (azul) y no osciladores (multicolores), de los dos motifs seleccionados (izquierda, repressilator y derecha, PPN). Podemos observar que los sets que oscilan pueden guardar memoria del estímulo, aun cuando los sistemas no cuentan con feedbacks positivos.

Resultados

Los resultados para ambas redes son concluyentes. Podemos observar, en los gráficos de la Figura 44 (paneles inferiores) que la memoria de todos los sets que no oscilan es nula, mientras que muchos de los sets que son oscilatorios, sí tienen memoria. Cabe destacar que hay grandes diferencias entre ambos osciladores. El repressilator (motif hallado en varios sistemas biológicos, incluso sintéticos⁷⁴) tiene menos sets que oscilan, probablemente porque el espacio de parámetros que da oscilaciones es más acotado⁷⁵. Sin embargo, la proporción de éstos que guarda memoria es mayor. Esto puede deberse a que el estímulo tiene más probabilidades de detener las oscilaciones temporalmente, mientras está presente, dada la mayor sensibilidad y el rango más acotado de parámetros que oscilan. Que las oscilaciones se den en una pequeña porción del espacio de parámetros significa que no son robustas antes los cambios, como es un estímulo. Si el estímulo logra detener las oscilaciones, es más probable luego que al finalizarlo, éstas vuelvan desfasadas.

Aporte de feedbacks positivos y negativos a la memoria

El objetivo de esta sección es ampliar lo encontrado anteriormente y probar nuevos métodos, usando nuestro cuantificador, para medir la memoria de algún motif de interés. Este nuevo método se basa en comparar cuánta memoria gana o pierde una red cuando le agregamos el motif de interés. La dificultad de esta medición radica en que un mismo motif puede ser agregado en una red cualquiera y su efecto sobre la memoria de ésta no va a ser siempre el mismo, sino que va a depender, a priori, de la red sobre la que se agregue, de sus enlaces y parámetros. Lo que buscamos es una medida global de cuánta memoria aporta un motif a una red en general. Con este objetivo en mente, diseñamos el algoritmo de comparación de vecinos.

Algoritmo de comparación de vecinos

El objetivo de este algoritmo es medir de manera global cuánta memoria suma o resta el agregado de un determinado motif de interés a una red. Para esto, compararemos la memoria de todas las *redes vecinas* con y sin el motif. Llamaremos *redes vecinas* a aquellas que se encuentren a una única interacción de distancia, es decir, que podamos llegar de una a la otra simplemente agregando o quitando una sola interacción entre nodos. A partir de esta definición, es posible armar una metarred que representa un espacio topológico donde podemos ubicar, de manera ordenada, todas las combinaciones de redes. Cada nodo de la metarred representa una de las combinaciones. Cada enlace une *redes vecinas*. Cabe destacar que, dado que no es lo mismo agregar que quitar una interacción, los enlaces de la metarred son dirigidos, y elegimos arbitrariamente que apunten en la dirección en que se agrega la interacción. En la Figura 45A (panel superior), ilustramos a modo de ejemplo cómo queda la metarred para las tres combinaciones de redes de 1 nodo. Podemos observar que es simplemente una línea donde hay una red central, la que no tiene ninguna autorregulación, unida a las dos restantes, con autorregulaciones positiva y negativa. Las flechas de la metarred indican el sentido en que se agregan las interacciones en las redes nodo, y las clasificamos en verdes y rojas para agregados de interacciones de activación e inactivación respectivamente.

En la Figura 45A (panel inferior), mostramos una parte de la metarred de redes combinación de 2 nodos. Recordemos que, como cada interacción puede ser de activación, inactivación, o no existir, por cada interacción existen 3 posibles combinaciones. Esto genera una línea de 3 combinaciones en nuestra metarred, por cada interacción. Para 2 nodos, las interacciones posibles son 4 (de *A* a *B*, de *B* a *A* y ambas autorregulaciones), de manera que la metarred tendrá la topología de un hipercubo de $3 \times 3 \times 3 \times 3$. En la Figura 45A, mostramos una porción de la metarred, solo las combinaciones que tienen fijas las interacciones de *A* a *B* (activación) y de autorregulación de *A* (nula). Cabe destacar que estas metarredes no tienen ciclos, y entonces podemos ordenar los nodos por su número de enlaces desde el primero (el que no tiene ninguna interacción) hasta los últimos, que tienen todas las interacciones posibles.

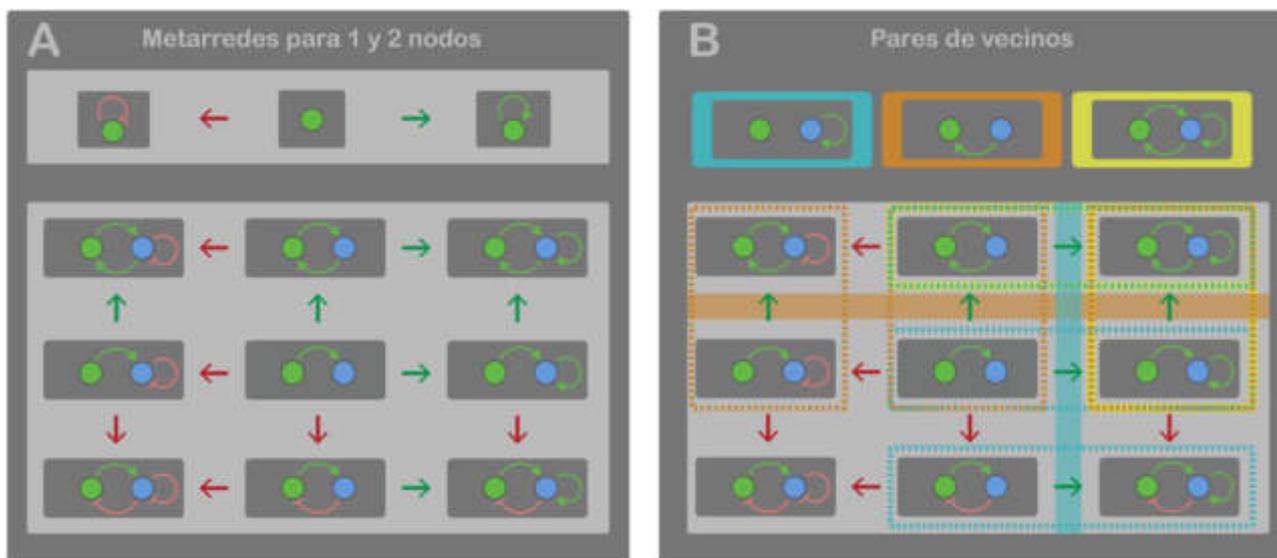


Figura 45

A: Metarredes para todas las combinaciones de redes de 1 (arriba) y 2 (abajo) nodos. Los nodos son las combinaciones de redes y los enlaces dirigidos significan que se agrega una activación (flecha verde) o inactivación (flecha roja). Cada fila o columna está formada por las 3 posibilidades (activación, inactivación, no interacción) de una única interacción. De esta forma, podemos construir un espacio topológico donde ubicar todas las combinaciones de redes y medir distancias entre éstas.

B: Método de identificación de pares de redes vecinas con y sin un motif de interés. Arriba, 3 motifs ejemplo, donde el amarillo contiene las interacciones del cyan y el naranja. Abajo, los pares de vecinos quedan delimitados por las líneas perpendiculares a la dirección en que se agrega cada interacción del motif de interés. En recuadros punteados, los pares de vecinos de los motifs cyan, naranja y amarillo.

Volviendo al algoritmo, imaginemos que deseamos medir cuánta memoria suma o resta agregar una autoactivación sobre *B* en redes de dos nodos, un motif simple formado por una sola interacción (Figura 45B arriba, red cyan). La idea consiste en detectar todos los pares de redes vecinas en la metarred, donde una vecina contiene el motif (la autoactivación) y la otra no. Estos

pares están marcados en la Figura 45B (abajo, recuadros punteados cyan), donde podemos apreciar que no son más que los que se encuentran a un lado y otro de la línea vertical (cyan) que divide a la red y es perpendicular a la dirección en que se agrega esta interacción (la autoactivación). Esto mismo es aplicable a cualquier motif de una interacción, por ejemplo, la activación de *B* a *A* (Figura 45B arriba, red naranja), donde obtenemos los pares de vecinos con y sin esta interacción (Figura 45B abajo, recuadros punteados naranjas), que se encuentran a un lado y otro de la línea horizontal (naranja). Si quisiéramos los pares de un motif con más de una interacción, la misma idea es generalizable. Pensemos, por ejemplo, en el motif compuesto por ambas interacciones anteriores (Figura 45B arriba, red amarilla). Notemos ahora que solo una red contiene este motif, la que resulta de la intersección entre los conjuntos de pares de vecinos de cada interacción. Esta red es la que se encuentra delimitada por ambas líneas divisorias. Cruzar de allí la línea horizontal es "romper" el motif por quitar la activación de *B* a *A*, y cruzarla por la línea vertical es por eliminar la autoactivación de *B*.

En la Figura 45, mostramos solo una parte de la metarred de 2 nodos que en su totalidad tiene 4 dimensiones. En la metarred entera, un motif cualquiera formado por 2 interacciones aparecerá en 9 combinaciones diferentes y tendrá dos vecinas por cada una, las que surjan de quitar cada interacción del motif, esto es, habrá 18 pares de vecinas. La ec. 33 expresa la manera en que cuantificamos cuánta memoria suma o resta el motif de interés, a través de un número que llamamos *MCV* por *Memoria Comparativa entre Vecinos*. Consiste en una sumatoria sobre todos estos pares de vecinas, donde cada sumando está compuesto por la diferencia de memorias entre las vecinas con y sin el motif, medida con nuestro cuantificador (ec. 22). Si la diferencia es positiva, la memoria de la combinación que contiene al motif es mayor que la que no lo tiene y si es negativa, vale la inversa. Esta diferencia se suma sobre los 10 mil sets de parámetros, para tener una medida global de la red que sea independiente de un set particular de parámetros.

Finalmente, sumamos sobre todos los pares de vecinas y normalizamos, dado que no todos los motifs tendrán la misma cantidad. Esto nos dará la idea global buscada sobre si el agregado del motif de interés mejora o empeora la memoria de un estímulo.

$$MCV = \frac{1}{\# \text{ pares}} \sum_{\text{pares}} \left[\sum_{10000 \text{ sets}} \left(\left| M_{C_{\text{con motif}}}(t \rightarrow \infty) \right| - \left| M_{C_{\text{sin motif}}}(t \rightarrow \infty) \right| \right) \right] \quad 33$$

Cabe destacar que para este análisis elegimos quedarnos solamente con la memoria del nodo *B* o *C* para 2 y 3 nodos respectivamente. Esto se debe a que, pensando que éstos son los nodos Output que conectarían con el resto de la red biológica, nos interesa medir cuánta memoria es capaz de transmitirle a este resto un determinado motif.

Motivación para aplicar en feedbacks

Aplicamos el algoritmo sobre todos los posibles feedbacks positivos y negativos en redes de 2 nodos. El interés en los feedbacks positivos está motivado en que son los motifs encontrados previamente, y ya podemos afirmar que les dan memoria a las redes. Funcionan, por un lado, como control positivo de este cuantificador y, por otro, los resultados permitirán comparar cuáles son los que suman más o menos memoria.

En cuanto a los feedbacks negativos, es sabido que éstos disminuyen el rango dinámico de las variables involucradas⁶⁹, siendo el rango dinámico el conjunto de valores que toma una variable en toda su dinámica. Comparemos por ejemplo los dos modelos de la Figura 46A, cuya única diferencia está en la autoinactivación del nodo. En la Figura 46B, podemos ver que la dinámica no alcanza valores altos al agregar la inactivación, dado que la misma no le permite al nodo activarse completamente. Si pensamos ahora en un sistema biestable, es posible que uno de los puntos fijos se encuentre justo dentro del espacio que queda fuera del rango dinámico, de manera que un feedback negativo podría romper el mecanismo de memoria por multiestabilidad. Con esta hipótesis, estudiamos los feedbacks negativos para ver si funcionan efectivamente restando memoria.

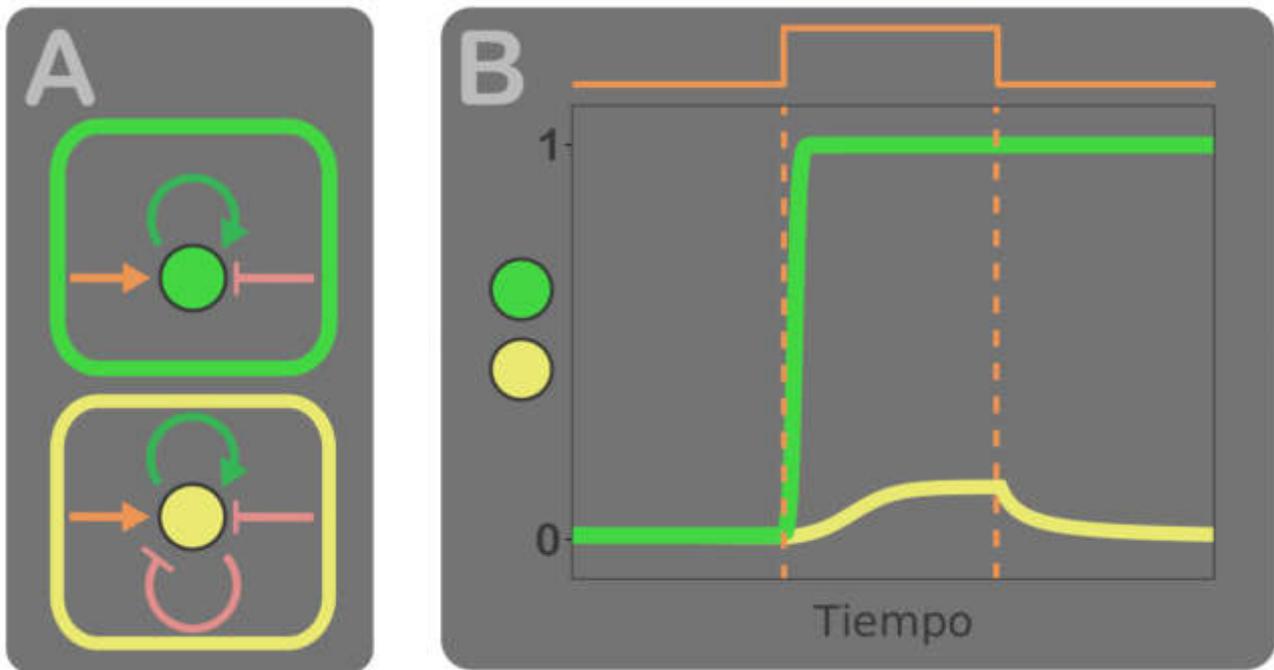


Figura 46. **A:** Redes a comparar, ambas tienen una autoactivación que les da biestabilidad y difieren en una autoinactivación. **B:** Dinámica de ambas redes cuando reciben el estímulo transitorio. La autoinactivación disminuye el rango dinámico y no permite alcanzar el punto fijo cercano a 1, quitándole memoria del estímulo a la red.

Resultados

Redes de 2 nodos

La Figura 47 muestra los resultados para los 4 feedbacks positivos (izquierda) y 4 negativos (derecha) posibles. Podemos ver que todos los positivos parecen aportar positivamente a la memoria, es decir, que agregar cualquier feedback positivo a una red de dos nodos va a contribuir, en general, a darle memoria, como era esperado. Para los feedbacks negativos, también obtuvimos los resultados que esperábamos, todos aportan negativamente a la memoria, es decir, que agregar cualquiera a una red de dos nodos, va a, en general, reducir su memoria.

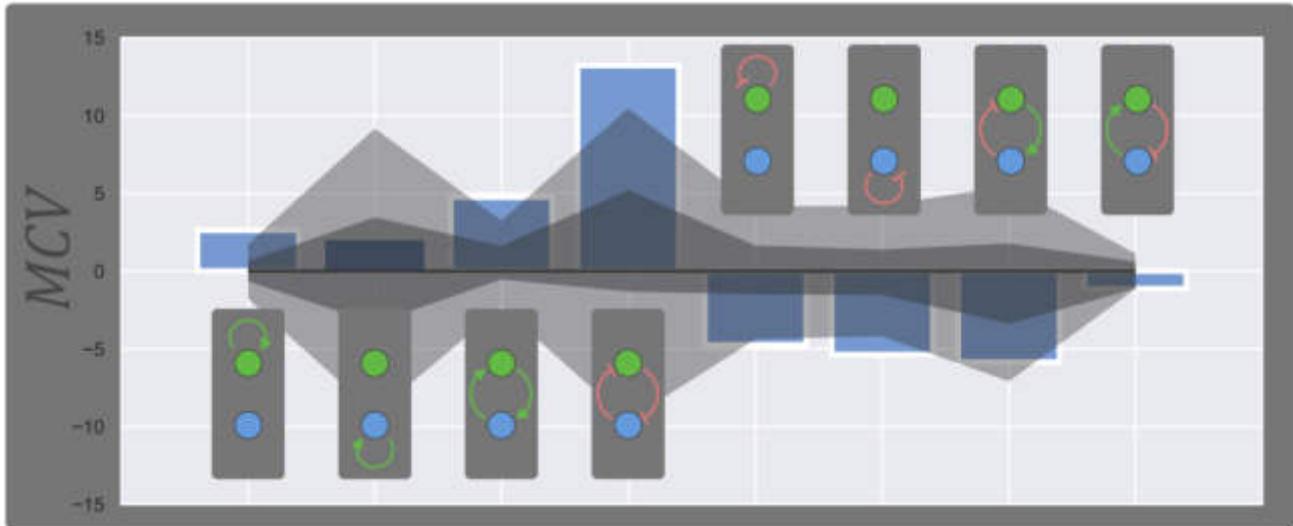


Figura 47: Memoria comparada entre vecinos para todos los feedbacks positivos (4 combinaciones de la izquierda) y negativos (4 combinaciones de la derecha) de redes de 2 nodos. Cada barra azul de cada columna representa la memoria comparada del motif dibujado. Las bandas de control representan los intervalos de confianza del 50% (gris oscuro) y 95% (gris claro). Podemos apreciar que, en todos los casos positivos, los motifs les otorgan memoria a sus redes, mientras que, para los casos negativos, restan. Sin embargo, cabe destacar que muchos valores no están muy por encima de las distribuciones control.

Analizando más en detalle, podemos comparar y ver que, de todos los feedbacks positivos, la doble inactivación es el motif que da más memoria a las redes. Esto podría deberse a que la competencia entre nodos aumenta el rango dinámico: el estímulo podría llevar a uno de los nodos de la total inactivación a la activación y viceversa con el otro, como analizamos en el modelo de la Figura 33. Mayor rango dinámico implica mayores chances de alcanzar un segundo punto fijo con el estímulo, aumentando la memoria. Esta doble inactivación funciona como *switch*, como una llave controlada por el estímulo que puede encender o apagar un programa entero de cambio en la expresión génica, como sucede en el mecanismo de diferenciación estudiado en los Capítulos 1 y 3 de esta tesis¹⁰ (Figura 5).

Es importante destacar que estos resultados no se alejan tanto de las bandas de control, que indican la distribución de valores de *MCV* que esperaríamos obtener simplemente por el azar en los parámetros elegidos, esto puede deberse a que la cantidad de pares de vecinos en redes de dos nodos no es tan alta, dado que no es tan grande la metarred. En la siguiente sección, detallaremos cómo fueron obtenidas las bandas.

Redes de 3 nodos

La Figura 48 muestra los resultados para los 17 feedbacks positivos (panel superior) y los 17 negativos (panel inferior) posibles. Nuevamente, vemos que en general los feedbacks positivos suman memoria a las redes, mientras que los negativos restan.

Analizando en detalle los feedbacks positivos, observamos que parecen dar más memoria aquellos en los que el nodo *C* forma parte. Probablemente, se deba a que elegimos que éste fuera el nodo output sobre el que aplicamos el cuantificador. Esto implica que, si un nodo en particular es el que está conectado a una red biológica mayor, es mejor que el mismo forme parte del feedback positivo que da memoria a la red.

Pasando a los feedbacks negativos, parece que aquellos donde el nodo *A*, que recibe el estímulo, forma parte del feedback siendo el nodo inactivado, son los mejores para restar o eliminar la memoria. Esta evidencia apoya la hipótesis de la reducción de memoria por disminución en el rango dinámico. Si el nodo que recibe el estímulo no alcanza un amplio rango dinámico, tendrá menos posibilidades de activar o inhibir considerablemente al resto, y de hacer que la red alcance otro punto fijo si lo hubiera, mecanismo ilustrado en la Figura 46.

Bandas de Control

Las bandas de control de las Figura 47 y la Figura 48 muestran la distribución para los valores de MCV esperados de cada motif en caso de que éste fuera completamente indiferente en términos de sumar o restar memoria. Para armar estas distribuciones, tomamos los dos términos dentro de la doble sumatoria de la ec. 33, el positivo que corresponde a un set de parámetros de la red con el motif y el negativo que corresponde al mismo set, pero de la red sin el motif. Del total de estos pares de términos de la sumatoria (dado por los 10mil sets de todos los pares de vecinos), elegimos la mitad de manera aleatoria y les cambiamos el signo. Esto genera que en la mitad de los pares de términos la red vecina con el motif suma y en la otra mitad reste, rompiendo el sesgo de memoria que genera el motif, lo que queríamos medir con el MCV. Repetimos esta medida 10 mil veces, siempre eligiendo mitades diferentes de manera aleatoria. Al final, obtenemos con este método distribuciones que representan los valores

esperados por azar si el motif de interés fuera indiferente en cuanto a sumar o restar memoria de la red. En la Figura 47 y la Figura 48, representamos estas distribuciones con sus medianas (las líneas negras), y sus intervalos de 50% y 95% de confianza (en gris oscuro y claro respectivamente). Cabe destacar que, para todos los motifs, las distribuciones generadas estuvieron centradas en 0, como era esperable de romper el sesgo de memoria de cada motif.

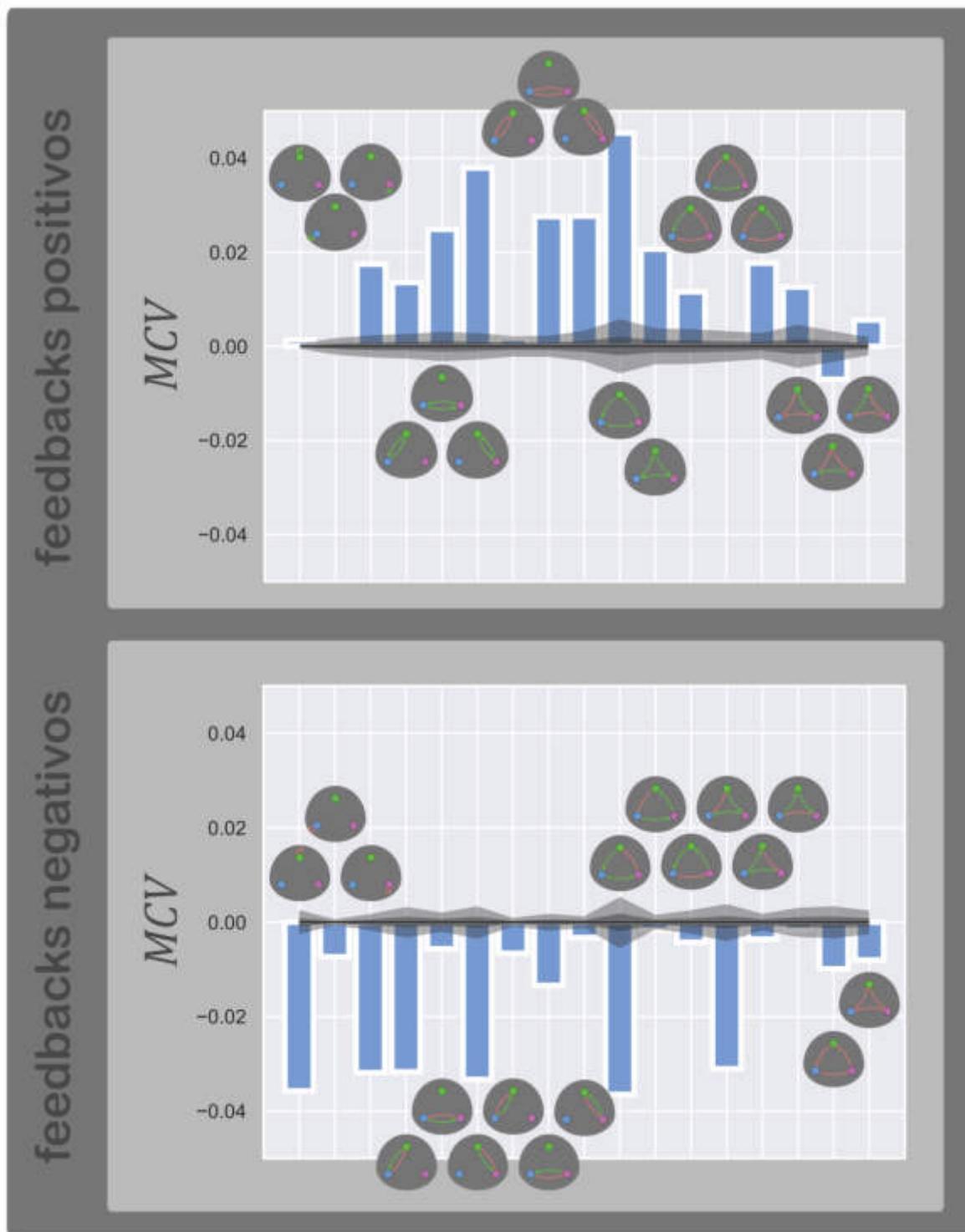


Figura 48. Memoria comparada entre vecinos para todos los feedbacks positivos (panel superior) y negativos (panel inferior) en redes de 3 nodos. Cada barra azul de cada columna representa la memoria comparada del motif dibujado. Las bandas de control representan los intervalos de confianza del 50% (gris oscuro) y 95% (gris claro). Podemos apreciar, al igual que en la Figura 47, que la mayoría de los casos positivos suman memoria a sus redes mientras que los negativos restan. Además, pareciera convenir que el nodo *C*, el output, forme parte del feedback positivo para dar memoria. Por otro lado, para restar memoria, el nodo *A*, que recibe el estímulo, preferentemente debería formar parte del feedback negativo.

Conclusiones y perspectivas

En resumen, en este capítulo definimos un cuantificador de memoria aplicable a cualquier sistema dinámico y que permite comparar la memoria de un estímulo en diferentes sistemas. Este cuantificador presenta ventajas frente a algunos propuestos anteriormente. Uno de estos⁴⁷, consiste en contar la cantidad de oscilaciones remanentes de un sistema a medida que se le va quitando un estímulo que decae exponencialmente. Este cuantificador tiene las desventajas de que no permite medir la memoria en sistemas que no oscilen y que ésta solo puede medirse en casos muy particulares, para sistemas que oscilan y que reciben un pulso que decae exponencialmente. No es una medida que permita comparar la memoria entre dos sistemas cualesquiera.

Un cuantificador propuesto en otro trabajo previo⁷⁶ consiste en medir el tiempo que un sistema tarda en alcanzar la mitad de su máxima respuesta, luego de un estímulo transitorio. Este cuantificador asume que el sistema va a dejar de responder y por ende no se puede aplicar a sistemas con memoria permanente o a largo plazo. Por otro lado, la medida dependerá de la variable que se elija como respuesta y no es una medida propia del sistema entero, como nuestro vector de memoria que incluye a todas las variables. Esto impide la comparación directa entre sistemas diferentes. En este mismo trabajo plantean medir más de un tiempo de decaimiento posterior al estímulo, con cada tiempo correspondiente a una variable diferente. Sin embargo, esto mantiene la asunción de que todas estas variables presentarán un decaimiento y no se unifica los valores en un vector, lo que impide poder hacer operaciones vectoriales que tengan en cuenta al sistema como un todo.

Usando este cuantificador, estudiamos qué motifs combinaciones de 1, 2 y 3 nodos pueden darle memoria a una red. En el primer estudio, nos limitamos a mirar la memoria a largo plazo y encontramos nada más que sistemas con feedbacks positivos que dan lugar a multiestabilidad. La memoria en estos casos se basa en que el estímulo lleva al sistema de un punto fijo estable a otro. Cabe destacar que este estudio depende en gran medida de los umbrales elegidos, los que se seleccionaron pensando en la memoria a largo plazo como consecuencia de la biestabilidad. Sería

interesante repetir el trabajo con mayor poder de cómputo y de manera de no tener que interrumpir la integración numérica.

Otro punto consiste en que solamente calculamos las memorias de cada combinación en el estado estacionario de un punto fijo, es posible que hayamos dejado fuera del estudio mecanismos que no se basen en la multiestabilidad y que tampoco oscilen, los que estudiamos posteriormente. Una pregunta válida es qué otras dinámicas pueden existir en sistemas de 3 nodos, acotados, que ni lleguen a un punto fijo estable ni se queden oscilando. Desde ya, sí sabemos que dejamos afuera del estudio mecanismos que cuentan con biestabilidad, pero transitoria. Son aquellos donde la biestabilidad no perdura una vez terminado el estímulo, y entonces el sistema vuelve siempre al punto fijo original.

Por otro lado, los resultados de los motifs encontrados muestran muchas simetrías, donde aparecen los mismos motifs que resultan tanto de intercambiar nodos, como de intercambiar la activación por inactivación de un nodo. Estas simetrías halladas son completamente independientes al algoritmo de búsqueda, e indican un grado importante de robustez en los resultados y en el protocolo utilizado. Sería interesante repetir el análisis dando un estímulo de inactivación en lugar de activación. Una predicción natural proveniente de nuestros resultados indicaría que encontraríamos los motifs análogos con la simetría de intercambio de activación por inactivación en el nodo que recibe el estímulo.

Otro punto interesante consiste en que sería posible armar un algoritmo de búsqueda de motifs en base a las simetrías halladas. La idea consistiría en, una vez encontrado un motif, aplicar todas las simetrías de intercambio y chequear si también cuentan o no con la propiedad de interés. La ventaja de este algoritmo consistiría en reducir la cantidad de redes a integrar y serviría para la búsqueda de otras propiedades dinámicas más allá de la memoria.

Para continuar y profundizar la búsqueda de motifs con memoria, sería interesante probar con otras dinámicas diferentes a las ecuaciones de Michaelis-Menten, que son más aplicables a cascadas de señalización. Un ejemplo podría consistir en linealizar las ecuaciones, método más utilizado en modelos de transcripción. Incluso cada nodo podría modelarse como un gen que produce el mRNA que sintetiza, a su vez, la proteína.

Otro trabajo adicional pendiente consiste en incluir al estímulo dentro de la exploración de parámetros. Cabe mencionar que un dado estímulo llevará a un sistema a otro punto fijo estable solamente si supera cierto umbral, dado por los parámetros de cada red. De esta forma, al no variar la amplitud del estímulo seguro quedaron redes que tenían memoria sin ser detectadas.

En la parte de la memoria en osciladores, encontramos nuevamente que el cuantificador funciona, incluso en sistemas de dinámica estacionaria más compleja que la de puntos fijos. Resulta particularmente interesante destacar el mecanismo de almacenamiento de información de un estímulo en la fase del oscilador. Este mecanismo bien podría representar lo que sucede con el reloj de segmentación del pez cebra⁷⁷, donde las células obtienen la información sobre cuándo segmentar de acuerdo con la fase del reloj, que es fijada por un estímulo previo. Por otro lado, también se ha demostrado que hay neuronas que se activan específicamente en una dada fase de oscilación⁷⁸, y que esto puede sincronizar varias neuronas y fijar una memoria específica a largo plazo que dependerá de la fase del oscilador.

Sin embargo, cabe destacar que fuimos a estudiar específicamente estos motifs, que no los habíamos encontrado en la búsqueda original, dado que no tienen estados estacionarios fijos. Por otro lado, son evidencia de que podría haber aún otros mecanismos dinámicos más complejos capaces de almacenar memoria, que aún no hemos encontrado. En este sentido, sería muy interesante repetir el estudio diseñando diferentes observables a partir de nuestro cuantificador que, al ser un vector que varía en el tiempo, contiene mucha información y da muchas posibilidades. En particular, podría utilizarse para estudiar memoria en sistemas caóticos donde la dinámica puede ser mucho más rica.

Actualmente, en el marco de la tesis de licenciatura de Paula Sánchez, estamos poniendo a prueba el cuantificador en sistemas oscilatorios y comparándolo con otros propuestos en la literatura^{47,76}.

Finalmente, armamos la metarred de combinaciones, un grafo que contiene a todas las combinaciones de redes de manera ordenada y nos permite ubicarlas dentro de un espacio topológico donde podemos definir distancias entre una y otra. En el estudio de comparación de la memoria entre pares de vecinas, mostramos que tiene mucha utilidad para comparar tanto la

memoria como cualquier otra propiedad dinámica que puedan tener los motifs. El análisis nos permitió concluir que la gran mayoría de los feedbacks positivos añaden memoria a una red, mientras que los negativos restan, probablemente por disminuir el rango dinámico de las redes.

Sería muy interesante pensar en este tipo de metarredes para catalogar y registrar otras propiedades dinámicas en este espacio topológico, como puede ser la adaptación o la detección de cambios en un estímulo en lugar de valores absolutos, recientemente estudiadas^{72,68,79,80}. Por otro lado, el ordenamiento y clasificación de las redes dentro de un espacio topológico abre las puertas para diseñar algoritmos de búsqueda de propiedades que no se basen en la exploración e integración completa de todas las combinaciones sino en exploraciones que vayan recorriendo la metarred, tomando decisiones de hacia dónde explorar en función de si mejora o no alguna propiedad de interés. Esto sería una manera de abordar este tipo de búsquedas en redes de más de 3 nodos, donde ya las combinaciones son tantas que se hace imposible integrarlas todas.

Capítulo 3

Biestabilidad y memoria en la diferenciación celular

Introducción

En este capítulo, buscaremos encontrar evidencia de memoria frente a diferentes estímulos en el mismo modelo biológico que el Capítulo 1: células madre pluripotentes de embrión de ratón, en estado naïve. El estímulo que daremos será un pulso de señales de diferenciación, que llevará las células al estado primed. Buscaremos medir la memoria aplicando nuestro cuantificador propuesto en el Capítulo 2.

La diferenciación es un proceso complejo que genera el cambio de identidad celular. Para que este cambio suceda, es necesario que la célula altere los niveles de expresión génica a escala global. Los estímulos de diferenciación gatillan programas enteros de cambio de expresión génica, muchas veces irreversibles.

Antes de presentar los experimentos diseñados, introduciremos mecanismos generales de regulación génica y algunos mecanismos particulares que controlan el paso del estado naïve al primed en estas células.

Redes de regulación génica

Las redes de regulación génica (GRN por *Gene Regulatory Networks*) han ido ganando mucho terreno en la literatura durante los últimos años^{81,82,83,84,85,86,87,88}. Consisten en redes cuyos nodos son los genes de un individuo y los enlaces, dirigidos, son las interacciones entre estos genes. Estas interacciones pueden darse a muchos niveles de regulación diferentes. Diremos que un gen *origen* regula a otro gen *target* si su expresión afecta de alguna manera la expresión del *target*. El mecanismo subyacente a la regulación podría variar. Por ejemplo, que la proteína que codifique el gen de origen se una al gen *target*, funcionando como factor de transcripción que activa o inhiba su expresión. También podría ser que el gen origen codifica un miRNA y éste inhibe la síntesis de la proteína del otro gen. Independientemente del mecanismo subyacente, estas regulaciones representarán enlaces dentro de las GRN.

La Figura 49 muestra un gráfico de una porción de la GRN que controla la formación de tejidos mesodermo y endodermo en el erizo de mar. Estas redes tienen una estructura jerárquica y modular⁸⁹. Contienen un núcleo

formado por genes que codifican factores de transcripción que regulan la expresión de otros factores de transcripción, siendo responsables de ejecutar programas enteros de cambio de expresión a nivel global, regulando qué camino de diferenciación seguirá la célula o cuál será el cambio de identidad. A su vez, estos módulos secundarios, asociados a cada linaje, regulan genes *efectores*, es decir, todos aquellos cuyas funciones son específicas de las tareas de cada tipo celular y no regulan la expresión de otros genes.

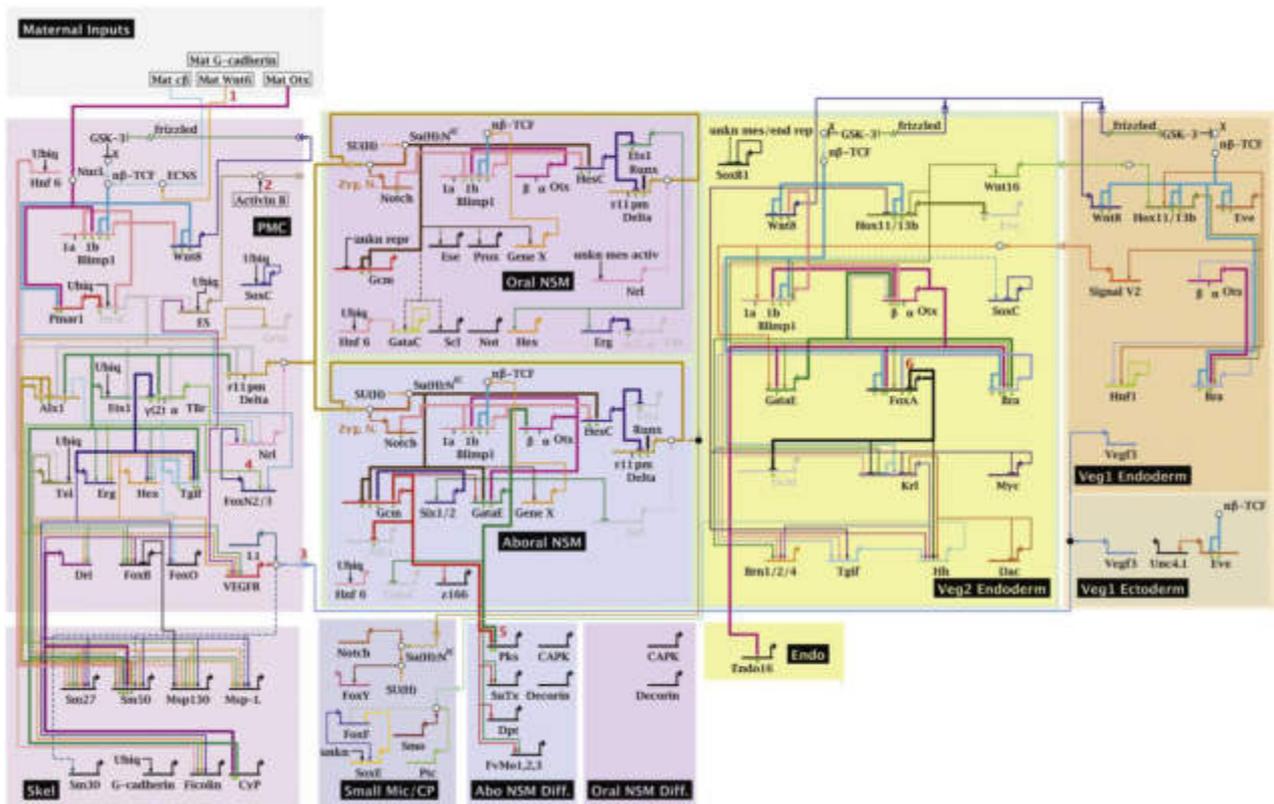


Figura 49. Imagen adaptada⁹⁰. Red de regulación génica (GRN) que controla la diferenciación a mesodermo y endodermo en el erizo de mar. Cada gen está simbolizado por su nombre y promotor y cada enlace de interacción dirigido es la flecha que sale de un gen origen y llega al promotor del gen target. Todos los genes representados codifican factores de transcripción o componentes de vías de señalización, es decir, controlan la expresión global. Solamente los paneles inferiores muestran genes efectores, que codifican proteínas cuyas funciones no son regular la expresión de otras. Los datos para construir esta red provienen de experimentos de perturbación, análisis dinámicos y patrones espaciales de expresión. Las flechas en negrita indican que la interacción fue validada por CHIP-seq. Esta red predice correctamente los patrones espaciales y temporales de casi la totalidad de los genes que contiene⁹¹.

La construcción de este tipo de redes es una tarea muy difícil a nivel experimental, ya que implica la demostración de causalidad en cada una de las interacciones, y no es suficiente con tener evidencia de que cambiando la expresión de un gen se modifica la de otro. Una ventaja a la

hora de encontrar estas interacciones es medir dinámicas de expresión, donde la diferencia temporal entre el cambio de expresión de un gen y otro puede ser evidencia de causalidad. Sin embargo, también existe la posibilidad de estar observando interacciones indirectas, donde hay genes intermedios en la red y no hay una interacción real entre los genes observados.

Construcción experimental

Una técnica muy efectiva para el armado de las GRN es el CHIP-seq⁹² (por *Chromatin Immuno Precipitation Sequencing*). El objetivo de este experimento es averiguar cuáles son los genes target de algún gen origen de interés. La técnica se desarrolla en dos pasos. Primero, se fijan las uniones entre la cromatina y todas las proteínas unidas a ésta y se separan de las proteínas citoplasmáticas. Luego, se trata con un anticuerpo específico para la proteína factor de transcripción que codifica el gen de interés y se aprovecha esta unión específica para separar la cromatina unida al factor de transcripción del resto del genoma. Finalmente, se secuencian la cromatina separada, obteniendo así el código de los genes target.

La ventaja de esta técnica está en que permite construir de una vez todos los enlaces salientes de un dado gen. Además, una vez que se conocen estos genes target, pueden detectarse las secuencias de los sitios de unión para el factor de transcripción, y buscarlos en otras secuencias de interés. Sin embargo, la secuenciación continúa siendo un proceso costoso y, para construir toda la GRN, es necesario repetir el CHIP-seq para cada gen involucrado.

Módulos de Regulación en Cis

Los Módulos de Regulación en Cis o CRMs (por *Cis Regulatory Modules*) son segmentos de la cromatina que se encuentran sobre la misma molécula de DNA, cerca de un gen codificante (unas decenas de kbp) y regulan su expresión. Entre éstos encontramos promotores y *enhancers*.

Los promotores son los sitios afines a la RNA polimerasa y sus cofactores, las proteínas encargadas de sintetizar el mRNA de los a partir

de los genes. Tienen entre 100 y 1000 pares de bases (bp) y se encuentran en general inmediatamente próximos al sitio de inicio de la transcripción.

Los enhancers son segmentos de la cromatina con gran densidad de sitios de pegado de factores de transcripción, de entre 50 y 1500 bp, que se encuentran cerca del promotor de uno o varios genes (hasta 1 Mbp de distancia). Algunas de las diferentes combinaciones de factores de transcripción pegados a los enhancers atraen otros cofactores que permiten la formación de un *loop* de ADN, que une a su vez al enhancer con el promotor y activa la transcripción, proceso denominado *looping*⁹³ (Figura 50). De esta manera, los enhancers funcionan como verdaderas compuertas lógicas⁹¹, donde solo la presencia o ausencia de determinados factores de transcripción activará o inactivará la transcripción de un gen. Se estima que al menos ocupan una región total del ADN equivalente a la codificante para proteína, y que incluso podría ser el doble. Estas compuertas lógicas son las que generan las prácticamente infinitas posibilidades de combinaciones para el control específico de la expresión génica, que requiere, para cada identidad celular, un patrón de expresión único.

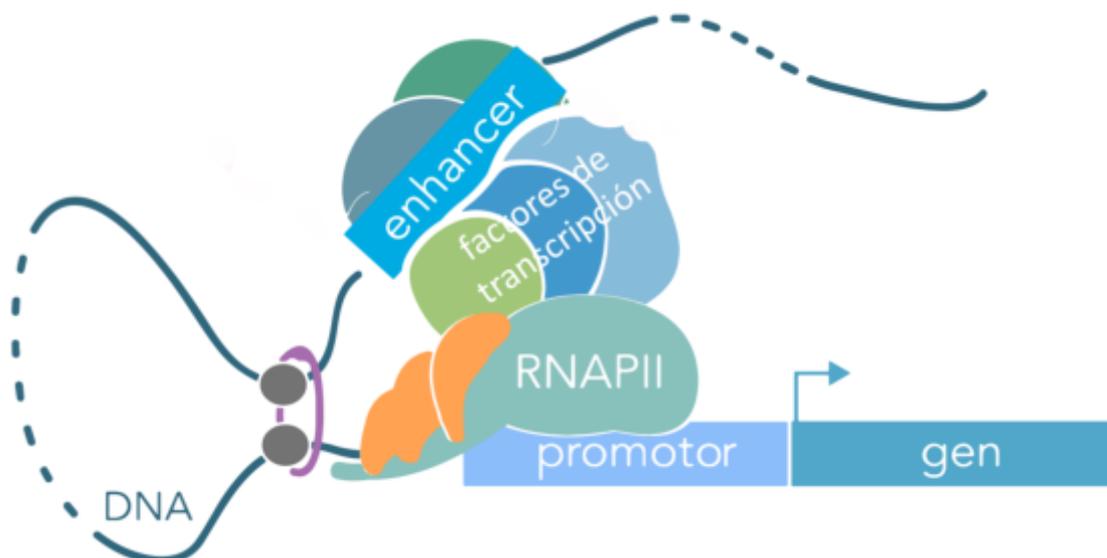


Figura 50: Imagen adaptada⁹³. Mecanismo de *looping* de DNA. Los factores de transcripción afines a sitios específicos en un enhancer se unen y activan la maquinaria proteica que los une al promotor de un gen target cercano. El *looping* se cierra y activa la RNAPII (RNA polimerasa), gatillando la expresión del gen.

Mecanismos de mantenimiento de la pluripotencia

En particular, volviendo a nuestro modelo biológico, las células madre pluripotentes de ratón, encontramos que la GRN que controla la salida del estado naïve aún está en construcción^{94,95}, dado que no se conoce la totalidad de los genes e interacciones que forman parte del proceso. Sin embargo, sí se están haciendo cada vez más estudios de esta transición a nivel local (midiendo la actividad de dos o tres genes)^{18,12} y global (midiendo toda la genómica)³. Es particularmente importante un trabajo³ donde lograron construir los perfiles de expresión de todas las proteínas (proteoma), mRNAs (transcriptoma), fosforilaciones (fosfoproteoma) e incluso marcadores epigenéticos (epigenoma). Sin embargo, esta información tan abundante resulta insuficiente para construir la GRN, dado que solo podemos establecer correlaciones entre los niveles de expresión génica pero no podemos asociarlas a causalidades directas, a las interacciones que formarían la GRN.

Más allá de la GRN, el epigenoma obtenido sí aporta información sobre qué enhancers se activan y se silencian en el pasaje de naïve a primed. Otro trabajo¹² estudió esta transición en detalle y detectaron cómo algunos factores de transcripción (como Oct4) asociados a la misma cambian los enhancers a los que se unen, lo que muestra que la epigenética cumple un rol fundamental. Esta idea fue retomada en un review reciente⁴⁹, donde agregan que Oct4 forma un complejo con Sox2 y Nanog (el complejo *OSN*)⁹⁶, siendo las tres proteínas factores de transcripción de la pluripotencia. Este complejo se une a sitios de estos mismos factores y otros asociados también al estado naïve (Figura 51C). Así, hay evidencia de que existe un triple feedback positivo de Oct4, Sox2 y Nanog⁴⁹ que promueve su propia síntesis y esto ayuda a mantener el estado pluripotente (Figura 51A y D). En la salida al estado primed, este complejo se separa y cambia los enhancers a los que se une⁹⁷, lo que esquematizamos en la Figura 51B.

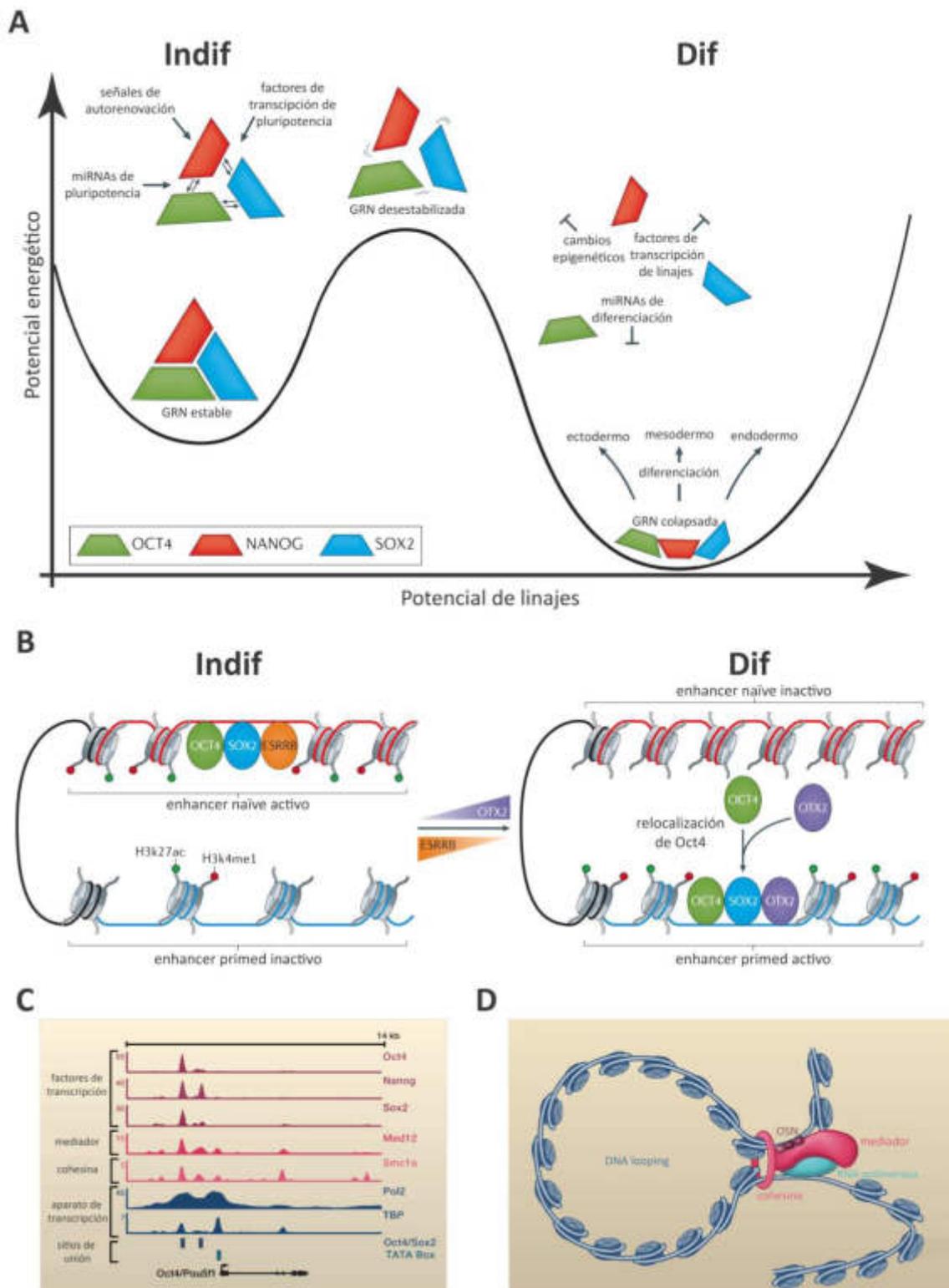


Figura 51: Imágenes adaptadas (A y B⁴⁹, C y D⁹⁵). A: Esquema que representa la salida del estado naïve como un diagrama de energía donde se desestabiliza el complejo OSN (por los factores de transcripción Oct4, Sox2 y Nanog). B: Relocalización del complejo Oct4-Sox2, guiado por la bajada de expresión de Esrrb y la subida de Otx2, y a partir del silenciamiento epigenético de los enhancers asociados al estado naïve y la activación de los asociados al primed. C: Sitios de unión de las proteínas listadas en las cercanías del gen Oct4. Data de Chip-seq. D: Modelo de DNA looping activado por la unión del complejo OSN a un enhancer. Múltiples loopings desde diferentes enhancers pueden darse en simultáneo para activar un promotor.

Por otro lado, también se ha detectado que existe una doble inhibición entre los genes que mantienen la pluripotencia y los que se expresan durante la diferenciación¹⁰. Uno de los mecanismos detectados⁹⁸ que subyacen esta doble inhibición está dado por el miRNA-145, que se une al mRNA de Oct4, Sox2 y Nanog e impide su traducción. A su vez, el complejo formado OSN inhibe la transcripción del miRNA-145, como esquematiza la Figura 52⁹⁹.

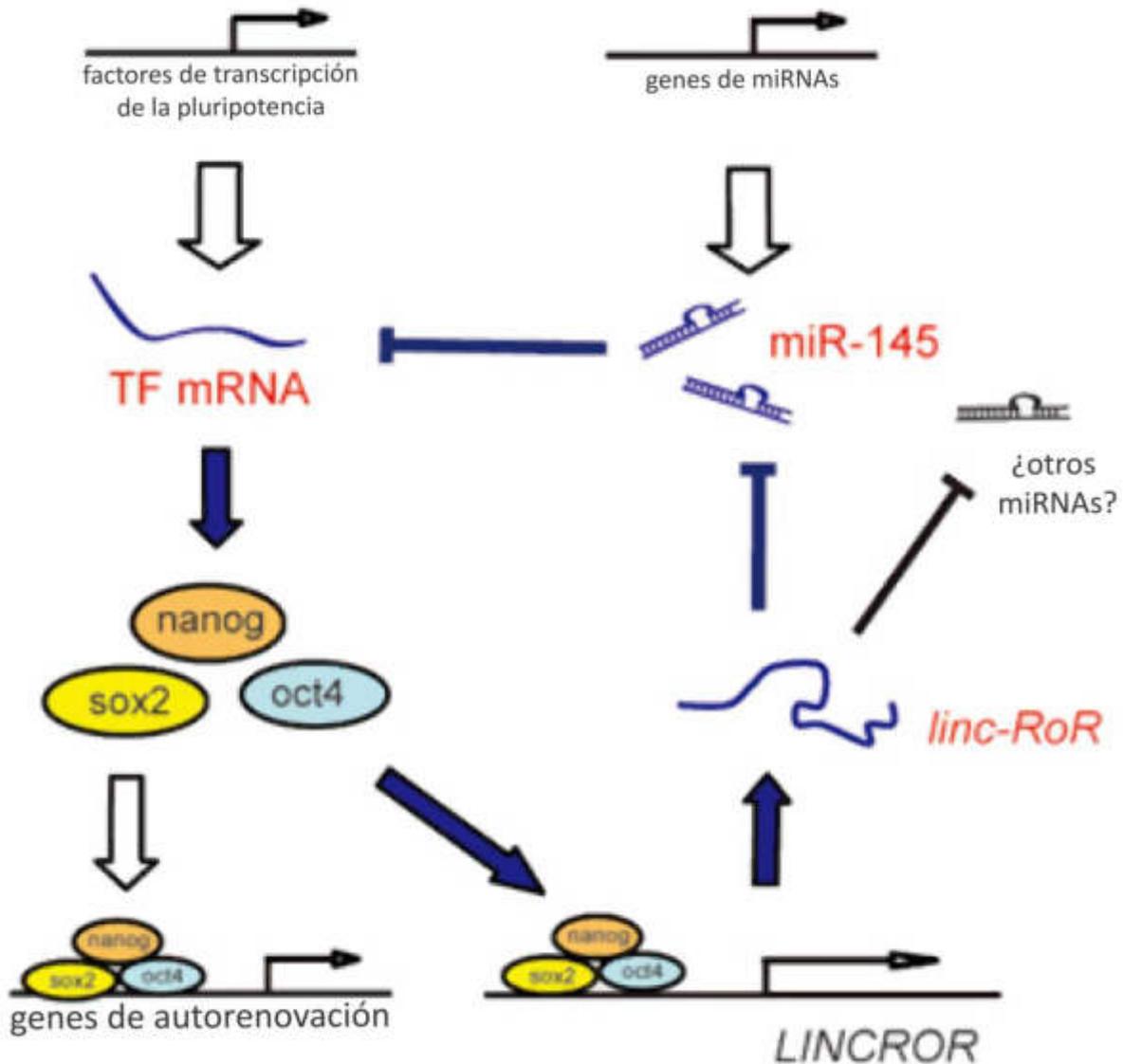


Figura 52: Imagen adaptada⁹⁹. Mecanismo de inhibición mutua. Feedback positivo entre los factores de transcripción del complejo OSN y el miRNA-145.

Feedbacks positivos y curvas dosis respuesta

Dado que existe evidencia de la existencia de mecanismos que cuentan con feedbacks positivos para mantener la pluripotencia (ver Mecanismos de mantenimiento de la pluripotencia), introducimos primero cómo dependen

cualitativamente las curvas dosis respuesta con estos feedbacks. Las curvas dosis respuesta son las que resultan de graficar la respuesta de un sistema complejo en función del input o del estímulo que se le aplique, en general en estado transitorio o de equilibrio. Un ejemplo, ilustrado en la Figura 53, consiste en aplicar un estímulo tipo escalón sobre un nodo sin ninguna regulación y medir su dinámica en el tiempo, esperando a que llegue a estado estacionario (Figura 53A, panel izquierdo). Una vez que llegó, se toma este valor y se repite el procedimiento para muchas intensidades del estímulo. Así, podemos graficar la curva dosis respuesta (Figura 53A panel derecho). Este tipo de gráficos dan una idea resumida de cómo se activa un sistema en función del estímulo de interés.

Si ahora agregamos un feedback positivo, la curva dosis respuesta puede ser biestable (Figura 53B), dependiendo del valor de los parámetros. Esto significa que, por cada valor de intensidad del estímulo, puede haber dos puntos fijos estables, dos respuestas diferentes. Qué respuesta tendrá el sistema dependerá únicamente de desde qué condición inicial partió.

Por otro lado, la biestabilidad podría ser irreversible. Imaginemos que vamos incrementando la intensidad del estímulo de forma mucho más lenta que la dinámica propia de la respuesta, lo que se conoce como aproximación cuasi-estacionaria. La idea consiste en que, para cada intensidad por la que vamos pasando, el sistema alcanza el equilibrio antes que la cambiemos, y es una forma de obtener las curvas dosis respuesta. Haciendo esto en una curva biestable (Figura 53B), llegaremos al punto donde la curva vuelve sobre sí misma y el sistema pasa de un punto fijo bajo a otro alto. Si repetimos el procedimiento, pero ahora bajando las intensidades del estímulo, podemos ver que el salto entre puntos fijos existe para un valor menor de estas intensidades, y de manera que las curvas serán diferentes si vamos subiendo o bajando el estímulo. Habrá histéresis. Notemos que, si bajamos la intensidad a cero, el sistema deja de responder, al igual que si nunca hubiéramos aplicado estímulo (Figura 53B panel izquierdo). A este caso lo llamaremos reversible, en contraposición al siguiente. Si el feedback fuera muy fuerte, existe la posibilidad de que aún sin estímulo el sistema no pueda volver al equilibrio inferior (Figura 53B panel derecho). Este sería el caso irreversible, dado que una vez que aplicamos un estímulo suficiente, es imposible volver atrás en la dinámica. Cabe mencionar que la aproximación cuasi-estacionaria no es necesaria para

obtener la curva dosis respuesta. Para esto solo es necesario hacer una exploración dando muchos estímulos al sistema y esperando que alcance el equilibrio en cada uno. Por otra parte, en sistemas complejos la reversibilidad seguramente dependa de más parámetros que la intensidad del feedback.

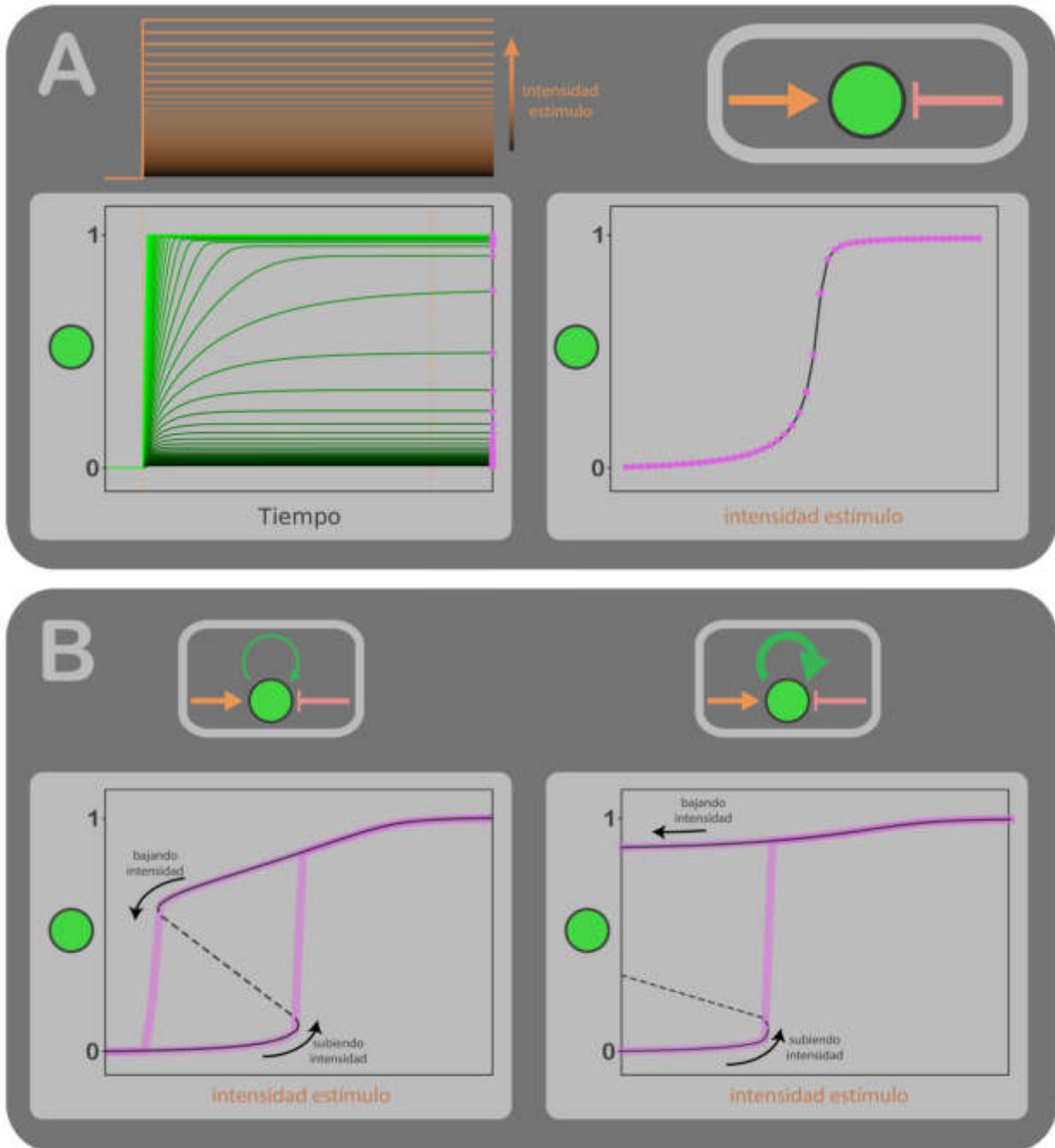


Figura 53. A: Dinámica para varias intensidades de un estímulo escalón (izquierda) y curvas dosis respuesta obtenida a partir de los puntos estacionarios (derecha), para el modelo de regulación simple (arriba a la derecha). B: Curvas dosis respuesta para los modelos con autorregulaciones positivas débil (izquierda) y fuerte (derecha). Ambas son curvas biestables y graficamos los puntos fijos estables en líneas negras continuas e inestables en líneas punteadas. Las curvas magentas representan la ida y vuelta en un experimento donde se sube y baja la intensidad del estímulo de manera cuasiestacionaria. Si la autorregulación es suficientemente fuerte (derecha), la activación se hace irreversible.

En uno de los experimentos que describiremos en este capítulo, buscaremos aplicar estímulos de diferenciación y medir la expresión de diferentes genes ligados al estado pluripotente naïve y al diferenciado primed como respuesta.

Objetivo

Cabe destacar que ambos mecanismos presentados incluyen un feedback positivo. El primero, formado por la autoactivación del complejo OSN. El segundo, por la doble inhibición entre este complejo y el miRNA-145. Estos dos mecanismos forman parte de los motifs capaces de guardar memoria encontrados en el Capítulo 2. Es por esto que aprovecharemos este modelo biológico para aplicar nuestro cuantificador y mostrar un ejemplo de células que pueden recordar un estímulo, en este caso, un pulso de señales de diferenciación.

Buscaremos, en primer lugar, detectar la existencia de algún feedback positivo en la expresión de factores de transcripción asociados a la pluripotencia. En segundo lugar, aplicaremos el cuantificador de memoria para ver si la respuesta al estímulo de diferenciación es irreversible.

Desarrollo Experimental

Este capítulo está compuesto por 3 experimentos principales, en todos ellos trabajaremos con la diferenciación de células madre pluripotentes de ratón, en estado naïve y diferenciándolas al estado primed. El mantenimiento de estas células en el estado naïve pluripotente se logra añadiendo dos moléculas inhibitoras al medio de cultivo usual (CHIR y PD). Estas interactúan cada una con una vía de señalización e inhiben la diferenciación a cualquier linaje. Así, utilizaremos dos tipos de medios de cultivo, que llamamos **Indif** y **Dif**, y que difieren exclusivamente en el agregado o no de estos dos inhibidores.

Para más detalle sobre los medios de cultivo y las células con las que trabajamos, ver las secciones Células Madre pluripotentes de ratón, Modelo de estudio, Vías de señalización y control, Control de la diferenciación naïve a primed y Medios de cultivo, escritas en la Introducción del Capítulo 1 de esta tesis.

Diluciones 2i

Hipótesis

Se han descrito diferentes mecanismos de mantenimiento de la pluripotencia que cuentan con feedbacks positivos (ver Mecanismos de mantenimiento de la pluripotencia). La hipótesis que planteamos consiste en que estos feedbacks deberían generar curvas dosis respuesta biestables, tomando como respuesta la expresión de los genes asociados a los estados naïve y primed.

Diseño experimental

El primer experimento, que llamamos "Diluciones 2i", consistió en cultivar a las células por 48hrs (el tiempo usual que lleva la diferenciación naïve a primed) diluyendo los dos inhibidores, desde la concentración utilizada usualmente hasta 1/32 de la misma, siempre bajando a la mitad, y agregamos la condición de control sin inhibidores (1, 1/2, 1/4, 1/8, 1/16, 1/32, 0). Es decir, largamos la diferenciación usual, pero en lugar de utilizar medio **Dif** puro lo hicimos para varias mezclas con

concentraciones intermedias de los medios **Indif** y **Dif**. Esta fue la manera de hacer una exploración en las intensidades del estímulo, en este caso, la señal de diferenciación.

El objetivo es medir la expresión de genes marcadores luego del tratamiento con estas concentraciones durante 48hrs y construir la curva dosis respuesta, donde la dosis es la concentración de los inhibidores y la respuesta es la expresión génica. Para esto elegimos dos genes, Nanog y Oct6. El primero, es un factor de transcripción asociado al estado naïve, y es sabido que su expresión disminuye más de 32 veces a nivel transcripcional durante las primeras 48hrs de esta diferenciación. El segundo está asociado al estado primed, y se conoce que su expresión aumenta más de 200 veces en este mismo intervalo.

Finalizado el tratamiento con las concentraciones, realizamos una inmunofluorescencia contra las proteínas Nanog y Oct6 (ver Métodos). Cabe destacar que esta técnica permite medir la expresión proteica de estos genes marcadores a nivel de célula única.

Resultados

La Figura 54 muestra los resultados de la inmunofluorescencia contra Nanog y Oct6. Podemos observar, como era esperable, que Nanog disminuye bajando la concentración de inhibidores mientras que Oct6 aumenta. Además, este cambio pareciera ser brusco, binario y mutuamente excluyente. Es decir, que existen solo dos poblaciones de células. Una caracterizada por alta expresión de Nanog y baja de Oct6 y otra por la inversa, y casi no existen células donde ambas proteínas se expresen simultáneamente.

Esta evidencia va en favor de una biestabilidad, donde cada población se correspondería con uno de los dos puntos fijos. Además, cada punto fijo estaría asociado a una identidad distinta: la población con alta expresión de Nanog corresponde a células naïve mientras que la otra corresponde a células que ya iniciaron la transición a primed.

Cabe mencionar que, de no haber medido la expresión de ambos genes a nivel de célula única, no habríamos podido identificar dos poblaciones de células, sino distintas cantidades de expresión a nivel global. Por ejemplo, si tuviéramos la mitad de las células expresando Nanog al 100% y la otra mitad al 0%, habiéramos obtenido que en ese tratamiento Nanog se expresaba

al 50%. Esto hubiera sido indistinguible de una curva dosis respuesta continua o sin feedback.

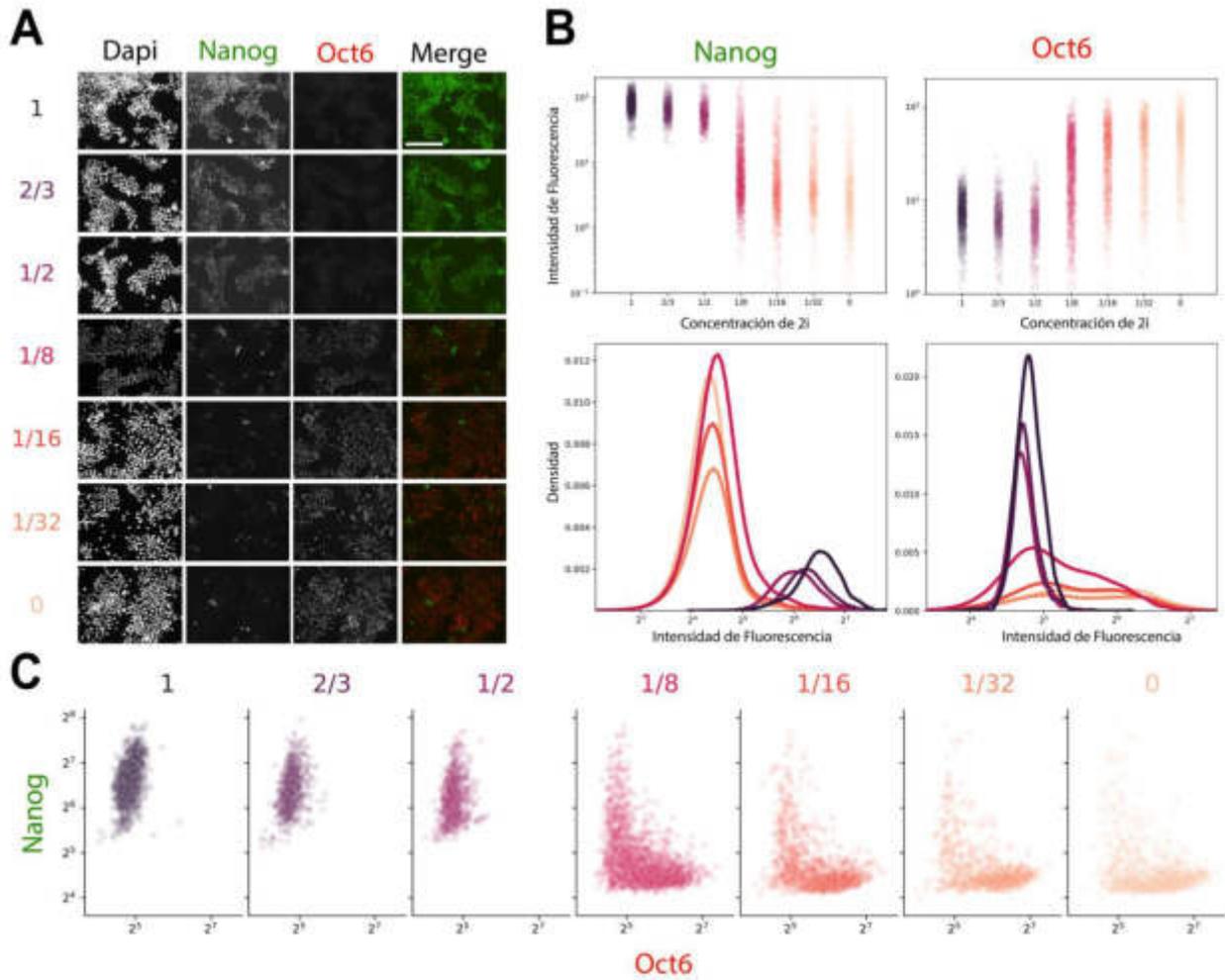


Figura 54: Inmunofluorescencia contra Nanog y Oct6, para muchas concentraciones de 2i.

A: Fotos en 20X de los preparados en los 3 canales de fluorescencia. En el canal azul, Dapi que marca el núcleo; en el verde, Nanog, marcador del estado naïve y en rojo, Oct6, marcador de la diferenciación a primed. Cada fila representa una condición experimental con una concentración diferente de 2i, los dos inhibidores de la diferenciación. 1 representa la concentración usual utilizada. Podemos apreciar que a medida que bajamos la concentración, disminuye Nanog y aumenta Oct6, como esperábamos comparando con el proceso natural de la diferenciación naïve a primed. Además, no hay células que expresen ambas proteínas simultáneamente.

B: Intensidad de fluorescencia en función de la concentración de 2i (arriba) para las proteínas Nanog (izquierda) y Oct6 (derecha). Cada punto representa la expresión en una única célula. Abajo, la densidad poblacional de la misma intensidad de fluorescencia para ambas proteínas. Las cuantificaciones confirman que Nanog baja y Oct6 sube a medida que disminuye la concentración de 2i. Además, parece ocurrir de manera binaria, ya que las densidades de Oct6 presentan bimodalidad y no hay muchas células en los estados de intensidad intermedios.

C: Scatter-plots de la intensidad de fluorescencia de Nanog vs Oct6, para cada concentración de 2i. Cada punto representa una célula única. Podemos verificar que no hay células que expresen ambas proteínas simultáneamente, lo que confirma lo visto en las fotos de **A** y es evidencia de una exclusión mutua entre estas proteínas.

Pulso cuadrado de Dif

Hipótesis

El siguiente experimento que diseñamos consistió en cultivar las células en estado naïve con medio **Dif** durante 24hrs y 48hrs y luego cambiarles el medio por **Indif**, intentando revertir el proceso de diferenciación. Siguiendo la línea de los resultados anteriores, buscamos medir la curva dosis respuesta nuevamente, pero suponiendo que es biestable, que tiene histéresis e intentando medir el segmento superior. Además, en caso de encontrarlo, buscaremos verificar si se trata de una curva reversible o irreversible.

Sin embargo, cabe mencionar que el diseño del protocolo experimental es muy similar al utilizado normalmente para continuar la diferenciación hacia los linajes de endodermo y mesodermo, un estadio celular aún más avanzado en la diferenciación con el que trabajamos también en el laboratorio. Para llevar las células naïve a este estadio, se las cultiva en medio **Dif** por 48hrs y luego se las trata con CHIR, uno de los dos inhibidores. La única diferencia con dar un pulso de 48hrs de diferenciación está en que, al volver al medio **Indif**, este cuenta con los dos inhibidores en lugar de uno solo. La pregunta que surge es si las células, en lugar de regresar al estado naïve cuando les agreguemos medio **Indif**, no avanzarán hasta diferenciarse a mesendodermo o algo similar.

En conclusión, la hipótesis es que las células, luego de un pulso de medio **Dif**, vuelven al estado naïve. En el siguiente experimento daremos este pulso y mediremos la expresión a nivel transcripcional (los mRNAs) de varios genes, con el objetivo de determinar la identidad celular y distinguir si vuelven o no.

Diseño experimental

Para dar el pulso de diferenciación, usaremos células en estado naïve cultivadas en medio **Indif** y lo cambiaremos por medio **Dif**, que no tiene los inhibidores. Dejaremos este medio por intervalos de 24hrs y 48hrs para luego volver al medio **Indif**, dado que sabemos que la diferenciación al estado primed lleva 48hrs. El objetivo de dejarlas 24hrs es buscar, en paralelo, si existe algún umbral temporal de reversibilidad.

En caso de que las células vuelvan al estado naïve, no sabemos cuánto tiempo tardarían en hacerlo y no hay razón para suponer que es el mismo

tiempo de ida que de vuelta. Por esto también exploramos con 24hrs y 48hrs (condiciones 24 **Dif** 24 **Indif**, 48 **Dif** 24 **Indif** y 48 **Dif** 48 **Indif**). Sumando los controles negativos 24 **Indif** (dejando siempre el medio **Indif** por 24hrs) y positivos 24 **Dif** y 48 **Dif** (24hrs y 48hrs de medio **Dif**, sin cambiarlo), esto nos da las 6 condiciones que ilustra la Figura 55A.

Para todas estas condiciones realizamos 3 réplicas biológicas. Para cada una, terminados los tratamientos con los medios, medimos la expresión a nivel transcripcional (los mRNAs) de dos genes asociados al estado naïve (Nanog y Klf4), dos al estado primed (Oct6 y Fgf5) y uno asociado a la diferenciación a mesodermo (Brachyury, de ahora en adelante, Bra). Estas mediciones las hacemos por RT-qPCR (ver Métodos) y nos permitirán distinguir la identidad de las células para cada tratamiento. Cabe mencionar que, en este caso, las medidas son a nivel global de la población. Sin embargo, el objetivo de este experimento es averiguar si las células vuelven al estado naïve o continúan a EpiSC, y no necesitamos cuantificar la proporción de cada población.

Resultados

La Figura 55B muestra los resultados de los análisis de RT-qPCR, para los 6 tratamientos distintos y para los 5 genes: Nanog y Klf4, marcadores del estado naïve; Oct6 y Fgf5 del estado primed y Bra de la diferenciación a mesodermo.

Podemos observar que tanto Nanog como Klf4 bajan su expresión durante las 24hrs ó 48hrs de diferenciación, así como Oct6 y Fgf5 la suben, como era de esperarse. Esto indica que la diferenciación se desarrolló normalmente sin imprevistos.

En cuanto a las condiciones de pulsos de diferenciación. Observamos que, para los tres tratamientos 24 **Dif** 24 **Indif**, 48 **Dif** 24 **Indif** y 48 **Dif** 48 **Indif**, Nanog y Klf4 vuelven a aumentar su expresión, aunque no alcanzan a los niveles iniciales. Este aumento es similar para 24 **Dif** 24 **Indif**, 48 **Dif** 24 **Indif** y algo mayor para 48 **Dif** 48 **Indif**, debido probablemente al mayor tiempo de tratamiento con los inhibidores. Para los genes Oct6 y Fgf5, el comportamiento es completamente inverso y análogo, como se esperaba dado que son marcadores del estado primed. Es importante notar que, si bien Nanog y Klf4 vuelven a expresarse y Oct6 y Fgf5 dejan de hacerlo, los

valores medidos no llegan a ser los mismos que en la condición de control 24 **Indif**. Esto indica que, si las células revierten el pulso de diferenciación, no lo hacen completamente.

Analizando ahora el gen *Bra*, que nos permite distinguir entre el estado naïve y la diferenciación a mesodermo, vemos que sí comienza a expresarse luego del pulso de diferenciación, y para todas las variantes de este (24 **Dif** 24 **Indif**, 48 **Dif** 24 **Indif** y 48 **Dif** 48 **Indif**). Nuevamente, es notable que en la condición de más tiempo de diferenciación y reversión (48 **Dif** 48 **Indif**), la expresión de *Bra* es mayor que en el resto de los tratamientos. Esto indica que las células no retornan al estado naïve, sino que continúan el camino de diferenciación a mesodermo.

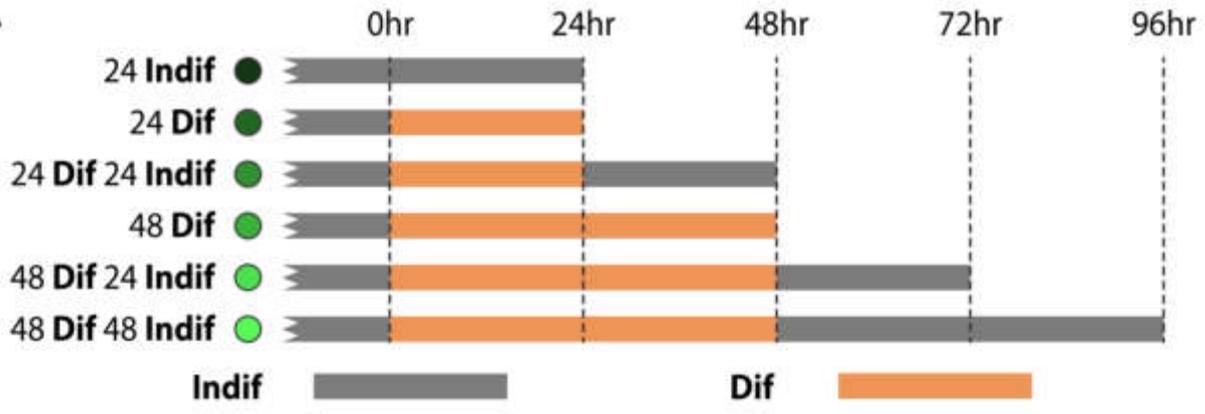
Aplicando el cuantificador de memoria

Este experimento permite ejemplificar una aplicación del cuantificador de memoria propuesto en el Capítulo 2 (ver Cuantificador de memoria molecular). Las condiciones con el pulso de diferenciación son el estímulo transitorio necesario. El sistema sin estimular se corresponde con la condición 24I. Para aplicar el cuantificador, dado por la ec. 22, basta con escribir la respuesta transcripcional como un vector de 5 dimensiones, cada una representando un gen, y tomar la diferencia en las respuestas medidas con y sin estímulo, como indican las ecs. 34.

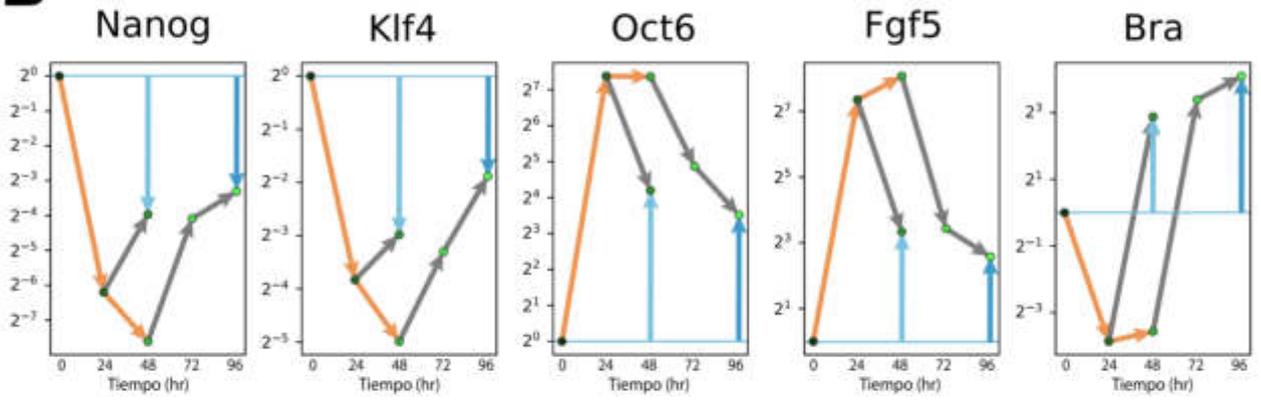
$$\begin{aligned} & (Nanog, Klf4, Oct6, Fgf5, Bra)_{24D24I} - (Nanog, Klf4, Oct6, Fgf5, Bra)_{24I} \\ & (Nanog, Klf4, Oct6, Fgf5, Bra)_{48D48I} - (Nanog, Klf4, Oct6, Fgf5, Bra)_{24I} \end{aligned} \quad 34$$

La Figura 55B muestra la componente del vector memoria obtenido, correspondiente a cada gen, para cada condición experimental. Cabe destacar que, a excepción de *Bra*, en todos los casos la memoria parece disminuir a medida que pasa más tiempo desde el cese del estímulo. Esto parecería indicar que las células tienen tendencia a olvidar el estímulo de diferenciación y revertir el cambio de identidad nuevamente al estado naïve. Sin embargo, comparando con la componente de memoria del gen *Bra*, vemos que allí aumenta con el tiempo desde el estímulo. Esta componente muestra que las células no olvidan el estímulo y retornan, sino que continúan el camino de la diferenciación hacia mesodermo. Incluso sugiere que, con 48hrs, las células quedan más preparadas para continuar la diferenciación que con solo 24hrs de **Dif**, ya que elevan más los niveles de *Bra*.

A



B



C

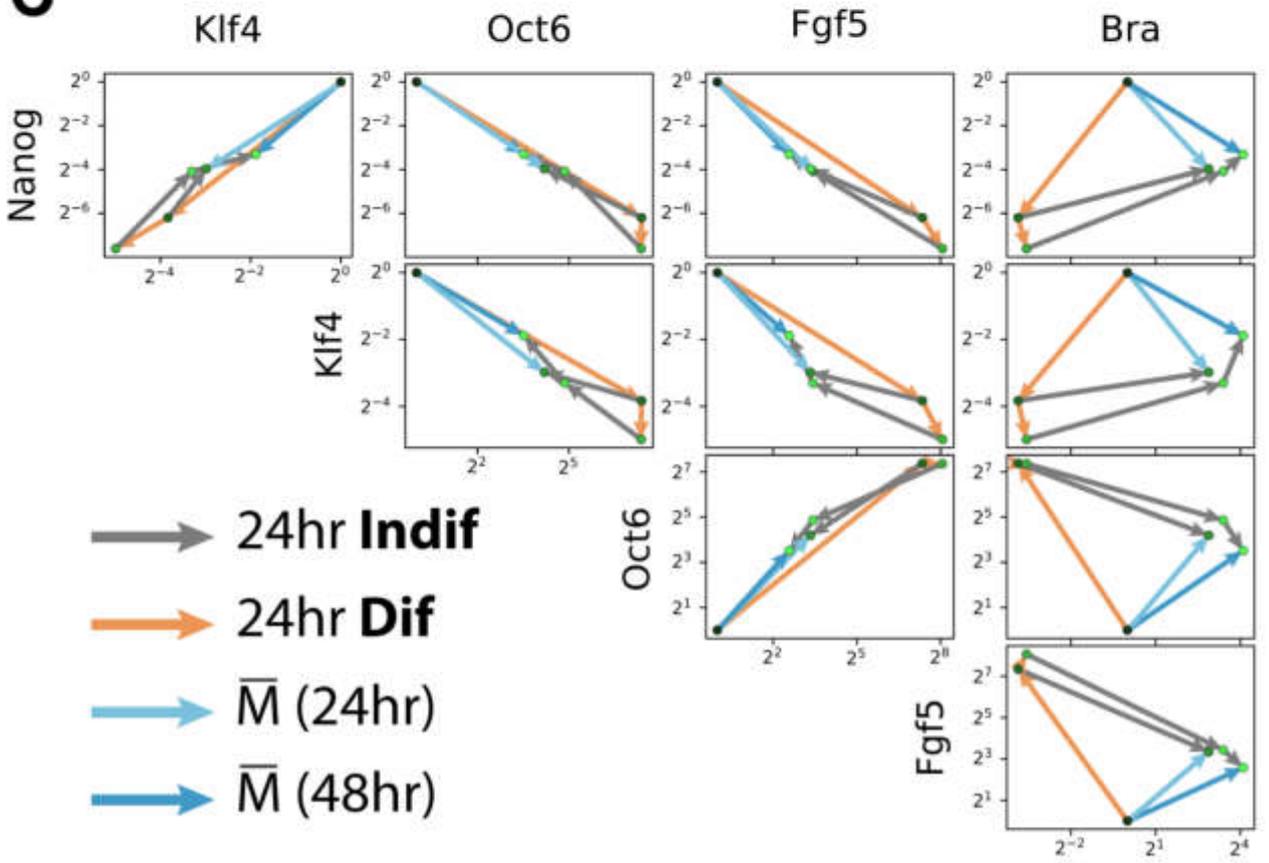


Figura 55: Pulso de diferenciación.

A: Esquema experimental. Cultivamos las células en medios **Indif** (gris) y **Dif** (naranja) alternando ambos medios durante distintos intervalos de tiempo. Las condiciones 24 **Dif** 24 **Indif**, 48 **Dif** 24 **Indif** y 48 **Dif** 48 **Indif** representan pulsos de diferenciación y reversión. El resto de las condiciones son controles negativos (24 **Indif**) y positivos (24 **Dif** y 48 **Dif**), que permiten comparar la expresión génica de las condiciones de reversión tanto con el estado naïve como con el primed.

B: Expresión génica a nivel transcripcional medida por RT-qPCR para los 5 genes estudiados, 2 marcadores del estado naïve (*Nanog* y *Klf4*), 2 marcadores del estado primed (*Oct6* y *Fgf5*) y un marcador de la diferenciación más allá del estado primed (*Bra*). Los puntos en gradiente negro-verde indican la condición experimental. Las flechas que los unen indican que están separados por 24hrs de tratamiento con medios **Indif** (flechas grises) y **Dif** (flechas naranjas). Las flechas azules indican el valor obtenido para las componentes del vector de memoria correspondientes a los genes de cada gráfico, tanto para el pulso de 24hrs en medio **Dif** (azul claro) como para el de 48hrs (azul oscuro). Como era esperable, *Nanog* y *Klf4* bajan mientras que *Oct6* y *Fgf5* suben. Sin embargo, la expresión nunca llega a los mismos valores luego del pulso y esto da una memoria no nula, que parece disminuir si aumenta el tiempo desde el cese del estímulo. Por otro lado, en el gráfico de *Bra* (derecha) podemos ver que la memoria aumenta, indicando que las células continúan el proceso de diferenciación más allá del estado primed en lugar de retornar al estado naïve.

C: Expresión génica a nivel transcripcional para todos los pares de combinaciones de los genes estudiados. Los puntos en gradiente negro-verde indican la expresión para las 6 condiciones experimentales. Las flechas que los unen indican que están separados por 24hrs de tratamiento con medios **Indif** (flechas grises) y **Dif** (flechas naranjas). Las flechas azules indican el valor obtenido para las componentes del vector de memoria correspondientes a los genes de cada gráfico, tanto para el pulso de 24hrs en medio **Dif** (azul claro) como para el de 48hrs (azul oscuro). Podemos apreciar que, en todos los gráficos, a excepción del gen *Bra*, la expresión parece revertirse volviendo por la misma dirección que la llevó del estado naïve al primed y todas las flechas se encuentran alrededor de la misma dirección. Sin embargo, la expresión nunca llega a los mismos valores luego del pulso y esto da una memoria no nula. Comparando con los gráficos de *Bra* (columna derecha), vemos que son cualitativamente diferentes, y que la expresión toma otra dirección luego del pulso, lo que se corresponde con que las células siguen adelante en el camino de la diferenciación en lugar de volver atrás al estado naïve.

La Figura 55C refuerza esta idea. Podemos ver que, en todos los gráficos, a excepción de los que incluyen a *Bra*, los caminos de ida con el estímulo de diferenciación y vuelta (habiéndolo quitado) parecen ser en la misma dirección, con diferentes sentidos, e indicando una tendencia a retornar al mismo estado de partida. Esto no ocurre en el caso de *Bra*, donde la dinámica no retorna al mismo punto, sino que toma otra dirección, mostrando finalmente que las células no retornan al estado naïve sino que continúan el proceso de diferenciación a mesodermo.

Reversión de Oct6

Línea celular con KO de Oct6 endógeno y agregado de Oct6 exógeno

Hasta ahora, encontramos evidencia de que existen feedbacks positivos que ayudan a mantener el estado pluripotente y a recordar el estímulo de diferenciación una vez que este se termina. Sin embargo, desconocemos cuáles son las interacciones que forman este feedback y entre qué genes se dan. Si bien contamos con evidencia de que *Oct6* y *Nanog* actúan como si se inhibieran mutuamente, esta evidencia es de correlación entre ambos y no demuestra ninguna causalidad. Ignoramos si la expresión excluyente de uno u otro se debe a alguna inhibición directa entre éstos o es consecuencia

de interacciones indirectas. Una pregunta que surge naturalmente es si podemos demostrar la interacción directa entre Oct6 y Nanog.

Una manera de acercarse a esta pregunta es haciendo un *knock out* de Oct6, esto es, eliminar su expresión por ingeniería genética. Si repetimos el experimento en células que no expresan Oct6 nunca, podría pasar tanto que Nanog siga bajando su expresión durante la diferenciación como que no. La hipótesis es que, si se deja de expresar, mostraría que Oct6 no es responsable único de la inhibición de Nanog. Si efectivamente Nanog se siguiera expresando, entonces podríamos decir que Oct6 juega un rol importante en su inhibición, aunque seguiríamos sin demostrar interacción directa.

Cabe destacar que, dentro de los genes que funcionan como factores de transcripción asociados a la diferenciación naíve a primed, Oct6 es uno de los que más rápido comienzan a expresarse durante este proceso (ya se expresa a las 2hrs de recibir el estímulo). La expresión temprana es necesaria para que sea el responsable directo y causal de la inhibición de Nanog. Sería imposible que Oct6 causara la bajada de expresión de Nanog si comenzara a expresarse una vez que Nanog ya bajó. Esto hizo que lo eligiéramos como candidato y realizáramos una línea de células con un knock out de Oct6, mediante la técnica de *CRISPR-Cas9*¹⁰⁰. Esta técnica permite eliminar un único par de bases en el código genético en un lugar específico, en este caso, elegimos una base cercana al sitio de inicio de transcripción en el gen de Oct6 (Figura 56A), de manera de cambiar el marco de lectura. Las células con este knock out, podrán sintetizar el mRNA de Oct6 con esta base modificada, pero este mRNA no codificará la proteína.

Por otro lado, a esta misma línea de células las volvimos a modificar genéticamente con el sistema *PiggyBack*¹⁰¹ de transposición para insertarles un nuevo gen de Oct6. La diferencia entre este gen insertado artificialmente y el Oct6 endógeno que tenían las células naturalmente, está en que podemos activar o desactivar al promotor del gen artificial simplemente agregando Doxiciclina (Figura 56B). El mecanismo permite regular la expresión de Oct6 como queramos (ver Sistema de Control de Oct6 exógeno). Además de Oct6, el gen insertado contiene mCherry, una proteína fluorescente roja, que en consecuencia se transcribe y traduce en conjunto con Oct6, separada solamente por un *linker* que codifica para un péptido de *autoclivaje*, diseñado para cortarse una vez que se formaron ambas proteínas. De esta

manera, midiendo la intensidad de brillo rojo sabremos cuál es la cantidad de proteína Oct6.

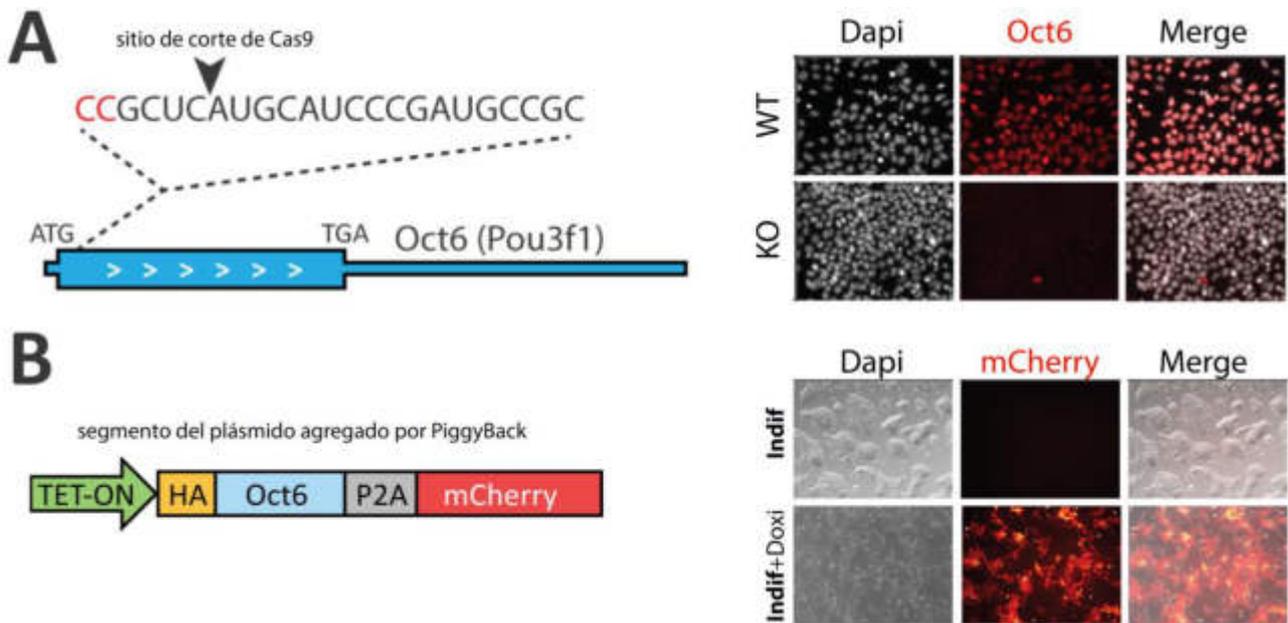


Figura 56. **A:** Sitio de corte donde modificamos el código del gen Oct6 utilizando la técnica de CRISPR-Cas9 (izquierda). Fotos de microscopía de una inmunofluorescencia contra Oct6 en células WT (por wild type) y KO de Oct6 (derecha). Podemos apreciar que no hay más proteína y que la modificación genética funcionó correctamente. **B:** Segmento del plásmido insertado por Piggyback¹⁰¹ en las células KO de Oct6 (izquierda). El promotor TET-ON se activa con el agregado de Doxyciclina, permitiendo controlar la expresión de Oct6. Además, al código de Oct6 le sigue un péptido de autoclivaje P2A y mCherry. Este péptido está diseñado para cortarse luego de la traducción a proteína. De esta manera, cada vez que se traduzca este segmento sintetizará una proteína de Oct6 y una de mCherry, que quedarán en la misma proporción. A la derecha, fotos de microscopía de fluorescencia de la línea celular modificada con el plásmido insertado. Comparamos a las células sin tratar en medio Indif contra las células tratadas con Doxyciclina. Podemos apreciar, en el canal rojo, que la línea modificada efectivamente expresa mCherry, indicando que el inserto funcionó correctamente.

Controles

La Figura 56 muestra los resultados de dos experimentos control para asegurarnos que la ingeniería genética funcionara. Primero, hicimos una inmunofluorescencia contra Oct6 en células diferenciadas durante 48hrs y verificamos que las células con el knock out efectivamente no lo expresaran. En paralelo, tratamos células mantenidas en estado indiferenciado con Doxyciclina, para inducir la expresión de Oct6 exógeno y les sacamos fotos de fluorescencia en el canal rojo de mCherry para verificar que estuvieran expresando mCherry y por ende Oct6.

Hipótesis

Luego de los resultados positivos de los controles, estudiamos la reversibilidad y la memoria de la sobreexpresión de Oct6. La hipótesis es que, si Oct6 inhibe a Nanog, su sobreexpresión artificial usando Doxiciclina también debería bajar la expresión de Nanog. Además, esta bajada podría ser tanto reversible como irreversible si luego lavamos la Doxiciclina para dejar de inducir Oct6.

Diseño Experimental y Resultados

Tomamos a las células con el knock out de Oct6 endógeno y el Oct6 exógeno agregado en estado naïve y las tratamos con un pulso de 72hrs de doxiciclina, donde evaluaremos indicios de diferenciación al estado primed. En paralelo, seguiremos el tratamiento con 0hrs, 24hrs, 48hrs y 72hrs de medio **Indif** común sin doxiciclina. Una vez tratadas, realizamos una inmunofluorescencia solo contra Nanog. Recordemos que estas células expresan mCherry en la misma proporción que Oct6, de manera que podemos medir su expresión simplemente por fotos de microscopía de fluorescencia en el canal rojo, sin hacer una inmunofluorescencia.

Los resultados, ilustrados en la Figura 57, muestran que la sobreexpresión de Oct6 baja la expresión de Nanog a nivel proteico, induciendo una diferenciación artificialmente. Esta evidencia va en favor de que Oct6 sea uno de los responsables directos de la inhibición de Nanog.

Luego, nos propusimos ver si esta bajada de Nanog era reversible, para lo cual removimos la doxiciclina, como ya mencionamos. Como era esperable, vimos que los niveles de Oct6 exógeno disminuyeron en tan solo 24hrs, así como también aumentaron los de Nanog hasta los niveles iniciales.

Esto muestra que la bajada de Nanog es completamente reversible, y las células no guardan memoria del estímulo de la doxiciclina. Esta evidencia apunta en contra de que Oct6 forme un feedback positivo capaz de recordar el estímulo de diferenciación y llevarla adelante una vez que cesó. Sin embargo, tampoco demostramos que Oct6 no forma parte de un feedback. Podría suceder que Oct6 sea solo uno de varios actores necesarios para superar el umbral necesario que activa el feedback positivo. Por otro lado, hay que tener en cuenta que el Oct6 exógeno tiene un promotor distinto

al endógeno. De manera que si existiera una autorregulación en las células *wild type* (sin modificaciones genéticas), podríamos haberla eliminado.

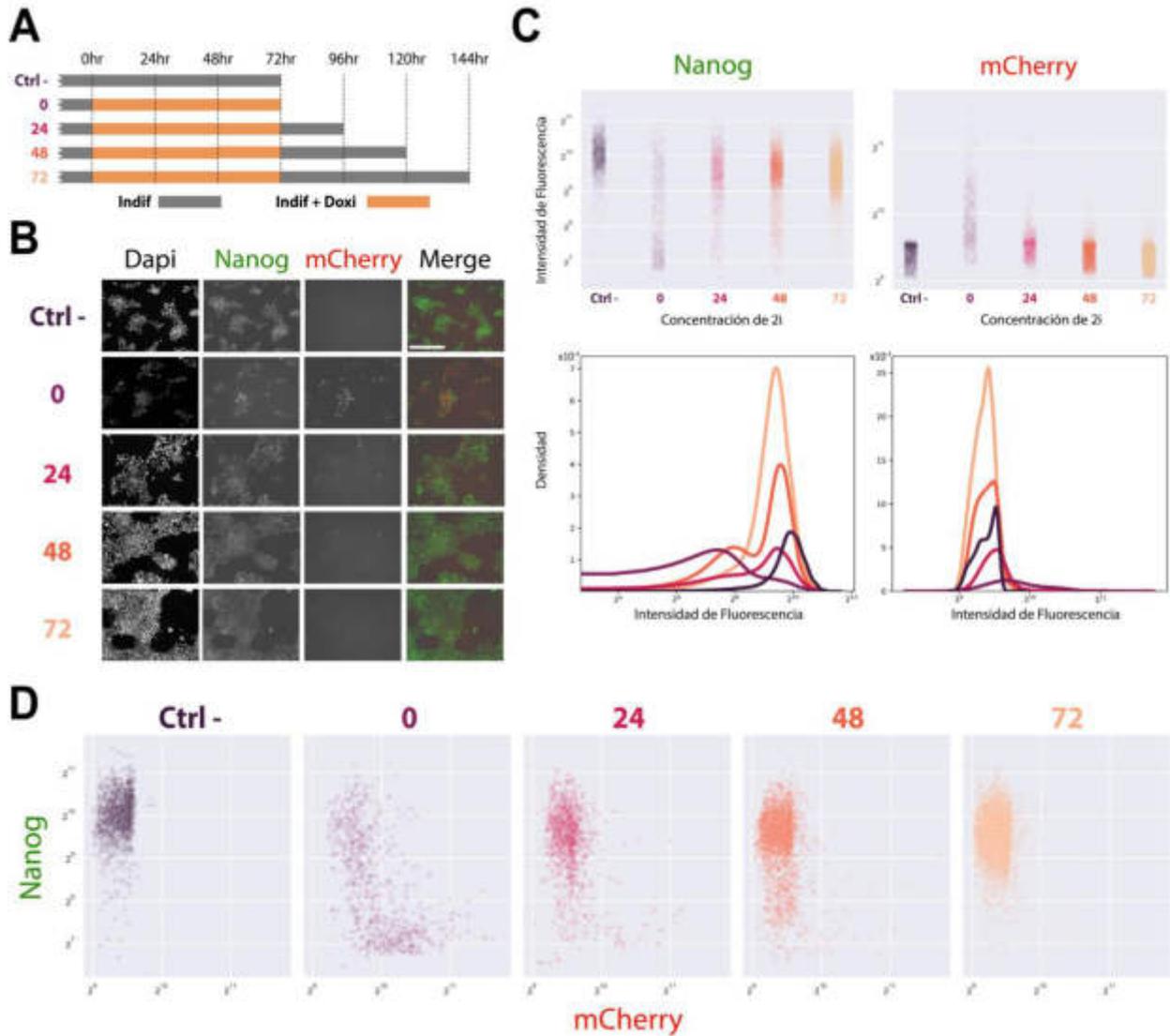


Figura 57: Resultados Reversión de Oct6.

A: Esquema experimental. Colocamos Doxiciclina durante 72hrs para inducir la expresión de Oct6 y luego la lavamos por 24hrs, 48hrs, y 72hrs para ver cómo es el proceso de reversión. Para comparar las condiciones sin lavado (0hrs) y sin haber puesto Doxiciclina (Ctrl-).

B: Fotos de microscopía de fluorescencia. El canal de Dapi muestra los núcleos celulares que utilizamos para segmentar las células. La inmunofluorescencia fue contra Nanog en el canal verde y, en el canal rojo, mCherry que sintetizan las células genéticamente modificadas, en la misma proporción que Oct6. Podemos apreciar que, luego del tratamiento de 72hrs con Doxiciclina, las células dejan de expresar Nanog y comienzan a expresar Oct6. Sin embargo, este proceso parece revertirse ya a las 24hrs de lavado. **C:** Intensidad de fluorescencia de Nanog (arriba derecha) y mCherry (arriba izquierda, indicando Oct6). Cada punto representa la intensidad media de una célula única. Podemos confirmar que la expresión de Nanog baja y la de Oct6 sube con el tratamiento de 72hrs de Doxiciclina, además de que el proceso se revierte casi totalmente a las 24hrs de lavado, y continúa revirtiéndose a las 48hrs y 72hrs. Los gráficos de abajo muestran las densidades de intensidad de fluorescencia. Nuevamente y al igual que en la Figura 54, vemos que aparecen dos poblaciones marcadas de expresión de Nanog.

D: Scatter-plots para cada condición experimental. Cada punto representa una célula única. Nuevamente, y al igual que en la Figura 54, vemos una exclusión mutua en la expresión entre Nanog y mCherry, y no hay células que expresen ambas proteínas simultáneamente.

Métodos

RT-qPCR

La RT-qPCR (por *Reverse Transcription Quantitative Polymerase Chain Reaction*) es una técnica experimental muy utilizada en el campo de la biología molecular, que permite medir la cantidad de mRNA de un gen, es decir, cuánto se está expresando un gen a nivel transcripcional.

El mecanismo deriva de una PCR tradicional. En esta técnica, se prepara una solución que contiene toda la maquinaria celular para copiar el DNA y se le añade la muestra de interés, el material genómico sobre el cual quiere conocerse si contiene o no alguna determinada secuencia. Para especificar cuál es esta secuencia de interés, se agregan también *primers forward* y *reverse*, que son oligonucleótidos o segmentos cortos (entre 10 y 50 bp) de secuencias de DNA. Cada primer contiene una secuencia complementaria a uno de los extremos de la secuencia de interés, y *forward* y *reverse* indican de qué extremo se trata con respecto a la dirección de polimerización. Una vez preparada la solución, se coloca en una máquina que la somete a ciclos de temperatura que oscila entre un valor alto ($\sim 95^{\circ}\text{C}$) y uno bajo ($\sim 50^{\circ}\text{C}$). La alta temperatura separa la doble hélice de la muestra de DNA de interés en dos cadenas simples. La baja temperatura permite que se unan los primers a las cadenas simples y la DNA polimeraza, presente en la solución, realice la copia complementaria de cada cadena simple. El ciclo de temperaturas se repite hasta 40 veces y el resultado, en caso de que esté presente la secuencia de interés, es duplicarla en cada ciclo (Figura 58). Al final, obtenemos una solución con altísima concentración de DNA de la secuencia de interés.

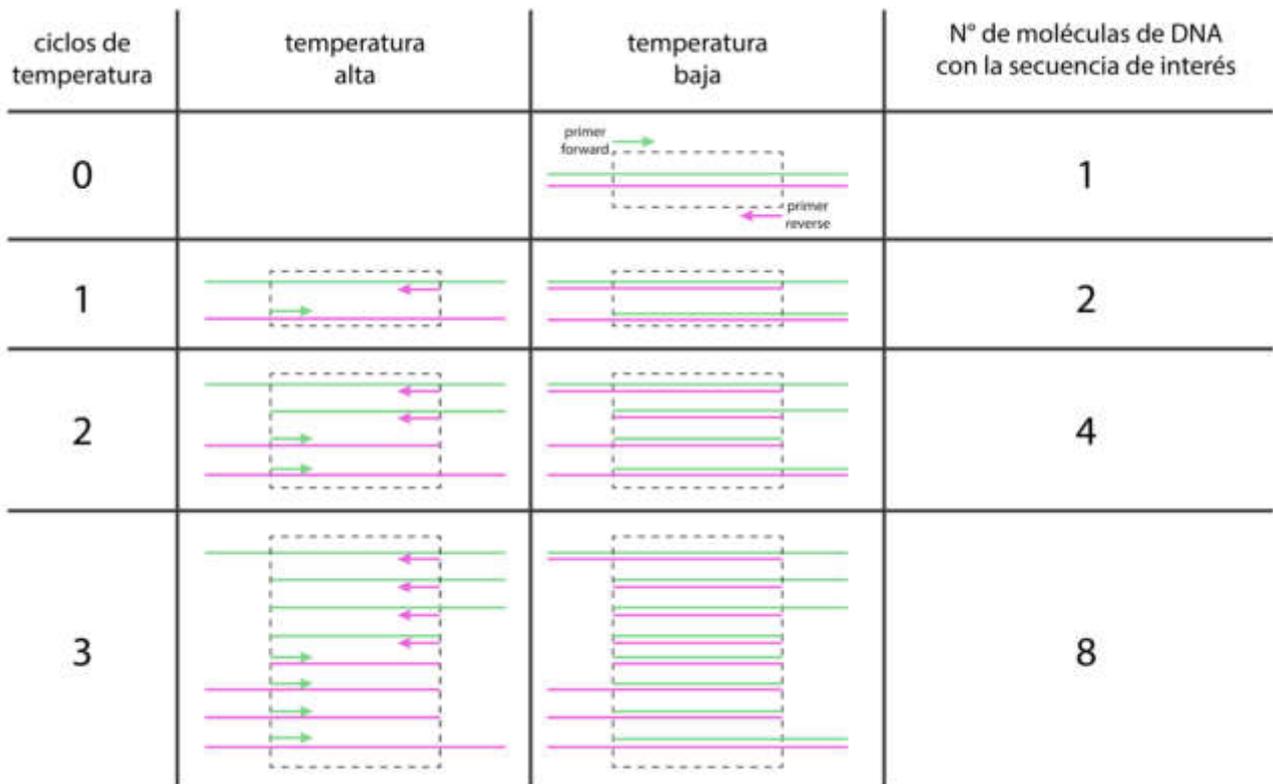


Figura 58: Esquema del mecanismo de duplicación del segmento del DNA de interés que se encuentra entre los primers forward (flechas verdes) y reverse (flechas magentas). Cada primer es complementario a la simple cadena de ADN opuesta, el verde con la cadena magenta y el magenta con la cadena verde. Así, la temperatura baja permite que cada primers se una al sitio ubicado en la secuencia complementaria. Luego, la maquinaria de replicación duplica las simples cadenas en dobles cadenas, siempre en la dirección que indican las flechas de los primers. La temperatura luego divide las dobles cadenas en simples cadenas y vuelve a comenzar el ciclo. Podemos notar que la cantidad de moléculas del ADN original no varían, la cantidad de cada molécula limitada solo por un primer aumenta, pero de a una por ciclo y la cantidad de las moléculas del segmento deseado se duplican en cada ciclo.

En la PCR tradicional, esta solución de alta concentración se coloca luego en un gel para correrla por electroforesis, técnica que permite medir el largo de las secuencias. En caso de que la secuencia de interés estuviera presente en la muestra de DNA, aparecerá una banda justo del largo de la secuencia de interés. Esta lectura final es binaria, y solo indica si estaba o no la secuencia en la muestra. La q-PCR se diferencia en que, en la solución primera, se agregan fluoróscoros que brillan al unirse a DNA de doble cadena, y permiten medir la cantidad de DNA de la solución en cada ciclo de temperatura. Esta cantidad, idealmente, se duplica con cada ciclo, y en la realidad aumenta exponencialmente a una tasa un poco menor que 2, que depende de la eficiencia de los primers. Un ajuste del gráfico de la intensidad de brillo en función del ciclo (ec. 35) permite calcular la cantidad inicial de la secuencia de interés.

$$N = N_0 2^{n^{\circ} \text{ciclo}}$$

Pasando finalmente a la RT-qPCR, esta consiste en, primero, purificar el RNA de las células de interés y separarlo de cualquier material genómico. Segundo, se convierte este RNA en cDNA (por *Complementary DNA*), esto es, RNA que tiene una cadena completamente análoga al RNA que había. Este paso se logra mezclando el RNA de la muestra con DNA polimerasa y muchos primers random (primers de secuencias random), que se unirán en muchos sitios del RNA y permitirán que se *retro-transcriba* completamente. Una vez obtenido el cDNA, se utiliza este para hacer una qPCR. Por cada gen de interés, se arma una solución diferente con primers forward y reverse específicos de ese gen y se obtiene, finalmente, la cantidad de mRNAs que había en la muestra.

Inmunofluorescencia

La inmunofluorescencia es una técnica experimental muy utilizada en el marco de la biología molecular. Su objetivo es medir la cantidad y ubicación de alguna molécula de interés, sea una proteína, un mRNA, la cromatina, etc., a nivel de células únicas.

Protocolo experimental

El mecanismo consiste en fijar a las células que ya hayan recibido algún tratamiento experimental de interés (por ejemplo, el pulso de diferenciación) para luego exponerlas a un anticuerpo *primario* específico contra la molécula de interés que se quiera medir y luego lavar el preparado. Este anticuerpo primario se unirá solamente a las moléculas de interés y, una vez lavado, solo quedarán, idealmente, las moléculas que se unieron, que funcionarán de guías del anticuerpo *secundario*. El siguiente paso consiste en tratar las células con el anticuerpo secundario, que es específico contra la especie en que se hizo el primario, y luego lavar nuevamente. Así, quedarán en el preparado solo las moléculas de anticuerpos secundarios que se hayan unido a las del primario que, a su vez, estarán unidas a las moléculas de interés. Los anticuerpos secundarios son además fluorescentes en algún rango de colores o longitudes de onda particular, de manera que el preparado brillará del color del anticuerpo secundario elegido allí donde haya moléculas de interés contra las que el primario era

específico. Los dos pasos de anticuerpos permiten combinar qué molécula se quiere medir con distintas posibilidades para el color del brillo.

Terminados los tratamientos con anticuerpos, el preparado se lleva a un *microscopio de fluorescencia*. Consiste en un microscopio óptico que tiene varios *canales de excitación y filtros de medición*, además de una cámara fotográfica. Cada canal emite un rango diferente de longitudes de onda, destinado a excitar uno solo de los muchos anticuerpos secundarios. Cabe destacar que podríamos tener en simultáneo, en un mismo preparado, varios anticuerpos secundarios, cada uno contra una molécula de interés distinta. Los filtros de medición se ubican entre la muestra y la cámara, y cada filtro está destinado a dejar pasar específicamente la luz de uno solo de los anticuerpos secundarios. De esta manera, ubicamos una muestra al microscopio y la fotografiamos en cada uno de los canales, donde para cada uno se elige el par de canal de excitación y filtro de medición destinados al mismo anticuerpo.

Análisis

En esta tesis, solamente hicimos inmunofluorescencias en dos canales, agregando DAPI, una molécula fluorescente que se une a la cromatina y permite marcar el núcleo. Así, sacamos 3 ó 4 fotografías para cada condición experimental y para cada canal. Es importante destacar que las fotos deben hacerse siempre con los mismos parámetros de la cámara y el microscopio (mismo tiempo de exposición, aumento, intensidad de excitación), de manera de obtener fotografías cuantitativamente comparables. Finalmente, utilizamos el software FIJI (ImageJ) para analizar las fotos. En particular, el plug in STARDIST⁴² que toma las fotos del canal de DAPI, reconoce los núcleos celulares y los segmenta utilizando estrategias de *machine learning*. Luego, medimos, en cada segmento de cada núcleo celular de cada foto, la intensidad de cada píxel de cada canal. Calculando un promedio de esta intensidad por cada núcleo celular y por cada canal, llegamos a una medida proporcional a la concentración de nuestras moléculas de interés. Todo este procedimiento lo hacemos de manera completamente automatizada.

Sistema de Control de Oct6 exógeno

La Figura 59 muestra un segmento del plásmido agregado por PiggyBack¹⁰¹ en las células con knock out de Oct6. Este segmento contiene la secuencia

de Oct6 exógeno unida a mCherry, que permite su medición por microscopía de fluorescencia. La síntesis de Oct6-mCherry exógenos está inhibida por TetR, un represor que forma parte del plásmido, se une a su promotor y silencia el gen. El agregado de Doxyciclina evita esta unión y activa así la síntesis de Oct6-mCherry.

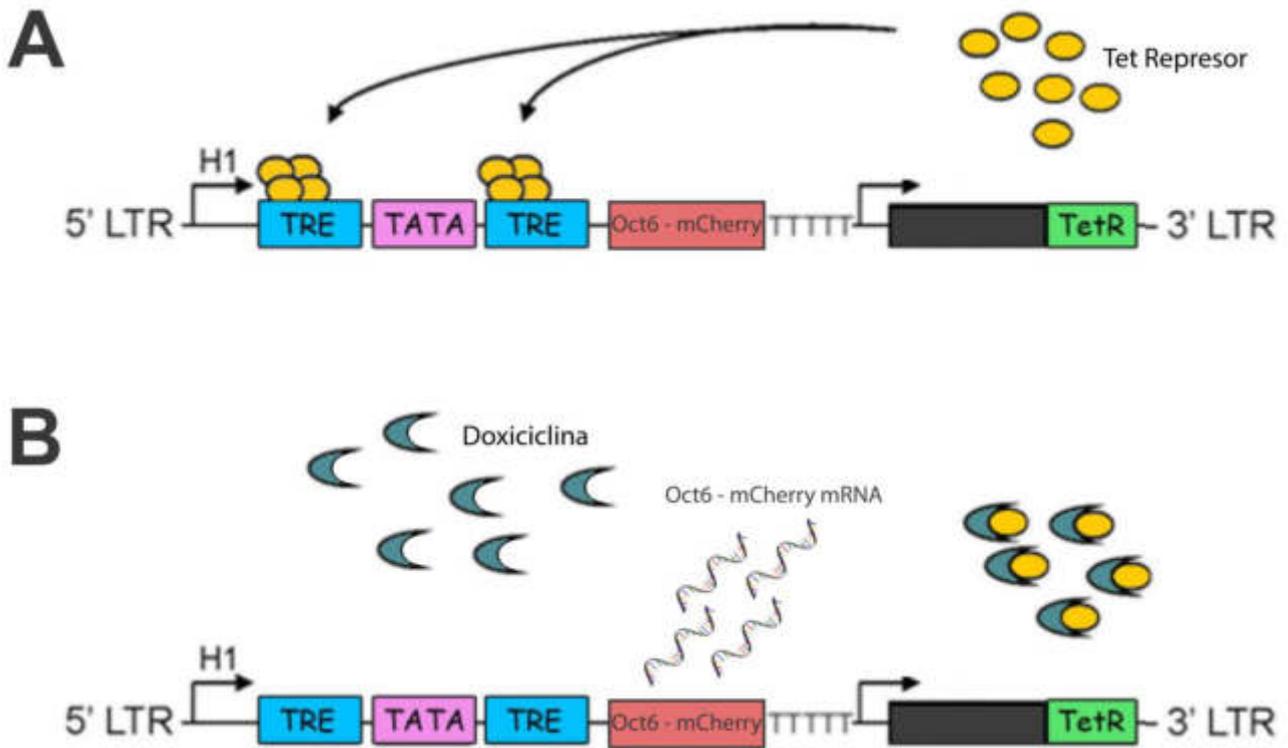


Figura 59: Imagen adaptada¹⁰². Secuencia dentro del plásmido insertado por PiggyBack en las células para controlar la expresión de Oct6-mCherry. El plásmido contiene un promotor que está siempre activado y sintetizando la molécula Tet que silencia al promotor para Oct6-mCherry (A). El agregado de Doxyciclina hace que esta se una a Tet e inhiba el pegado de este al promotor de Oct6-mCherry, activando su síntesis (B).

Discusiones y Perspectivas

En resumen, en este capítulo encontramos, primero, evidencias de la existencia de feedbacks positivos en las curvas dosis respuesta de la expresión proteica de Nanog y Oct6, los dos marcadores de los estados naïve y primed. Además, la inmunofluorescencia y su cuantificación a nivel de célula única nos permitieron encontrar bimodalidad en las distribuciones poblacionales de expresión, lo que también va en favor de la existencia de una biestabilidad donde ambos puntos fijos estables se corresponderían con los estados naïve y primed. También vemos evidencia de una mutua exclusión entre estos estadios, dado que en las fotos de inmunofluorescencia no vemos células con doble expresión positiva de Nanog y Oct6. O expresan uno u el otro, apoyando la hipótesis de que, de haber un feedback positivo, debe estar formado por una doble inhibición mutua.

Lo siguiente que estudiamos fue la reversibilidad de esta transición naïve a primed, donde encontramos que las células parecían revertir en cuanto a la expresión, a nivel transcripcional, de Nanog y Klf4 (marcadores de estado naïve) y de Oct6 y Fgf5 (marcadores de la diferenciación a primed). Sin embargo, esta reversión no es total, ya que ninguna de las expresiones vuelve a los niveles iniciales, sino que permanecen en valores intermedios. En paralelo, decidimos medir la expresión de Bra, el marcador de diferenciación a mesodermo, donde sí observamos que hay una diferencia cualitativa, y se expresa únicamente luego de la reversión. A estos resultados, les aplicamos el cuantificador de memoria y concluimos que las células están efectivamente recordando el estímulo de diferenciación, y continúan el proceso de transformación al linaje mesodérmico incluso con el estímulo quitado.

Para continuar el trabajo de este experimento, sería interesante analizar más tiempos de reversión, por ejemplo, dar un estímulo más corto que 24hrs y buscar en qué tiempo está el umbral donde las células no regresan más al estado naïve.

En cuanto al último experimento, el de las células con Oct6 exógeno, vimos que la expresión proteica de Nanog efectivamente se encuentra afectada por Oct6, lo que indica que existe alguna interacción, ya sea directa o indirecta. Además, también siguen estando los indicios de feedbacks

positivos en la bimodalidad de las distribuciones y en la mutua exclusión de las expresiones en las células únicas de Nanog y mCherry, lo que va en favor de que este feedback involucra una doble inhibición mutua. Un experimento interesante, en el que estamos pensando a futuro, consiste en construir una línea celular con Nanog exógeno, también de expresión controlable, y ver cómo varía la expresión de Oct6 cuando sobre-expresamos Nanog. En paralelo, también pensamos en hacer un Chip-seq que nos indique sobre qué lugares del genoma se unen las proteínas de Nanog y Oct6, lo que indicaría con qué genes interactúan de manera directa.

Para terminar, sería ideal hacer un Chip-seq masivo, que abarque muchos factores de transcripción ligados a la salida del estado naïve, y en vistas de poder construir más fielmente la parte de la GRN que controla este proceso. El armado de esta red no sólo permitiría estudiar dónde y cómo son los feedbacks positivos, sino que posibilitaría estudiar y modelar la dinámica de esta transición como un todo, más allá de qué motifs la componen. Poder aplicar la teoría de redes o grafos arrojaría una gran variedad de predicciones independientes y testeables experimentalmente, como ya se ha realizado para el modelo del erizo de mar⁹⁰. Es particularmente interesante que el caso del ratón es mucho más cercano al humano, donde ya se han estudiado en detalle muchas similitudes entre las GRN de ambas especies⁸.

Conclusiones, discusiones y perspectivas generales

Si bien esta tesis se divide en 3 capítulos, es importante destacar algunos aspectos generales del trabajo como un todo. En el primer capítulo, mostramos que, en nuestro modelo biológico, las células son capaces de transmitir información a su progenie que determina en parte la rapidez con que atravesarán el ciclo celular. Si bien ignoramos completamente cuál es el soporte físico de esta información, hay antecedentes parecidos, como es el caso de Far1 en levaduras⁶², que indican que podría ser la concentración de alguna proteína en el citoplasma. En este ejemplo, la célula hija hereda la misma concentración de la proteína Far1 citoplasmática. Esta proteína se inactiva mutuamente con las ciclinas y, en alta concentración, es capaz de detener el ciclo celular. Si una célula recibe un estímulo para activar Far1, pero que no es suficiente para detener su ciclo celular, igualmente transmitirá la información de este estímulo a su progenie, que heredará una cantidad de Far1 significativa. Al final, la célula hija será más sensible a un nuevo estímulo que otra célula cuya madre no había recibido estímulo alguno. Volviendo a las células madre pluripotentes de ratón, es posible que, en la mitosis, dos células hermanas hereden una dada concentración de alguna proteína que también regule la duración del ciclo celular. Esto además es coherente con la alta correlación encontrada para el largo de la fase G1 en células hermanas.

En el capítulo 2, tomamos la idea de mecanismos capaces de almacenar memoria, que tengan como soporte físico la concentración de una o varias proteínas. Con esto en mente, encontramos fundamentalmente dos tipos de mecanismos dinámicos: la multiestabilidad y las oscilaciones desfasadas. El primero se basa en que el estímulo lleva al sistema de un punto fijo estable a otro. El segundo, en que un estímulo sobre un sistema que oscila autónomamente puede generar una diferencia de fase. Una pregunta que quedó inconclusa es si puede haber algún mecanismo adicional. Lo que sí podemos descartar son algunas posibilidades para estos mecanismos. Sabemos que, si se basaran en dinámicas que terminen en puntos fijos estables, deberíamos haberlos encontrado en la integración masiva de todas las combinaciones de redes. Por otro lado, si la dinámica no tendiera a puntos fijos estables, no existe otra posibilidad que las oscilaciones. Recordemos que las variables están acotadas, de manera que tampoco existen dinámicas con variables que crezcan hasta infinito.

Una continuación muy interesante para este trabajo sería embarcarse en estudiar combinaciones de más de 3 nodos. Aquí, las posibilidades dinámicas son mayores, pudiendo haber incluso variables caóticas que permitan nuevos mecanismos de almacenamiento de estímulos. Si bien es un trabajo de mucha más complejidad (la cantidad de combinaciones asciende a más de medio millón para 4 nodos), lo aprendido en esta tesis podría simplificar este análisis. Las simetrías encontradas en los motifs con memoria deberían respetarse para redes más grandes, reduciendo mucho el espacio de combinaciones de búsqueda. Por otro lado, la metarred donde ordenamos las combinaciones da un espacio topológico que permitiría diseñar una estrategia de movimiento en ese espacio para una búsqueda dirigida en vez de exhaustiva.

Lo que sí encontramos en el capítulo 2, son todas las combinaciones posibles de feedbacks positivos que dan lugar a biestabilidades y memoria. Esta es la conexión con el capítulo 3, donde buscamos este tipo de feedbacks dentro del mecanismo que regula la salida del estado naïve pluripotente. Si bien encontramos evidencias que apoyan la existencia de estos feedbacks, como las curvas dosis respuesta de Nanog y Oct6 y la irreversibilidad de la transición naïve a primed, es difícil detectar interacciones directas entre genes que, a priori, no sabemos si son los que interactúan. Sería interesante continuar este trabajo en esa dirección, algo en lo que posiblemente nos embarquemos en la etapa futura. Por otro lado, hemos podido aplicar nuestro cuantificador de memoria y cuantificarla para algunos genes cuya expresión cambia de manera irreversible en la salida del estado naïve. Esta es una de las conexiones más grandes entre la parte puramente teórica como es el diseño del cuantificador y la parte más experimental de esta tesis.

Referencias

1. Mackie, S. A. & Bates, G. W. Contribution of the doctoral education environment to PhD candidates' mental health problems: a scoping review. *High. Educ. Res. Dev.* **38**, 565–578 (2019).
2. Waisman, A. *et al.* Cell cycle dynamics of mouse embryonic stem cells in the ground state and during transition to formative pluripotency. *Sci. Rep.* **9**, (2019).
3. Yang, P. *et al.* Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. *Cell Syst.* **8**, 427-445.e10 (2019).
4. Alberts. *Biología molecular de la célula.* (Omega, 1983).
5. Thomas D. Pollard, William C. Earnshaw, Jennifer Lippincott-Schwartz, G. J. *Cell Biology.* (2016).
6. Ferrell, J. E., Tsai, T. Y. C. & Yang, Q. Modeling the cell cycle: Why do certain circuits oscillate? *Cell* **144**, 874–885 (2011).
7. Ferrell, J. E., Tsai, T. Y. C. & Yang, Q. Modeling the cell cycle: Why do certain circuits oscillate? *Cell* **144**, 874–885 (2011).
8. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: Naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).
9. Saiz, N. & Plusa, B. Early cell fate decisions in the mouse embryo. *Reproduction* **145**, (2013).
10. Hackett, J. A. & Azim Surani, M. Regulatory principles of pluripotency: From the ground state up. *Cell Stem Cell* vol. 15 416–430 (2014).
11. Hirai, H., Karian, P. & Kikyo, N. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. *Biochem. J.* **438**, 11–23 (2011).
12. Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
13. Dalton, S. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol.* **25**, 592–600 (2015).
14. Waisman, A. Relevancia del Ciclo Celular en la Definición del Destino Celular durante la Diferenciación de Células Madre Embrionarias de Ratón. (Universidad de Buenos Aires, 2017).
15. Simunovic, M. *et al.* A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. *Nat. Cell Biol.* **21**, 900–910 (2019).

16. Morgani, S. M., Metzger, J. J., Nichols, J., Siggia, E. D. & Hadjantonakis, A. K. Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. *Elife* **7**, 1–35 (2018).
17. Etoc, F. *et al.* A Balance between Secreted Inhibitors and Edge Sensing Controls Gastruloid Self-Organization. *Dev. Cell* **39**, 302–315 (2016).
18. Gunne-Braden, A. *et al.* GATA3 Mediates a Fast, Irreversible Commitment to BMP4-Driven Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell* **26**, 693–706.e9 (2020).
19. Scherf, N. *et al.* Imaging, quantification and visualization of spatio-temporal patterning in mESC colonies under different culture conditions. *Bioinformatics* **28**, 556–561 (2012).
20. Mojtahedi, M. *et al.* Cell Fate Decision as High-Dimensional Critical State Transition. *PLoS Biol.* **14**, 1–28 (2016).
21. Noor, N. *et al.* 3D Printing of Personalized Thick and Perfusable Cardiac Patches and Hearts. *Adv. Sci.* **6**, (2019).
22. Xia, Z., Jin, S. & Ye, K. Tissue and Organ 3D Bioprinting. *SLAS Technol.* **23**, 301–314 (2018).
23. Sakaue-Sawano, A. *et al.* Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132**, 487–498 (2008).
24. Downey, M. J. *et al.* Extracting fluorescent reporter time courses of cell lineages from high-throughput microscopy at low temporal resolution. *PLoS One* **6**, (2011).
25. Kishore, J., Goel, M. & Khanna, P. Understanding survival analysis: Kaplan-Meier estimate. *Int. J. Ayurveda Res.* **1**, 274 (2010).
26. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap. An introduction to Bootstrap* (1993).
27. Sandler, O. *et al.* Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519**, 468–471 (2015).
28. Richard Cowan and Robert Staudte. The Bifurcating Autoregression Model in Cell Lineage Studies. *Biometrics* **42**, 769–783 (1986).
29. Grassberger, P. & Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349 (1983).
30. Sandler, O. *et al.* Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519**, 468–471 (2015).

31. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The Quickhull Algorithm for Convex Hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
32. Lü, D., Luo, C., Zhang, C., Li, Z. & Long, M. Differential regulation of morphology and stemness of mouse embryonic stem cells by substrate stiffness and topography. *Biomaterials* **35**, 3945–3955 (2014).
33. Gothard, D., Dawson, J. I. & Oreffo, R. O. C. Assessing the potential of colony morphology for dissecting the CFU-F population from human bone marrow stromal cells. *Cell Tissue Res.* **352**, 237–247 (2013).
34. Boward, B., Wu, T. & Dalton, S. Concise Review: Control of Cell Fate Through Cell Cycle and Pluripotency Networks. *Stem Cells* vol. 34 1427–1436 (2016).
35. Martin Gonzalez, J. *et al.* Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. *Stem Cell Reports* **7**, 177–191 (2016).
36. Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–485 (2015).
37. Cannon, D., Corrigan, A. M., Miermont, A., McDonel, P. & Chubb, J. R. Multiple cell and population-level interactions with mouse embryonic stem cell heterogeneity. *Dev.* **142**, 2840–2849 (2015).
38. Edel, M. J., Menchon, C., Vaquero, J. M. A. & Izpisua Belmonte, J. C. A protocol to assess cell cycle and apoptosis in human and mouse pluripotent cells. *Cell Commun. Signal.* **9**, 1–11 (2011).
39. Coronado, D. *et al.* A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10**, 118–131 (2013).
40. Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J. & Rathjen, P. D. Transient pluripotent cell populations during primitive ectoderm formation: Correlation of in vivo and in vitro pluripotent cell development. *J. Cell Sci.* **115**, 329–339 (2002).
41. Snow, M. H. L. Gastrulation in the mouse: growth and regionalization of the epiblast. *J. Embryol. Exp. Morphol.* **Vol. 42**, 293–303 (1977).
42. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
43. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).

44. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Dev.* **144**, 1221–1234 (2017).
45. Angeli, D., Ferrell, J. E. & Sontag, E. D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1822–1827 (2004).
46. Han, Z. *et al.* Signal transduction network motifs and biological memory. *J. Theor. Biol.* (2007) doi:10.1016/j.jtbi.2007.01.022.
47. Mönke, G. *et al.* Excitability in the p53 network mediates robust signaling with tunable activation thresholds in single cells. *Sci. Rep.* **7**, 1–14 (2017).
48. Chakrabarti, B. K. & Acharyya, M. Dynamic transitions and hysteresis. *Rev. Mod. Phys.* **71**, 847–859 (1999).
49. Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* **18**, 180–191 (2017).
50. Yoney, A. *et al.* WNT signaling memory is required for ACTIVIN to function as a morphogen in human gastruloids. *Elife* **7**, (2018).
51. Jiang, Y., AkhavanAghdam, Z., Li, Y., Zid, B. M. & Hao, N. A Protein Kinase A-Regulated Network Encodes Short- And Long-Lived Cellular Memory. *bioRxiv* **3585**, (2019).
52. Harmon, B. *et al.* Timescale Separation of Positive and Negative Signaling Creates History-Dependent Responses to IgE Receptor Stimulation. *Sci. Rep.* (2017) doi:10.1038/s41598-017-15568-2.
53. Xiong, W. & Ferrell, J. E. A positive-feedback-based bistable ‘memory module’ that governs a cell fate decision. *Nature* **426**, 460–465 (2003).
54. Burrill, D. R. & Silver, P. A. Making Cellular Memories. *Cell* **140**, 13–18 (2010).
55. Stan, R. C., Bhatt, D. K. & de Camargo, M. M. Cellular Adaptation Relies on Regulatory Proteins Having Episodic Memory: Proteins Modulate Cell Metabolism and Reproduction by Remembering, Transmitting, and Using Data on the Environment. *BioEssays* **42**, 1–7 (2020).
56. Atay, O., Doncic, A. & Skotheim, J. M. Switch-like Transitions Insulate Network Motifs to Modularize Biological Networks. *Cell Syst.* **3**, 121–132 (2016).
57. Ferrell, J. E. Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology* vol. 14 140–148 (2002).

58. Johnston, R. J., Chang, S., Etchberger, J. F., Ortiz, C. O. & Hobert, O. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12449–12454 (2005).
59. Gordley, R. M. *et al.* Engineering dynamical control of cell fate switching using synthetic phosphoregulons. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13528–13533 (2016).
60. Jiang, Y. & Hao, N. Memorizing environmental signals through feedback and feedforward loops. *Curr. Opin. Cell Biol.* **69**, 96–102 (2021).
61. Mitrophanov, A. Y. & Groisman, E. A. Positive feedback in cellular control systems. *BioEssays* vol. 30 542–555 (2008).
62. Doncic, A. *et al.* Compartmentalization of a bistable switch enables memory to cross a feedback-driven transition. *Cell* **160**, 1182–1195 (2015).
63. Keener, J. P. & Sneyd, J. *Mathematical Physiology. Book 580* (2008) doi:10.1007/978-0-387-79388-7.
64. Michaelis, L., Menten, M. L., Goody, R. S. & Johnson, K. A. Die Kinetik der Invertinwirkung/ The kinetics of invertase action. *Biochemistry* **49**, 333–369 (1913).
65. Ferrell, J. E. & Ha, S. H. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in Biochemical Sciences* vol. 39 496–503 (2014).
66. Gomez-Urbe, C., Verghese, G. C. & Mirny, L. A. Operating regimes of signaling cycles: Statics, dynamics, and noise filtering. *PLoS Comput. Biol.* **3**, 2487–2497 (2007).
67. Alon, U. Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
68. Adler, M. & Alon, U. Fold-change detection in biological systems. *Curr. Opin. Syst. Biol.* **8**, 81–89 (2018).
69. Alon, U. *An Introduction to Systems Biology. An Introduction to Systems Biology* (2006). doi:10.1201/9781420011432.
70. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* (2009) doi:10.1016/j.cell.2009.06.013.
71. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**, 55–61 (2000).
72. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* **138**, 760–773 (2009).

73. Ananthasubramaniam, B. & Herzog, H. Positive feedback promotes oscillations in negative feedback loops. *PLoS One* **9**, (2014).
74. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
75. Tsai, T. Y. C. *et al.* Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* (80-.). (2008) doi:10.1126/science.1156951.
76. Mitra, T., Menon, S. N. & Sinha, S. Emergent memory in cell signaling: Persistent adaptive dynamics in cascades can arise from the diversity of relaxation time-scales. *Sci. Rep.* **8**, 1–16 (2018).
77. Schröter, C. *et al.* Topology and dynamics of the zebrafish segmentation clock core circuit. *PLoS Biol.* **10**, 11 (2012).
78. Fell, J. & Axmacher, N. The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* **12**, 105–118 (2011).
79. Adler, M., Szekely, P., Mayo, A. & Alon, U. Optimal Regulatory Circuit Topologies for Fold-Change Detection. *Cell Syst.* **4**, 171-181.e8 (2017).
80. Ferrell, J. E. Perfect and near-perfect adaptation in cell signaling. *Cell Syst.* **2**, 62–67 (2016).
81. Coyne, J. A. Comment on ‘Gene regulatory networks and the evolution of animal body plans’. *Science* (80-.). **313**, 796–801 (2006).
82. Davidson, E. H. Gene Regulatory Networks for Development : What They Are , How They Work , and What They Mean General Structural Properties of Developmental Gene Gene Regulatory Networks for Embryonic. in *The regulatory genome: gene regulatory networks in development and evolution.* 125–185 (2006).
83. Cheng, C. *et al.* Construction and analysis of an integrated regulatory network derived from High-Throughput sequencing data. *PLoS Comput. Biol.* **7**, (2011).
84. Bolouri, H. & Davidson, E. H. Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc. Natl. Acad. Sci. U. S. A.* (2003) doi:10.1073/pnas.1533293100.
85. Cary, G. A. *et al.* Systematic comparison of sea urchin and sea star developmental gene regulatory networks explains how novelty is incorporated in early development. *Nat. Commun.* **11**, 1–9 (2020).
86. Dunn, S. J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription

- factor program for naïve pluripotency. *Science (80-.)*. **344**, 1156–1160 (2014).
87. Pfeuty, B., Kress, C. & Pain, B. Network Features and Dynamical Landscape of Naive and Primed Pluripotency. *Biophys. J.* **114**, 237–248 (2018).
 88. Pinto, J. P. *et al.* StemCellNet: An interactive platform for network-oriented investigations in stem cell biology. *Nucleic Acids Res.* **42**, 154–160 (2014).
 89. Davidson, E. H. & Erwin, D. H. Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science (80-.)*. **311**, 796–800 (2006).
 90. Peter, I. S. & Davidson, E. H. *Genomic Control Process: Development and Evolution. Genomic Control Process: Development and Evolution* (2015). doi:10.1016/C2012-0-02817-7.
 91. Peter, I. S., Faure, E. & Davidson, E. H. Predictive computation of genomic logic processing functions in embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16434–16442 (2012).
 92. Das, P. M., Ramachandran, K., VanWert, J. & Singal, R. Chromatin immunoprecipitation assay. *Biotechniques* **37**, 961–969 (2004).
 93. Carullo, N. V. N. & Day, J. J. Genomic enhancers in brain health and disease. *Genes (Basel)*. **10**, (2019).
 94. MacArthur, B. D., Maayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* **10**, 672–681 (2009).
 95. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
 96. Tapia, N. *et al.* Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci. Rep.* **5**, 1–9 (2015).
 97. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
 98. Wang, Y. *et al.* Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Dev. Cell* **25**, 69–80 (2013).
 99. Wang, Y. *et al.* Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Dev. Cell* **25**, 69–80 (2013).
 100. Pennisi, E. The CRISPR Craze. *Science (80-.)*. **341**, 833–836 (2013).
 101. Grabundzija, I. *et al.* Comparative analysis of transposable element vector systems in human cells. *Mol. Ther.* **18**, 1200–1209 (2010).
 102. Waisman, A. Estudio de remodeladores de cromatina en el mantenimiento del estado indiferenciado

y la diferenciación de células madre embrionarias murinas. (Universidad de Buenos Aires, 2012).